



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	Σημσιολογικά Κατευθυνόμενη Εξόρυξη Γνώσης με Χρήση Οντολογιών
Όνοματεπώνυμο Φοιτητή	Πετράκης Ανδρέας του Δημητρίου
Αριθμός Μητρώου	ΜΠΣΠ/09060
Κατεύθυνση	Συστήματα Υποστήριξης Αποφάσεων
Επιβλέπων	Ιωάννης Θεοδωρίδης, Αναπληρωτής Καθηγητής
Συνεπιβλέπων	Ευάγγελος Κοτσιφάκος, Διδάκτωρ

Πανεπιστήμιο Πειραιώς-Τμήμα Πληροφορικής
Πρόγραμμα Μεταπτυχιακών Σπουδών στα
Προηγμένα Συστήματα Πληροφορικής

Ημερομηνία Παράδοσης **Σεπτέμβριος 2011**

Περίληψη

Η εξέλιξη του Διαδικτυακού Ιστού σε Σημασιολογικό Ιστό έχει δώσει ώθηση στην ανάπτυξη προτύπων και τεχνολογιών για την έκφραση και τη διάδοση των Οντολογιών. Αυτό τις καθιστά άμεσα ένα ισχυρό εργαλείο, του οποίου η χρήση πρέπει να επανεξετασθεί σε όλα τα επιστημονικά πεδία όπως και σε αυτό της Εξόρυξης Γνώσης από Δεδομένα. Κάνοντας χρήση της αναπαράστασης μίας αντίληψης/ έννοιας μέσω των OWL Οντολογιών, των Μηχανών Συλλογιστικής Ανάλυσης για την εκμετάλλευση της εκφραστικότητας των Οντολογιών και των Μετρικών για την ανάδειξη σημασιολογικών μέτρων απόστασης/ ομοιότητας, είναι δυνατόν να επιτευχθεί η ενσωμάτωση των οντολογιών στις τεχνικές εξόρυξης γνώσης για την εξαγωγή ενδιαφερόντων προτύπων. Στην παρούσα εργασία, αυτή η ενσωμάτωση εξετάζεται μέσα από τον αλγόριθμο εξαγωγής Κανόνων Συσχετίσεων Apriori του Weka, και αξιολογείται σε δεδομένα από το πεδίο της Τρομοκρατίας. Η προσέγγιση αυτή καταφέρνει αυτόματα να υποστηρίξει υποθέσεις σε μορφή κανόνων που ειδικοί είχαν υιοθετήσει μέσω διαφόρων στατιστικών προσεγγίσεων.

Λέξεις κλειδιά: Αναπαράσταση Γνώσης, Οντολογία, Εξόρυξη Γνώση από Δεδομένα, Κανόνες Συσχετίσεων, Apriori, Τρομοκρατία

Abstract

The evolution of the Web in Web Semantic has motivated the development of standards and technologies for the expression and distribution of Ontologies. This makes them immediately a powerful tool that should be reviewed for use in various scientific fields, and in particular in that of Data Mining. Using a conception of representation through OWL Ontologies, using Reasoning Engines Analysis for operating the expressiveness of Ontologies and metrics for the emergence of semantic measures of distance/ similarity, it is possible to achieve the integration of the data mining techniques for mining interestingness patterns. This integration is examined through the Association Rules algorithm Apriori in Weka and evaluated in the field of terrorism. This approach is able to support cases in the form of specific rules that have been adopted by the experts using various statistical approaches.

Keywords: Knowledge Representation, Ontology, Data Mining, Association Rules, Apriori, Terrorism

Πανεπιστήμιο Πειραιώς

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Ι. Θεοδωρίδης (επιβλέπων)
Αν. Καθηγητής

Ι. Σίσκος
Καθηγητής

Ν. Πελέκης
Λέκτορας

Πίνακας Περιεχομένων

Περίληψη	2
Abstract.....	2
Λίστα Διαγραμμάτων	6
Λίστα Εικόνων	7
Λίστα Κώδικα	8
Λίστα Παραδειγμάτων.....	9
Λίστα Πινάκων	10
1. Εισαγωγή.....	12
ΜΕΡΟΣ Α - ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	13
2. Σχετικές εργασίες.....	13
3. Γνωσιακή Επιστήμη.....	22
3.1 Αναπαράσταση Γνώσης.....	24
3.2 Μέθοδοι Αναπαράστασης Γνώσης	25
4. Οντολογίες.....	28
4.1 Ορισμός Οντολογίας	29
4.2 Χαρακτηριστικά Οντολογίας.....	31
4.3 Κατηγορίες Οντολογιών	32
5. Αξιοποίηση Οντολογιών	38
5.1 OWL 1 και OWL 2.....	40
5.2 Μηχανές Συλλογιστικής Ανάλυσης	49
5.3 Σημασιολογική Συνάφεια	51
6. DM και Οντολογία.....	56
ΜΕΡΟΣ Β – ΑΝΑΛΥΣΗ, ΣΧΕΔΙΑΣΗ & ΥΛΟΠΟΙΗΣΗ.....	58
7. Εισαγωγή.....	58
8. Ανάλυση απαιτήσεων	58
8.1 Σκοπός συστήματος.....	59
8.2 Παρούσα κατάσταση.....	59
8.3 Προσδιορισμός συστήματος	61
8.4 Μη λειτουργικές απαιτήσεις	61
8.5 Λειτουργικές απαιτήσεις.....	62
8.6 Επιλογή εργαλείων	63
9. Ανάλυση και Σχεδιασμός	63
9.1 Διάγραμμα Τάξεων	64
9.2 Διαγράμματα Καταστάσεων	68
9.3 Διαγράμματα Ακολουθίας	68
9.4 Διαγράμματα Πακέτων	70
10. Υλοποίηση	71
10.1 GUIdm.java	72
10.2 tabOAssociationRule.java.....	72
10.3 oApriori.java	74
10.4 MappingFile.java.....	77
10.5 OntologyReasoner.java	79
10.6 JenaReasoner.java	80
10.7 PelletReasoner.java.....	80
10.8 Metrics.....	81
10.9 JenaMertrics.java	81
10.10 Σύνοψη.....	82
11. Έλεγχος (Εφαρμογή – Αποτίμηση)	83
11.1 Πρώτο Σενάριο.....	84
11.2 Δεύτερο Σενάριο	85

11.3 Τρίτο Σενάριο	85
11.4 Τέταρτο Σενάριο	86
11.5 Σύνοψη.....	87
12. Επεκτάσεις.....	88
12.1 Μετρική TaxonomyMetrics.java	88
12.2 OAprioriWithoutSupport.java	90
ΜΕΡΟΣ Γ – ΠΕΡΙΠΤΩΣΗ ΧΡΗΣΗΣ.....	93
13. Εισαγωγή.....	93
14. Σενάριο χρήσης	96
15. Δεδομένα εκπαίδευσης.....	97
16. Οντολογία πεδίου	99
17. Συσχέτιση δεδομένων εκπαίδευσης & εννοιών Οντολογίας.....	101
18. Σημαιολογικές αποστάσεις.....	102
19. Ανάλυση αποτελεσμάτων	104
19.1 Στατιστικά στοιχεία	104
19.2 Πρώτη περίπτωση ανάλυσης	113
19.3 Δεύτερη περίπτωση ανάλυσης	114
19.4 Τρίτη περίπτωση ανάλυσης.....	115
19.5 Τέταρτη περίπτωση ανάλυσης.....	117
19.6 Πέμπτη περίπτωση ανάλυσης.....	117
19.7 Έκτη περίπτωση ανάλυσης	118
20. Σύνοψη	121
21. Μελλοντική μελέτη και βελτιώσεις	123
22. Επίλογος.....	123
Γλωσσάρι.....	125
Βιβλιογραφία.....	130
ΠΑΡΑΡΤΗΜΑΤΑ.....	138
Α Παράρτημα - Σύμβολα εκφραστικότητας DL & Πολυπλοκότητα	138
Β Παράρτημα – Διαφορές μεταξύ OWL 1 Lite- DL- Full	139
Γ Παράρτημα – Διαφορές μεταξύ OWL 2 EL- QL- RL	141
Δ Παράρτημα – Βάσεις Δεδομένων Τρομοκρατίας.....	144
Ε Παράρτημα – Οντολογίες Τρομοκρατίας	145
Ζ Παράρτημα – Εγκατάσταση Reasoner ή Metrics	145
Η Παράρτημα – Εγκατάσταση PatternMinner.....	146
Θ Παράρτημα – Οδηγός Χρήσης	149
Ι Παράρτημα – Scripts για την δημιουργία των δεδομένων εκπαίδευσης.....	152
Κ Παράρτημα – Αποτελέσματα σεναρίων	157

Λίστα Διαγραμμάτων

Διάγραμμα 8-1. Διάγραμμα Περιπτώσεων Χρήσης.....	62
Διάγραμμα 9-1. Διάγραμμα Τάξεων	64
Διάγραμμα 9-2. Διάγραμμα Καταστάσεων Συστήματος.....	68
Διάγραμμα 9-3. Μη ενσωμάτωση Οντολογιών.....	69
Διάγραμμα 9-4. Ενσωμάτωση Οντολογιών	70
Διάγραμμα 9-5. Γενικό Διάγραμμα Πακέτων	70
Διάγραμμα 9-6. Διάγραμμα Πακέτων Source Packages.....	71
Διάγραμμα 10-1. Έννοιες – Συσχετίσεις Mapping File	77
Διάγραμμα 19-1. Τρομοκρατικές επιθέσεις ανά έτος.	104
Διάγραμμα 19-2. Τρομοκρατικές επιθέσεις ανά μήνα	105
Διάγραμμα 19-3. Τρομοκρατικές επιθέσεις ανά περιοχή	107
Διάγραμμα 19-4. Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης.....	107
Διάγραμμα 19-5. Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών	111
Διάγραμμα 19-6. Τρομοκρατικές επιθέσεις ανά αριθμό τραυματιών.....	112
Διάγραμμα 19-7. Τρομοκρατικές επιθέσεις ανά μέγεθος υλικών ζημιών	112

Λίστα Εικόνων

Εικόνα 2-1. Διαδικασία Εξόρυξης Γνώσης από Δεδομένα	14
Εικόνα 2-2. Σημεία εισαγωγής μέτρων ενδιαφέροντος.....	15
Εικόνα 2-3. Επίπεδα σημασιολογικού δικτύου	22
Εικόνα 3-1. Πυραμίδα αναπαράστασης γνώσης.....	24
Εικόνα 4-1. Είδη Οντολογιών	33
Εικόνα 4-2. Ταξινόμηση Οντολογιών με βάση το επίπεδο γενικότητας τους	36
Εικόνα 4-3. Βαθμός Αποδοχής Οντολογίας	37
Εικόνα 5-1. Επίπεδα σημασιολογικού ιστού.....	38
Εικόνα 5-2. Πρόγονοι τις OWL	40
Εικόνα 5-3. Υπογλώσσες OWL 1 και OWL 2	49
Εικόνα 6-1. Εξόρυξη γνώσης από δεδομένα.....	56
Εικόνα 8-1. Αρχιτεκτονική για PBMS ενισχυμένο με οντολογίες	60
Εικόνα 8-2. Στιγμιότυπο από την εφαρμογή Pattern-Miner ver. 2.0.....	60
Εικόνα 8-3. Διεπαφή συστήματος.....	62
Εικόνα 11-1. Μέρος αρχείου δεδομένων εκπαίδευσης Apriori	83
Εικόνα 11-2. Μέρος OWL Οντολογίας AKTiveSAOntology.....	84
Εικόνα 12-1. Ιεραρχίες πριν και μετά την χρήση Reasoner	88
Εικόνα 12-2. Νέα διεπαφή συστήματος.....	91
Εικόνα 16-1. Μικροθεωρίες στην OpenCyc.....	100
Εικόνα Θ-1. Διεπαφή Ontology Association Rules.....	150
Εικόνα Θ-2. Ενεργά πεδία για παραγωγή κανόνων χωρίς support	151

Λίστα Κώδικα

Κώδικας 10-1. Επέκταση GUIdm.java.....	72
Κώδικας 10-2. Γ' μέρος της tabOAssociationRule.java.....	73
Κώδικας 10-3. Δ' μέρος της tabOAssociationRule.java.....	74
Κώδικας 10-4. Προσθήκη στην μέθοδο oApriori.buildAssociations.....	74
Κώδικας 10-5. Μέθοδος oApriori.removeNoMappingColumns.....	75
Κώδικας 10-6. Προσθήκη στην μέθοδο oApriori.findLargestItemSets.....	75
Κώδικας 10-7. Μέθοδος oApriori.deleteItemSetsByDistances.....	76
Κώδικας 10-8. Μέθοδος oApriori.checkDistance.....	76
Κώδικας 10-9. Προσθήκη στην μέθοδο oApriori.findRulesQuickly.....	77
Κώδικας 10-10. Μέθοδος oApriori.deleteRulesBySubDistances.....	77
Κώδικας 10-11. Μέθοδος MappingFile.readMappingFile.....	78
Κώδικας 10-12. Μέθοδος MappingFile.checkMappingFile.....	78
Κώδικας 10-13. Μέθοδος MappingFile.get.....	79
Κώδικας 10-14. Μέθοδος MappingFile.getHashtable.....	79
Κώδικας 10-15. Μέθοδος MappingFile.toString.....	79
Κώδικας 10-16. Μέθοδος OntologyReasoner.readOntology.....	79
Κώδικας 10-17. Μέθοδος OntologyReasoner.readOntologySchemaModel.....	80
Κώδικας 10-18. Μέθοδος OntologyReasoner.getResource.....	80
Κώδικας 10-19. Μέθοδος OntologyReasoner.getInfModel.....	80
Κώδικας 10-20. Μέθοδος OntologyReasoner.validation.....	80
Κώδικας 10-21. Μέθοδος JenaReasoner.validation.....	80
Κώδικας 10-22. Μέθοδος PelletReasoner.validation.....	81
Κώδικας 10-23. Μέθοδος Metrics.distance.....	81
Κώδικας 10-24. Μέθοδος Metrics.isDuplex.....	81
Κώδικας 10-25. Μέθοδος JenaMetrics.distance.....	81
Κώδικας 10-26. Μέθοδος JenaMetrics.getFilter.....	82
Κώδικας 10-27. Μέθοδος JenaMetrics.isDuplex.....	82
Κώδικας 12-1. Μέθοδος TaxonomyMetrics.distance().....	89
Κώδικας 12-2. Μέθοδος TaxonomyMetrics.isDuplex().....	89
Κώδικας 12-3. Μέθοδος TaxonomyMetrics.taxonomySimilarity().....	89
Κώδικας 12-4. Μέθοδος TaxonomyMetrics.getHierarchySimilarity().....	90
Κώδικας 12-5. Μέθοδος TaxonomyMetrics.getHierarchyDeference().....	90
Κώδικας 12-6. Μέρος μεθόδου oAprioriWithoutSupport.buildAssociations().....	92
Κώδικας 12-7. Μέθοδος oAprioriWithoutSupport.meanRuleNoDuplexDistance().....	92
Κώδικας 12-8. Μέθοδος oAprioriWithoutSupport.deleteRulesWithNaNConf().....	93

Λίστα Παραδειγμάτων

Παράδειγμα 4-1. Οντολογία και στιγμιότυπο της.....	31
Παράδειγμα 4-2. Τυπικότητα Οντολογιών.....	33
Παράδειγμα 4-3. Κατάλογος - Γλωσσάρι.....	33
Παράδειγμα 4-4. Θησαυρός.....	34
Παράδειγμα 4-5. Θησαυρός, Ταξινόμια.....	35
Παράδειγμα 4-6. Πλαίσια, Περιορισμό τιμών.....	35
Παράδειγμα 4-7. Περιορισμοί Γενικής Λογικής.....	36
Παράδειγμα 4-8 Τρισδιάστατη κατηγοριοποίηση Οντολογιών.....	37
Παράδειγμα 5-1. RDF graph, RDF/XML Syntax.....	40
Παράδειγμα 5-2. OWL απαρίθμηση.....	41
Παράδειγμα 5-3. Ορισμός περιορισμού ιδιότητας.....	42
Παράδειγμα 5-4. Παράδειγμα δήλωσης ατόμων στην OWL.....	44
Παράδειγμα 10-1. Αποδοχή Στοιχειοσυνόλου.....	75
Παράδειγμα 10-2. Αποδοχή Κανόνα.....	76
Παράδειγμα 10-3. Mapping File.....	78
Παράδειγμα 12-1. Μέση σημασιολογική απόσταση κανόνα.....	91

Λίστα Πινάκων

Πίνακας 2-1. Μέτρα ενδιαφέροντος	15
Πίνακας 2-2. Εργαλεία διαχείρισης οντολογιών	19
Πίνακας 2-3 Γλώσσες Οντολογίας και Αναπαράστασης Γνώσης.....	20
Πίνακας 2-4. Μηχανές συλλογιστικής ανάλυσης	21
Πίνακας 5-1. Περιορισμοί ιδιοτήτων	42
Πίνακας 5-2. OWL τομή, ένωση, συμπλήρωμα.....	42
Πίνακας 5-3. Αξιώματα ιδιοτήτων	43
Πίνακας 5-4. Γεγονότα για την ταυτοποίηση των ατόμων	44
Πίνακας 5-5. Συντακτικές διευκολύνσεις OWL 2	45
Πίνακας 5-6. Δομές αυξημένης εκφραστικότητας.....	45
Πίνακας 5-7. Δομές για τύπους δεδομένων.....	46
Πίνακας 5-8. Τύποι δεδομένων	46
Πίνακας 5-9. Δυνατότητες σχολιασμού	47
Πίνακας 5-10. Πολυπλοκότητες σε προβλήματα συλλογισμού	50
Πίνακας 5-11. Χαρακτηριστικά μετρικών	55
Πίνακας 8-1. Πακέτα υλοποίησης Pattern-Minner	59
Πίνακας 8-2. Περιγραφή διεπαφής	63
Πίνακας 9-1. Τάξη tabOAssociationRule	64
Πίνακας 9-2. Τάξη MappingFile	65
Πίνακας 9-3. Τάξη OntologyReasoner {αφηρημένη}	65
Πίνακας 9-4. Τάξη oApriori	66
Πίνακας 9-5. Τάξη Metrics {αφηρημένη}.....	67
Πίνακας 10-1. Μέθοδοι της OntologyReasoner.java	79
Πίνακας 11-1. Αποτελέσματα Apriori με min Conf: 0.9.....	84
Πίνακας 11-2. Αποτελέσματα Apriori με min Conf: 0.3.....	84
Πίνακας 11-3. Δεύτερο Σενάριο Περίπτωση α'	85
Πίνακας 11-4. Δεύτερο Σενάριο Περίπτωση β'	85
Πίνακας 11-5. Δεύτερο Σενάριο Περίπτωση γ'	85
Πίνακας 11-6. Τρίτο Σενάριο Περίπτωση α'	85
Πίνακας 11-7. Τρίτο Σενάριο Περίπτωση β'	86
Πίνακας 11-8. Τρίτο Σενάριο Περίπτωση γ'	86
Πίνακας 11-9. Τέταρτο Σενάριο, Περίπτωση α'	86
Πίνακας 11-10. Τέταρτο Σενάριο, Περίπτωση β'	86
Πίνακας 11-11. Τέταρτο Σενάριο, Περίπτωση γ'	86
Πίνακας 11-12. Σύγκριση Σεναρίων	87
Πίνακας 12-1. Μέθοδοι της TaxonomyMetrics.java.....	88
Πίνακας 15-1. Πολιτική ιδεολογία και οργανώσεις.....	97
Πίνακας 15-2. Χρήσιμες μεταβλητές από GTD	98
Πίνακας 16-1. Υποσύνολο εννοιών που αποτελούν την Οντολογία Τρομοκρατίας.....	100
Πίνακας 17-1. Αντιστοίχιση μεταβλητών δεδομένων εκπαίδευσης & εννοιών Οντολογίας... ..	101
Πίνακας 17-2. Mapping File Οντολογία τρομοκρατίας.....	101
Πίνακας 18-1. Σημσιολογικές αποστάσεις εννοιών Οντολογίας.....	102
Πίνακας 18-2. Αύξουσα διάταξη αποστάσεων εννοιών.....	103
Πίνακας 19-1. Τρομοκρατικές επιθέσεις ανά έτος	104
Πίνακας 19-2. Τρομοκρατικές επιθέσεις ανά μήνα	105
Πίνακας 19-3. Τρομοκρατικές επιθέσεις ανά περιοχή, χώρα	105

Πίνακας 19-4. Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης	107
Πίνακας 19-5. Τρομοκρατικές επιθέσεις ανά τύπο στόχου	108
Πίνακας 19-6. Τρομοκρατικές επιθέσεις ανά εθνικότητα στόχου	109
Πίνακας 19-7. Τρομοκρατικές επιθέσεις ανά Ιδεολογία, Οργάνωση	110
Πίνακας 19-8. Τρομοκρατικές επιθέσεις ανά είδος όπλου	111
Πίνακας 19-9. Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών	111
Πίνακας 19-10. Τρομοκρατικές επιθέσεις ανά αριθμό τραυματιών	112
Πίνακας 19-11. Τρομοκρατικές επιθέσεις ανά μέγεθος υλικών ζημιών	112
Πίνακας 19-12. Αποτελέσματα Arriori με min Conf: 0.9.....	113
Πίνακας 19-13. Αποτελέσματα Arriori με ενσωμάτωση Οντολογίας.....	115
Πίνακας 19-14. Αποτελέσματα Arriori με ενσωμάτωση Οντολογίας για υπόθεση 2	117
Πίνακας 19-15. Αποτελέσματα Arriori με ενσωμάτωση Οντολογίας για υπόθεση 5	118
Πίνακας 19-16. Αποτελέσματα με ενσωμάτωση Οντολογίας χωρίς support για υπόθεση 5	118
Πίνακας 19-17. Σχέση χωρών με τύπο στόχου	119
Πίνακας 19-18. Σχέση εθνικότητας στόχου με τύπο στόχου	120
Πίνακας 19-19. Σχέση «καταγωγής» οργάνωσης με τύπο στόχου	120
Πίνακας 20-1. Απόδοση επεκτάσεων Arriori	122
Πίνακας Α-1. Σύμβολα εκφραστικότητας DL	138
Πίνακας Β-2. Διαφορές μεταξύ OWL 1 Lite- DL- Full	139
Πίνακας Γ-3. Χαρακτηριστικά OWL 2 EL- QL- RL.....	141
Πίνακας Γ-4. Τύποι δεδομένων OWL 2 EL- QL- RL.....	143
Πίνακας Η-5. Απαιτήσεις για την εγκατάσταση του συστήματος PatternMiner v.2	146
Πίνακας Η-6. Οι εντολές του αρχείου «run_me_first.bat»	147
Πίνακας Η-7. Οι εντολές του αρχείου «Run_PatternMiner.bat»	149
Πίνακας Θ-8. Μηνύματα ορθής εκτέλεσης διαδικασίας	150
Πίνακας Κ-9. Αποτελέσματα σεναρίων δεύτερης ενότητας	157
Πίνακας Κ-10. Αποτελέσματα περιπτώσεων ανάλυσης τρίτης ενότητας	157

1. Εισαγωγή

Η παρούσα Μεταπτυχιακή διατριβή πράγματευεται το θέμα των οντολογιών (Ontologies) ως εργαλείο σημασιολογικής αξιολόγησης των προτύπων (patterns), τα οποία εξάγονται από την Διαδικασία της Εξόρυξης Γνώσης (Knowledge Discovery of Data -KDD). Ο απώτερος στόχος είναι η προώθηση ενδιαφέρουσας (interestingness) γνώσης στους ειδικούς που εξετάζουν τις σχέσεις των δεδομένων για να εντοπίσουν νέα ή κρυμμένη γνώση. Η καινοτομία έγκειται στο ότι η διαδικασία της KDD μπορεί να εκμεταλλευτεί τη γνώση ενός πεδίου (Domain Knowledge), που έχει προσδιοριστεί από τις εμπειρίες των ειδικών μέσω οντολογιών, αντί της βασισμένης σε δεδομένα στατιστικής ανάλυσης, που συνήθως χρησιμοποιείται. Βιβλιογραφικά δεν είναι ένας νέος τρόπος προσέγγισης, αλλά αυτός ο ερευνητικός στόχος φαίνεται να είναι ρεαλιστικότερος από ποτέ, τώρα που η πρόοδος του σημασιολογικού ιστού (Semantic Web) έχει δώσει ώθηση στην ανάπτυξη προτύπων και τεχνολογιών για την έκφραση και τη διανομή των οντολογιών.

Όπως σε κάθε διαδικασία έτσι και στις τεχνικές της εξόρυξης γνώσης δύο είναι τα βασικά προβλήματα που παρουσιάζονται. Το ένα έχει να κάνει με την αξιοπιστία των αποτελεσμάτων και το άλλο με την απόδοση των τεχνικών. Στο πρώτο κομμάτι παρουσιάζονται θέματα όπως η εξόρυξη ενδιαφέρουσας γνώσης που εξαρτάται από τις μετρικές που χρησιμοποιούνται μέσα στις τεχνικές της εξόρυξης γνώσης. Για παράδειγμα, οι Οντολογίες προσανατολίζονται κυρίως στην ενσωμάτωση των επιθυμιών ή της γνώσης των ειδικών, έχοντας σκοπό να εξορύξουν ποικιλόμορφα, καινοτόμα, μη αναμενόμενα πρότυπα. Φυσικά, από τη μεριά της αξιολόγησης πάντα γίνεται προσπάθεια για απομόνωση λανθασμένων, τριμμένων προτύπων, ή προτύπων που επαναλαμβάνονται χωρίς να προσφέρουν νέα γνώση. Επίσης, είναι πολύ σημαντικό τα αποτελέσματα να είναι κατανοητά και διαχειρίσιμα, χωρίς να απαιτείται ιδιαίτερη γνώση από τη μεριά του χρήστη. Από την άποψη της απόδοσης, οι τεχνικές επιδιώκουν να προ επεξεργάζονται κατάλληλα τα δεδομένα για να τα μειώσουν και να τα εξορθολογήσουν. Ένας ακόμα στόχος των τεχνικών είναι να μειώσουν εσωτερικά τα βήματα εκτέλεσής τους, ώστε να βελτιστοποιήσουν τα αποτελέσματα και να επιταχύνουν τις διαδικασίες.

Οι Οντολογίες παρουσιάζονται ως μια ολοκληρωμένη λύση σε όλα αυτά τα προβλήματα. Θεωρούνται ως η καταλληλότερη δομή για την αναπαράσταση της γνώσης ενός πεδίου και είναι απόλυτα κατανοητές από τις υπολογιστικές μηχανές λόγω του ρητού και τυπικού τους ορισμού. Το παραπάνω τους δίνει το πλεονέκτημα ότι μπορούν να μεταλαμπαδεύουν τη γνώση στις τεχνικές εξόρυξης γνώσης ώστε να αποφεύγεται η λανθασμένη, τριμμένη γνώση. Επίσης παρέχουν τη δυνατότητα στον χρήστη να ορίζει ποικίλες μετρικές προσδιορίζοντας ο ίδιος το τι θεωρεί ενδιαφέρον. Ένα ακόμη πλεονέκτημα των οντολογιών είναι ότι προσδίδουν σημασιολογική υπόσταση στα αποτελεσμάτα, έτσι ώστε να μπορούν να ερμηνευθούν και να διαχειρισθούν ευκολότερα. Μεγάλο ενδιαφέρον παρουσιάζει η ευελιξία που τους επιτρέπει τη χρήση σε όλα τα στάδια της εξόρυξης γνώσης, αλλά και το συνδυασμό τους με άλλα μέτρα ενδιαφέροντος αυξάνοντας την αξιοπιστία και την απόδοση των αποτελεσμάτων.

Οι Οντολογίες από μόνες τους δεν μπορούν να προσφέρουν τίποτε άλλο εκτός από την αναπαράσταση μίας σύλληψης. Για να γίνει πλήρης εκμετάλλευση των δυνατοτήτων τους απαιτούνται άλλα δύο εργαλεία. Το πρώτο και το σημαντικότερο είναι οι Μηχανές Συλλογιστικής Ανάλυσης, που ουσιαστικά διαβάζουν και ελέγχουν τις Οντολογίες, ενώ έχουν την ικανότητα να παράγουν συμπεράσματα και να αξιοποιούν κρυμμένη γνώση. Εδώ, δύο έννοιες συγκρούονται, η εκφραστικότητα και η πολυπλοκότητα των Οντολογιών. Όταν μία Οντολογία υποστηρίζει μεγάλη εκφραστικότητα τότε δυσκολεύει το έργο των Μηχανών Συλλογιστικής Ανάλυσης, τόσο που πολλές φορές αδυνατούν να δώσουν αποτελέσματα. Το δεύτερο εργαλείο, το οποίο έχει ήδη αναφερθεί είναι οι μετρικές. Οι μετρικές δεν είναι τίποτε άλλο από ορισμούς ή αλγόριθμους που επιτρέπουν να καθορίζονται ομοιότητες και αποστάσεις μέσα στις έννοιες μίας Οντολογίας. Επίσης, αξιοποιούνται από τις τεχνικές εξόρυξης γνώσης για να αξιολογήσουν τα δεδομένα, να παραχθούν ενδιαφέροντα πρότυπα ώστε οι ειδικοί να λάβουν σωστές αποφάσεις.

Πάνω σε αυτές τις τρεις έννοιες, Οντολογίες, Μηχανές Συλλογιστικής Ανάλυσης και Μετρικές υλοποιείται μία επέκταση στον αλγόριθμο Apriori του Weka για την τεχνική των Κανόνων Συσχετίσεων. Ο στόχος εδώ είναι να ελαχιστοποιηθούν οι παραγόμενοι κανόνες, διατηρώντας αυτούς που έχουν αυξημένο σημασιολογικό ενδιαφέρον. Οι επεκτάσεις που πραγματοποιούνται στον Apriori επιτρέπουν τέσσερις εναλλακτικές χρήσης της ενσωμάτωσης

μιας Οντολογίας μέσα στον αλγόριθμο. Οι τρεις πρώτες σχετίζονται με το ποια στιγμή γίνεται η ενσωμάτωση της γνώσης από τις Οντολογίες, δηλαδή κατά την διάρκεια εκτέλεσης του *Argioi*, μετά το πέρας του ή και κατά τη διάρκεια αλλά και κατά το πέρας του αλγόριθμου. Η τέταρτη εναλλακτική στοχεύει να προσδώσει στους παραγόμενους κανόνες ακόμα μεγαλύτερη σημασιολογική εξάρτηση, καθώς η διαγραφή των στοιχειοσυνόλων δεν γίνεται πλέον με τη μετρική της εμπιστοσύνης αλλά με τη μετρική που ορίζεται πάνω στην Οντολογία.

Για τη χρησιμότητα αλλά και για την απόδειξη της βελτίωσης των παραγόμενων κανόνων γίνεται έλεγχος των επεκτάσεων σε πραγματικά δεδομένα πάνω στον τομέα της τρομοκρατίας. Έτσι επιλέγονται κατάλληλα δεδομένα εκπαίδευσης, συγκεκριμένες Οντολογίες πάνω στο συγκεκριμένο πεδίο γνώσης και πραγματοποιείται μία απεικόνιση των στοιχείων των δεδομένων εκπαίδευσης με τις έννοιες της Οντολογίας. Έπειτα γίνεται προσπάθεια για να αποδειχθούν έξι υποθέσεις σε μορφή κανόνων που έχουν υποστηριχθεί από ειδικούς του χώρου. Ως αποτέλεσμα έχουμε τη βελτίωση της σημασιολογικής πλευράς των παραγόμενων κανόνων με μείωση της τετριμμένης γνώσης, την αυξημένη κατανόηση κανόνων και δυνατότητας ερμηνείας, την ανάδειξη κανόνων με μικρή υποστήριξη ή εμπιστοσύνη, τη δυνατότητα αναζήτησης στοχευόμενων κανόνων. Από τη μεριά της απόδοσης βέβαια τα αποτελέσματα δεν ήταν εντυπωσιακά –αλλά ούτε απογοητευτικά– καθώς οι επεκτάσεις χρησιμοποιούσαν λίγο περισσότερο μνήμη, επιβάρυναν το χρόνο εκτέλεσης με πρόσθετους ελέγχους και μερικές φορές παρουσίαζαν αυξημένο αριθμό στοιχειοσυνόλων. Το προηγούμενο αποδεικνύει για ακόμα μία φορά ότι η βελτίωση της αξιοπιστίας μίας διαδικασίας συνήθως συνεπάγεται και ελαχιστοποίηση της απόδοσής της.

Σχετικά με την δομή της διπλωματικής, γίνεται διαχωρισμός της σε τρεις ολοκληρωμένες ενότητες. Η πρώτη σχετίζεται με το θεωρητικό υπόβαθρο, έτσι ξεκινάει με μία γενικότερη περιγραφή της βιβλιογραφίας που υπάρχει στο πεδίο της Εξόρυξης Γνώσης και των Οντολογιών, ενώ συνεχίζει προσεγγίζοντας γενικά το θέμα της αναπαράστασης γνώσης και καταλήγοντας σε πιο συγκεκριμένες έννοιες και μεθόδους. Η δεύτερη ενότητα έχει να κάνει με την ανάλυση, το σχεδιασμό και την υλοποίηση των επεκτάσεων συν τον έλεγχο τους σχετικά με την απόδοση πάνω σε ένα περιορισμένο σύνολο εκπαίδευσης δεδομένων και τη χρήση μίας μικρής Οντολογίας περί τρομοκρατίας. Στην τελευταία ενότητα προσεγγίζεται το θέμα της τρομοκρατίας και αναδεικνύεται η αναγκαιότητα συνεργασίας της με το πεδίο των Οντολογιών και του DM, ενώ παράλληλα αποδεικνύεται η χρησιμότητα και η αξιοπιστία των επεκτάσεων της προηγούμενης ενότητας. Τέλος, ακολουθούν συνοπτικά προτάσεις βελτίωσης της τρέχουσας διπλωματικής, τα συμπεράσματα, το γλωσσάρι, η βιβλιογραφία και τα παραρτήματα που περιέχουν υποστηρικτικό υλικό.

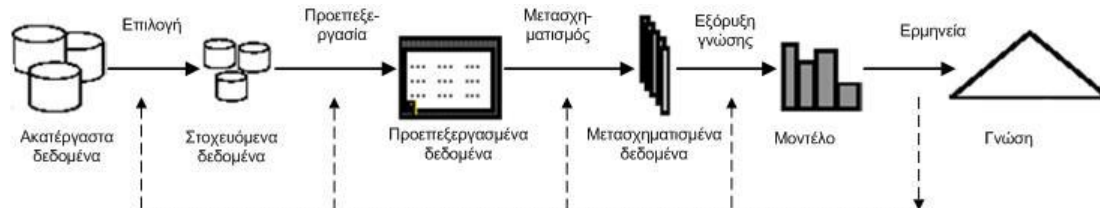
ΜΕΡΟΣ Α - ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2. Σχετικές εργασίες

Εισαγωγή. Σε αυτή την ενότητα γίνεται μία γενική αναφορά στη βιβλιογραφία του διαπραγματευόμενου θέματος. Ο στόχος του είναι να παρουσιαστούν όλες οι βασικές λειτουργικές δομές που περικλείουν το DM και τις οντολογίες, ώστε να επιτευχθεί το κατάλληλο υπόβαθρο για περαιτέρω ανάλυση συγκεκριμένων κομματιών τους. Έτσι, διαχωρίζεται σε δύο τμήματα, την εξόρυξη γνώσης από δεδομένα και το πεδίο γνώσης. Στο πρώτο γίνεται αναφορά στην KDD, τη διαχείριση προτύπων, την έννοια του ενδιαφέροντος προτύπου, τα υποκειμενικά μέτρα και τις σημασιολογικές αποστάσεις. Στο δεύτερο προσεγγίζεται το θέμα της αναπαράστασης γνώσης και κατ' επέκταση των οντολογιών, της οντολογικής μηχανικής και της συνεισφοράς του Σημασιολογικού Διαδικτύου (*Semantic Web*) στις οντολογίες. Σημειώνεται ότι η προσέγγιση που ακολουθείται είναι απαγωγική ξεκινώντας από γενικές έννοιες και καταλήγοντας σε πράγματα που θα αναλυθούν περαιτέρω στη συνέχεια.

Εξόρυξη Γνώσης από Δεδομένα. Το 1996 οι Fayyad, Piatesky-Shapiro και Smyth ορίζουν τι είναι η εξόρυξη γνώσης και προτείνουν τη διαδικασία της KDD, η οποία γίνεται ευρέως αποδεκτή. Ως εξόρυξη γνώσης από δεδομένα ορίζουν *“τη ντετερμινιστική διαδικασία αναγνώρισης έγκυρων, καινοτόμων, ενδεχομένως χρήσιμων και εν τέλει κατανοητών προτύπων στα δεδομένα”*. Για την εξόρυξη γνώσης παρουσιάζονται πέντε διαδοχικά βήματα που πρέπει να πραγματοποιηθούν, αυτά της επιλογής δεδομένων από τα ακατέργαστα, της

προ-επεξεργασίας των στοχευόμενων δεδομένων, του μετασχηματισμού των προεπεξεργασμένων δεδομένων, της εξόρυξη γνώσης από τα μετασχηματισμένα δεδομένα και τέλος της ερμηνεία, αξιολόγησης των προτύπων που παρήχθησαν από το προηγούμενο βήμα για την εξαγωγή της γνώσης.



Εικόνα 2-1. Διαδικασία Εξόρυξης Γνώσης από Δεδομένα

Πηγή: Fayyad και συν. (1996:41)

Η τέταρτη φάση της διαδικασίας περιλαμβάνει όλες τις τεχνικές του DM που μπορούν να χρησιμοποιηθούν για να γίνει η ανακάλυψη χρήσιμων και ενδιαφερόντων πληροφοριών, οι οποίες κρύβονται στη βάση δεδομένων (Chen & συν., 1996; Fayyad & συν., 1996). Τέτοιες τεχνικές είναι η συσταδοποίηση (clustering), η κατηγοριοποίηση (classification), οι κανόνες συσχετίσεων (Association Rules- AR), οι τεχνικές εξόρυξης χρονικής και χωρικής γνώσης (temporal, spatial data mining). Μία από τις σημαντικότερες εφαρμογές του DM είναι αυτή της εξαγωγής κανόνων συσχετίσεων. Η AR εισάγεται για πρώτη φορά το 1993 από τον Agrawal (Agrawal και συν., 1993) και χρησιμοποιείται για να προσδιορίσει τις σχέσεις μεταξύ ενός συνόλου στοιχείων μίας βάσης.

Η βασική δομή που παράγεται από τις τεχνικές του DM και μπορούμε να επεξεργαστούμε, είναι τα πρότυπα. Ουσιαστικά, αν τα δεδομένα είναι ένα σύνολο γεγονότων που αποθηκεύονται σε αρχεία ή βάσεις δεδομένων, τότε τα πρότυπα είναι μια έκφραση σε μία γλώσσα που περιγράφουν ένα υποσύνολο των δεδομένων ή ένα μοντέλο εφαρμόσιμο στο συγκεκριμένο υποσύνολο (Fayyad & συν., 1996). Η διαχείριση τους γίνεται πλέον από τα Συστήματα Διαχείρισης Προτύπων που επιδιώκουν να επιτύχουν την ευελιξία, τη διαλειτουργικότητα, την ομοιομορφία και την διαχειριστική ισχύ των Συστημάτων Διαχείρισης Δεδομένων (Database Management System- DBMS). Παράλληλα, μπορούν να παραχθούν από εργαλεία που υποστηρίζουν τις αντίστοιχες τεχνικές του DM, όπως είναι το Weka (2011), KNIME (2011), RapidMiner (2011), Orange (2011).

Διαχείριση Προτύπων. Για τη διαχείριση προτύπων (pattern management) υπάρχουν δύο κυρίαρχες τάσεις που ακολουθούν τα συστήματα. Η πρώτη είναι οι βάσεις επαγωγικής προσέγγισης (inductive databases approach), όπως το πρόγραμμα CINQ (2011) (Consortium on Discovering Knowledge with Inductive Queries), το οποίο επιδιώκει να επεκτείνει τα ήδη υπάρχοντα συστήματα διαχείρισης δεδομένων, καθώς τα πρότυπα αποθηκεύονται μαζί με τα ακατέργαστα δεδομένα. Η δεύτερη προσέγγιση, όπως αυτή που ακολουθεί το PANDA (2011) (Patterns for Next-generation Database Systems) είναι η εξαρχής κατασκευή ενός συστήματος που είναι εξ' ολοκλήρου προσανατολισμένο στα πρότυπα και διαφοροποιημένο από τα ακατέργαστα δεδομένα, τα επονομαζόμενα και Συστήματα Διαχείρισης Βάσεων Προτύπων (Pattern Based Management Systems- PBMS) (Theodoridis και συν., 2003). Περισσότερες πληροφορίες παρουσιάζονται από τους Catania και Maddalena (2006), Ntoutsis και Theodoridis (2007). Ένα τέτοιο σύστημα είναι ο PatterMiner (Kotsifakos & συν., 2008).

Οι Yang και Wu (2006) παρουσιάζουν δέκα από τα κύρια προβλήματα που θα έπρεπε να απασχολήσουν την ερευνητική κοινότητα μετά το 2005. Μέσα σε αυτά ήταν η ανάπτυξη μιας ενοποιημένης θεωρίας για το DM το οποίο κατ' επέκταση επηρέαζε και την ανάγκη για μία ενιαία θεωρία για τη διαχείριση των προτύπων. Βέβαια, η βιομηχανική κοινότητα είχε κινηθεί από πολύ νωρίς για την παραγωγή προτύπων ώστε να επιτύχουν την διαλειτουργικότητα μεταξύ συστημάτων διαφορετικής αρχιτεκτονικής. Τέτοια πρότυπα είναι το PMML (Predictive Model Markup Language) (2011), SQL/MM part 6 (2006), CWM (Common Warehouse Model) (2011), JDM API (Java Data Mining API) (2011). Σύμφωνα με τον Kotsifakos και συν. (2005), τα PBMS για να μπορέσουν να είναι αποτελεσματικά, ολοκληρωμένα και γενικού σκοπού θα πρέπει να στηριχθούν πάνω στην ημι-δομημένη προσέγγιση (semi-structured Approach). Γεγονός που κάνει την PMML μια από τις πιο δημοφιλείς προσεγγίσεις, αφού στηρίζεται στην XML.

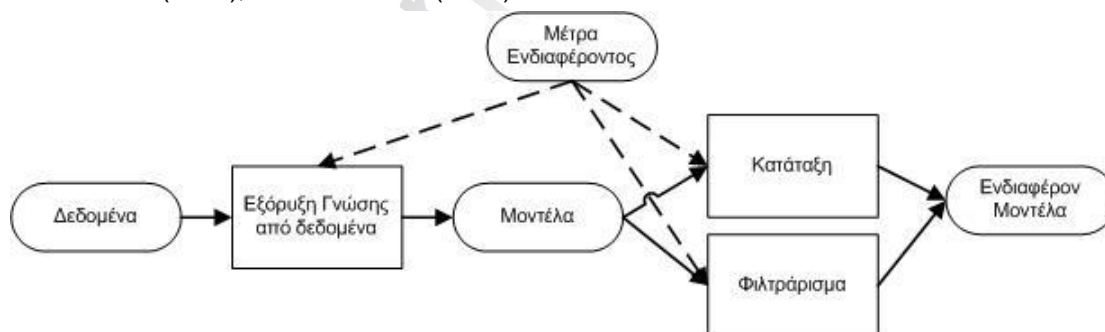
Interestingness. Μια κομβική, αλλά συνάμα αρκετά αφηρημένη έννοια της διαδικασίας KDD είναι η ενδιαφέρουσα γνώση. Οι Geng και Hamilton (2006) εξετάζουν μέτρα ενδιαφέρουσας γνώσης που έχουν παρουσιαστεί στη βιβλιογραφία. Ορίζουν εννέα κριτήρια τα οποία καθορίζουν πότε ένα πρότυπο είναι ή όχι ενδιαφέρον. Τα εννέα αυτά κριτήρια είναι :

Πίνακας 2-1. Μέτρα ενδιαφέροντος

Αντικειμενικά μέτρα	Υποκειμενικά μετρα	Σημασιολογικά μέτρα
περιεκτικότητα (conciseness)	καινοτομία (novelty)	χρησιμότητα (utility),
γενικότητα (generality)	απροσδόκητο/ μη αναμενόμενο (surprisingness/unexpectedness)	δυνατότητα εφαρμογής (applicability).
αξιοπιστία (reliability)		
ιδιαιτερότητα (peculiarity)		
ποικιλομορφία (diversity)		

Έπειτα, τα κατατάσσουν σε τρεις γενικότερες κατηγορίες και παρουσιάζουν τα αντικειμενικά, τα υποκειμενικά και τα σημασιολογικά μέτρα, τα οποία μπορούν να χρησιμοποιηθούν πριν, μετά ή κατά τη διάρκεια της διαδικασίας εξόρυξης γνώσης ώστε να φιλτράρουν, να ταξινομήσουν ή να επιλέξουν ενδιαφέροντα πρότυπα (Εικόνα 2-2. Σημεία εισαγωγής μέτρων ενδιαφέροντος).

Τα αντικειμενικά μέτρα είναι αυτά που προσανατολίζονται με βάση τα δεδομένα και καμιά γνώση του χρήστη δεν απαιτείται. Τα περισσότερα από αυτά τα μέτρα χρησιμοποιούν τη θεωρία πιθανοτήτων, τη στατιστική και τη θεωρία πληροφορίας. Τα πιο γνωστά και διαδεδομένα μέτρα που χρησιμοποιούνται είναι η *υποστήριξη* και η *εμπιστοσύνη* μέσα από το άρθρο του Agrawal, Imielinski και Swami (1993) καθώς μελετούσαν το πρόβλημα της παραγωγής κανόνων από μεγάλες Βάσεις Δεδομένων. Άλλο ένα πολύ δημοφιλές μέτρο ενδιαφέροντος είναι του Piatetsky-Shapiro (1991). Στη βιβλιογραφία παρουσιάζεται ένα σύνολο εργασιών που προσπαθούν να βάλουν τάξη σε αυτά τα μέτρα και να εξετάσουν πια είναι τα κατάλληλα για κάθε περίπτωση, ενδεικτικές εργασίες είναι αυτές του Freitas (1999), Tan και συν. (2002), Lenca και συν. (2004).



Εικόνα 2-2. Σημεία εισαγωγής μέτρων ενδιαφέροντος
Πηγή: Geng και Hamilton (2006:5)

Τα υποκειμενικά μέτρα συνυπολογίζουν τα δεδομένα και τη γνώση του χρήστη. Η γνώση μπορεί να ληφθεί είτε με την αλληλεπίδραση του χρήστη κατά τη διαδικασία της εξόρυξης γνώσης είτε με ρητή αναπαράσταση της γνώσης ή των προσδοκιών του χρήστη. Ενδεικτικές εργασίες είναι του Sahar (1999) όπου ο χρήστης ανατροφοδοτεί το σύστημα για να απομακρύνει μη ενδιαφέροντα πρότυπα, του Liu και συν. (1997, 1999), Silberschtz και Tuzhilin (1995, 1996) οι οποίοι χρησιμοποιούν λογικές μεθόδους αναπαράστασης γνώσης ώστε το σύστημα από μόνο του να εξάγει μη αναμενόμενα ή χρησιμοποιήσιμα πρότυπα. Επίσης, μία ακόμα ενδιαφέρουσα προσέγγιση παρουσιάζεται από τον Padmanabhan και Tuzhilin (1998) οι οποίοι χρησιμοποιούν κανόνες για να περιγράψουν τις πεποιθήσεις των χρηστών και να εξάγουν μη αναμενόμενα πρότυπα από τα δεδομένα που έρχονται σε αντίθεση με αυτές.

Υπάρχει μια τρίτη κατηγορία, τα σημασιολογικά μέτρα τα οποία ουσιαστικά είναι μια πιο εξειδικευμένη υποκατηγορία των υποκειμενικών μέτρων, καθώς εκτός από την χρήση ενός πεδίου γνώσης, ενδιαφέρονται για την σημασιολογική και επεξηγηματική πλευρά των εξαγόμενων προτύπων (Yao και συν 2006). Εδώ οι ερευνητές εμπνέονται κυρίως από την Θεωρία Αποφάσεων όπως οι Cai και συν. (1998), Lu και συν. (2001), Shen και συν. (2002), Carter και συν (1997) για να εξάγουν ωφέλιμα πρότυπα. Άλλοι πάλι χρησιμοποιούν ασαφείς κανόνες (Liu και συν., 1997) ή την έννοια του κέρδους (Wang και σύν., 2002) για να προωθήσουν πρότυπα τα οποία προσφέρουν περεταίρω δυνατότητες εφαρμογή. Αντιλαμβανόμαστε ότι οι δύο κύριες έννοιες που προσπαθούν να διαχειριστούν τα σημασιολογικά μέτρα είναι η ωφελιμότητα και η ικανότητα περεταίρω εφαρμογής των προτύπων.

Υποκειμενικά μέτρα. Ακολουθώντας την υποκειμενική προσέγγιση, ένα σύνολο εργασιών μπορούν να χαρακτηριστούν ως πρόγονοι των οντολογιών. Οι Srikant και Agrawal (1997) χρησιμοποιούν τις ταξινομίες για να περιγράψουν το πεδίο γνώσης. Σκοπός τους είναι να παράγουν γενικότερους κανόνες, για παράδειγμα τους ενδιαφέρει περισσότερο να εξάγουν το συμπέρασμα ότι οι καταναλωτές που αγοράζουν «ενδυμασία για την ύπαιθρο» τείνουν να αγοράζουν «μπότες πεζοπορίας» αντί ενός κανόνα «Μπουφάν => Μπότες Ορειβασίας», όπου το μπουφάν συνδέεται με μια σχέση *is-a* με την έννοια «ενδυμασία για την ύπαιθρο». Επίσης, εισάγουν ένα νέο μέτρο ενδιαφέροντος που καταφέρνει να απορρίψει κανόνες που δεν προσφέρουν νέα, ενδιαφέρουσα γνώση από αυτή που μας δίνουν οι γενικότεροι κανόνες.

Οι Chen και συν. (2003) χρησιμοποιούν ιεραρχημένη γνώση πριν την διαδικασία του DM για να διαχειριστούν τα δεδομένα τους, ώστε να αυξήσουν την απόδοση του support στον αριθμό αλγόριθμο και να εξάγουν γενικότερους κανόνες. Ουσιαστικά, μέσα από την μέθοδο τους, καταφέρνουν να εξάγουν κανόνες που η δυναμική των αρχικών δεδομένων, κυρίως για δεδομένα που ανήκουν στα κατώτερα επίπεδα της ιεραρχίας, δε θα τους το επέτρεπε.

Οι Hotho και συν. (2003) εξετάζουν τις οντολογίες ως ενισχυτικό εργαλείο πριν, μετά ή κατά την διάρκεια των τεχνικών εξόρυξης γνώσης. Ενθαρρύνουν τη χρήση τους καθώς βοηθάνε στην εξαγωγή γενικών και βασισμένων στο πεδίο γνώσης αποτελεσμάτων, τα οποία μπορεί ο εκάστοτε χρήστης να τα διαχειριστεί με μεγαλύτερη οικειότητα. Παρουσιάζουν μέτρα ομοιότητας για το clustering και εφαρμόζουν τις τεχνικές association rules και clustering για να υλοποιήσουν text mining.

Σημασιολογική απόσταση. Αντιλαμβανόμαστε ότι μαζί με τη χρήση των οντολογιών είναι σημαντικό να υπάρχει ένας τρόπος υπολογισμού των σχετικών αποστάσεων μεταξύ των εννοιών (Roddik και συν, 2003). Οι Saruladha και συν (2010) παρουσιάζουν συνοπτικά ένα σύνολο μεθόδων που σχετίζονται με την σημασιολογική ομοιότητα εννοιών μέσα σε μια οντολογία. Εδώ γίνεται διάκριση μεταξύ τεσσάρων προσεγγίσεων. Η πρώτη έχει να κάνει με τον υπολογισμό της απόστασης του μονοπατιού μεταξύ των εννοιών (path length approaches). Η δεύτερη ενδιαφέρεται και δίνει βάρος, στο οποίο βρίσκονται οι έννοιες μέσα στην οντολογία (depth relative approaches). Η τρίτη προσέγγιση προσπαθεί να εισάγει το περιεχόμενο των πληροφοριών που ήδη παρουσιάζεται μέσα σε μια οντολογία (corpus based approaches) και τέλος, η προσέγγιση των πολλαπλών μονοπατιών που προσπαθούν να λύσουν τα προβλήματα αλλά και να διευρύνουν την πρώτη κατηγορία που αναφέραμε (multiple paths approaches).

Ενδεικτικές εργασίες της πρώτης προσέγγισης είναι οι ακόλουθες δύο. Οι Kotsifakos και συν. (2007) χρησιμοποιούν τις οντολογίες για να απομονώσουν τους 'θόρυβους' που παράγονται από την εξαγωγή κανόνων συσχέτισης. Ο ειδικός εισάγει ένα όριο το οποίο είναι αντιστρόφως ανάλογο με το να εμφανιστούν αναμενόμενα αποτελέσματα. Όσο μεγαλύτερο το όριο τόσο μεγαλύτερη απόσταση (μονοπάτι) μεταξύ των εννοιών στους κανόνες, οπότε θα έχουμε αύξηση των μη αναμενόμενων αποτελεσμάτων. Οτιδήποτε είναι εκτός ορίου θεωρείται ως θόρυβος και είναι στο χέρι του ειδικού αν θα τον αξιοποιήσει ή όχι.

Οι Garcia, Ferraz και Vivasqua (2009) θεωρούν τις οντολογίες ως ένα γράφο στον Ευκλείδειο χώρο, έτσι καθορίζουν μια μετρήσιμη απόσταση, βασισμένη σε βάρη, μεταξύ των όρων που συμμετέχουν σε έναν κανόνα συσχέτισης. Αυτή η απόσταση απεικονίζει πως συσχετίζονται δύο όροι μέσα σε μια συγκεκριμένη περιοχή γνώσης, και πόση νέα πληροφορία προσθέτει ο ένας όρος όταν εμφανίζεται από κοινού με έναν άλλο.

Πεδία εφαρμογής. Η χρήση των οντολογιών στις τεχνικές του DM έχουν εφαρμοστεί σε ένα ευρύ φάσμα γνωστικών πεδίων. Οι αρχικές έρευνες προσανατολίζονταν στην περιοχή του

μάρκετινγκ (Zhou & Geller, 2007) αλλά και στο Text Mining (Hotho και συν., 2003). Φυσικά, όλες αυτές οι έρευνες ήταν και είναι η βάση σε έναν γενικότερο τομέα των Συστημάτων Υποστήριξης Αποφάσεων (Decision Support System). Ακολουθούν ενδεικτικές εργασίες οι οποίες άπτονται με την παρούσα.

Οι Kotsifakos και συν. (2007) ενθαρρύνουν την ενσωμάτωση των οντολογιών σε όλα τα στάδια του DM, αλλά και για την αποδοτικότερη διαχείριση των προτύπων. Εφαρμόζουν αυτή την προσέγγιση στο πεδίο της σεισμολογίας χρησιμοποιώντας μια πρότυπη οντολογία της περιοχής, ενώ οι κανόνες εξάγονται από μια πραγματική βάση δεδομένων η οποία διατηρεί ιστορικά σεισμολογικά δεδομένα.

Μια ακόμα εφαρμογή που έχει παρουσιαστεί είναι για την αποφυγή ατυχημάτων σε πλατφόρμες εξόρυξης πετρελαίου. Εδώ, ένα υπάρχον αυτοματοποιημένο σύστημα παρήγαγε εβδομαδιαία μερικές χιλιάδες αναφορές σχετικά με προβλήματα που δημιουργούνταν στις πλατφόρμες. Λόγω του ρίσκου και του μεγάλου όγκου των δεδομένων μια ευφυή και γρήγορη διαδικασία έπρεπε να χρησιμοποιηθεί. Έτσι, δημιουργήθηκε με τη βοήθεια των ειδικών μια οντολογία πάνω στις ανωμαλίες που παρουσιάζονται στην πλατφόρμα, και συνδυάστηκε με την τεχνική εξαγωγής κανόνων συσχέτισης. Αυτό είχε ως αποτέλεσμα τα αρμόδια στελέχη να μπορούν να αξιολογήσουν πιο αποδοτικά τα προβλήματα και τις ενέργειες που είχαν πραγματοποιηθεί για την επίλυσή τους (Garcia και συν., 2009).

Αρκετά ενδιαφέρον πεδίο εφαρμογής είναι αυτό της τρομοκρατίας. Κυρίως μετά το 2001 και τα τρομοκρατικά χτυπήματα στις ΗΠΑ η ερευνητική κοινότητα προσπάθησε να εξελίξει τις διαδικασίες για την αποτροπή ή την ελαχιστοποίηση των συνεπειών τους. Οι Gruenwald και συν. (2003) παρουσίασαν τη χρησιμότητα των οντολογιών πάνω στο τομέα, και τις χρησιμοποίησαν για να επεκτείνουν την βάση γνώση της Oracle Text ώστε να βελτιώσουν την αναζήτηση και την ανάκτηση εγγράφων. Ενώ η DeRose (2004) εξέφρασε την αναγκαιότητα συνεργασίας του DM με το πεδίο της τρομοκρατίας. Εύλογη είναι η ενθάρρυνση εφαρμογών που να αξιοποιούν συνδυαστικά αυτές τις δύο τεχνικές.

Οι Huang και συν. (2010) επισημαίνουν τις οντολογίες ως μέσο για την ύπαρξη καινοτομίας στις επιστήμες της υγείας και της βιολογίας. Η πληθώρα των δεδομένων αλλά και οι εξαντλητικές έρευνες οδηγούν στην εξειδικευμένη γνώση, με αποτέλεσμα να υποβαθμίζεται η αποτελεσματική πληροφορία, κάτι που μπορεί να ισοσταθμιστεί με την συνδυαστική χρήση των οντολογιών και άλλων τεχνικών. Ένα σύνολο αυτών είναι το DM όπου οι οντολογίες μπορούν να χρησιμοποιηθούν πριν, μετά ή κατά την διάρκεια χρήσης τους (Sebastian & συν., 2009; Kuo & συν., 2007).

Πεδίο Γνώσης. Ο απώτερος στόχος στα σύγχρονα συστήματα είναι η εισαγωγή νοημοσύνης καθ' όλη τη διάρκεια των διαδικασιών που ακολουθούν, ώστε να είναι όσο το δυνατό αυτοματοποιημένα και πειθαρχημένα. Οι Cao και συν. (2010) διακρίνουν πέντε γενικές κατηγορίες εισαγωγής και διαχείρισης της νοημοσύνης, οι οποίες έχουν διαφορετικές αφετηρίες και προφανώς διαφορετικούς στόχους. Αυτές είναι η νοημοσύνη από τα δεδομένα (Data Intelligence), από το πεδίο (Domain Intelligence), τα δίκτυα (Network and web Intelligence) και την οργανωτική, κοινωνική νοημοσύνη (Organizational and social Intelligence).

Ένας από τους τρόπους εφαρμογής της νοημοσύνης πεδίου είναι το πεδίο γνώσης, το οποίο εκμεταλλεύεται τις συσχετίσεις των ιδιοτήτων που το αποτελούν. Στόχος του είναι να αποδώσει μια έγκυρη και ολοκληρωμένη γνώση για τον πραγματικό κόσμο. Υπεύθυνη για την ανάπτυξή τους είναι οι ειδικοί του εκάστοτε αντικείμενου. Μια περισσότερο διεξοδική ανάλυση παρουσιάζεται στους Hjørland και Albrechtsen (1995).

Ένα εκτεταμένο θέμα που προκύπτει είναι πως θα υλοποιηθεί πλέον η αναπαράσταση της γνώσης (Knowledge Representation -KR) (Brachman, 1992). Από την μεριά της Τεχνητής Νοημοσύνης (Artificial Intelligent- AI) έχουμε ένα μεγάλο σύνολο εργαλείων, μεταξύ αυτών βρίσκουμε την Λογική όπως για παράδειγμα την λογική πρώτης τάξης, τις δομημένες αναπαραστάσεις, όπως τα σημασιολογικά δίκτυα, τα πλαίσια και φυσικά την αναπαράσταση με κανόνες (Dragan & συν, 2009). Με την ανάπτυξη του σημασιολογικού διαδικτύου, μια νέα τεχνική βασισμένη στις προηγούμενες αναπτύχθηκε και έγινε ευρέως αποδεκτή ως η καλύτερη για την αναπαράσταση ενός πεδίου γνώσης, αυτή είναι οι Οντολογίες.

Οντολογίες. Οι οντολογίες κάνουν την εμφάνισή τους από τα πρώτα κίολας χρόνια της φιλοσοφικής σκέψης (Smith, 2003). Πολύ ορισμοί έχουν δοθεί και ποικίλουν ανάλογα με την προσέγγιση που τυγχάνουν (Gómez-Pérez & συν., 2005). Στην επιστήμη της πληροφορικής

ο ορισμός του Gruber (1993) έμενε να εδραιωθεί, σύμφωνα με τον οποίο η *‘Οντολογία είναι ρητή προδιαγραφή μιας σύλληψης’*. Βέβαια ο απώτερος στόχος όλων είναι να διερευνηθούν δύο βασικοί προβληματισμοί: ποιες τάξεις οντοτήτων χρειάζονται για μια πλήρη περιγραφή και εξήγηση του κόσμου και τι τάξεις οντοτήτων χρειάζονται για να δώσουν μια περιγραφή στο τι καθίστα αληθή την αλήθεια (Smith, 2003).

Εκτός της ποικιλίας των ορισμών, οι Fridman και Hafner (1997) παρουσιάζουν ένα σύνολο διαφορών που υφίστανται μεταξύ των οντολογιών είτε ίδιου ή διαφορετικού επιπέδου (Semy & συν., 2004). Δηλαδή, μεταξύ των οντολογιών υψηλού επιπέδου (upper ontology or top-level ontology) και των εξειδικευμένων σε ένα πεδίο (domain specific ontology). Παρόλα αυτά, όπως είναι φυσικό, δεν υπάρχουν μόνο ανομοιότητες μεταξύ τους, αλλά και πολλές κοινές παραδοχές. Όπως ότι ο κόσμος μπορεί να αναπαρασταθεί με αντικείμενα (objects), όπου τα αντικείμενα έχουν ιδιότητες (properties or attributes) τα οποία λαμβάνουν τιμές (value). Επίσης, συμφωνούν ότι τα αντικείμενα μπορούν να δημιουργήσουν ποικίλες σχέσεις (relations) μεταξύ τους. Οι ιδιότητες και οι σχέσεις μπορούν να αλλάξουν με την πάροδο του χρόνου. Υπάρχουν γεγονότα (events) που εμφανίζονται σε διαφορετικές χρονικές στιγμές. Ακόμα υπάρχουν διαδικασίες (procedure) στις οποίες τα αντικείμενα συμμετέχουν και αυτό συντελείται κατά την πάροδο του χρόνου. Ο κόσμος και τα αντικείμενά του μπορούν να βρίσκονται σε διαφορετικές καταστάσεις, και τέλος τα αντικείμενα μπορούν να αποτελούνται από μέρη (parts) (Chandrasekaran & Josephson, 1999).

Οι οντολογίες ειδικευμένες σε ένα πεδίο, όπως μας προδιαθέτει η ονομασία τους, προσανατολίζονται στην αντιπροσώπηση των εννοιών των όρων του συγκεκριμένου πεδίου. Τέτοια παραδείγματα είναι η e-Business Model Ontology (e-BMO) (Osterwalder, 2002) που αναφέρεται στον επιχειρησιακό τομέα, η Cell-Cycle Ontology (CCO, 2011) που είναι μια οντολογία εφαρμογής πάνω στο κύκλο ζωής των κυττάρων, το Profiles in Terror (2011) το οποίο αποτελεί μια εργασία του πανεπιστημίου του Maryland πάνω στην τρομοκρατία, η Travel Ontology (2011) που σχετίζεται με τον ταξιδιωτικό τομέα. Από την άλλη μεριά έχουμε τις οντολογίες υψηλού επιπέδου που έχουν μια γενικότερη προσέγγιση της γνώσης από τις προηγούμενες. Παραδείγματα αυτών είναι η Basic Formal Ontology (BFO) (2011), η Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (2011), η Suggested Upper Merged Ontology (SUMO, 2011; Niles & Pease, 2001) όπου η τελευταία αποτελεί και πρότυπο σύμφωνα με την IEEE (IEEE P1600.1, 2011).

Ontological Engineering. Η ραγδαία ανάπτυξη των οντολογιών κυρίως μέσα στον 20^ο αιώνα είχε ως αποτέλεσμα την δημιουργία ενός νέου τομέα της Οντολογικής Μηχανικής. Ο σκοπός του είναι να καθορίσει τις δραστηριότητες οι οποίες σχετίζονται με τις διαδικασίες, τις μεθόδους και τις μεθοδολογίες ανάπτυξης οντολογιών, τον κύκλο ζωής μιας οντολογίας, όπως επίσης τον προσδιορισμό εργαλείων και γλωσσών που τις υποστηρίζουν. Μια γενική αναφορά παρουσιάζεται από τους Corcho, Fernández-López και Gómez-Pérez (2007), ενώ ο Devedzic (2002) επισημαίνει την αναγκαιότητα της ύπαρξης ενός τέτοιου τομέα.

Ένα σύνολο μεθόδων και μεθοδολογιών έχουν προταθεί για να καλύψουν τις ανάγκες των ειδικών και για να μπορούν να κατηγοριοποιηθούν σε ευρύτερες ομάδες (Gómez-Pérez, 2005). Χαρακτηριστικές είναι αυτές οι οποίες σχετίζονται με την δημιουργία οντολογιών από το μηδέν, όπως η Methodology (Gómez-Pérez & συν., 1996) και η On-To-Knowledge (Staab & συν., 2001). Άλλες σχετίζονται με τον ανασχεδιασμό οντολογιών (ontology reengineering) (Gómez-Pérez & συν., 1999), την αξιολόγηση οντολογιών (ontology evaluation) (Gómez-Pérez, 2001; Garcia και συν., 2010), την εξέλιξη οντολογιών και υποστήριξη πολλαπλών εκδόσεων (ontology evolution, versioning) (Stojanovic & Motik, 2002; Klein & Fensel, 2001), και φυσικά αυτές που σχετίζονται με την ευθυγράμμιση και συγχώνευση οντολογιών (alignment and merging ontology) (Euzenat & Shvaiko, 2007).

Όπως είναι φυσικό, το ενδιαφέρον για τις οντολογίες επεκτάθηκε και στο να δημιουργηθούν εργαλεία. Αρχικά σκοπός τους ήταν να υποστηρίξουν την ανάπτυξη νέων οντολογιών, σιγά-σιγά να υλοποιήσουν τις εκάστοτε μεθόδους και εν τέλει να προσθέσουν νέες χρησιμότητες του παρουσιάζονταν στη βιβλιογραφία. Τα κυριότερα κριτήρια πάνω στα οποία μπορούν να αξιολογηθούν είναι για την ευκολία διαχείρισης, αποθήκευσης, τον διαφορετικό τρόπο εισαγωγής και εξαγωγής οντολογιών, οπτικοποίησης, αναζήτησης, δυνατότητας συμπερασμού, δυνατότητα συγχώνευσης, υλοποίησης κάποιας μεθοδολογίας ανάπτυξης οντολογιών. Ένα ενδιαφέρον σύνολο κριτηρίων χρησιμοποιείται από τους Corcho, Fernández-López και Gómez-Pérez (2003). Ενδεικτικά εργαλεία είναι το Protégé (2011), OntoStudio (2011), Ontolingua Server (2011), WebODE (Corcho & συν., 2002) (Πίνακας 2-2. Εργαλεία διαχείρισης οντολογιών).

Πίνακας 2-2. Εργαλεία διαχείρισης οντολογιών

Σχόλια	Υποστηρίζει δύο τρόπους μοντελοποίησης Οντολογιών μέσω των Protégé-Frames και Protégé-OWL		Πρόσβαση μέσω διαδικτύου	Πρόσβαση μέσω διαδικτύου
Υποστηρίζει Μεθοδολογίες	Όχι	Όχι	Όχι	Ναι (Methodology)
Δυνατότητα αποθήκευσης	text, DBMS	text, DBMS	text	DBMS
Υπηρεσίες Συμπερασμού	Ναι	Ναι (OntoBroker)	Όχι	Ναι
Υποστηρίζει format	XML, XMLs, RDFs, F-Logic, CLIPS	OWL, RDFs, RIF, F-Logic, N-Triples, N3	KIF, LOOM, CLIPS	XML, XCARIN, RDFs, DAML+OIL, OWL, F-LOGIC
Αρχιτεκτονική	Αυτόνομο (plug-in, Java ανοιχτός Κώδικας)	Αυτόνομο (plug-in)	Πελάτη/Εξυπηρετητή	Πελάτη/Εξυπηρετητή
Άδεια	Ελεύθερο, Ανοιχτός Κώδικας	Εμπορικό	Ελεύθερο	Ελεύθερο
Από	U. Stanford	Ontoprise	U. Stanford	U. Madrid
Αναφορά	(Protégé, 2011)	(OntoStudio, 2011)	(Ontolingua Server 2011)	(Corcho & συν, 2002)
Έκδοση	4.1	3.0	-	-
Ονομασία	Protégé-2000	OntoStudio	Ontolingua	WebODE

Ένα από τα βασικά χαρακτηριστικά των οντολογιών είναι η εκφραστικότητά τους, η οποία έχει να κάνει με τις δυνατότητες της γλώσσας πάνω στην οποία στηρίζεται. Σύμφωνα με τον Fensel και συν. (2001) μια αντίστοιχη γλώσσα θα πρέπει να έχει συμπαγή σύνταξη, να είναι κατανοητή από τον άνθρωπο, να είναι καλά ορισμένη με μία επίσημη σημασιολογία, να είναι ικανή να αντιπροσωπεύσει την ανθρώπινη γνώση και φυσικά να εξασφαλίζει την διαλειτουργικότητα με τα υπάρχοντα πρότυπα. Η ποικιλία αυτών των χαρακτηριστικών αλλά και ο τρόπος σύνταξης, η ορολογία που χρησιμοποιείται, η εκφραστικότητα και η σημασιολογία της κάθε γλώσσας είναι οι αιτίες για τις διαφορές που υπάρχουν μεταξύ τους. Ενδεικτικά, κάποιες από τις σημαντικότερες γλώσσες οντολογιών και αναπαράστασης γνώσης παρουσιάζονται στο πίνακα που ακολουθεί (Πίνακας 2-3 Γλώσσες Οντολογίας και Αναπαράστασης Γνώσης).

Πίνακας 2-3 Γλώσσες Οντολογίας και Αναπαράστασης Γνώσης

Όνομασία	Έκδοση	Γλώσσα	Αναφορές	Διάθεση	Άδεια	Αιτίο	Στηρίζεται σε	Εκφραστικότητα	Συμπέρασμα	Πρότυπο	Σχόλια
OWL	2	RDF/XML	(Dean & Schreider, 2004)	2002	Ελεύθερο	W3C OWL Working Group	Περιγραφική Λογική	Υψηλή	Επιπλέον	Ναι	Επέκταση της DAML+OIL. Επεκτείνεται σε OWL 2
DAML+OIL	-	RDF/XML	(Harmelen & συν., 2001)	2000	Ελεύθερο	Επιτροπή από της US και EU	Βασισμένο σε Πλαίσια	Υψηλή	Επιπλέον	Όχι	Συνένωση DAML και OIL
RDF/ RDFs	1.0	XML	(RDF, 2011)	1998	Ελεύθερο	Στα πλαίσια του προγράμματος DARPA	-	Χαμηλή	Επιπλέον	Ναι	-
XML/ XMLs	1.0	Markup Language	(XML, 2011)	1996	Ελεύθερο	World Wide Web Consortium	-	Χαμηλή	Όχι	Ναι	-
SHOE	-	Markup Language	(Heflin & Hendler, 2000)	1996	Ελεύθερο	University of Maryland.	Περιγραφική Λογική	Χαμηλή	Όχι	Όχι	Επισκιάστηκε από την XML
Cycl	-	LISP Like	(Cycorp, 2009)	-	Ελεύθερο	Cycorp company	Επεκτείνει την Πρώτης Τάξης Λογική	Υψηλή	ΝΑΙ	Όχι	Χρησιμοποιείται από το Cyc Knowledge Base. Επίσης αναφέρεται και ως MELD
F-Logic	-	Δεν έχει σύνταξη Lisp-like	(Kifer και συν., 1995)	1995	Ελεύθερο	University of Karlsruhe	Συνδυασμός Πλαίσιας και 1 ^{ης} Τάξης Λογικής	Υψηλή	Όχι	Όχι	Βασίστηκε η FLORA-2
PowerLoom	3.2.0	STELLA	(PowerLoom Documentation 2006)	-	Ελεύθερο	University of Southern California	Λογική Πρώτης Τάξης	Υψηλή	ΝΑΙ	Όχι	Διάδοχος της Loom, Επηρεασμένο από KIF 3.0
KIF	3.0	Lisp-Like	(Genesereth, 1991; KIF, 2011)	1991	Ελεύθερο	KSL at Stanford University	Λογική Πρώτης Τάξης	Υψηλή	Όχι	Όχι	Βασίστηκε πάνω της το Ontolingua. Επέκταση της SUO-KIF SUMO
LOOM	4.0	-	(Loom Documentation, 2009)	1986	Ελεύθερο	University of Southern California	Περιγραφική Λογική και Παραγωγή Κανόνων	Υψηλή	LOOM Classifier	Όχι	-

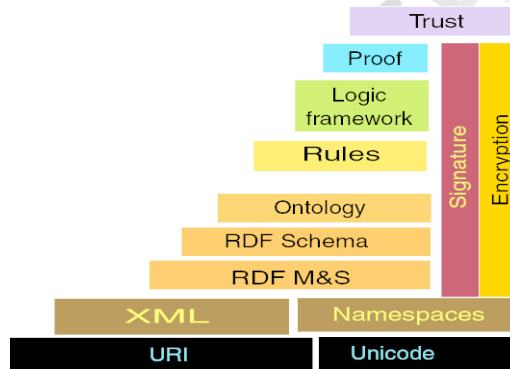
Οι οντολογίες εκτός της απαίτησης για ακριβή αναπαράσταση της γνώσης θα πρέπει να είναι ικανές να συνδυαστούν με κατάλληλους Μηχανισμούς Εξαγωγής Συμπερασμάτων (Inference Mechanisms) (Βλαχαβάς & συν, 2002; Dragan & συν, 2009). Σκοπός αυτών των μηχανών συλλογιστικής ανάλυσης (reasoner) είναι να συμπεραίνουν τις λογικές συνέπειες από ένα σύνολο βέβαιων γεγονότων ή αξιωματών (Velásquez & Jain, 2010). Αυτοί οι μηχανισμοί, υλοποιούνται από τη στρατηγική αναζήτησης, όπως η Πρόσθια Σύνδεση (Forward Chaining) και η Προς τα Πίσω Σύνδεση (Backward Chaining), αλλά και από τη συλλογιστική (reasoning) που αναφέρεται στην παραγωγή γνώσης από είδη υπάρχουσας, όπως η παραγωγή (deduction), επαγωγή (induction) και απαγωγή (abduction). Ο Πίνακας 2-4. *Μηχανές συλλογιστικής ανάλυσης*, παρουσιάζει κάποιες από τις πιο γνωστές μηχανές συλλογιστικής ανάλυσης.

Πίνακας 2-4. Μηχανές συλλογιστικής ανάλυσης

Όνομασία	Έκδοση	Αναφορές	Γλώσσα ανάπτυξης	Εκφραστικότητα	Μηχανισμός Συμπερασμού	Υποστήριξη Κανόνων	Διεπαφή	Άδεια	Από	Σχόλια
Pellet	2.2.2	(Pellet Documentation, 2011; Srin, 2007)	Java	OWL-DL/SROIQ(D)	Tableau	SWRL, DL Safe Rules	OWL-API, DIG-API, Jena	Ελεύθερο, Ανοιχτού Κώδικα	Clark and Parsia	Χρησιμοποιείται στο Swoop.
OWLIM	3.4	(OWLIM Documentation, 2011)	Java	RDFS, OWL DLP, OWL Horst, OWL 2 RL	Forward chaining on RDF graphs, Rule-based	Own format	-	Ελεύθερο, Ανοιχτού κώδικα & Μη ελεύθερου κλειστού Κώδικα	Ontotext	SwiftOWLIM ελεύθερη έκδοση, BigOWLIM εμπορική έκδοση. Ανήκει στην οικογένεια των Semantic Repository.
FaCT++	1.5.1	(FaCT++ Documentation, 2005)	C++	OWL-DL/SROIQ(D)	Tableau	No	Lisp-API, DIG-API	Ελεύθερο, Ανοιχτού Κώδικα	U. Manchester	Βασίστηκε πάνω στην FaCT
Racer-Pro2.0		(RacerPro, 2010)	Lisp	OWL Lite, OWL-DL, /SHIQ(D)	Tableau	SWRL & Own format	OWL-API, DIG-API	Εμπορικό	Racer Systems	
Jena	2.6.4	(Jena Documentation, 2010)	Java	Ελλιπή υποστήριξη της OWL, DAML+OIL, RDFs	Rule-based	Own rule format	OWL-API, DIG-API, DAML-API	Ελεύθερο, Ανοιχτού Κώδικα	HP Labs at Bristol	Μπορεί να κάνει χρήση διάφορων εσωτερικών reasoners, αλλά και εξωτερικών όπως η Pellet
KAON2	2	(Kaon2 Documentation, 2011)	Java	OWL-DL, F-Logic, /SHIQ(D)	Resolution & Datalog	SWRL, DL Safe Rules	DIG-API	Ελεύθερο, Κλειστού Κώδικα	U. Karlsruhe & FZI	
Hermit	1.3.2	(Hermit, 2011)	Java	SROIQ(D)	Hypertableau	SWRL, DL Safe Rules	OWL-API	Ελεύθ., Av. Κωδ.	U. Oxfordy	

Semantic Web. Αμέσως μετά την εξάπλωση του διαδικτύου παρουσιάστηκε η ανάγκη για την εξέλιξη του σε ένα δίκτυο όπου οι πληροφορίες θα ήταν σαφώς ορισμένες και θα επέτρεπαν την ομαλή συνεργασία των πρακτόρων (agent) (Berners-Lee & συν., 2001). Έτσι ξεκίνησε η επέκτασή του στο σημερινό Σημασιολογικό Δίκτυο (Semantic Web), που τείνει να αυξήσει τη σημασιολογική διαλειτουργικότητα, επενδύοντας στο διαμοιρασμό κοινής γνώσης. Επομένως, ο πυρήνας του σημασιολογικού δικτύου δε θα μπορούσε να ήταν άλλος από τις οντολογίες (Ding & συν., 2007), με τις οποίες παρουσιάζει μια αμοιβαία και παράλληλη εξέλιξη.

Προφανώς, αυτή η εξέλιξη οδήγησε στην ανάπτυξη ενός νέου μοντέλου αντιμετώπισης του διαδικτύου από πλευράς γλωσσών. Η καθιερωμένη γλώσσα HTML (HyperText Markup Language) θεωρήθηκε ανεπαρκής για να ανταποκριθεί στις νέες προσδοκίες. Έτσι, αναπτύχθηκε μια αδελφική Γλώσσα Σήμανσης (Markup Language) η XML (Extensible Markup Language, 2011), η οποία επέτρεπε την αναπαράσταση πληροφορίας με τρόπο προσπελάσιμο από τις μηχανές. Έπειτα ακολούθησε η RDF, RDFs (Resource Description Framework, RDF schema, 2011), η οποία προσέφερε ένα απλό λεξιλόγιο και αξιώματα κάτω από έναν αντικειμενοστραφή σχεδιασμό. Σύμμαχος της RDF υπήρξε η προσπάθεια για έναν μοναδικό μηχανισμό ονοματολογίας, τον Παγκόσμιο Προσδιοριστή (Universal Resource Locator- URI) (Thompson, 2010) που χρησιμοποιείται από τους πόρους και τις ιδιότητες της RDF για αποφυγή ασαφειών. Το επόμενο βήμα ήταν να δημιουργηθούν γλώσσες όπως η OWL (Ontology Web Language), οι οποίες προσέφεραν μεγαλύτερη εκφραστικότητα και περισσότερα αξιώματα. Μια αναλυτικότερη παρουσίαση για αυτά τα μεταβατικά βήματα δίνεται από τους Antoniou και Harmelen (2004).



Εικόνα 2-3. Επίπεδα σημασιολογικού δικτύου

Πηγή: Berners-Lee, <http://www.w3.org/2002/Talks/04-sweb/slide12-0.html>

Καθώς επεκτείνεται ο παγκόσμιος ιστός είναι λογικό να εμφανίζονται νέες ή προσαρμοσμένες εφαρμογές. Μια κατηγορία που αρχίζει να προσαρμόζεται στα νέα δεδομένα είναι οι μηχανές αναζήτησης. Πλέον οι αναζητήσεις θα πρέπει να γίνονται σε αρχεία του σημασιολογικού ιστού (Semantic Web Document- SDW), τα οποία δεν είναι τίποτε άλλο από αρχεία που υπόκεινται κάτω από μία σημασιολογική γλώσσα και τα οποία είναι προσβάσιμα από τους χρήστες και τους πράκτορες του διαδικτύου (Ding & συν., 2004). Ενδεικτικά παραδείγματα είναι η Swoogle (2011), SchemaWeb (2011).

3. Γνωσιακή Επιστήμη

Με αυτό το κεφάλαιο γίνεται προσπάθεια να τοποθετηθούν οι Οντολογίες μέσα στο ερευνητικό πεδίο της Πληροφορικής. Αρχικά θα γίνει αναφορά στο ευρύτερο ερευνητικό περιβάλλον, έπειτα θα εξεταστεί το τι είναι τελικά η αναπαράσταση της γνώσης και θα καταλήξει με την παρουσίαση μεθόδων πάνω στις οποίες βασίζονται ουσιαστικά οι δομές και οι γλώσσες περιγραφής Οντολογιών.

Το γενικότερο ερευνητικό πεδίο που εμφανίζεται να συσχετίζεται με το αντικείμενο είναι η Γνωσιακή Επιστήμη (Cognitive Knowledge). Στόχος της είναι να μελετήσει τη φύση του μυαλού και ειδικότερα να εξετάσει τη νόηση, τις νοητικές καταστάσεις και διεργασίες, όπως η σκέψη (Dragan 2009). Μια τέτοια επιστήμη ήταν αναγκαία για να θεμελιώσει τις αρχές πάνω στις οποίες η Τεχνητή Νοημοσύνη (Artificial Intelligent- AI) θα βασιζόταν για να επιτύχει τους στόχους της, που δεν ήταν άλλοι από την απόκτηση, αποθήκευση, ανάκτηση της γνώσης αλλά και την δυνατότητα εξαγωγής συμπερασμάτων και επεξηγήσεων κάνοντας χρήση μιας στρατηγικής επίλυσης προβλήματος πάνω στην υπάρχουσα γνώση.

Η Τεχνητή Νοημοσύνη στράφηκε και μελέτησε κυρίως την ανθρώπινη φύση και σκέψη με αποτέλεσμα να παραχθούν διάφοροι τύποι και μοντέλα αναπαράστασης της Ανθρώπινης Γνώσης (Human Knowledge). Σε αυτά, τα μοντέλα κωδικοποίησης της γνώσης συσχετίζονταν με δομές αποθήκευσης δεδομένων, όπως πίνακες, δέντρα, συνδέσμους. Παρόλα αυτά καμία από αυτές δεν έχει καταφέρει μέχρι σήμερα να αναπαραστήσει πλήρως όλα τα είδη της γνώσης. Αυτό καθιστά κάθε μοντέλο επαρκές για την αντιμετώπιση συγκεκριμένων προβλημάτων, ενώ για την αντιμετώπιση περίπλοκων απαιτήσεων είναι αναγκαίος ο συνδυασμός τους.

Στην εικόνα που ακολουθεί παρουσιάζονται βασικοί τύποι γνώσης που οι άνθρωποι χρησιμοποιούν. Τέτοιοι είναι (Durkin, 1994):

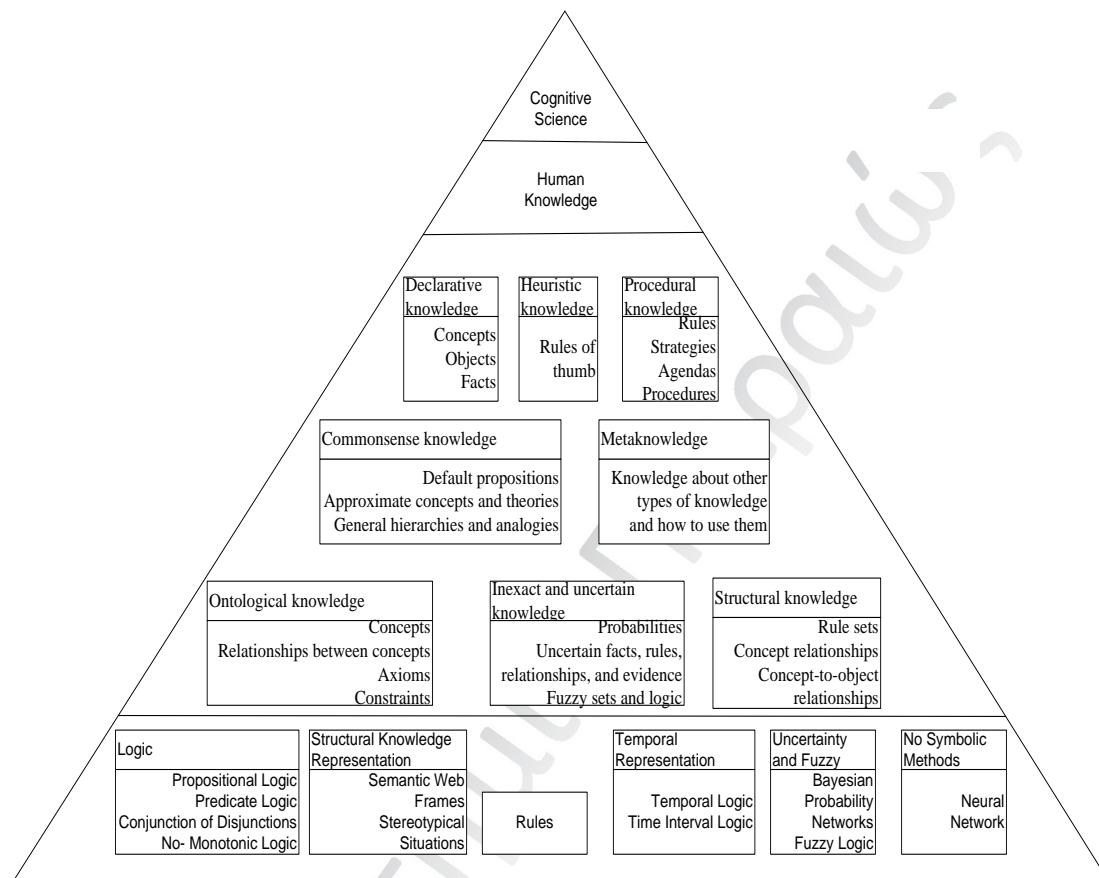
- *Διαδικαστική Γνώση (Procedural knowledge)*, που αναφέρεται στο πώς να γίνει κάτι βάση μιας διαδικασίας. Εδώ χρησιμοποιούνται κανόνες, στρατηγικές προβλήματος-επίλυσης, διαδικασίες.
- *Δηλωτική Γνώση (Declarative knowledge)*, όπου περιγράφει τι είναι γνωστό για ένα πεδίο ή για ένα πρόβλημα. Εδώ χρησιμοποιούνται εργαλεία όπως οι έννοιες, αντικείμενα, γεγονότα.
- *Μεταγνώση (Metaknowledge)*, γνώση για άλλους τύπους γνώσης και πως αυτοί χρησιμοποιούνται.
- *Ευρετική Γνώση (Heuristic knowledge)*, περιλαμβάνει εμπειρικούς κανόνες που καθοδηγούν την επίλυση προβλημάτων με βάση προηγούμενες εμπειρίες, ατομική διαίσθηση, δεξιότητες, καθώς και επαρκή κατανόηση του προβλήματος.
- *Δομημένη Γνώση (Structural Knowledge)*, διαθέτει μια δομημένη οργάνωση του προβλήματος και της λύσης, εδώ χρησιμοποιούνται σύνολα κανόνων, σχέσεις μεταξύ εννοιών ή εννοιών και αντικειμένων, όπως kind-of, part-of, grouping και sets.
- *Ανακριβής και Αβέβαιη Γνώση (Inexact and uncertain knowledge)*, χαρακτηρίζει προβλήματα, θέματα και καταστάσεις στις οποίες οι πληροφορίες είναι εγγενώς ασαφείς, ελλιπείς ή τυχαίες. Εδώ οι έννοιες που επικρατούν είναι οι πιθανότητες, τα αβέβαια γεγονότα, οι κανόνες, οι σχέσεις ενώ χρησιμοποιούνται εργαλεία όπως τα ασαφής σύνολα και η ασαφής λογική.
- *Γνώση Κοινής Λογικής (Commonsense knowledge)*, χρησιμοποιείται κυρίως για να υποδηλώσει ένα μέρος της ανθρώπινης γνώσης για τον κόσμο που δεν μπορεί να τεθεί εύκολα υπό μορφή συγκεκριμένων θεωριών. Εδώ εφαρμόζονται γενικές ιεραρχίες, αναλογίες, έννοιες με ασαφή ορισμό.
- *Οντολογική Γνώση (Ontological knowledge)*, αυτό κυρίως που ενδιαφέρει την τρέχουσα εργασία και περιγράφει τις κατηγορίες των πραγμάτων και των συσχετίσεων του μέσα σε ένα πεδίο. Εδώ τα βασικά στοιχεία που την αποτελούν είναι οι έννοιες, οι σχέσεις μεταξύ των εννοιών, τα αξιώματα και οι περιορισμοί.

Γίνεται αντιληπτό ότι η χρήση πολλαπλών τύπων αναπαράστασης γνώσης κάθε άλλο παρά αρνητικό μπορεί να θεωρηθεί, καθώς αυτή η ικανότητα των ανθρώπων να οργανώνουν την γνώση σε διαφορετικές δομές και να τις χρησιμοποιούν τους επιτρέπει να επιλύουν αποτελεσματικά τα προβλήματα.

Στη βάση της πυραμίδας εμφανίζονται οι τεχνικές της αναπαράστασης της γνώσης οι οποίες μπορούν να συνδυαστούν για να καλύψουν τις απαιτούμενες ανάγκες. Σύμφωνα με τους Βλαχαβάς & συν (2002) μπορούν να ομαδοποιηθούν ακολούθως:

- Λογική (Logic)
 - Προτασιακή Λογική (Propositional Logic)
 - Κατηγορηματική Λογική (Predicate Logic)
 - Διαζευκτική Μορφή της Λογικής (Clausal Form of Logic)
- Δομημένες αναπαραστάσεις γνώσεις
 - Σημασιολογικά Δίκτυα (Semantic Networks)
 - Πλαίσια (Frames)
 - Εννοιολογική Εξάρτηση (Conceptual Dependency)
 - Σενάρια (Scripts)
- Κανόνες (Rules)
- Αναπαράσταση του Χρόνου (Temporal Representation)
 - Χρονική Λογική (Temporal Logic)
 - Λογική Χρονικών Διαστημάτων (Time Interval Logic)
- Αναπαράσταση Αβέβαιης και Ασαφούς Γνώσης (Uncertainty and Fuzzy Knowledge)

- Δίκτυα Πιθανοτήτων (Probability Networks)
- Ασαφής Λογική (Fuzzy Logic)
- Ασαφής Περιγραφική Λογική (Fuzzy Description Logic)
- Μη συμβολικές Μέθοδοι (No Symbolic Methods)
- Νευρωνικά Δίκτυα (Neural Network)



Εικόνα 3-1. Πυραμίδα αναπαράστασης γνώσης

3.1 Αναπαράσταση Γνώσης

Το θέμα που πραγματεύεται αυτή η ενότητα είναι η γνώση και ουσιαστικά η αναπαράστασή της. Ένα μεγάλο σύνολο συστημάτων, τα Βασισμένα στη Γνώση Συστήματα (Knowledge Based Systems) όπως τα Ευφυή Συστήματα (Expert Systems), προσπαθούν να περιγράψουν τη γνώση του πεδίου ορισμού τους και να εξάγουν συμπεράσματα. Η διαδικασία αυτή ορίζεται και καθοδηγείται από το πεδίο της Αναπαράστασης Γνώσης (Knowledge Representation- KR) (Brachman, 1992).

Ουσιαστικά η Αναπαράσταση Γνώσης είναι ο ερευνητικός τομέας που εξετάζει τις γλώσσες αναπαράστασης και την εξαγωγή συμπερασμάτων που πηγάζουν από αυτές (Dragan & συν, 2009). Όπως αντιλαμβανόμαστε, υπάρχουν δύο βασικοί άξονες. Ο πρώτος είναι η γλώσσα αναπαράστασης γνώσης που προσδιορίζει σύμβολα και δομές ενός κόσμου που στο σύνολό τους αποτελούν την Βάση Γνώσης του, και ο δεύτερος άξονας έχει να κάνει με την ικανότητα παραγωγής γνώσης μέσω των μηχανισμών εξαγωγής συμπερασμάτων.

Από τα πρώτα βήματα της Τεχνητής Νοημοσύνης (Russell & Norvig, 1995), η Αναπαράσταση Γνώσης υπήρξε ένα από τα βασικά της εργαλεία. Ο κύριος στόχος τους ήταν να υπάρξει ένας τρόπος μετάδοσης της γνώσης του κόσμου σε ένα υπολογιστικό σύστημα, προσδίδοντάς του μια κατάλληλη συλλογιστική ικανότητα ώστε να μπορεί να χρησιμοποιήσει αυτή τη γνώση για να προσαρμοστεί και να εκμεταλλευτεί το περιβάλλον του. Βέβαια, οι ερευνητές διαπίστωσαν από πολύ νωρίς πως η υλοποίηση ενός τέτοιου συλλογιστικού συστήματος ήταν αρκετά περίπλοκη για να παραγάγει γρήγορα και ορθά συμπεράσματα.

Παρόλο που για τον άνθρωπο η αναπαράσταση της γνώσης του κόσμου γίνεται με έναν σχετικά εύκολο τρόπο, με απλή περιγραφή σε φυσική γλώσσα, για τα υπολογιστικά συστήματα κάτι τέτοιο είναι σχεδόν αδύνατο. Οι κυριότεροι λόγοι είναι η πολύ-σημαντικότητα της γλώσσας, δηλαδή μια φράση αποδέχεται πολλές ερμηνείες, και η ύπαρξη συμφραζομένων τα οποία αυξάνουν την πολυπλοκότητα του εγχειρήματος. Γι' αυτό από πολύ νωρίς, σιωπηρά υπονοήθηκε ότι πρέπει να χρησιμοποιηθεί ένας ρητός και τυποποιημένος συμβολισμός (Βλαχαβάς & συν, 2002).

Η αναπαράσταση γνώσης πάντα είχε σχέση με τις μορφές έκφρασης των πληροφοριών, από τα άτυπα πρότυπα μνήμης μέχρι και πιο σύνθετες επίσημες γλώσσες. Μεταξύ των βασικών τεχνολογιών που έχουν χρησιμοποιηθεί ευρέως, βρίσκεται η Λογική όπως για παράδειγμα η λογική πρώτης τάξης, οι δομημένες αναπαραστάσεις, όπως τα σημασιολογικά δίκτυα, τα πλαίσια και φυσικά η αναπαράσταση με κανόνες. Είναι λογικό ότι κάθε προσέγγιση είναι κατάλληλη για επίλυση διαφορετικών προβλημάτων. Σε πολλές περιπτώσεις, το ενδιαφέρον εστιάζεται στο πώς θα γίνει μια αντιπαρασταση της γνώσης σε μια τυπική δομή, αδιαφορώντας για το πώς θα χρησιμοποιηθεί.

3.2 Μέθοδοι Αναπαράστασης Γνώσης

Συνοψίζοντας, γίνεται αντιληπτό ότι μια μέθοδος αναπαράστασης γνώσης θα πρέπει να χρησιμοποιεί έναν μονοσήμαντο και τυποποιημένο συμβολισμό έτσι ώστε να ορίζει μια σαφή και ακριβή γλώσσα αναπαράστασης γνώσης που θα επιτρέπει την εξαγωγή συμπερασμάτων. Για να επιτευχθεί η απαίτηση της σαφούς και ακριβούς αναπαράστασης θα πρέπει να έχει καλά ορισμένη σύνταξη και σημασιολογία. Δηλαδή, θα πρέπει να καθορίζει ποιες είναι οι αποδεκτές εκφράσεις (σύμβολα και κανόνες) της γλώσσας, για να κατανοεί ένα υπολογιστικό σύστημα την εκάστοτε αναπαράσταση, αλλά θα πρέπει συνάμα να ορίζει σαφώς τη σημασία κάθε έκφρασης μέσα σε αυτή τη γλώσσα. Όσο μεγαλύτερος είναι ο βαθμός της σαφήνειας και της ακρίβειας της γλώσσας τόσο η τυπικότητά της αυξάνεται με αποτέλεσμα να μπορούν να συνδυαστούν με κατάλληλους μηχανισμούς εξαγωγής συμπερασμάτων και να παράγονται περίπλοκοι συλλογισμοί, συμπεράσματα, γνώση.

Με βάση τα παραπάνω χαρακτηριστικά οι Rich και Knight (1991) διακρίνουν τέσσερα κριτήρια αξιολόγησης των μεθόδων αναπαράστασης γνώσης. Το πρώτο κριτήριο έχει να κάνει με την επάρκεια αναπαράστασης, δηλαδή με την ικανότητά τους να αναπαριστούν όσο το δυνατόν περισσότερα είδη γνώσης. Το δεύτερο κριτήριο σχετίζεται με την επάρκεια συνεπαγωγής, δηλαδή αν μπορούν να χρησιμοποιηθούν μηχανισμοί που επεξεργάζονται ήδη υπάρχουσες δομές γνώσης με σκοπό να δημιουργηθούν νέες δομές που αντιστοιχούν σε νέα γνώση. Το τρίτο κριτήριο εξετάζει την αποδοτικότητα της συνεπαγωγής και το κατά πόσο είναι ικανές να εισάγουν επιπλέον πληροφορία στις δομές γνώσης ώστε να κατευθύνονται οι μηχανισμοί εξαγωγής συμπερασμάτων γρηγορότερα προς τη λύση. Το τελευταίο κριτήριο που διακρίνεται είναι η αποδοτικότητα τους στην απόκτηση νέας γνώσης εύκολα και γρήγορα.

Παρακάτω θα εξετασθούν κλασικές μέθοδοι αναπαράστασης γνώσης που έχουν χρησιμοποιηθεί για να υποστηρίξουν διάφορους τύπους Οντολογιών. Μέσα σε αυτές οι πιο δημοφιλείς είναι η Λογική πρώτης τάξης, η Περιγραφική Λογική, τα Πλάισια και οι Κανόνες. Οι περιγραφές βασίζονται κατεξοχήν από τους Βλαχαβάς και συν. (2002).

Προτασιακή Λογική. Αποτελεί την απλούστερη μορφή λογικής και προσεγγίζει φορμαλιστικά απλές μορφές συλλογιστικής. Σε αυτή γεγονότα του πραγματικού κόσμου αναπαρίστανται με μια λογική πρόταση η οποία μπορεί να είναι είτε αληθής είτε ψευδής, αλλά όχι και τα δύο. Οι λογικές προτάσεις αναπαριστώνται με λατινικούς χαρακτήρες P, Q, R κλπ και ονομάζονται άτομα (atoms) ή ατομικοί τύποι (atomic formulas). Τα άτομα μπορούν να συνδυαστούν με την χρήση λογικών συμβόλων ή συνδετικών (connectives), όπως η άρνηση (\neg), η σύζευξη (\wedge), η διάζευξη (\vee), η συνεπαγωγή (\rightarrow) και η ισοδυναμία (\leftrightarrow).

Η προτασιακή λογική διακρίνεται για την απλότητα της σύνταξης της και για το ότι μπορεί να καταλήξει πάντα σε συμπεράσματα, μέσω των πινάκων αληθείας ή των αποδείξεων εφαρμόζοντας τους κανόνες που αυτή ορίζει. Παρόλα αυτά είναι δύσκολη λόγω της έλλειψης γενικότητας που οδηγεί σε ογκώδεις αναπαραστάσεις.

Λογική Πρώτης Τάξης ή Κατηγορηματική Λογική. Είναι μία επέκταση της προτασιακής λογικής που εισάγει όρους, κατηγορήματα και τους δύο ποσοδείκτες, τον καθολικό (\forall) και τον υπαρξιακό (\exists). Κατά τα άλλα και αυτή χρησιμοποιεί σύμβολα για την αναπαράσταση γνώσης και λογικούς τελεστές για την κατασκευή των προτάσεων. Τα σύμβολα εδώ μπορούν να

εκπροσωπήσουν σταθερές, μεταβλητές ή συναρτησιακούς όρους. Η γλώσσα KIF είναι μία από αυτές που ακολουθεί τη λογική της λογικής πρώτης τάξης.

Σε αντίθεση με την προτασιακή λογική, η κατηγορηματική έχει αυξημένη εκφραστική ικανότητα λόγω της δυνατότητά της να περιλαμβάνει ποσοτικοποιημένες έννοιες κάνοντας χρήση των κατάλληλων ποσοδεικτών αλλά και την ικανότητά της να συλλαμβάνει τη γενικότητα. Παρόλα αυτά αδυνατεί να αναπαραστήσει διάφορα αντικείμενα όπως κατηγορίες, χρόνο, συμβάντα τα οποία εμφανίζονται σε διάφορα γεγονότα του κόσμου, όπως επίσης αδυνατεί να εκφράσει, μέσα από τις προτάσεις της, την ασάφεια καθώς οι μοναδικές τιμές που μπορεί να λάβει είναι αληθής ή ψευδής. Ένα ακόμα αρνητικό που προσάπτεται στην κατηγορηματική λογική είναι η αδυναμία της, λόγω της μονοτονικότητά της, να αναθεωρήσει διάφορα συμπεράσματα που προστίθενται στη γνώση σε περίπτωση που κριθούν εσφαλμένα.

Όπως στην προτασιακή λογική έτσι και εδώ ο βασικός μηχανισμός εξαγωγής συμπερασμάτων είναι η απόδειξη. Οι διαδικασίες ελέγχου ή οι κανόνες εξαγωγής συμπερασμάτων, είναι παρόμοιοι με αυτούς της προτασιακής λογικής με μία μόνο μικρή τροποποίηση για να μπορούν να διαχειριστούν εκφράσεις που περιέχουν μεταβλητές και τελεστές. Μια κλασική διαδικασία ελέγχου είναι αυτή της *αρχής της ανάλυσης*, όπου από δύο προτάσεις $(P \vee Q) \wedge (R \vee \neg Q)$ μπορεί να εξαχθεί το συμπέρασμα $P \vee R$. Αυτός ο κανόνας είναι τόσο σημαντικός καθώς είναι ο μοναδικός που χρειάζεται μια αποδεικτική διαδικασία για να εξάγει όλα τα σωστά συμπεράσματα, χρησιμοποιώντας τη μέθοδο της *εις άτοπο απαγωγής*.

Περιγραφική Λογική (Description Logic- DL). Είναι ένα υποσύνολο της κατηγορηματικής λογικής, που την κάνει πιο ευέλικτη και αποδοτική στην παραγωγή συμπερασμάτων αλλά και πιο εκφραστική από την προτασιακή λογική. Βασικό της αρνητικό είναι ότι απαιτείται πλήρης περιγραφή του κόσμου που αναπαριστά με αποτέλεσμα περίπλοκα προβλήματα να αδυνατεί να τα δηλώσει ή να χρειάζεται πολύ μεγάλη περιγραφή.

Τα βασικά στοιχεία διαμόρφωσης της DL είναι οι ατομικές έννοιες (atomic concepts), οι ατομικοί ρόλοι (atomic roles) ή σχέσεις και τα άτομα (individuals). Το σύνολο των ατομικών εννοιών και ρόλων αποτελούν το λεξιλόγιο της DL το οποίο αντιστοιχεί σε ένα μέρος της βάσης γνώσης, το TBox. Ενώ οι ισχυρισμοί σχετικά με τα άτομα δηλαδή τα στιγμιότυπα και τις σχέσεις τους με τις έννοιες διατηρούνται στο σώμα ισχυρισμών ABox. Για την σύνταξη της χρησιμοποιεί τους λεγόμενους κατασκευαστές $\{ \neg, \sqcap, \sqcup, \exists \}$ που δηλώνουν μια λειτουργία (operation) μεταξύ δύο ή περισσότερων εννοιών ή ρόλων. Ο συνδυασμός τους με τα σύμβολα του συντακτικού και τα στοιχεία του λεξιλογίου οδηγεί στην κατασκευή των αξιωμάτων (axioms). Ουσιαστικά, η γνώση που μοντελοποιεί η DL είναι ένα σύνολο αξιωμάτων.

Το αλφάβητο της DL, αντίθετα με τις άλλες γλώσσες δεν είναι σταθερό αλλά μπορεί να οριστεί από τον χρήστη ανάλογα με τις απαιτήσεις του πεδίου εφαρμογής. Γενικά, υπάρχει μια άτυπη σύμβαση ονομασίας, που αντιστοιχίζει σε συγκεκριμένα σύμβολα ένα σύνολο λειτουργιών και ο συνδυασμός τους δίνει την τελική εκφραστικότητα της γλώσσας. Στο Παράρτημα Α παρουσιάζονται τα σύμβολα που χρησιμοποιούνται για τον λεκτικό προσδιορισμό αυτής της εκφραστικότητας. Έτσι διάφορες περιγραφικές γλώσσες έχουν οριστεί κατά καιρούς με διαφορετική εκφραστικότητα, όπως η OWL2 με εκφραστικότητα *SROIQ^(D)*, η OWL-DL με *SROIN^(D)*, η OWL-Lite με *SRIF^(D)*, η DALM+OIL με *SRIQ^(D)*

Σχετικά με τους μηχανισμούς συλλογιστικής ανάλυσης στην DL κάνουν χρήση των TBox και ABox. Με βάση αυτά μπορούν να αποφανθούν αν μια περιγραφή είναι ικανοποιήσιμη μέσα στο TBox ή αν μια περιγραφή είναι υποσύνολο μιας άλλης μέσω του ορισμού τους. Άλλα δύο σημαντικά καθήκοντα των μηχανισμών συλλογιστικής ανάλυσης, σχετικά με τους ισχυρισμούς του ABox είναι να προσδιορίσουν αν οι ισχυρισμοί που υπονοούνται για ένα συγκεκριμένο άτομο είναι ένα στιγμιότυπο μιας συγκεκριμένης έννοιας, και κατά πόσον ένα σύνολο ισχυρισμών είναι συνεπές. Οι στρατηγικές συλλογιστικής που χρησιμοποιούνται εδώ είναι η Πρόσθια Σύνδεση (Forward Chaining), η Προς τα Πίσω Σύνδεση (Backward Chaining) και οι tableau αλγόριθμοι. Υλοποιήσεις τέτοιων μηχανισμών είναι οι Pellet, Fact++, KAON2, RacerPro, HermiT.

Πλαίσια. Αντιστοιχούν περισσότερο σε μια δομή δεδομένων για την οργάνωση στερεότυπων γεγονότων ή καταστάσεων, ενώ θεωρούνται επέκταση των σημασιολογικών δικτύων. Με τα πλαίσια περιγράφονται τα αντικείμενα του κόσμου, οι σχισμές ή ιδιότητες τους και οι τιμές που λαμβάνουν. Αυτή είναι και η ουσιαστική τους διαφορά με τα σημασιολογικά δίκτυα, καθώς οι κόμβοι των δικτύων αναπαριστούν μονό τις έννοιες, ενώ τα πλαίσια περιέχουν όλη την

πληροφορία για την έννοια που αναπαριστούν. Ιδιαίτερο χαρακτηριστικό τους είναι η αντικειμενοστραφής προσέγγιση στην αναπαράσταση της γνώσης και η δυνατότητά τους να εκφράζουν γενικευμένες ή εξειδικευμένες εκφράσεις, οργανώνοντας τα αντικείμενα σε μια ιεραρχική δομή.

Τα δυνατά σημεία της χρήσης των πλαισίων είναι, κατ' αρχάς, ότι διατηρούν περισσότερη πληροφορία για τις έννοιες, σε σχέση με τις άλλες προσεγγίσεις. Δεύτερον, ότι οι ιδιότητες μπορούν να οριστούν είτε περιγραφικά είτε διαδικαστικά με αποτέλεσμα να προσαρμόζονται ευκολότερα στις αλλαγές των πληροφοριών. Τρίτον, δίνεται η δυνατότητα ορισμού σταθερών σχέσεων, οι οποίες επιτρέπουν την εύκολη εξαγωγή συμπερασμάτων και την αναπαράσταση της γνώσης ανεξάρτητα από τη γλώσσα διατύπωσης των προτάσεων. Στον αντίποδα αυτών βρίσκεται η υψηλή πολυπλοκότητά που διακρίνει τα πλαίσια.

Η συλλογιστική στα πλαίσια μπορεί να γίνει με οποιονδήποτε μηχανισμό εξαγωγής συμπερασμάτων που αναφέρονται στην λογική αρκεί να ενσωματωθούν δύο διαδικασίες, μία για την ανάκληση των τιμών των ιδιοτήτων μέσα από τα πλαίσια και μία για την πλοήγηση μέσα στο ιεραρχημένο γράφο δίνοντας προσοχή στο φαινόμενο της πολλαπλής κληρονομικότητας που τον διακρίνει.

Κανόνες. Οι κανόνες αποτελούν μία από τις πιο δημοφιλείς μεθόδους αναπαράστασης γνώσης στα έμπειρα συστήματα και ίσως ο πρακτικότερος τρόπος εξαγωγής συμπερασμάτων. Αποτελείται συντακτικά από δύο μέρη, τις *προϋποθέσεις ή αριστερό μέρος του κανόνα* που είναι μια ακολουθία από κατηγορήματα τα οποία συνδέονται μεταξύ τους με τους λογικούς τελεστές ΚΑΙ/Η (AND/OR) και το δεύτερο μέρος που είναι τα *επακόλουθα ή δεξιό μέρος του κανόνα*. Τα *επακόλουθα* με την σειρά τους διακρίνονται σε δύο κατηγορίες τις *ενέργειες* όπου οι κανόνες εκφράζουν μια διαδικαστική γνώση και τα *συμπεράσματα* που σε αυτή την περίπτωση εκφράζουν μια δηλωτική γνώση.

Τα πλεονεκτήματα που τους διακρίνουν είναι ότι μπορούν να διασπάσουν ένα πρόβλημα σε μικρά ανεξάρτητα κομμάτια και να προστίθενται σε ένα σύνολο κανόνων ανεξάρτητο των υπαρχόντων, ενώ αλλάζοντας η γνώση μπορεί να γίνεται εφικτή η τροποποίηση των αντίστοιχων κανόνων.

Στους κανόνες ο μηχανισμός εξαγωγής συμπερασμάτων διακρίνεται σε δύο μέρη, το μέρος που περιέχει τους κανόνες και αποτελεί τη *βάση κανόνων* και τον *έλεγχο*, ο οποίος καθορίζει τον τρόπο με τον οποίο θα εκτελεστούν οι κανόνες για να εξαχθούν τα συμπεράσματα. Ο έλεγχος δεν είναι τίποτε άλλο από έναν αλγόριθμο που αποφασίζει ποιοι κανόνες και με ποιο τρόπο θα αξιοποιηθούν για να αναζητηθεί η λύση και να εξαχθούν τα κατάλληλα συμπεράσματα.

Χρονική Λογική. Πολλά συστήματα απαιτούν δυνατότητες συμπερασμού με βάση το χρόνο, όπως τα συστήματα διάγνωσης ή πλάνων. Έτσι η τεχνητή νοημοσύνη δε θα μπορούσε να αποφύγει τη σχεδίαση μιας αντίστοιχης μεθόδου αναπαράστασης γνώσης μέσα στο χρόνο. Η χρονική λογική συμπληρώνει την κλασική, στο ότι μια πρόταση μπορεί να είναι αληθής σε μια συγκεκριμένη χρονική στιγμή αλλά σε μια άλλη να μην είναι. Η εκφραστικότητα αυτή μπορεί να αποδοθεί με δύο νέους τελεστές, τους τροπικούς τελεστές (modal operators), όπως τον \square που συμβολίζει ότι μια πρόταση είναι αληθής πάντα και τον \diamond που συμβολίζει ότι μία πρόταση είναι αληθής στο μέλλον. Ο κυρίως λόγος χρήσης αυτών των τελεστών είναι ο διαχωρισμός της γνώσης από την πεποίθηση, όπου η γνώση καθορίζεται ως μια αντίληψη που ισχύει πάντα ενώ σε αντίθετη περίπτωση αποτελεί μία πεποίθηση.

Μια ακόμα πιο εκφραστική μέθοδος που ενσωματώνει το χρόνο είναι η *λογική χρονικών διαστημάτων*, η οποία σε αντίθεση με την χρονική λογική, ασχολείται με τις σχέσεις που μπορεί να έχουν γεγονότα μεταξύ τους μέσα στο χρόνο. Εδώ το βασικό στοιχείο είναι τα χρονικά διαστήματα στα οποία ορίζονται δύο χρονικά σημεία, αυτά της έναρξης και της λήξης, ενώ μεταξύ των διαστημάτων καθορίζονται επτά βασικές δυαδικές σχέσεις αυτές της επικάλυψης (overlaps), προηγείται η μία της άλλης (precedes), συναντά (meets), αρχίζει (starts), τελειώνει (ends), στη διάρκεια (during) και ισοδυναμία (equals).

Η χρησιμότητα της χρονικής λογικής είναι η εξαγωγή συμπερασμάτων που αφορούν την εξέλιξη του κόσμου ενός προβλήματος. Εδώ τα εργαλεία που χρησιμοποιούνται για τον συμπερασμό είναι η παραγωγική συλλογιστική, η επαγωγική συλλογιστική και απαγωγική συλλογιστική.

Ασαφής Λογική. Αυτές οι μέθοδοι προσπαθούν να καλύψουν περιπτώσεις που η πληροφόρηση είναι ατελής είτε επειδή αυτή δεν είναι διαθέσιμη είτε επειδή απλά δε χρειάζεται. Η ασαφής λογική είναι ένα υπερσύνολο της κλασικής λογικής, η οποία έχει

επεκταθεί ώστε να μπορεί να χειριστεί τιμές αληθείας μεταξύ του απολύτως αληθούς και του απολύτως ψευδούς. Τα εργαλεία που χρησιμοποιεί είναι τα ασαφές σύνολα που περιέχουν όλα τα αντικείμενα στα οποία μπορεί να γίνει αναφορά και τον βαθμό αληθείας που αντιστοιχεί στο βαθμό συγγένειας ενός αντικειμένου με το ασαφές σύνολο. Όπως και στην κλασική λογική έτσι και εδώ ισχύουν ιδιότητες και ορίζονται πράξεις όπως η ένωση και η τομή, απλά οι μεταβλητές και οι κανόνες προσαρμόζονται στα νέα δεδομένα και ορίζονται ως ασαφείς μεταβλητές και ασαφείς κανόνες.

Η εξαγωγή συμπερασμάτων πλέον γίνεται μέσα από τους ασαφείς κανόνες βάσει του αρμόδιου μηχανισμού που ονομάζεται ασαφής συλλογιστική. Γενικά, η χρήση της ασαφούς συλλογιστικής απαιτεί την ύπαρξη μιας ασαφούς λεκτικής περιγραφής του προβλήματος και περιλαμβάνει τέσσερα στάδια. Στο πρώτο γίνεται ο υπολογισμός της συνάρτησης συνεπαγωγής για κάθε εμπλεκόμενο κανόνα. Δεύτερον, γίνεται η παραγωγή επιμέρους αποτελεσμάτων μέσω κάποιας συλλογιστικής διαδικασίας. Τρίτον, πραγματοποιείται συνάθροιση των επιμέρους αποτελεσμάτων και τέλος η αποσαφήνιση των αποτελεσμάτων.

Όπως έχει επισημανθεί μια μέθοδος αναπαράστασης δεν μπορεί να καλύψει όλα τα είδη της γνώσης. Γι' αυτό πολλές φορές πραγματοποιείται η συνένωσή τους, ώστε να επιτευχθεί η κάλυψη ένα ευρύτερου πλαισίου απαιτήσεων. Τέτοια παραδείγματα είναι αυτά των γλωσσών Ontolingua, OCML, FLogic που συνδυάζουν τα πλαίσια με την κατηγορηματική λογική αλλά και της SHOE που συνδυάζει πάλι τα πλαίσια με τους κανόνες.

4. Οντολογίες

Η Οντολογία εμφανίστηκε αρχικά ως κλάδος της φιλοσοφίας, για την κατανόηση του φυσικού ή του μεταφυσικού. Η ιδέα ήταν να αναζητηθούν οι οντότητες και οι σχέσεις που τις διέπουν, τα οποία αποτελούν ή θα ήταν εν δυνάμει στοιχεία της πραγματικότητας. Υπήρξαν και υπάρχουν σχολές που προσεγγίζουν και αντιλαμβάνονται την Οντολογία με διαφορετικό τρόπο. Αυτό οδήγησε από πολύ νωρίς στην δημιουργία πολλών ορισμών και εργαλείων προσέγγισης. Παρόλα αυτά όλες προσπαθούν να πλαισιώσουν τις Οντολογίες με ένα οριστικό και πλήρες ορισμό. Χρησιμοποιούνται διάφορα εργαλεία όπως η άλγεβρα, η θεωρία συνόλων, η τοπολογία αλλά και διάφορες γλώσσες όπως η λογική. Η χρήση τέτοιων εργαλείων επιτρέπουν σε όσους ασχολούνται με αυτή να εκφράσουν διαισθητικές αρχές και ορισμούς μέσα σε ένα σαφή και αυστηρό πλαίσιο, στην εφαρμογή τους, αλλά και τον έλεγχο των θεωριών για τη συνοχή και την πληρότητά τους. Όλα αυτά δείχνουν το πόσο περίπλοκο είναι το πεδίο της Οντολογίας αλλά και πόσο μείζονος σημασίας είναι για πολλά πεδία εφαρμογής (Smith, 2003).

Οι Οντολογίες στην Πληροφορική δεν παύουν να δεσμεύονται από τις ίδιες «ανησυχίες» με αυτές της φιλοσοφικής σκέψης. Και εδώ υπάρχει μια πληθώρα διαφορετικών προσεγγίσεων με διαφορετικούς ορισμούς και στόχους. Βέβαια, η κεντρική ιδέα της κατηγοριοποίησης, της εύρεσης εννοιών και συσχετίσεων παραμένει σταθερή. Στην Επιστήμη της Πληροφορικής, οι Οντολογίες χρησιμοποιούνται ως μέσα αναπαράστασης γνώσης, τα οποία μπορούν να υποστηρίξουν διαδικασίες συμπερασμού, αναπαραγωγής και μεταλαμπάδευσης γνώσης. Έχουν αναπτυχθεί μεθοδολογίες και εργαλεία τα οποία περισσότερο φαίνεται να προσαρμόζουν τα ήδη υπάρχοντα στις νέες απαιτήσεις παρά να δημιουργούνται καινούριες. Η χρήση των Οντολογιών εμφανίζεται σε διάφορες επιστήμες όπως η Τεχνητή Νοημοσύνη, Τεχνολογία Λογισμικού αλλά και σε πιο εξειδικευμένα πεδία όπως η Διαχείριση Γνώσης (Knowledge Management), Ανάκτηση Πληροφορίας (Information Retrieval), Σημασιολογικό Δίκτυο, Βιολογία, Ασφάλεια, Τρομοκρατία.

Το δυνατό σημείο μιας Οντολογίας είναι ο τυπικός και σαφής ορισμός της. Αυτό της επιτρέπει να είναι αξιόπιστη, κατανοητή και αποδεκτή κατά γενική ομολογία. Είναι φιλική προς τους ανθρώπους αλλά συνάμα λειτουργική για την επικοινωνία μεταξύ υπολογιστικών συστημάτων. Ορίζεται αυστηρά, έτσι είναι εύκολο να βρεθεί τι δεν συμπεριλαμβάνει και κατ'επέκταση να το ενσωματώσει και να επεκταθεί. Επίσης, επιτρέπει την εύκολη ανάλυση και διαχείριση της γνώσης, καθώς ειδικά υπολογιστικά εργαλεία μπορούν να την εκμεταλλευτούν.

Βέβαια η πιο βασική έμμεση χρησιμότητα που προσφέρει είναι η διαλειτουργικότητα (interoperability) μεταξύ των συστημάτων. Είναι τόσο τα διαφορετικού τύπου δεδομένα όσο και η διαφορετική αρχιτεκτονική μεταξύ των συστημάτων που τα αναγκάζουν να αναζητήσουν μία καλά ορισμένη δομή που θα τους βοηθήσει να επικοινωνήσουν. Μια τέτοια δομή μπορεί να είναι οι Οντολογίες που είτε μπορούν να χρησιμοποιηθούν άμεσα μεταξύ των δύο

συστημάτων, γεγονός το οποίο σημαίνει ότι θα διατηρούν την ίδια ορολογία και τη γνώση, είτε κάθε σύστημα να χρησιμοποιεί την δική του Οντολογία. Οι Οντολογίες αυτές θα πρέπει να αντιστοιχίζουν τις έννοιες που ενσωματώνουν, ώστε να αποφευχθεί τυχόν σύγχυση κατά τη διάρκεια της συναλλαγής τους.

Φυσικά η Οντολογία ως μέσο αναπαράστασης της γνώσης χρησιμεύει για την επαναχρησιμοποίηση (reusability) και τη διάδοση της. Μεγάλες εταιρείες με πολυετή εμπειρία και γνώση θα ήταν ασύμφορο να ανακαλύπτουν από το μηδέν όλη αυτή την γνώση που αποκομίζουν όλα αυτά τα έτη. Με τις Οντολογίες, αυτή η γνώση διαμοιράζεται, διατηρείται, εμπλουτίζεται και πολλές φορές γενικεύεται για να καλύπτει περισσότερες και πιο αφηρημένες ανάγκες. Μιλώντας για την πρακτική εφαρμογή των Οντολογιών, άλλο ένα χρόνιο πρόβλημα, κυρίως για τις μεγάλες εταιρείες, είναι ο μεγάλος όγκος δεδομένων και πληροφοριών που έχουν στην κατοχή τους. Καθώς μετά από ένα σημείο αντί να βοηθούν να λαμβάνονται αποφάσεις, δημιουργούν πρόβλημα στην κατανόηση τους και παράγουν άχρηστη ή τετριμμένη γνώση.

Άλλες δύο άμεσες χρησιμότητες των Οντολογιών είναι ότι προσδίδουν σημασιολογία και την δυνατότητα συμπερασμού. Δύο συστήματα που χρησιμοποιούν την ίδια ορολογία δεν είναι αναγκαίο ότι τους προσδίδουν την ίδια σημασιολογία, αυτό το κενό καλύπτουν οι Οντολογίες για την αποφυγή σημασιολογικής σύγχυσης. Βέβαια, ένα καλύτερο παράδειγμα που αξιοποιεί πλήρως αυτές τις δύο ιδιότητες είναι οι διαδικασίες αναζήτησης δεδομένων, πληροφοριών ή γνώσης όπου η μεν σημασιολογία πραγματοποιεί στοχευόμενη εξόρυξη με δυνατότητα σημασιολογικής ερμηνείας του αποτελέσματος, ενώ ο συμπερασμός μπορεί και ανακαλύπτει κρυμμένες πτυχές της ερώτησης.

Ένα ακόμα κλασικό παράδειγμα χρήσης των Οντολογιών είναι το διαδίκτυο ή καλύτερα η εξέλιξη του, ο Σημασιολογικός Ιστός. Εκεί ένας μεγάλος όγκος πληροφορίας ζητά να κατανοηθεί και να διαμοιρασθεί. Πολύπλοκα συστήματα με διαφορετικούς τύπους δεδομένων συνδέονται, ενώ άνθρωποι και υπολογιστικά συστήματα αναζητούν συγκεκριμένες πληροφορίες. Οι Οντολογίες παίζουν τον ρόλο του καθοδηγητή, όπως στην περίπτωση των πρακτόρων, που τους βοηθούν να επικοινωνήσουν, να αναζητήσουν και να αναλύσουν συγκεκριμένες πληροφορίες. Επίσης βοηθούν τους ανθρώπους στο να αναζητήσουν πληροφορία, όχι μόνο με βάση κάποια συγκεκριμένη λέξη αλλά με την σημασιολογία της. Αυτό δίνει την δυνατότητα για ακόμα πιο στοχευόμενη αναζήτηση πληροφορίας και το σημαντικότερο για εύρεση και προώθηση πληροφορίας με αυξημένο ενδιαφέρον, ανάλογα με τις εκάστοτε απαιτήσεις.

4.1 Ορισμός Οντολογίας

Όπως έχει ειπωθεί, πολύ ορισμοί έχουν παρουσιαστεί για το τι τελικά είναι οι Οντολογίες. Σύμφωνα με τον Corcho και συν. (2003) αλλά και τους Guarino και Giaretta (1995) υπάρχει συναίνεση μεταξύ της ερευνητικής κοινότητας και έτσι δεν υπάρχει σύγχυση σχετικά με τη χρήση τους. Διαφορετικοί ορισμοί παρέχουν διαφορετικές και συμπληρωματικές οπτικές γωνίες της ίδιας πραγματικότητας. Κάποιοι ορισμοί δίνουν σημασία στο τι περιλαμβάνει μια Οντολογία και αναφέρονται στις σχέσεις, στο πεδίο ενδιαφέροντος, στη γλώσσα που θα τις καθορίσει. Άλλοι πάλι προσανατολίζονται στο τρόπο κατασκευής μιας Οντολογίας όπως ότι γίνεται εξαγωγή της από μια Βάση Γνώσης ή ότι έχουν έναν ιεραρχικό σκελετό που μπορούν να χρησιμοποιήσουν οι Βάσεις Γνώσης. Άλλοι πάλι είναι πιο ελαστικοί ώστε να διευρύνεται ο όρος και να συμπεριλαμβάνει και άλλες δομές όπως οι Ταξινομίες. Επίσης πολλοί προσπαθούν να κάνουν έναν πιο γενικό ορισμό για να συμπεριλάβουν και άλλα πεδία έρευνας όπως οι Βάσεις Δεδομένων όπου οι όροι: έννοιες, σχέσεις, ιδιότητες δεν είναι ξένοι.

Ο πιο ευρέως διαδεδομένος και χρησιμοποιούμενος ορισμός τυγχάνει να είναι αυτός του Gruber (2003). Ο Gruber αντιλαμβάνεται την Οντολογία ως μια ρητή προδιαγραφή μιας σύλληψης. Όπου ο όρος *ρητή* σημαίνει ότι οι έννοιες και οι περιορισμοί που χρησιμοποιούνται είναι σαφώς καθορισμένες. Ενώ η *σύλληψη* δεν είναι τίποτε άλλο από μια αφηρημένη, απλοποιημένη άποψη του κόσμου που θέλει κάποιος να αντιπροσωπεύσει για έναν σκοπό, που μπορεί να περιγραφεί μέσω αντικειμένων, εννοιών ή οντοτήτων καθώς και σχέσεων που συνδέονται, και υποτίθεται φυσικά ότι, βρίσκονται μέσα σε μια περιοχή ενδιαφέροντος. Έπειτα στον ορισμό προστέθηκε ακόμα μία έννοια αυτή της *τυπικότητας* δηλαδή ότι η Οντολογία πρέπει να είναι αναγνώσιμη από μία υπολογιστική μηχανή (Corcho και συν., 2003).

Ο Gruber (2003) εκτός του ορισμού καθορίζει και πέντε βασικά κριτήρια που θα πρέπει να καλύπτουν οι Οντολογίες, αυτά είναι:

- Σαφήνεια (Clarity): Δυνατότητα φορμαλιστικού ορισμού της Οντολογίας και αποτελεσματική σύνδεση των όρων με την αντικειμενική τους έννοια.
- Συνοχή (Coherence): Οι ορισμοί να συνάδουν με τους περιορισμούς.
- Επεκτασιμότητα (Extendibility): Δυνατότητα για να ορισθούν νέοι όροι με βάση το υπάρχον λεξιλόγιο.
- Ελαχιστοποίηση της μεροληψίας στην κωδικοποίηση (minimal coding bias): Η σύλληψη δεν θα πρέπει να βασίζεται στην συμβολική κωδικοποίηση των όρων.
- Ελαχιστοποίηση της οντολογικής δέσμευσης (minimal ontological commitment): Η Οντολογία θα πρέπει να αποτελείται από τις ελάχιστες δυνατές έννοιες και δεσμεύσεις για να επιτευχθεί η μέγιστη δυνατή απεικόνιση της γνώσης μέσα σε μία σύλληψη.

Ενώ οι Ding και συν. (2005) συμπληρώνουν:

- Ορατότητα (Visibility): Απαιτείται μια κατάλληλη σύνταξη που να είναι κατανοητή από τις υπολογιστικές μηχανές.
- Ικανότητα συμπερασμού (Inferenceability): Να επιτρέπει λογικούς συμπερασμούς μέσω των αξιωμάτων και να εκμεταλλεύεται την εκφραστικότητα (expressiveness) της εκάστοτε γλώσσας.

Γίνεται αντιληπτό ότι πολλά από τα κριτήρια είναι αντικρουόμενα και εδώ τονίζεται ο σημαντικός ρόλος του σχεδιαστή που θα πρέπει να λάβει κάποιες αποφάσεις για το που θα υπερτερεί ή όχι μία Οντολογία (Gruber, 2003; Mizoguchi, 2004).

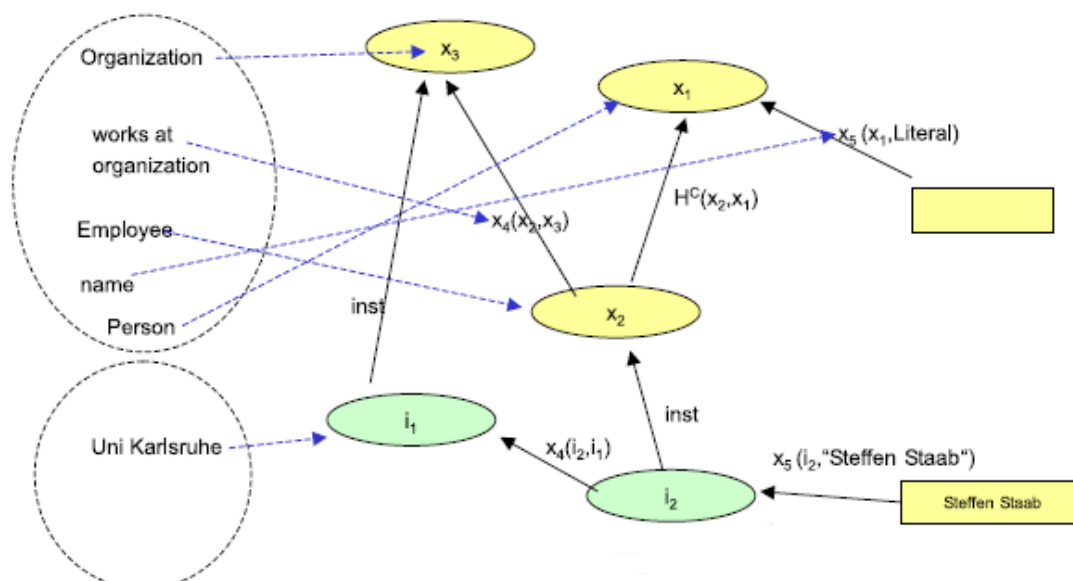
Προσεγγίζοντας φορμαλιστικά τον ορισμό της Οντολογίας αποτελείται από έξι στοιχεία $O := \{C, P, A, H^{C,P}, prop, att\}$ (Hotho και συν., 2003). Από αυτά τα σύνολα C, P και A είναι μεταξύ τους ξένα. Τα στοιχεία του C αποτελούν τις έννοιες, του P τις σχέσεις και του A τις ιδιότητες της Οντολογίας, πάνω στα οποία ορίζονται τέσσερις σχέσεις:

- Ιεραρχία εννοιών ή ταξινόμια H^C : είναι μια άμεση και θετική σχέση $H^C \subseteq C \times C$, όπου η δήλωση $H^C(C_1, C_2)$ εκφράζει ότι το C_1 είναι υπο- έννοια του C_2 .
- Η συνάρτηση $prop: P \rightarrow C \times C$ η οποία συσχετίζει (rel) τις έννοιες μη ταξινομικά και η οποία περιλαμβάνει τις συναρτήσεις:
 - Πεδίο ορισμού $dom: P \rightarrow C$ όπου το $dom(P) := \Pi_1(rel(P))$ δίνει το πεδίο ορισμού του P .
 - Πεδίο τιμών $range: P \rightarrow C$ όπου το $range(P) := \Pi_2(rel(P))$ δίνει το πεδίο τιμών του P .
- Ιεραρχία σχέσεων H^P : είναι μια άμεση και θετική σχέση $H^P \subseteq P \times P$, όπου η δήλωση $H^P(P_1, P_2)$ εκφράζει ότι το P_1 είναι υπο- σχέση του P_2 .
- και η συνάρτηση $att: A \rightarrow C$ που συνδέει τις έννοιες με κυριολεκτικές τιμές, αυτό υπονοεί ότι $range(A) := STRING$.

Πάνω σε μία Οντολογία πλέον μπορεί να οριστεί ένα στιγμιότυπο το οποίο καθορίζεται από έξι στοιχεία $MD := \{O, I, L, inst, instr, instl\}$, τα οποία αποτελούνται από μία Οντολογία O , ένα σύνολο αναγνωριστικών στιγμιότυπων I , όπου τα C, P, I είναι ξένα μεταξύ τους και ένα σύνολο κυριολεκτικών τιμών L . Με βάση αυτά ορίζονται και οι ακόλουθες συναρτήσεις:

- Στιγμιότυπο εννοιών, $inst: C \rightarrow 2^I$, όπου η συνάρτηση $inst(C) = I$ μπορεί να γραφεί ως $C(I)$.
- Στιγμιότυπο σχέσεων, $instr: P \rightarrow 2^{I \times I}$, όπου η συνάρτηση $instr(P) = \{(I_1, I_2)\}$ μπορεί να γραφεί $P(I_1, I_2)$.
- Στιγμιότυπο χαρακτηριστικών, $instl: P \rightarrow 2^{I \times L}$, όπου συνδέει τα στιγμιότυπα με τις κυριολεκτικές τιμές, και εδώ η συνάρτηση $instr(A) = \{(I_1, L)\}$ μπορεί να γραφεί $A(I_1, L)$.

**Παράδειγμα 4-1. Οντολογία και στιγμιότυπο της
Πηγή: Hotho και συν. (2003:5)**



Με βάση τους παραπάνω ορισμούς και το παράδειγμα της εικόνας μπορεί να θεωρηθεί ένα σύνολο εννοιών $C := \{x_1, x_2, x_3\}$, που μεταξύ των x_1, x_2 υφίστανται μια συνάρτηση ιεραρχίας εννοιών, όπου η x_2 είναι υπο-έννοια της x_1 , μία συνάρτηση πεδίου τιμών $R := \{x_4\}$, όπου η ιδιότητα x_4 ορίζεται για την κλάση x_2 και μία συνάρτηση $A := \{x_5\}$ που καθορίζει την τιμή της έννοιας x_1 ως λεκτική. Με βάση αυτή την Οντολογία μπορεί να καθοριστούν τα στιγμιότυπα $I := \{i_1, i_2\}$ των εννοιών x_1, x_2 με τη βοήθεια της συνάρτησης $inst$, και το στιγμιότυπο της ιδιότητας που τα συνδέει $x_4(i_1, i_2)$. Τέλος, εμφανίζεται το στιγμιότυπο της ιδιότητας x_5 που δίνει ως τιμή της i_1 το λεκτικό "Steffen Staab".

4.2 Χαρακτηριστικά Οντολογίας

Ο Gómez-Pérez (1999) εξετάζοντας τον ορισμό του Gruber περί Οντολογίας διακρίνει πέντε βασικά χαρακτηριστικά που την αποτελούν. Αυτά είναι οι έννοιες (concepts), οι σχέσεις (relations), οι συναρτήσεις (functions), τα αξιώματα (axioms) και τα στιγμιότυπα (instances), και ο κατάλληλος συνδυασμός τους είναι ικανός να περιγράψει μία σύλληψη. Οι Chandrasekaran και Josephson (1999) προχωρούν λίγο πιο πέρα και διακρίνουν άλλο ένα χαρακτηριστικό, αυτό του χρόνου.

Οι έννοιες τυχάνουν να είναι η πρώτη ύλη μέσα σε μία Οντολογία και αποτελούν ουσιαστικά το λεξιλόγιο της. Βέβαια, εδώ το σημαντικό στοιχείο δεν είναι το λεξιλόγιο που δημιουργείται αλλά η σύλληψη που οι όροι του λεξιλογίου στοχεύουν να περιγράψουν (Chandrasekaran & Josephson, 1999). Κατά κανόνα οι έννοιες ακολουθούν τη λογική της ταξινόμιας (taxonomies) (Gómez-Pérez, 1999) όπου συμμετέχουν σε μία ιεραρχική δομή, που πολλές φορές όμως έχει τη μορφή της πολλαπλής κληρονομικότητας (multiple inheritance) (Noy & McGuinness, 2001), δηλαδή μία έννοια μπορεί να έχει δύο ή περισσότερους προγόνους. Άλλες ονομασίες που τις έχουν αποδοθεί είναι οι κλάσεις (classes) ή αντικείμενα (objects).

Οι σχέσεις συμβολίζουν έναν τύπο αλληλεπίδρασης μεταξύ των εννοιών που αποτελούν την σύλληψη. Ένα κλασικό παράδειγμα είναι οι σχέσεις κληρονομικότητας μεταξύ των εννοιών. Οι σχέσεις μπορούν να αναφέρονται ως ιδιότητες (properties), ρόλοι (role) ή σχισμές (slots) ανάλογα με την ορολογία της γλώσσας που έχει επιλεγεί για την αναπαράσταση της σύλληψης. Πάνω στις ιδιότητες ορίζονται περιορισμοί (restrictions), όπως οι τύποι τιμών, οι περιορισμοί ποσοτικοποίησης (cardinality restrictions), ο καθολικός ποσοδείκτης (universal quantifier- \forall), ο υπαρξιακός ποσοδείκτης (existential restriction- \exists) και το πεδίο τιμών (Range) ή ορισμού (Domain). Άλλες ονομασίες που μπορούν να βρεθούν για του περιορισμούς είναι εκφάνσεις (facets) ή περιορισμοί ρόλων (role restrictions). Ενδιαφέρον παρουσιάζει ότι είναι δυνατή η δημιουργία ιεραρχικής δομής και για τον ορισμό των ιδιοτήτων.

Οι συναρτήσεις είναι ειδικές περιπτώσεις σχέσεων πάνω στις οποίες το n -οστό στοιχείο της σχέσης είναι μοναδικό για τα $n-1$ προηγούμενα στοιχεία, φORMALISτικά μπορεί να οριστεί ως $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$. Μεγάλο βάρος έχουν οι ιδιότητες των συναρτήσεων που παίζουν σημαντικό ρόλο στην εκφραστικότητα της Οντολογίας και τις δυνατότητες συμπερασμού από τις Μηχανές Συλλογιστικής Ανάλυσης. Τέτοιες ιδιότητες είναι της αντιστροφής (inverse), συναρτησιακές (functional), μεταβατικότητας (transitive), συμμετρική (symmetric) (Horridge & συν., 2009).

Τα αξιώματα χρησιμοποιούνται για να μοντελοποιηθούν προτάσεις που αληθεύουν πάντα. Καθορίζουν τους ορισμούς των εννοιών της Οντολογίας και των περιορισμών σχετικά με την ερμηνεία τους. Αποτελούνται από την σύνδεση των εννοιών, των ιδιοτήτων αλλά και από συλλογές περιορισμών. Οι Uschold και Gruninger (1996) επισημαίνουν την χρησιμότητα και την αναγκαιότητα για τον πλήρες και αυστηρό ορισμό τους, παρόλη τη δυσκολία του εγχειρήματος, καθώς μπορούν και αυξάνουν σημαντικά την εκφραστικότητα της Οντολογίας.

Τέλος, τα στιγμιότυπα χρησιμοποιούνται για την αναπαράσταση συγκεκριμένων στοιχείων. Δηλαδή, είναι στιγμιότυπα των εννοιών που λαμβάνουν συγκεκριμένες τιμές και ακολουθούν τους περιορισμούς και τα αξιώματα που έχουν οριστεί. Ο συνδυασμός της Οντολογίας με τα στιγμιότυπα μπορούν να δώσουν πλέον μία ολοκληρωμένη Βάση Γνώσης (Noy & McGuinness, 2001).

Με την έννοια του χρόνου οι Chandrasekaran και Josephson θέλουν να τονίσουν τις αλλαγές που μπορούν να συμβούν αλλά και τις διαδικασίες που λαμβάνουν χώρα μέσα σε ένα χρονικό διάστημα. Για παράδειγμα, αναφέρουν ότι οι σχέσεις μπορούν να αλλάζουν κατά τη διάρκεια του χρόνου, ότι υπάρχουν γεγονότα που συμβαίνουν σε διαφορετική χρονική στιγμή, υπάρχουν διαδικασίες που τα αντικείμενα συμμετέχουν κατά τη διάρκεια του χρόνου, ότι ο κόσμος και τα αντικείμενα μπορούν να έχουν διαφορετική κατάσταση, ενώ μπορεί να υπάρξει μια σύνδεση αιτίας αποτελέσματος, δηλαδή ένα γεγονός μπορεί να προκαλέσει ένα νέο ή μία μετάβαση σε μία νέα κατάσταση.

4.3 Κατηγορίες Οντολογιών

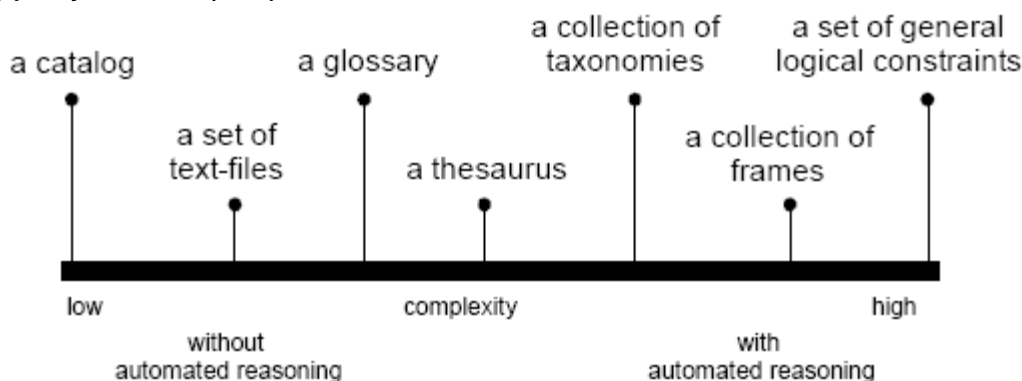
Ο Gómez-Pérez (1999) παρουσιάζει ένα μικρό σύνολο των τύπων των Οντολογιών που έχουν παρουσιαστεί στη βιβλιογραφία, λόγω του μεγάλου αριθμού τους. Οι Mizoguchi και συν. (1995) επεξηγούν ότι κάτι τέτοιο είναι λογικό καθώς δεν μπορεί να υπάρξει μόνο ένα είδος Οντολογιών, πράγμα που φαίνεται και από το πλήθος των διαφορετικών ορισμών που υπάρχουν. Αφορμή για την ομαδοποίηση των Οντολογιών μπορεί να σταθεί το μέγεθος, η δομή, το αντικείμενο και ο σκοπός που καλύπτει η σύλληψη (Mizoguchi & συν., 1995; Van Heijst & συν., 1997). Παραδείγματα αυτών είναι οι Οντολογίες Αναπαράστασης Γνώσης (Knowledge Representation Ontologies), Οντολογίες Εφαρμογής (Application Ontologies), Μετα- Οντολογίες (Meta-Ontologies). Αυτή η ενότητα θα σταθεί σε τέσσερις κατηγορίες που έχουν να κάνουν με την τυπικότητα, τις γλώσσες, την εσωτερική δομή και την γενικότητα των Οντολογιών.

Βαθμός τυπικότητας. Αναφέρεται στο βαθμό τυπικότητας με το οποίο ένα λεξιλόγιο δημιουργείται (Uschold & Gruninger, 1996) και παίζει πολύτιμο ρόλο στην περίπτωση που οι όροι του χρειάζεται να επαναχρησιμοποιηθούν ή να επεκταθούν αλλά και στην περίπτωση συμπερασμού. Θεωρώντας την τυπικότητα ως ένα συνεχές ευθύγραμμο τμήμα τα δύο άκρα του θα ήταν υψηλά άτυπη (highly informal) και αυστηρά τυπική (rigorously formal). Έτσι μπορούν να καθοριστούν:

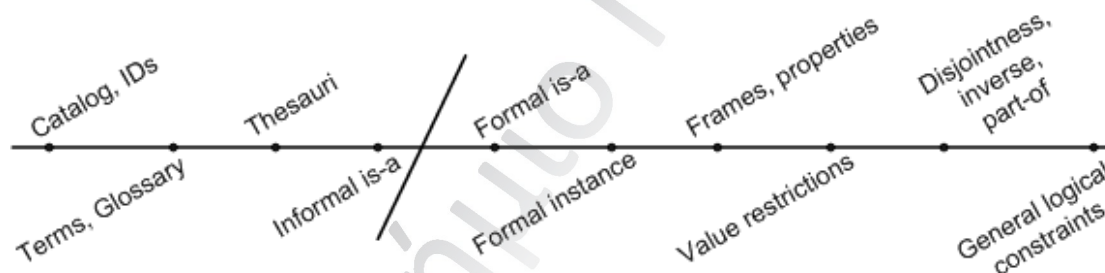
- **Υψηλά άτυπη:** εκφράζεται με μία χαλαρή φυσική γλώσσα, όπως καταλόγους, γλωσσάρια, θησαυρούς.
- **Ημι- άτυπη (semi- informal):** εκφράζεται σε μια κλειστή και δομημένη φυσική γλώσσα. Χαρακτηριστικό της είναι ότι προσπαθεί να αυξήσει τη σαφήνεια μειώνοντας την ασάφεια. Τέτοια παραδείγματα είναι οι κατάλογοι του Yahoo, όπου παρέχουν γενίκευση και εξειδίκευση στους όρους αλλά δεν ακολουθούν μία αυστηρή ιεραρχία.
- **Ημι- τυπική (semi- formal):** εκφράζεται μέσω μιας τεχνητής τυπικά ορισμένης γλώσσας, όπως με αυστηρές ιεραρχικές υποκλάσεις
- **Αυστηρά τυπική:** εκφράζεται με σχολαστικό καθορισμό όρων και ιδιοτήτων, με τυπική σημασιολογία, θεωρήματα και αποδείξεις για την ορθότητα και την πληρότητά τους, όπως τον καθορισμό ιδιοτήτων των εννοιών, περιορισμούς και αξιώματα.

Στην εικόνα που ακολουθεί παρουσιάζονται παραδείγματα τυπικών δομών. Στα αριστερά ξεκινάνε οι υψηλά άτυπες, οι οποίες έχουν χαμηλή πολυπλοκότητα αλλά είναι αδύναμες να παράγουν αυτόματο συμπερασμό, ενώ στο αντίθετο άκρο είναι οι αυστηρά τυπικές με πολύ μεγάλο βαθμό πολυπλοκότητας αλλά και με αυξημένες δυνατότητες συμπερασμού.

Παράδειγμα 4-2. Τυπικότητα Οντολογιών
 Πηγή: Floyd & Ukena (2005)



Εσωτερική δομή. Με βάση την κατηγοριοποίηση κατά τυπικότητα οι Lassila και Swick (1999) εξετάζουν συγκεκριμένες δομές που έχουν υλοποιηθεί. Εδώ παρόλο που η κατηγοριοποίηση συνάδει πλήρως με τις παραπάνω περιγραφές, δίνεται βάρος στις διαφορετικές δομές που υπάρχουν.



Εικόνα 4-1. Είδη Οντολογιών
 Πηγή: Lassila & Swick (1999:4)

- **Ελεγχόμενο Λεξιλόγιο (Controlled Vocabulary):** Είναι μία λίστα από όρους που έχουν απαριθμηθεί ρητά και χρησιμοποιούνται για επισήμανση, ευρετηρίαση ή κατηγοριοποίηση. Χρησιμοποιείται ο όρος ελεγχόμενο, γιατί μπορούν να χρησιμοποιηθούν μόνο οι όροι του λεξιλογίου σε ένα πεδίο εφαρμογής, ενώ η διαχείριση τους γίνεται κάτω από συγκεκριμένες ελεγχόμενες διαδικασίες. Ένα παράδειγμα αυτών είναι οι κατάλογοι (Catalog), οι οποίοι μπορούν να παρέχουν μια σαφή και ενιαία ερμηνεία σε συγκεκριμένους όρους. Μία πιο εξελιγμένη μορφή είναι το **γλωσσάρι (Glossary)**, που μαζί με τη λίστα με τους όρους περιέχει και τη σημασία τους. Η σημασιολογία τους προσδιορίζεται μέσω της φυσικής γλώσσας και αυτό τους δίνει το πλεονέκτημα για πλήρη κατανόηση τους από τον άνθρωπο αλλά είναι αδυνατούν να ερμηνευθούν από τις υπολογιστικές μηχανές.

Παράδειγμα 4-3. Κατάλογος - Γλωσσάρι

Κατάλογος	Γλωσσάρι
Al-Qa`ida ⇔ 63	<i>al-Qaeda, Qaeda, al-Qa'ida, al-Qaida, al Qaeda, al Qaida, Base (a terrorist network intensely opposed to the United States that dispenses money and logistical support and training to a wide variety of radical Islamic terrorist groups; has cells in more than 50 countries)</i> πηγή: (Wordnet, 2011)

- **Θησαυρός Γνώσης (Thesaurus):** Είναι ένα ελεγχόμενο λεξιλόγιο, το οποίο παρέχει μία επιπρόσθετη σημασιολογία στις σχέσεις μεταξύ των όρων του. Τυπικά οι θησαυροί δεν παρέχουν ρητή ιεραρχία, παρόλα αυτά πολλές φορές θεωρούνται ότι είναι οργανωμένοι με μία ορισμένη τάξη και δομή έτσι ώστε να υφίστανται συγκεκριμένες συσχετίσεις, ιεραρχίες και ισοδυναμίες, μεταξύ των διάφορων όρων που παρατίθενται σαφώς και προσδιορίζονται από συγκεκριμένους, τυποποιημένους δείκτες (ANSI/NISO Z39.19 – 2003). Από τον ορισμό γίνεται αντιληπτό ότι μπορούν να υπάρξουν θησαυροί οι οποίοι είναι δυνατόν να ερμηνευθούν σαφώς από υπολογιστικές μηχανές. Η εκφραστικότητα των σχέσεων μέσα στον θησαυρό ποικίλουν και μία από τις πιο γνωστές είναι οι σχέσεις συνωνύμου (Synonym), όπως φαίνεται και στο παράδειγμα που ακολουθεί.

Παράδειγμα 4-4. Θησαυρός

Σχέση	Ορισμός	Παράδειγμα
Συνώνυμο	Ένας όρος X έχει σχεδόν την ίδια σημασία με τον όρο Y	«Αναφορά» είναι ένα συνώνυμο για το «Έγγραφο»
<p><i>Main Entry:</i> report <i>Part of Speech:</i> verb <i>Definition:</i> communicate information, knowledge <i>Synonyms:</i> account for, advise, air, announce, bring word, broadcast, cable, circulate, cover, declare, describe, detail, disclose, document, enunciate, give an account of, give the facts, impart, inform, inscribe, itemize, list, make known, make public, mention, narrate, note, notify, pass on, present, proclaim, promulgate, provide details, publish, recite, record, recount, rehearse, relate, relay, retail, reveal, set forth, spread, state, summarize, telephone, tell, trumpet, wire, write up <i>Antonyms:</i> conceal, suppress</p>		

- **Ταξινομία (Taxonomy):** Είναι μία συλλογή από ένα ελεγχόμενο λεξιλόγιο όρων οργανωμένο σε ιεραρχική δομή. Κάθε όρος σε μια ταξινομία μπορεί να συμμετέχει σε μία ή περισσότερες σχέσεις γονέα- παιδιού με έναν άλλο όρο, ενώ οι σχέσεις γονέα-παιδιού μπορεί να λαμβάνουν διάφορους τύπους μέσα στην ταξινομία, όπως γένος-ειδος, τύπος- στιγμιότυπο. Βέβαια σε τέτοιες σχέσεις ο γονέας περιορίζεται να ανήκει σε έναν μόνο τύπο. Μερικές ταξινομίες επιτρέπουν την πολλαπλή ιεραρχία, πράγμα που σημαίνει ότι ένας όρος μπορεί να έχει πολλούς γονείς. Γενικά, η χρήση της είναι δημοφιλής για ιεραρχικές κατατάξεις ή συστήματα κατηγοριοποίησης, όπως παραδοσιακά εφαρμόζεται για την ταξινόμηση ζώων ή φυτών.

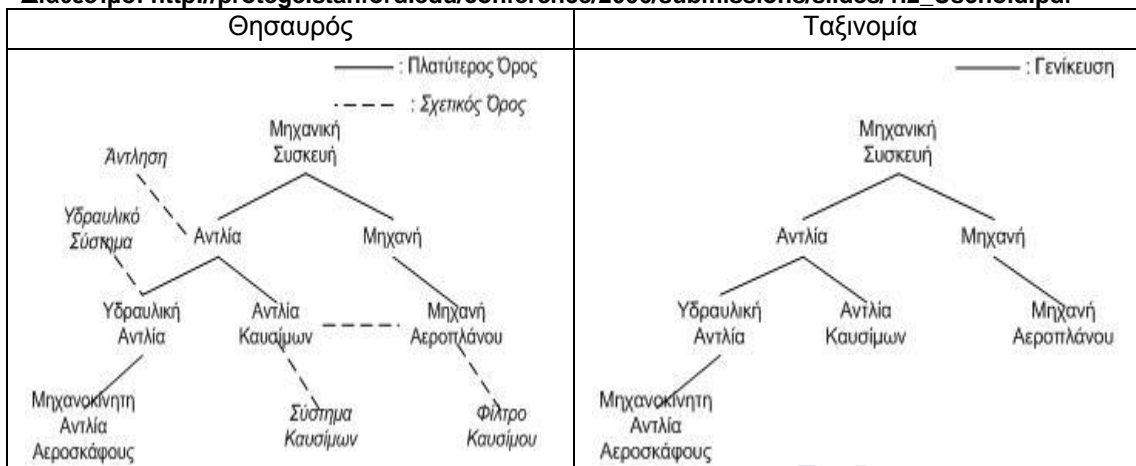
Μεταξύ των ταξινομιών υπάρχουν δύο μεγάλες κατηγορίες, αυτή της άτυπης και αυτή της τυπικής ταξινομίας. Και οι δύο έχουν μία σαφώς ορισμένη ιεραρχία αλλά στην πρώτη δεν υφίσταται η έννοια της κληρονομικότητας (is-a), δηλαδή ένα στιγμιότυπο μίας έννοιας δεν είναι απαραίτητα στιγμιότυπο του υπερώνυμους της, αντίθετα στη δεύτερη περίπτωση η κληρονομικότητα επιβάλλεται. Το επόμενο βήμα για να υπάρξει ακόμα μεγαλύτερη δυνατότητα συμπερασμού είναι να ενσωματωθούν τα στιγμιότυπα μέσα στην ήδη υπάρχουσα ιεραρχία (formal instance), σε αντίθεση με άλλες περιπτώσεις όπου οι ιεραρχίες κατηγοριοποιούν μόνο την ονομασία των κλάσεων.

Η διαφορά της με τους θησαυρούς είναι ότι πλέον επιβάλλεται η ιεραρχική δομή αλλά δεν συνηθίζεται να περιέχει σχέσεις ιδιοτήτων μεταξύ των όρων. Οι θησαυροί φαίνεται να διαθέτουν ένα πιο πλούσιο λεξιλόγιο για την περιγραφή των όρων, το οποίο αν συνδυαστεί με μια ιεραρχική δομή γίνεται ένα πάρα πολύ ισχυρό εργαλείο.

Παράδειγμα 4-5. Θησαυρός, Ταξινόμια

Πηγή: Uschold M (2006), διαφάνειες συνεδρίου

Διαθέσιμο: http://protege.stanford.edu/conference/2006/submissions/slides/1.2_Uschold.pdf



- Πλαίσια (Frames):** Αντιστοιχούν σε αντικείμενα που αναπαριστούν μία έννοια σε ένα πεδίο ενδιαφέροντος. Τα πλαίσια μοιράζονται πολλά κοινά χαρακτηριστικά με τους άλλους τρόπους αναπαράστασης γνώσης και δεδομένων και αποτελούν την αντικειμενοστραφή προσέγγιση στην αναπαράσταση γνώσης. Τεχνικά, αποτελούνται από την ονομασία τους και μια σειρά από ιδιότητες που συνδέονται άμεσα με τις τιμές τους. Ουσιαστικά, επιτρέπουν την ύπαρξη μιας ιεραρχικής δομής με όλες τις πληροφορίες για τις συγκεκριμένες έννοιες που αναπαριστούν ένα πεδίο ενδιαφέροντος. Παραδείγματα προτύπων πλαισίων είναι το XML, RDFs.

Παράδειγμα 4-6. Πλαίσια, Περιορισμοί τιμών



- Περιορισμοί Τιμών (Value Restrictions):** Όπως προδιαθέτει η ονομασία τους, είναι περιορισμοί στις τιμές και τους τύπους που μπορούν να λάβουν οι ιδιότητες των κλάσεων. Βάσει του παραπάνω παραδείγματος, θα μπορούσε να θεωρηθεί ότι το πεδίο τιμών της ιδιότητας *Τύπος υγρού* της κλάσης *Αντλία* είναι *(νερό, καύσιμα)*, ενώ ο τύπος των τιμών είναι αλφαριθμητικός. Εδώ επισημαίνεται ότι μπορεί στη δομή ταξινόμησης να εμπεριέχονται αυστηρές σχέσεις is-a ή υποκλάσεων.
- Περιορισμοί Γενικής Λογικής (General Logical Constraints):** Εδώ αυξάνεται κατά πολύ η ανάγκη για αυστηρότερη και πλουσιότερη αναπαράσταση της πληροφορίας. Έτσι χρησιμοποιούνται λογικές ή μαθηματικές εκφράσεις και περιορισμοί για να εμπλουτιστεί η δομή με εκφραστικότητα. Το πόσο αυστηρή, καλά ορισμένη θα είναι αλλά και οι δυνατότητες συμπερασμού που θα έχει μία δομή εξαρτάται πλέον από τις μεθόδους αναπαράστασης γνώσης.

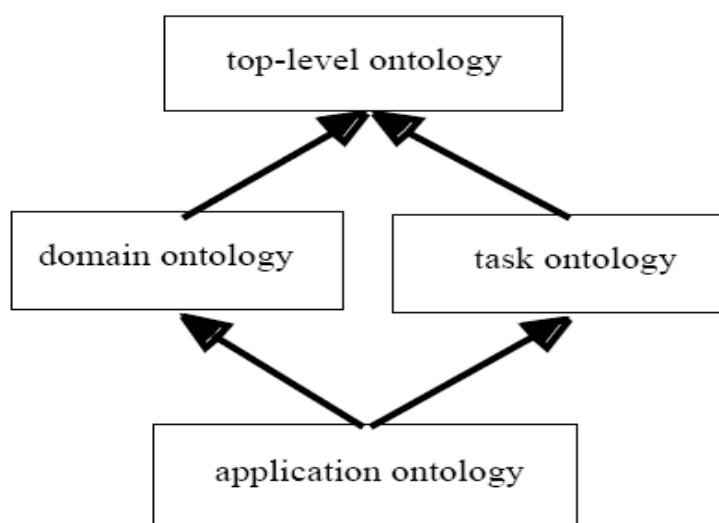
Παραδείγματα αυτών είναι οι Περιορισμοί Λογικής Πρώτης Τάξης που μέσα σε αυτούς εμπεριέχονται και πιο λεπτομερείς σχέσεις όπως disjoint, inverse, part-

whole. Άλλη μέθοδος αναπαράστασης είναι η Περιγραφική Λογική όπου χρησιμοποιεί ένα περιορισμένο σύνολο της Λογικής Πρώτης Τάξης, αλλά έχει καλύτερη υπολογιστική δυνατότητα και ικανότητα εξαγωγής συμπερασμάτων.

Παράδειγμα 4-7. Περιορισμοί Γενικής Λογικής



Επίπεδα Γενικότητας της Σύλληψης. Ο Guarino (1998) παρουσιάζει ένα σύστημα ταξινόμησης των Οντολογιών με βάση το επίπεδο γενικότητας των αντικειμένων που διαπραγματεύονται. Έτσι διακρίνει τρία βασικά επίπεδα:



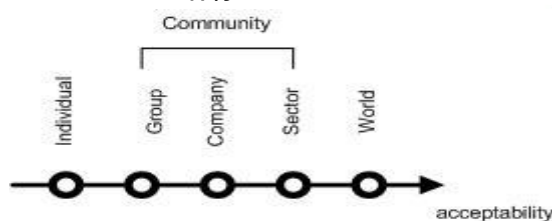
Εικόνα 4-2. Ταξινόμηση Οντολογιών με βάση το επίπεδο γενικότητας τους
 Πηγή: Guarino N (1998:7)

- **Οντολογίες Υψηλού Επιπέδου (Top Level Ontology ή Upper Ontology ή Foundation Ontology):** Είναι Οντολογίες οι οποίες περιγράφουν πολύ γενικές έννοιες και είναι κοινές για όλα τα πεδία γνώσης. Τέτοιες μπορεί να είναι ο χρόνος, χώρος, γεγονότα, ενέργειες, αντικείμενα αλλά και άλλες αφηρημένες έννοιες οι οποίες είναι ανεξάρτητες του προβλήματος ή του πεδίου εφαρμογής.
- **Οντολογίες Πεδίου (Domain Ontology ή Domain Specific Ontology) και Οντολογίες Διεργασιών (Task Ontology):** Περιγράφουν, αντίστοιχα, το λεξιλόγιο ή τις διεργασίες και τις δραστηριότητες μέσα σε ένα συγκεκριμένο τομέα, οι οποίες ουσιαστικά είναι μια εξειδίκευση των εννοιών από τις Οντολογίες Υψηλότερου Επιπέδου. Πολλές φορές ανάμεσα στις Οντολογίες Υψηλού Επιπέδου και τις Οντολογίες Πεδίου υπάρχει ένα επιπλέον υποστηρικτικό επίπεδο αυτό των επονομαζόμενων Οντολογιών Μεσαίου Επιπέδου (Mid-Level Ontology) οι οποίες προσπαθούν να γεφυρώσουν τις άλλες δύο. Ένα τέτοιο παράδειγμα είναι η Οντολογία Υψηλού επιπέδου SUMO (2011)

και η αντίστοιχη Οντολογία Μεσαίου επιπέδου Mid-Level Ontology (MILO) που γεφυρώνει την SUMO με Οντολογίες Πεδίων σχετικά με την επικοινωνία, τα οικονομικά, τη γεωγραφία και ούτω καθεξής.

- **Οντολογίες Εφαρμογών (Application Ontology):** Περιγράφουν έννοιες που εξαρτώνται τόσο από ένα συγκεκριμένο πεδίο αλλά και από συγκεκριμένες διαδικασίες, που ουσιαστικά είναι μια ειδικότερη περιγραφή των άλλων δύο σχετικών Οντολογιών. Εδώ το βασικό στοιχείο είναι οι ρόλοι, οι οποίοι αντιστοιχούν σε έννοιες του πεδίου όταν εκτελείται μια συγκεκριμένη διαδικασία.

Βάση Αποδοχής. Οι Schaffert και συν. (2005) εισάγουν ακόμα ένα μοντέλο με το οποίο μπορούν οι Οντολογίες να κατηγοριοποιηθούν, αυτό της αποδοχής. Αυτό που προσπαθούν να επιτύχουν είναι να ταξινομήσουν τις Οντολογίες με βάση τις ομάδες στόχου που θέλουν να ικανοποιήσουν. Το κριτήριο στο οποίο βασίζονται είναι ότι η επιτυχία μιας εφαρμογής συνάδει πάντα με την αποδοχή της από τις ομάδες στόχου που προσπαθεί να προσεγγίσει. Έτσι διακρίνουν τρία βασικά επίπεδα αποδοχής:

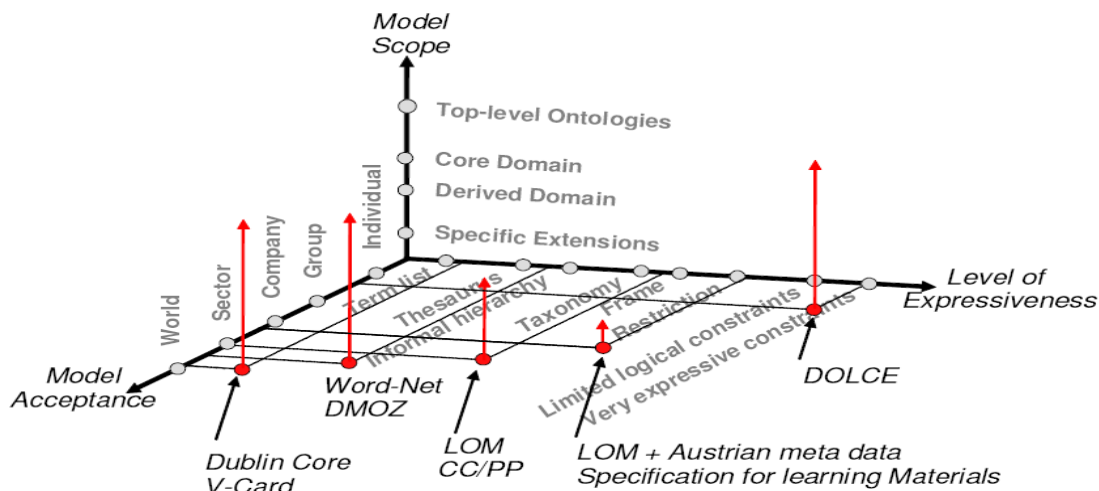


Εικόνα 4-3. Βαθμός Αποδοχής Οντολογίας
 Πηγή: Schaffert & συν. (2005:9)

- **Μεμονωμένη αποδοχή (individual):** Εδώ η ανάπτυξη και η χρήση των Οντολογιών γίνεται για να καλύψει μεμονωμένες περιπτώσεις χρήσης ή ανάγκες, ενώ υπάρχει ένας μόνο αποδέκτης.
- **Κοινότητα (Community):** Καθώς οι στόχοι είναι πιο περίπλοκοι οι Οντολογίες θα πρέπει να επιτρέπουν αυξημένη δυνατότητα συνεργασίας μεταξύ διαφορετικών ομάδων και μοντέλων, ενώ όλες οι διαδικασίες θα πρέπει να υποστηρίζονται από υψηλή τυποποίηση.
- **Κόσμος (World):** Εδώ οι ανάγκες που πρέπει να καλυφθούν είναι αυτές της κοινότητας αλλά και των μηχανισμών της αγοράς.

Τρισδιάστατη Κατηγοριοποίηση Οντολογιών. Όμως το πιο ενδιαφέρον στην εργασία των Schaffert και συν. (2005) είναι οι τρισδιάστατη κατηγοριοποίηση των Οντολογιών που προτείνουν και συνδυάζουν τους τρεις τύπους που περιγράφηκαν παραπάνω. Οι τρεις διαστάσεις είναι ο σκοπός (scope), δηλαδή το πεδίο που προσπαθούν οι Οντολογίες να καλύψουν, η εκφραστικότητα, δηλαδή ο βαθμός της τυπικότητάς της και τέλος η αποδοχή, δηλαδή ποια είναι ουσιαστικά η ομάδα στόχος. Στην παρακάτω εικόνα παρουσιάζεται ένα παράδειγμα ταξινόμησης Οντολογιών με βάση αυτήν την τρισδιάστατη προσέγγιση.

Παράδειγμα 4-8 Τρισδιάστατη κατηγοριοποίηση Οντολογιών
 Πηγή: Schaffert & συν. (2005:9)



Γίνεται αντιληπτό επομένως ότι κάτω από τον όρο Οντολογία μπορεί να εισαχθεί ένα μεγάλο σύνολο διαφορετικών τυπικών και ρητών προδιαγραφών που επιδιώκουν να αναπαραστήσουν μία ειδικευμένη ή αφηρημένη αντίληψη, για να ικανοποιήσουν τις ανάγκες μιας συγκεκριμένης κοινότητας. Έτσι καθορίζουν ένα συγκεκριμένο λεξιλόγιο με τόση εκφραστικότητα όσο χρειάζεται για να μπορούν οι μονάδες μιας κοινότητας να αλληλεπιδρούν, να αλληλοκατανοούνται, να επαναχρησιμοποιούν την γνώση τους πάνω σε μια σύλληψη και να τη διαμοιράζονται.

5. Αξιοποίηση Οντολογιών

Το πιο κλασικό παράδειγμα χρήσης των Οντολογιών είναι ο Σημασιολογικός Ιστός, που όχι μόνο τις αξιοποιεί αλλά τις οδηγεί στο να βρίσκονται σε ένα κοινό δρόμο συνεχούς ανάπτυξης και αλληλεπίδρασης. Αυτός είναι ο κυριότερος λόγος για τον οποίο οι Οντολογίες, παρόλο που δεν είναι κάτι καινούριο, εξακολουθούν να απασχολούν την ερευνητική κοινότητα ξανά και ξανά σε διάφορα επιστημονικά πεδία όπως στην DM. Μέσα από αυτόν το νέο τύπο διαδικτύου αναπτύσσονται αθόρυβα τα πρότυπα που θα απαρτίζουν τα επίπεδα του σημασιολογικού ιστού αλλά και της αναπαράστασης γνώσης. Αυτά τα επίπεδα δεν είναι τίποτε άλλο από διαφορετικές δομές αναπαράστασης γνώσης. Όσο υψηλότερα αναφέρεται κάποιος σε αυτά τα επίπεδα, τόσο μεγαλύτερη εκφραστικότητα αλλά και πολυπλοκότητα παρουσιάζουν. Δεν είναι καθόλου τυχαίο που πλέον οι πιο δυνατές και χρησιμοποιούμενες γλώσσες αναπαράστασης γνώσης, όπως η OWL, αναφέρονται ως γλώσσες του σημασιολογικού ιστού. Σε αυτή την ενότητα θα εξεταστούν τα επίπεδα του σημασιολογικού ιστού και θα υπενθυμίζονται οι ανάγκες που προκάλεσαν τη σταδιακή εξέλιξη των γλωσσών.

Μία από τις αφορμές για την εξέλιξη του διαδικτύου υπήρξε το πρόβλημα επιλογής πληροφοριών. Το διαδίκτυο μέσα σε λίγα χρόνια υποστηρίχθηκε από ένα τεράστιο όγκο δεδομένων όπου είτε ήταν δύσκολο να αναζητηθεί χρήσιμη πληροφορία είτε ήταν αδύνατον να επεξεργασθούν και να διαβιβαστούν οι πληροφορίες. Έτσι θα έπρεπε να δημιουργηθούν οι κατάλληλες δομές ώστε οι άνθρωποι αλλά και τα υπολογιστικά συστήματα να μπορούν να θέτουν σημασιολογικά ερωτήματα σε δεδομένα με σημασιολογική υπόσταση. Δίνοντας σημασιολογική υφή στα υπάρχοντα δεδομένα, οι υπολογιστικές μηχανές μπορούν να ξεπεράσουν τα εμπόδια της ετερογένειας και να εξορύξουν χρήσιμη γνώση με βάση τις πραγματικές ανάγκες.



Εικόνα 5-1. Επίπεδα σημασιολογικού ιστού

Το πρώτο βήμα ήταν να προσδιοριστεί μια κοινή αλφάβητος και ένα κοινό λεξιλόγιο. Η κοινή αλφάβητος επιτυγχάνεται με την κωδικοποίηση των αλφαριθμητικών, όπως προσπαθεί να κάνει το διεθνές πρότυπο Unicode, το οποίο στοχεύει να κωδικοποιήσει όλα τα συστήματα γραφής που χρησιμοποιούνται, ώστε να γίνει δυνατή η αποθήκευση στη μνήμη ενός υπολογιστικού συστήματος τα κείμενα οποιασδήποτε γλώσσας συμπεριλαμβανομένου

και επιστημονικών συμβόλων. Το κοινό λεξιλόγιο επιτυγχάνεται με την χρήση παγκόσμιων προσδιοριστών URI ή IRI (Internationalized Resource Identifiers), οι οποίοι προσφέρουν ένα παγκόσμιο μοναδικό μηχανισμό ονοματολογίας που βοηθάει στην αποφυγή του προβλήματος της συνωνυμίας, αλλά και στην κατανομημένη αναπαράσταση των δεδομένων. Παράδειγμα των URI είναι οι διευθύνσεις URL (Universal Resource Locator) στο διαδικτυακό ιστό.

Στο επόμενο επίπεδο αντιστοιχούν οι βασικές δομές αναπαράστασης περιεχομένου. Η πιο απλή και αντιπροσωπευτική γλώσσα είναι η HTML, η οποία ως μια γλώσσα σήμανσης επιτρέπει την εισαγωγή περιεχομένου και την παροχή πληροφοριών σχετικά με το ρόλο του περιεχομένου. Βέβαια η απλότητά της δεν της επιτρέπει να διατηρεί ρητή πληροφορία και σχέσεις που διέπουν τα περιεχόμενα, ενώ ενδιαφέρεται περισσότερο για την μορφοποίηση τους. Για να καλυφθεί το τελευταίο κενό υπήρξε η ανάπτυξη πιο δομημένων γλωσσών όπως η XML, η οποία αναφέρεται και ως μεταγλώσσα σήμανσης καθώς δεν απαιτείται η χρήση ενός προκαθορισμένου λεξιλογίου όπως στην HTML αλλά μπορεί να το ορίσει μέσω της XMLs. Αυτό της επιτρέπει να περιγράφει κάθε τμήμα της πληροφορίας αλλά και τις σχέσεις που υφίστανται μεταξύ τους, ενώ μπορεί πλέον να καθορίζει περιορισμούς που θα πρέπει να ακολουθούνται. Ο συνδυασμός της XML και της HTML δίνουν την γλώσσα XHTML που επιτρέπει στην HTML να είναι επεκτάσιμη και περισσότερο διαλειτουργική, ενώ δεν ενδιαφέρεται μόνο για τη μορφή των περιεχομένων αλλά και τη δομή τους, επιτρέποντας την ανάκτηση περισσότερων πληροφοριών.

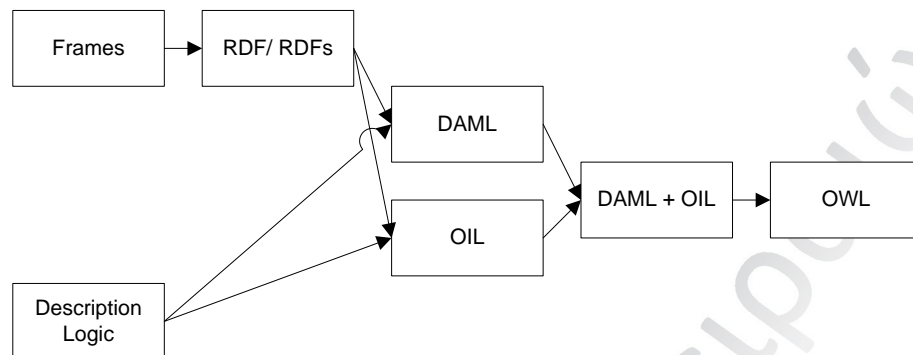
Παρόλο που η XML δίνει ένα σαφή ορισμό της δομής των δεδομένων, αδυνατεί να προσδιορίσει ένα τρόπο για την απόδοση νοήματος σε αυτά, εισάγοντας για παράδειγμα την έννοια της ιδιότητας ενός αντικειμένου. Αυτή την επιπλέον λειτουργία προσπαθεί να την καλύψει το επίπεδο αναπαράστασης νοήματος, με γλώσσες όπως η RDF. Σε αυτή ορίζεται μια βασική δομή *αντικείμενο- χαρακτηριστικό- τιμή*, η οποία καλείται *δήλωση*, και προσπαθεί να καθορίσει τις ιδιότητες που ισχύουν για συγκεκριμένα αντικείμενα και τις τιμές που μπορούν να πάρουν, γι αυτό κατά κόρον αναφέρεται και ως μοντέλο δεδομένων (data-model) αντί για γλώσσα. Υποστηρικτικό της εργαλείο είναι η RDFs η οποία καθορίζει το λεξιλόγιό της και τις σχέσεις που μπορούν να υπάρξουν, σε αντίθεση με την XMLs που ορίζει κυρίως την δομή των XML αρχείων. Οι βασικές έννοιες που μπορούν να αναπαρασταθούν με τη νέα δομή είναι οι κλάσεις, οι σχέσεις υποκλάσεων, ιδιότητες, σχέσεις υπο-ιδιοτήτων και περιορισμού του εύρους- περιοχής.

Το επίπεδο της αναπαράστασης νοήματος μπορεί να διαθέσει καλές γλώσσες μοντελοποίησης, αλλά όχι τόσο ισχυρές σε εκφραστικότητα και συλλογιστική ικανότητα. Έχει μεν μια επαρκή περιγραφική δύναμη αλλά όχι αρκετή για το ρητό και τυπικό ορισμό μιας αντίληψης ενός μοντέλου περιοχής. Έτσι πάνω σε αυτές βασίζονται πλουσιότερες γλώσσες με επιπρόσθετες ιδιότητες και περιορισμούς. Ενδεικτικές δυνατότητες που επιτρέπονται σε αυτό το επίπεδο σε σχέση με την εκφραστικότητα είναι η τυπική εμβέλεια ιδιοτήτων, ο ορισμός ξένων κλάσεων, ο συνδυασμός λογικών μεταβλητών μεταξύ κλάσεων, οι περιορισμοί μεγέθους πεδίου τιμών ιδιοτήτων. Σε σχέση με τις συλλογιστικές ικανότητες που επιτρέπονται λόγω των νέων ειδικών χαρακτηριστικών των ιδιοτήτων όπως της μεταβατικότητας, της μοναδικότητας, της αντιστροφής, είναι η εξαγωγή συμπερασμάτων για ισοδυναμία κλάσεων, για την συνέπεια και την κατηγοριοποίηση τους. Παραδείγματα γλωσσών που ανήκουν σε αυτό το επίπεδο και εμφανίζονται ως πρότυπα από την W3C είναι η OWL, η οποία χρησιμοποιείται για την περιγραφή και ανταλλαγή οντολογιών, όπως και η RIF η οποία είναι ένα πρότυπο για την ανταλλαγή κανόνων μέσα στο Διαδικτυακό Ιστό.

Το επόμενο επίπεδο προσπαθεί να ενοποιήσει τα λογικά μοντέλα και μέσα από πιο πολύπλοκους συσχετισμούς να προσφέρει δυνατότερη συλλογιστική ικανότητα, αλλά και να επιτρέψει την απεικόνιση διαφορετικών δομών μέσω των κανόνων. Στο προηγούμενο θα βασιστεί το επίπεδο απόδειξης το οποίο το χρησιμοποιεί για την εξαγωγή συμπερασμάτων τόσο φιλικά προς τους χρήστες, είτε μέσω φυσική γλώσσας είτε μέσω οπτικοποίησης, όσο και της εύκολης κατανόησης, μέσω λογικών διαδοχικών βημάτων, το οποίο θα προσδίδει στη διαδικασία μεγάλη εμπιστοσύνη. Το τελευταίο επίπεδο είναι αυτό της αξιοπιστίας, το οποίο διαπραγματεύεται τα θέματα δικαιωμάτων, όπως για το ποιος και πότε θα πρέπει να έχει πρόσβαση στα δεδομένα- γνώση. Η απόδειξη της αξιοπιστίας, θα ενθαρρύνει την εχεμύθεια μεταξύ συστημάτων και επιτυγχάνεται μέσω κανόνων που αξιολογούν τους λόγους πρόσβασης σε σχέση με τα περιεχόμενα των δεδομένων.

5.1 OWL 1 και OWL 2

Η OWL είναι ένα από τα πρόσφατα αλλά όχι και το τελευταίο επίτευγμα μέσα στο γενικότερο πεδίο αναπαράστασης της γνώσης. Αν ήθελε κάποιος να εξερευνήσει τις ρίζες της θα έπρεπε να γυρίσει στις δεκαετίες του 70^α και 80^α κατά τις οποίες εξελίσσονται οι μέθοδοι αναπαράστασης των πλαισίων και της περιγραφικής λογικής. Βασιζόμενοι στα πλαίσια στις επόμενες δεκαετίες θα εξελιχθούν οι βασικές γλώσσες RDF/RDFs πάνω στις οποίες, και συνδυαστικά με την περιγραφική λογική, θα δημιουργηθούν δύο ανεξάρτητες γλώσσες, οι DALM και OIL οι οποίες θα ενοποιηθούν στην DALM+OIL και θα μετεξελιχθούν τελικά στην OWL, όπου με την σειρά της επεκτείνεται από το 2009 στην OWL 2.

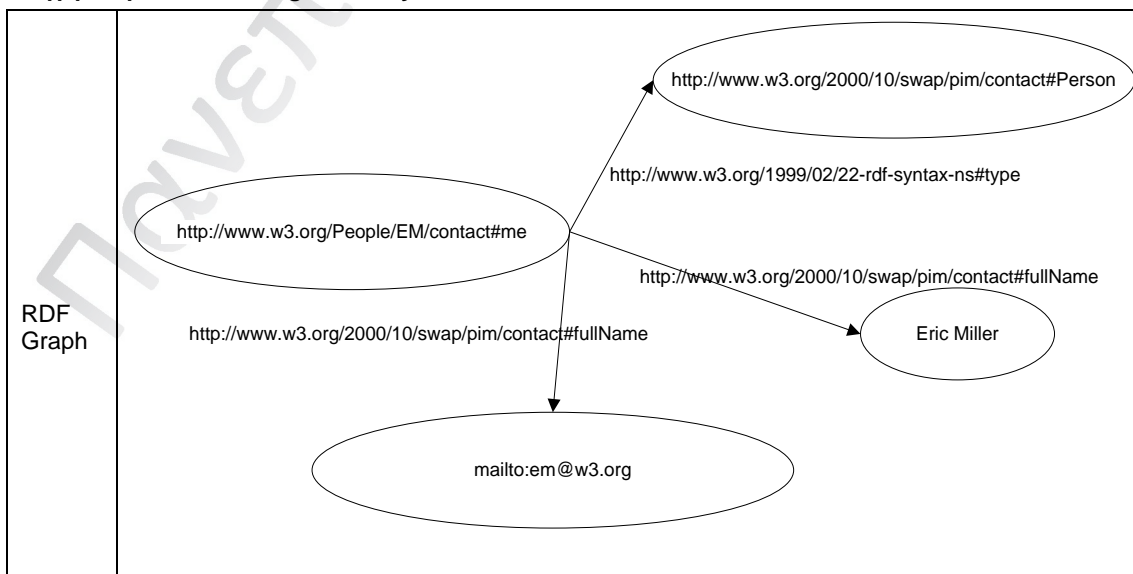


Εικόνα 5-2. Πρόγονοι τις OWL

Χαρακτηριστικά της OWL. Η Web Ontology Language (OWL) όπως προδιαθέτει και η ονομασία της είναι μια γλώσσα του Σημασιολογικού Ιστού σχεδιασμένη να αναπαριστά πλούσια και πολύπλοκη γνώση σχετικά με πράγματα, ομάδες πραγμάτων και τις μεταξύ τους σχέσεις. Ουσιαστικά η OWL είναι μια υπολογιστική γλώσσα βασισμένη στη λογική. Πάνω στη γνώση που διαθέτει μπορεί να εκτελεστούν διαδικασίες συλλογιστικής από υπολογιστικές μηχανές είτε για να ελεγχθεί η συνέπεια της γνώσης είτε να κάνει ρητή την υπονοούμενη γνώση. Τα OWL κείμενα είναι γνωστά ως οντολογίες και μπορούν να δημοσιευθούν στον παγκόσμιο ιστό. Επίσης η OWL αποτελεί πρότυπη και μέρος των τεχνολογιών της W3C’s Semantic Web (Hitzler & συν., 2009).

Κατά κανόνα μια OWL οντολογία είναι ένα RDF γράφημα, όπου με τη σειρά του αποτελεί ένα σύνολο RDF τριάδων. Όπως κάθε γράφημα RDF, έτσι και ένα γράφημα OWL οντολογίας μπορεί να γραφεί σε πολλές διαφορετικές συντακτικές μορφές όπως XML Syntax. Ένα απλό παράδειγμα αυτών των δύο μορφών παρουσιάζεται στον ακόλουθο πίνακα. Σε αυτό πόροι συνδυάζονται με άλλους πόρους ή με δηλώσεις μέσω συγκεκριμένων ιδιοτήτων.

Παράδειγμα 5-1. RDF graph, RDF/XML Syntax
 Πηγή: <http://www.w3.org/TR/rdf-syntax/>



RDF/ XML	<pre><?xml version="1.0"?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#"> <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me"> <contact:fullName>Eric Miller</contact:fullName> <contact:mailbox rdf:resource="mailto:em@w3.org"/> <contact:personalTitle>Dr.</contact:personalTitle> </contact:Person> </rdf:RDF></pre>
-------------	--

Η OWL για την αναπαράσταση μίας σύλληψης χρησιμοποιεί τρεις διαφορετικές συντακτικές κατηγορίες. Η πρώτη κατηγορία είναι οι *έννοιες* (entities), δηλαδή στοιχεία που χρησιμοποιούνται για να γίνει αναφορά στα αντικείμενα του πραγματικού κόσμου. Τέτοια στοιχεία μπορεί να είναι οι κλάσεις, οι ιδιότητες και τα άτομα. Η δεύτερη κατηγορία είναι οι *εκφράσεις* (expressions), σύνθετες έννοιες για την περιγραφή του πεδίου. Πολλές φορές συναντώνται στην βιβλιογραφία ως περιγραφές ή σύνθετες έννοιες. Κλάσεις και εκφράσεις ιδιοτήτων χρησιμοποιούνται για την κατασκευή εκφράσεων κλάσεων. Ουσιαστικά οι εκφράσεις κλάσεων περιγράφουν το σύνολο των ατόμων, με τυπικές προσδιοριστικές συνθήκες, που βασίζονται στις ιδιότητες των συγκεκριμένων ατόμων. Δηλαδή, τα άτομα που ικανοποιούν τις συγκεκριμένες συνθήκες, θεωρούνται ότι είναι περιπτώσεις των αντίστοιχων εκφράσεων κλάσης. Παραδείγματα αυτών είναι οι προτασιακές συνδέσεις, η απαρίθμηση ατόμων, οι περιορισμοί αντικειμένων ιδιοτήτων. Η τελευταία κατηγορία είναι τα *αξιώματα* (axioms) που αποτελούν τις δηλώσεις που βεβαιώνουν για το τι αληθεύει μέσα στο πεδίο περιγραφής. Παραδείγματα αυτών είναι οι δηλώσεις των υποκλάσεων, η ισοδυναμία κλάσεων και ιδιοτήτων.

Όπως έχει αναφερθεί ένα από τα σημαντικότερα χαρακτηριστικά μίας οντολογίας είναι οι κλάσεις. Στην OWL οι κλάσεις περιγράφονται μέσω των *περιγραφών των τάξεων* οι οποίες μπορούν να συνδυαστούν με *αξιώματα τάξεως*. Οι περιγραφές των τάξεων μπορούν να γίνουν μέσω έξι τύπων:

- Έναν προσδιοριστή της τάξης, μέσω μίας URI αναφοράς (για την OWL 1) η οποία καθορίζει το όνομα της κλάσης. Βέβαια, μία κλάση μπορεί να παραμείνει ανώνυμη και θα πρέπει να περιγραφεί μέσω των παρακάτω τύπων. Η OWL ορίζει και δύο προκαθορισμένες κλάσεις την owl:Thing και owl:Nothing οι οποίες είναι και οι γενικότερες τάξεις που μπορούν να οριστούν. Γενικό παράδειγμα αυτής της κατηγορίας μπορεί να είναι το ακόλουθο: <owl:Class rdf:ID="Human"/>.
- Μία εξαντλητική *απαρίθμηση* (enumeration) των ατόμων που αποτελούν τα στιγμιότυπα των κλάσεων. Όπως για παράδειγμα από τις ηπείρους που αποτελείται η γη:

Παράδειγμα 5-2. OWL απαρίθμηση

```
<owl:Class>
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing rdf:about="#Eurasia"/>
    <owl:Thing rdf:about="#Africa"/>
    <owl:Thing rdf:about="#NorthAmerica"/>
    <owl:Thing rdf:about="#SouthAmerica"/>
    <owl:Thing rdf:about="#Australia"/>
    <owl:Thing rdf:about="#Antarctica"/>
  </owl:oneOf>
</owl:Class>
```

- *Περιορισμούς ιδιοτήτων* (Property Restrictions), οι οποίοι περιγράφουν μία ανώνυμη κλάση ως μία κλάση από όλα εκείνα τα άτομα που ικανοποιούν συγκεκριμένους περιορισμούς. Υπάρχουν δύο κυρίως τύποι περιορισμών ιδιοτήτων:

Πίνακας 5-1. Περιορισμοί ιδιοτήτων

Κατηγορία	Περιγραφή	Κατασκευαστής	Περιγραφή
Περιορισμοί Τιμών (Value Constraint)	Περιορισμοί στο πεδίο τιμών	owl:allValuesFrom	προσδιορίζει μια κλάση ατόμων x για την οποία υποστηρίζει ότι, αν το ζεύγος (x, y) είναι ένα στιγμιότυπο της P, τότε το y πρέπει να είναι ένα στιγμιότυπο της κλάσης περιγραφής ή μίας τιμής στην περιοχή δεδομένων
		owl:someValuesFrom	προσδιορίζει μια κλάση ατόμων x για τις οποίες υπάρχει τουλάχιστον ένα y ώστε το ζεύγος (x, y) είναι στιγμιότυπο του P.
		owl:hasValue	περιγράφει μια τάξη όλων των ατόμων για τα οποία η εν λόγω ιδιότητα έχει τουλάχιστον μια τιμή σημασιολογικά ίση με V. Όπου V μία συγκεκριμένη τιμή.
Περιορισμοί Ποσοτικοποίησης	Περιορίζει την τιμή της αξίας της ιδιότητας	owl:maxCardinality	περιγράφει μια κλάση όλων των ατόμων που έχουν το πολύ N σημασιολογικά διακριτές τιμές για την εν λόγω ιδιότητα, όπου N είναι η αξία των περιορισμών ποσοτικοποίησης
		owl:minCardinality	περιγράφει μια κλάση όλων των ατόμων που έχουν τουλάχιστον N σημασιολογικά διακριτές τιμές για την εν λόγω ιδιότητα, όπου N είναι η αξία των περιορισμών ποσοτικοποίησης
		owl:cardinality	περιγράφει μια κλάση όλων των ατόμων που έχουν ακριβώς N σημασιολογικά διακριτές τιμές για την εν λόγω ιδιότητα, όπου N είναι η αξία των περιορισμών ποσοτικοποίησης

Οι περιορισμοί αυτοί ορίζονται μέσα στην κλάση owl: Restriction που ουσιαστικά είναι υποκλάση της owl:Class και συνδέονται με την ιδιότητα owl:onProperty όπως φαίνεται παρακάτω.

Παράδειγμα 5-3. Ορισμός περιορισμού ιδιότητας

```
<owl:Restriction>
  <owl:onProperty rdf:resource="#hasParent" />
  <owl:allValuesFrom rdf:resource="#Human" />
</owl:Restriction>
```

- *Τομή (Intersection,)* *Ένωση (Union)* και *Συμπλήρωμα (Complement)*, θεωρούνται ότι αντιπροσωπεύουν την πιο προηγμένη κατηγορία κατασκευαστών στην περιγραφική λογική και αντιπροσωπεύουν τους τελεστές ΚΑΙ, Ή και ΟΧΙ. Αυτά στην OWL ορίζονται ως:

Πίνακας 5-2. OWL τομή, ένωση, συμπλήρωμα

Κατηγορία	Κατασκευαστής	Περιγραφή
Τομή	owl:intersectionOf	περιγράφει μια κλάση που αποτελείται από τα κοινά στοιχεία των συμπεριλαμβανόμενων κλάσεων
Ένωση	owl:unionOf	περιγράφει μια κλάση που αποτελείται από την ένωση των στοιχείων των συμπεριλαμβανόμενων κλάσεων
Συμπλήρωμα	owl:complementOf	περιγράφει μια κλάση που αποτελείται από το συμπλήρωμα των στοιχείων των συμπεριλαμβανόμενων κλάσεων

Τα αξιώματα κλάσεων χρησιμοποιούνται κατά κόρον ως δομικά στοιχεία στις περιγραφές των κλάσεων, παρόλα αυτά συνήθως περιέχουν πρόσθετα συστατικά που είναι απαραίτητα και/ ή επαρκή χαρακτηριστικά μίας κλάσης. Η OWL περιλαμβάνει τρεις κατασκευαστές για να συνδυάζει τις περιγραφές κλάσεων μέσα στα αξιώματα κλάσεων:

- *dfs: subclassOf*, όπου επιτρέπει να οριστεί ότι μία περιγραφή κλάσης είναι υποσύνολο μίας άλλης περιγραφής κλάσης.
- *owl:equivalentClass*, όπου επιτρέπει να οριστεί ότι μία περιγραφή κλάσης είναι ισοδύναμη με μία άλλη περιγραφή κλάσης
- *owl:disjointWith*, όπου επιτρέπει να οριστεί ότι μία περιγραφή κλάσης είναι ξένη με μία άλλη περιγραφή κλάσης.

Το αμέσως επόμενο ενδιαφέρον χαρακτηριστικό των οντολογιών και επομένως και της OWL είναι οι σχέσεις- ιδιότητες που σχηματίζονται μεταξύ των εννοιών- κλάσεων. Η OWL διακρίνει δύο βασικές κατηγορίες ιδιοτήτων, τις οποίες ένας κατασκευαστής οντολογιών θα θελήσει να ορίσει. Αυτές είναι οι *ιδιότητες αντικειμένων* (object properties), που ορίζονται με την κλάση *owl:objectProperty* και τις *ιδιότητες τύπων τιμών* (datatype properties) που ορίζονται με την κλάση *owl:DatatypeProperty*. Και οι δύο κλάσεις που αναφέρθηκαν είναι υποκλάσεις της RDF κλάσης *rdf:Property*. Φυσικά και εδώ ορίζονται αξιώματα τα οποία εμπλουτίζουν τα χαρακτηριστικά των ιδιοτήτων και ονομάζονται *αξιώματα ιδιοτήτων* (property axioms). Σε αυτά υπάρχουν τέσσερις βασικές κατηγορίες κατασκευαστών:

Πίνακας 5-3. Αξιώματα ιδιοτήτων

Κατηγορία	Περιγραφή	Κατασκευ.	Περιγραφή
RDF Schema	Χαρακτηριστικά από το RDF σχήμα	<i>rdfs:subPropertyOf</i>	Ορίζει ότι μία ιδιότητα είναι υπο-ιδιότητα μίας άλλης. Ορίζεται και για ιδιότητες αντικειμένων και τύπων τιμών
		<i>rdfs:domain</i>	Συνδέει μια ιδιότητα, δηλαδή κάποια στιγμιότυπα της κλάσης <i>rdf:Property</i> , με μία περιγραφή κλάσης. Έτσι όταν ορίζεται σε μία κλάση υποδηλώνει ότι τα στιγμιότυπά της ανήκουν στα στιγμιότυπα της αναφερόμενης κλάσης περιγραφής.
		<i>rdfs:range</i>	Συνδέει μια ιδιότητα, δηλαδή κάποια στιγμιότυπα της κλάσης <i>rdf:Property</i> , με μία περιγραφή κλάσης ή ένα πεδίο τιμών. Έτσι υποδηλώνει ότι οι τιμές των ιδιοτήτων ανήκουν σε στιγμιότυπα μίας κλάσης περιγραφής ή σε τιμές δεδομένων μέσα σε ένα καθορισμένο πεδίο τιμών.
Σχέσεις με άλλες ιδιότητες	Ορίζονται μεταξύ ιδιοτήτων	<i>owl:equivalentProperty</i>	Δηλώνει ότι δύο ιδιότητες έχουν το ίδιο σύνολο στιγμιότυπων. Ορίζεται και για ιδιότητες αντικειμένων και τύπων τιμών
		<i>owl:inverseOf</i>	Εκφράζεται ως $P1 \text{ owl:inverseOf } P2$ και δηλώνει ότι για κάθε ζευγάρι (x,y) στιγμιότυπο της $P1$, υπάρχει ένα ζευγάρι (y,x) στιγμιότυπο της $P2$ και αντίστροφα. Ουσιαστικά είναι μία συμμετρική ιδιότητα.
Γενικοί περιορισμοί Ποσοτικοποίησης	Μπορούν να οριστούν σε οποιαδήποτε κλάση και είναι τελείως διαφορετικής λογικής από τους περιορισμούς ποσοτικοποίησης των περιορισμών ιδιοτήτων στην περιγραφή κλάσεων, που εκεί ορίζονται πάνω στις ιδιότητες που εφαρμόζονται για την περιγραφή μίας τάξης.	<i>owl:FunctionalProperty</i>	Ορίζει ότι ένα στιγμιότυπο x έχει μία μοναδική τιμή y . Βέβαια μία τιμή y μπορεί να αντιστοιχεί σε περισσότερα στιγμιότυπα x_1, x_2 . Επίσης, μπορεί να περιγράψει τις ιδιότητες αντικειμένων και πεδίου τιμών
		<i>owl:InverseFunctionalProperty</i>	Ορίζει ότι δύο στιγμιότυπα x_1, x_2 δεν μπορούν να έχουν την ίδια τιμή y . Ουσιαστικά είναι μία 1-1 σχέση

Κατηγορία	Περιγραφή	Κατασκευ.	Περιγραφή
Λογικά χαρακτηριστικά ιδιοτήτων	Πολύ χρήσιμο εργαλείο για την εξαγωγή συμπερασμάτων	owl:SymmetricProperty	Ορίζει τη μεταβατικότητα μίας ιδιότητας. Δηλαδή αν για μία μεταβατική ιδιότητα P υπάρχουν τα ζευγάρια (x,y) και (y,z) τότε μπορεί να συμπεραθεί ότι το ζευγάρι (x,z) είναι στιγμιότυπο της P
		owl:TransitiveProperty	Ορίζει τη συμμετρικότητα μιας ιδιότητας. Δηλαδή αν ένα ζευγάρι (x, y) είναι στιγμιότυπο της ιδιότητας P τότε μπορεί να συμπεραθεί ότι και το (y, x) είναι στιγμιότυπο της P

Εκτός από τις κλάσεις σε μία οντολογία είναι σημαντικό να περιγράφονται και τα μέλη που την αποτελούν. Η OWL χαρακτηρίζει ως άτομα τα μέλη μίας κλάσης, τα οποία αποτελούν τα πράγματα μίας σύλληψης, γι αυτό και όλα τα άτομα μέσα σε ένα OWL έγγραφο είναι μέλη της κλάσης owl:Thing. Για τον ορισμό των ατόμων χρησιμοποιούνται τα αξιώματα ατόμων (individual axioms) ή διαφορετικά γεγονότα (facts). Γενικά υπάρχουν δύο τύποι γεγονότων:

- Γεγονότα με τη μορφή δηλώσεων που συσχετίζονται με τις ιδιότητες μίας κλάσης και τις συγκεκριμένες τιμές των ιδιοτήτων που λαμβάνουν για ένα άτομο. Για παράδειγμα μπορεί να οριστεί ένα άτομο ως στιγμιότυπο μίας κλάσης που λαμβάνει συγκεκριμένες τιμές για τις ιδιότητες που την περιγράφουν, όπως φαίνεται παρακάτω:

Παράδειγμα 5-4. Παράδειγμα δήλωσης ατόμων στην OWL

```
<Opera rdf:ID="Tosca">
  <hasComposer rdf:resource="#Giacomo_Puccini"/>
  <hasLibrettist rdf:resource="#Victorien_Sardou"/>
  <premiereDate rdf:datatype="&xsd:date">1900-01-14</premiereDate>
  <premierePlace rdf:resource="#Roma"/>
  <numberOfActs rdf:datatype="&xsd:positiveInteger">3</numberOfActs>
</Opera>
```

- Γεγονότα για την ταυτοποίηση των ατόμων. Είναι σύνηθες διαφορετικά ονόματα αναφοράς να ορίζονται για διαφορετικά πράγματα, αυτό όμως δε μπορεί να πραγματοποιηθεί στον παγκόσμιο ιστό, οπότε ούτε και στην OWL, καθώς ένα πράγμα μπορεί να προσδιοριστεί με πολλούς διαφορετικούς τρόπους μέσω URI. Γι' αυτό η OWL ορίζει τρεις βασικούς κατασκευαστές ως γεγονότα για τις σχέσεις μεταξύ ατόμων:

Πίνακας 5-4. Γεγονότα για την ταυτοποίηση των ατόμων

Κατασκευαστές	Περιγραφή
owl:sameAs	Ορίζει ότι δύο αναφορές URI αναφέρονται στο ίδιο άτομο
owl:differentFrom	Ορίζει ότι δύο αναφορές URI αναφέρονται σε διαφορετικό άτομο
owl:AllDifferent	Ορίζεται για μια λίστα από άτομα που είναι διαφορετικά μεταξύ τους

Επέκταση της OWL. Η επέκταση της OWL 1 στην OWL 2 έχει να κάνει με την προσθήκη νέων λειτουργιών, όπως την ένωση ξένων κλάσεων, αλλά και τον εμπλουτισμό της στο θέμα της εκφραστικότητας. Οι δύο εκδόσεις συνεχίζουν να καλύπτουν τις ίδιες ανάγκες του σημασιολογικού ιστού και να βασίζονται στα ίδια εργαλεία, την RDF/XML. Σημαντικό είναι ότι κάθε OWL 1 Οντολογία είναι και OWL 2, ακολουθώντας την ίδια σύνταξη και εξάγοντας τα ίδια συμπεράσματα, λόγω της προς τα πίσω συμβατότητας που έχει επιτευχθεί.

Με βάση τους Golbreich και Wallace (2009) έξι είναι οι βασικές διαφορές σε αυτές τις δύο εκδόσεις:

- Πλέον παρέχονται συντακτικές διευκολύνσεις, για να μπορούν να ορισθούν κάποιες δηλώσεις ευκολότερα, χωρίς να αλλάξει η εκφραστικότητα, η σημασιολογία και η πολυπλοκότητα της γλώσσας.

Πίνακας 5-5. Συντακτικές διευκολύνσεις OWL 2

Συντακτική διευκόλυνση	OWL 2	OWL 1
DisjointUnion	Ορίζει μία κλάση ως ένωση άλλων κλάσεων, οι οποίες ανά δύο είναι ξένες	Οριζόταν πολύπλοκα με συνδυασμό αρκετών αξιωμάτων
DisjointClasses	Ορίζεται για ένα σύνολο υποκλάσεων και τις θέτει ξένες.	Μέχρι τώρα οριζόταν για ένα ζευγάρι τάξεων και όχι σύνολο αυτών
NegativeObjectPropertyAssertion/ NegativeDataPropertyAssertion	Αναφέρει ότι μία ιδιότητα δεν αντιστοιχεί σε συγκεκριμένα στιγμιότυπα	Παρέχει μέσα για την βεβαίωση τιμών μιας ιδιότητας για ένα στιγμιότυπο, αλλά δεν διαθέτει για την άμεση επιβεβαίωση τιμών που δεν μπορεί να λάβει

- Νέοι κατασκευαστές που αυξάνουν την εκφραστικότητα. Στην OWL 1 παρατηρείται ότι έχει δοθεί βάρος κυρίως σε δομές σχετικά με την έκφραση των πληροφοριών που έχουν να κάνουν με τις κλάσεις και τα στιγμιότυπα, και παρουσιάζει μια αδυναμία όσον αφορά την εκφραστικότητα των ιδιοτήτων. Σε αντίθεση με την OWL 2 που προσφέρει νέες δομές όπως τους περιορισμούς στις ιδιότητες, νέα χαρακτηριστικά των ιδιοτήτων, την μη συμβατότητα των ιδιοτήτων, τις ιδιότητες αλυσίδας και τα κλειδιά.

Πίνακας 5-6. Δομές αυξημένης εκφραστικότητας

Κατηγορία	Ονομασία	OWL 2	OWL 1
Αυτοπεριορισμοί	ObjectHasSelf	Επιτρέπει σε μία ιδιότητα να συσχετίζει μία κλάση με τον εαυτό της. Γενικά οι αυτοπεριορισμοί αποτελούν μέρος της SROIQ, μια επέκταση της περιγραφικής λογικής OWL-DL που ακολουθεί SROIQ εκφραστικότητα.	Δεν επιτρέπει τον ορισμό κλάσεων που συσχετίζονται με τον εαυτό τους με χρήση μίας ιδιότητας
Περιορισμοί Ποσοτικοποίησης	ObjectMinCardinality/ DataMinCardinality	Επιτρέπει τον ορισμό ελάχιστων, μέγιστων και με ακρίβεια, ειδικών περιορισμών ποσοτικοποίησης σε αντικείμενα ή δεδομένων ιδιοτήτων. Οι ειδικοί αυτοί περιορισμοί παρουσιάζονται μέσα στην SROIQ. Π.χ. OWL 2: ένα άτομο έχει τρία παιδιά που είναι κορίτσα.	Επιτρέπει περιορισμούς σχετικά με τον αριθμό των στιγμιότυπων μίας ιδιότητας, αλλά δεν διαθέτει ένα μέσο για να προσδιορίζει την κλάση ή το πεδίο τιμών των στιγμιότυπων που καταμετρούνται. (Ειδικοί περιορισμοί ποσοτικοποίησης) Π.χ. OWL 1: ένα άτομο έχει τρία παιδιά.
	ObjectMaxCardinality/ DataMaxCardinality		
	ObjectExactCardinality/ DataExactCardinality		
Ανακλαστικές, μη-ανακλαστικές, ασύμμετρες ιδιότητες αντικειμένων	ReflexiveObjectProperty IrreflexiveObjectProperty AsymmetricObjectProperty	Υλοποιούνται ως μέρη της SROIQ και αντιστοιχούν στην ανακλαστική, μη ανακλαστική και ασύμμετρη ιδιότητα των σχέσεων	Αν και επιτρέπει σε ένα αντικείμενο ιδιότητας να είναι συμμετρικό ή μεταβατικό, δεν του επιτρέπει να είναι ανακλαστικό, μη-ανακλαστικό ή ασύμμετρο
Ξένες ιδιότητες	DisjointObjectProperties/ DisjointDataProperties	Με αυτό δύο άτομα δεν μπορούν να συνδεθούν δύο διαφορετικές ιδιότητες που ανήκουν σε ένα σύνολο ιδιοτήτων ξένων μεταξύ τους. Είναι μέρος της SROIQ	Αν και επιτρέπει να ορισθούν ξένες κλάσεις, δεν επιτρέπει τον ορισμό ξένων ιδιοτήτων
Ιδιότητες αλυσίδα	ObjectPropertyChain	Επιτρέπει μέσω της SubObjectPropertyOf να οριστεί ένα αξίωμα το οποίο θα ορίζει μία ιδιότητα ως σύνθεση διάφορων άλλων. Είναι μέρος της SROIQ, μέσα στην οποία αυτά τα αξιώματα αναφέρονται ως περίπλοκες περιλήψεις όρων.	Δεν παρέχει ένα μέσο για να καθορίσει τις ιδιότητες ως σύνθεση άλλων ιδιοτήτων
Κλειδιά	HasKey	Επιτρέπει τον ορισμό κλειδιών σε κλάσεις. Έτσι τα άτομα μιας συγκεκριμένης κλάσης μπορούν να προσδιορίζονται μοναδικά από την τιμή των ιδιοτήτων κλειδιών. Είναι μια μορφή της DL Safe rule (DL-Safe)	Δεν παρέχει ένα μέσο για τον καθορισμό κλειδιών

- Διευρύνει τις δυνατότητες των τύπων δεδομένων. Τους εμπλουτίζει με νέους ή εκμεταλλεύεται τους ήδη υπάρχοντες από το XMLs, ενώ παράλληλα ορίζει νέους περιορισμούς.

Πίνακας 5-7. Δομές για τύπους δεδομένων

Κατηγορία	Ονομασία	OWL 2	OWL 1
Επιπλέον τύποι δεδομένων και περιορισμοί τύπων	*Επόμενος πίνακας	Επιτρέπει α) διάφορους τύπους αριθμών μέσω της XMLs Datatype(real, double, float, decimal, positiveInteger) αλλά και δικούς της (real) β) αλφαριθμητικά σε συγκεκριμένη γλώσσα γ)boolean values, binary data, IRIs, time instants	Επιτρέπει μόνο ακέραιους και αλφαριθμητικά ως τύπους δεδομένων, ενώ δε μπορεί να οριστεί υποσύνολο αυτών των τύπων. Π.χ. ανήλικος πάνω από 18
	DatatypeRestriction	Επιτρέπει τον ορισμό περιορισμών στους διάφορους τύπους δεδομένων ορίζοντας μέγιστο/ ελάχιστο μήκος ή τιμή	-
N-αδικοί τύποι δεδομένων	-	Δεν υπάρχει ρητός προσδιορισμός ακόμα, αλλά δημιουργήθηκαν συντακτικές δομές που χρειάζονται για να οριστεί αυτή η σχέση	Δεν επιτρέπει τησχέση μεταξύ τιμών ενός αντικειμένου. Π.χ. Το τετράγωνο είναι ένα ορθογώνιο όπου οι πλευρές του είναι ίσες.
Ορισμό τύπων δεδομένων	DatatypeDefinition	Ορίζει νέους τύπους δεδομένων	Παρόλο που επιτρέπει τον ορισμό μίας κλάσης μέσω της περιγραφής μίας άλλης, δεν επιτρέπει τον άμεσο ορισμό ενός νέου τύπου δεδομένων
Συνδυασμό εύρος δεδομένων	DataIntersectionOf	Επιτρέπει τον συνδυασμό ευρέως δεδομένων χρησιμοποιώντας δομές όπως τομή, ένωση και συμπλήρωμα	Παρόλο που επιτρέπει τον ορισμό μίας κλάσης μέσω τον συνδυασμό άλλων, δεν επιτρέπει τον ορισμό νέων δομών δεδομένων με συνδυασμό άλλων
	DataUnionOf		
	DataComplementOf		

Πίνακας 5-8. Τύποι δεδομένων

Αριθμητικά δεδομένα		Αλφαριθμητικά	Άλλοι τύποι
owl:real owl:rational xsd:decimal xsd:integer xsd:nonNegativeInteger xsd:nonPositiveInteger xsd:positiveInteger xsd:negativeInteger xsd:int	xsd:long xsd:short xsd:byte xsd:unsignedLong xsd:unsignedInt xsd:unsignedShort xsd:unsignedByte xsd:double xsd:float	xsd:string xsd:normalizedString xsd:token xsd:language xsd:Name xsd:NCName xsd:NMTOKEN	xsd:boolean xsd:hexBinary xsd:base64Binary sd:dateTime xsd:dateTimeStamp

- Απλές δυνατότητες μεταμοντελοποίησης. Σε αυτή την κατηγορία ανήκει η τεχνική της παρονομασίας (punning). Η OWL 1 απαιτεί ένα αυστηρό διαχωρισμό των ονομάτων παραδείγματός χάρη μεταξύ κλάσεων και ιδιοτήτων. Σε αντίθεση η OWL 2 χαλαρώνει αυτόν τον διαχωρισμό ώστε να επιτρέπει διαφορετικές χρήσεις του ίδιου όρου. Παρόλα αυτά η OWL 2 διατηρεί ακόμα κάποιους κύριους περιορισμούς όπως ότι ένα όνομα δεν μπορεί να χρησιμοποιείται από κλάσεις και τύπους δεδομένων, όπως επίσης ότι το ίδιο όνομα δεν μπορεί να χρησιμοποιηθεί για ένα είδος ιδιοτήτων.

- Επέκταση των δυνατοτήτων σχολιασμού. Η βασική χρήση των σχολιασμών στην OWL 1 ήταν να διατηρούνται ετικέτες, σχόλια, σημειώσεις, παραδείγματα μέσα στις έννοιες της Οντολογίας αλλά δεν επιτρεπόταν για τα αξιώματα. Η OWL 2 επιτρέπει τον σχολιασμό αξιωμάτων δίνοντας πληροφορίες για το πώς και πότε μπορούν να χρησιμοποιηθούν. Γενικά η χρήση σχολίων μπορεί να γίνει πλέον σε οντολογίες, έννοιες, ανώνυμα άτομα, αξιώματα και για να σχολιάσουν τον εαυτό τους.

Πίνακας 5-9. Δυνατότητες σχολιασμού

Κατηγορία	Ονομασία	OWL 2
Σχολιασμοί	AnnotationAssertion	Χρησιμοποιείται για τον σχολιασμό Οντολογικών εννοιών όπως κλάσεις και ιδιότητες, αλλά και ανώνυμων ατόμων
	Annotation	Χρησιμοποιείται για τον σχολιασμό αξιωμάτων, οντολογιών και άλλων σχολιασμών
Αξιώματα σχετικά με τις ιδιότητες σχολίων	SubAnnotationPropertyOf	Τα σχόλια μπορούν να συμμετέχουν σε ιεραρχίες και να τους ορίζεται πεδίο ορισμού και τιμών
	AnnotationPropertyDomain	
	AnnotationPropertyRange	

- Άλλες καινοτομίες, και δευτερεύουσες λειτουργίες.

Κατηγορία	Ονομασία	OWL 2	OWL 1
Δήλωση	Declaration	Μπορεί να γίνει δήλωση μιας οντότητας για το σε ποια κατηγορία ανήκει (class, datatype, object property, data property, annotation property, or individual)	
Ιδιότητες υψηλού και χαμηλού επιπέδου	owl:topObjectProperty	Διαθέτει εκτός τις δύο προκαθορισμένες κλάσεις και άλλες τέσσερις σχετικά με τις ιδιότητες των αντικείμενων και των δεδομένων: α) όλα τα ζευγάρια των ατόμων είναι συνδεδεμένα με το owl:topObjectProperty, β) κανένα άτομο δεν είναι συνδεδεμένο με το owl:bottomObjectProperty γ) όλα τα πιθανά άτομα είναι συνδεδεμένα με όλα τα λεκτικά με το owl:topDataProperty, δ) κανένα άτομο δεν είναι συνδεδεμένο μέσω του owl:bottomDataProperty με ένα λεκτικό.	Διαθέτει μόνο δύο πάνω και κάτω προκαθορισμένες οντότητες για κλάσεις αυτές της owl:Thing και owl:Nothing
	owl:bottomObjectProperty		
	owl:topDataProperty		
	owl:bottomDataProperty		
Διεθνοποιημένοι Προσδιοριστές Πόρων	-	Χρησιμοποιεί IRIs προσδιοριστές, καθώς το URI περιλαμβάνει περιορισμένη αλφάβητο	Χρησιμοποιεί URI προσδιοριστές
Εισαγωγές και εκδόσεις	Ontology	Κυρίως αποσαφηνίζει την διαδικασία της εισαγωγής Οντολογιών σε μία Οντολογία και τον μηχανισμό δημιουργίας έκδοσης, ο οποίος μπορεί να περιλαμβάνει μία IRI ονομασία.	Επιτρέπει την αποθήκευση Οντολογιών ως έγγραφα του σημασιολογικού ιστού και την εισαγωγή Οντολογιών σε άλλες Οντολογίες

Άλλες τρεις μεγάλες διαφορές που μπορούν να διακριθούν εύκολα σε αυτές τις δύο εκδόσεις είναι πρώτον ότι στην OWL 2 ορίζονται τρία νέα χαρακτηριστικά προφίλ σε σχέση με αυτά της OWL 1 (Lite, DL, Full). Δεύτερον, ότι ορίζεται ένα νέο συντακτικό που την υποστηρίζει, το OWL 2 Manchester Syntax (Horridge & Patel-Schneider, 2009), παρόλα αυτά το RDF/XML παραμένει η καθιερωμένη σύνταξη της OWL και την οποία πρέπει να υποστηρίζουν όλες οι εφαρμογές. Τρίτον, ορισμένοι από τους περιορισμούς που ισχύουν για την OWL DL έχουν γίνει ελαστικότεροι, ως αποτέλεσμα το σύνολο των οντολογιών που μπορεί να διαβαστούν από DL Reasoners να είναι ελαφρώς μεγαλύτερο στην OWL 2.

Ο ορισμός των τριών νέων υπογλωσσών της OWL 2 παρουσιάστηκε ως ανάγκη, καθώς μέσα από τις εφαρμογές που χρησιμοποιούσαν τα προφίλ της OWL 1 αποδείχθηκε ότι δεν ήταν επαρκείς για να ανταποκριθούν στις απαιτήσεις, που αργότερα ταυτίστηκαν με τα κίνητρα ανάπτυξης της γλώσσας. Τρία βασικά προβλήματα αναδύθηκαν. Πρώτον, ότι οι περισσότερες εφαρμογές χρησιμοποιούσαν μεγάλες Οντολογίες με ένα τεράστιο αριθμό κλάσεων και συσχετίσεων, με αποτέλεσμα να προκαλούνται προβλήματα στην εξελιξιμότητά τους αλλά και στην απόδοση των συλλογισμών. Αυτό που συμπέραναν τελικά, ήταν η διάθεση τους να χάσουν από την εκφραστικότητα της γλώσσας ως αντίκτυπο να αυξήσουν την απόδοση και εγκυρότητα της συλλογιστικής τους. Το δεύτερο πρόβλημα ήταν αυτό της διαλειτουργικότητας κυρίως με τεχνολογίες και εργαλεία των Βάσεων Δεδομένων καθώς χρησιμοποιούσαν ελαφριές Οντολογίες συνδυαστικά με άλλες γλώσσες όπως SQL για βελτιστοποιήσουν ερωτήσεις σε πολύ μεγάλα σύνολα αποθηκευμένων δεδομένων. Τρίτον, ότι έπρεπε να υπάρξει διαλειτουργικότητα με τις γλώσσες και τις μηχανές κανόνων (Golbreich & Wallace, 2009). Για να αντιμετωπίσουν λοιπόν αυτά τα προβλήματα ορίστηκαν τρία διαφορετικά προφίλ: OWL 2 EL, OWL 2 QL, OWL 2 RL.

Όπως έχει ξαναειπωθεί ο διαχωρισμός της OWL 2 γίνεται ως μια προσπάθεια να ζυγοσταθμιστεί η δυναμική της εκφραστικότητας μίας Οντολογίας με την αποδοτικότητα της συλλογιστικής, ανάλογα με τις ανάγκες της εφαρμογής, το μέγεθος και τους τύπους δεδομένων που διαπραγματεύεται. Αυτό κάνει τα τρία προφίλ να είναι ανεξάρτητα το ένα από το άλλο.

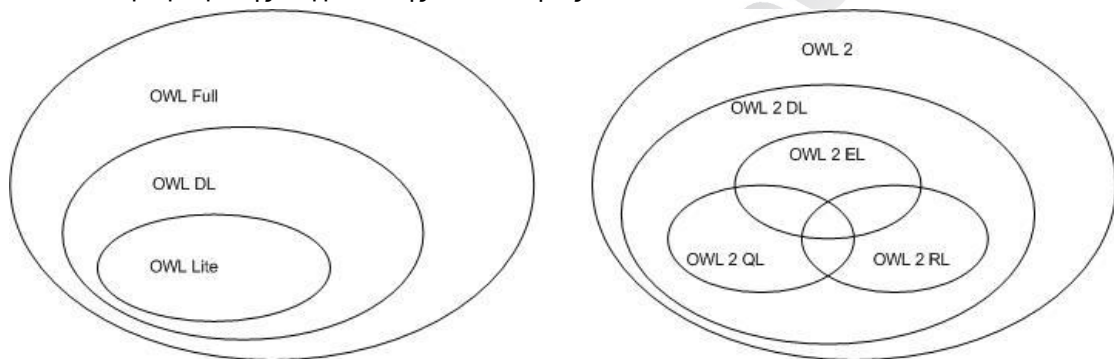
- *OWL 2 EL*. Είναι ιδιαίτερα χρήσιμο σε εφαρμογές που χρησιμοποιούν οντολογίες οι οποίες περιέχουν μεγάλο αριθμό κλάσεων και ιδιοτήτων. Είναι φυσικά ένα υποσύνολο της OWL 2 που προσπαθεί και εκτελεί βασικά προβλήματα συλλογιστικής σε πολυωνυμικό χρόνο εξαρτώμενο από το μέγεθος της οντολογίας. Ουσιαστικά ανήκει στην EL, γι αυτό έχει και το αντίστοιχο ακρωνύμιο, εκφραστικότητα της περιγραφικής λογικής, η οποία περικλείει τη διατομή και τους υπαρξιακούς περιορισμούς.
- *OWL 2 QL*. Χρησιμοποιείται για εφαρμογές που έχουν πολύ μεγάλο όγκο δεδομένων και η απάντηση ερωτημάτων είναι μια πολύ σημαντική διαδικασία συλλογιστικής. Μπορεί να συνδυαστεί με άλλες γλώσσες, όπως η SQL, σε συμβατικά συστήματα σχεσιακών βάσεων δεδομένων ώστε να βελτιστοποιήσουν τα ερωτήματα. Ανάλογα με το μέγεθος των δεδομένων, μπορεί να γίνει χρήση κατάλληλων τεχνικών συλλογισμού και να παράγει ορθά και πλήρες αποτελέσματα σε χρόνο LONGSPACE. Επίσης, αλγόριθμοι πολυωνυμικού χρόνου μπορούν να χρησιμοποιηθούν για να εξακριβωθεί η συνέπεια μιας οντολογίας και σχέσεις υπαγωγής μεταξύ κλάσεων. Η εκφραστική δύναμη του προφίλ είναι αρκετά περιορισμένη, αν και περιλαμβάνει τα περισσότερα από τα κύρια χαρακτηριστικά των εννοιολογικών μοντέλων όπως διαγράμματα κλάσεων UML και διαγράμματα ER. Το ακρωνύμιο QL αντικατοπτρίζει το γεγονός ότι οι απαντήσεις ερωτημάτων σε αυτό το προφίλ μπορούν να υλοποιηθούν ξαναγράφοντας ερωτήματα σε μια τυπική σχεσιακή γλώσσα ερωτήσεων (Query Language).
- *OWL 2 RL*. Το προφίλ OWL 2 RL έχει σχεδιαστεί ώστε να μπορεί να επιτευχθεί κλιμάκωση στη διαδικασία συλλογιστικής, χωρίς να χρειάζεται να θυσιάστούν πολύ οι εκφραστικές δυνατότητες της γλώσσας. Τα συστήματα συλλογιστικής για το προφίλ OWL 2 RL μπορούν να βασιστούν σε συστήματα κανόνων. Ο έλεγχος της συνέπειας της οντολογίας και της ικανοποιησιμότητας των κλάσεων, ο υπολογισμός της ιεραρχίας υπαγωγής, ο έλεγχος των αντικειμένων και η απάντηση συζευκτικών ερωτημάτων μπορούν να γίνουν σε πολυωνυμικό χρόνο, αναφορικά με το μέγεθος της Οντολογίας. Το ακρωνύμιο RL αντανάκλα το γεγονός ότι η συλλογιστική σε αυτό το προφίλ μπορεί να εφαρμοστεί χρησιμοποιώντας μία τυποποιημένη γλώσσα κανόνων.

Σε αντίθεση με την OWL 2 η OWL 1 ορίζει τρεις διαφορετικές υπογλώσσες οι οποίες η μία είναι υποσύνολο της άλλης. Δηλαδή καθορίζει τις γλώσσες OWL Lite, OWL DL και OWL Full, οι οποίες έχουν βαθμιαία αυξημένης εκφραστικότητα και συλλογιστικής πολυπλοκότητα, ενώ κάθε νόμιμη οντολογία ή έγκυρο συμπέρασμα μια υπογλώσσας, ξεκινώντας από τα αριστερά είναι νόμιμη οντολογία ή έγκυρο συμπέρασμα για τις υπογλώσσες στα δεξιά. Σχετικά έχουμε:

- *OWL Full*: Χρησιμοποιεί όλες τις θεμελιώδεις αρχές των υπογλωσσών OWL γι αυτό αποκαλείται πλήρης γλώσσα. Επιτρέπει το συνδυασμό των αρχών αυτών με τη RDF

και το RDFs. Η OWL Full είναι απόλυτα συμβατή με την RDF τόσο συντακτικά όσο και σημασιολογικά, οπότε οποιοδήποτε νόμιμο RDF κείμενο ή έγκυρο RDF/RDFs συμπέρασμα είναι ένα νόμιμο OWL Full κείμενο ή έγκυρο OWL Full συμπέρασμα. Αν και διαθέτει μια πολύ πλούσια εκφραστική ικανότητα, πολλές φορές αδυνατεί να δώσει αποτελέσματα λόγω της μεγάλης υπολογιστικής πολυπλοκότητας της συλλογιστικής.

- *OWL DL*: Υποστηρίζει εκείνους τους χρήστες που θέλουν τη μέγιστη εκφραστικότητα διατηρώντας την υπολογιστική πολυπλοκότητα της συλλογιστικής σε χαμηλά επίπεδα. Βασίζεται στην περιγραφική λογική SHOIN (D) και υποστηρίζει σχεδόν όλες τις δυνατότητες της OWL. Το δυνατό της σημείο είναι ότι επιτρέπει την αποδοτική υποστήριξη της συλλογιστικής, αλλά από την άλλη χάνει την πλήρη συμβατότητα με την RDF.
- *OWL Lite*: Περιορίζει επιπλέον τους γλωσσικούς κατασκευαστές της OWL DL, βασιζόμενη στην εκφραστικότητα SHIN (D) της περιγραφική λογικής, η οποία αποτελεί μια επέκταση της ALC. Η OWL Lite αποκλείει τις απαριθμημένες κλάσεις, τις δηλώσεις ασυμβατότητας και το αυθαίρετο μέγεθος πεδίου τιμών ιδιοτήτων. Αυτό έχει ως αποτέλεσμα να είναι μία γλώσσα εύκολη στην εκμάθηση, απλή στη χρήση αλλά περιορισμένης εκφραστικής δυνατότητας.



Εικόνα 5-3. Υπογλώσσες OWL 1 και OWL 2

5.2 Μηχανές Συλλογιστικής Ανάλυσης

Με τις Οντολογίες έχει επιτευχθεί η αναπαράσταση μία σύλληψης, παρόλα αυτά από μόνες τους δεν αρκούν για την εξόρυξη και επαναχρησιμοποίηση της γνώσης. Αυτό το κενό προσπαθούν να καλύψουν οι Μηχανές Συλλογιστικής Ανάλυσης που δεν είναι τίποτε άλλο από λογισμικά εργαλεία τα οποία επιτρέπουν να συναχθούν λογικές συνέπειες από το σύνολο των αξιωμάτων ή γεγονότων που διατηρούνται σε μία Οντολογία.

Γίνεται αντιληπτό ότι υπάρχει μία αλληλεπίδραση μεταξύ των Οντολογιών και των μηχανών συλλογιστικής ανάλυσης. Παραδείγματος χάρη, η τυπικότητα των Οντολογιών πηγάζει από την αναγκαιότητα να διαβάζουν οι μηχανές συλλογιστικής ανάλυσης γλώσσες με αυστηρότητα και όσο το δυνατόν λιγότερη ασάφεια. Από την άλλη μεριά, η αναγκαιότητα των μηχανών συλλογιστικής ανάλυσης να διαθέτουν πλούσιες διαδικασίες συμπερασμού πηγάζει από τη διάθεση των Οντολογιών να είναι αυξημένης εκφραστικότητας. Έτσι κάθε αναπαράσταση γνώσης διαθέτει τους δικούς της μηχανισμούς για την παραγωγή συμπερασμάτων, οι οποίοι δεν είναι αναγκαστικά διαφορετικοί.

Οι μηχανές συλλογιστικής ανάλυσης που χρησιμοποιούνται για την συλλογιστική ανάλυση OWL κειμένων ακολουθούν δύο κυρίως αλγόριθμους. Ο πρώτος είναι η *πρόσθια σύνδεση*, η οποία ξεκινάει με τα διαθέσιμα γεγονότα και χρησιμοποιεί τους κανόνες για την παραγωγή καινούριων. Η διαδικασία πραγματοποιείται σε βήματα μέχρι να επιτευχθεί ο στόχος. Η δεύτερη και η κυρίαρχη προσέγγιση είναι αυτή των *αλγορίθμων ταμπλό* (tableau). Εδώ η βασική ιδέα είναι να αποδειχτεί η συνέπεια μίας έννοιας δημιουργώντας ένα μοντέλο που την υποστηρίζει, σε περίπτωση που το μοντέλο δεν μπορεί να κατασκευαστεί τότε συμπεραίνεται ότι η έννοια είναι ασυνεπής. Ο αλγόριθμος ξεκινάει με ένα σύνολο ισχυρισμών και προχωράει προσθέτοντας επιπλέον περιορισμούς στο μοντέλο που προέρχονται είτε από τον συνδυασμό των αρχικών ισχυρισμών ή από ορισμένους περιορισμούς, ενώ τελειώνει όταν επιτευχθεί η δημιουργία του επιθυμητού μοντέλου ή αν δημιουργηθούν αντιφάσεις μεταξύ των περιορισμών.

Τα εργαλεία που κάνουν χρήση οι συλλογιστικές μηχανές στην OWL είναι οι κατασκευαστές της εκάστοτε υπογλώσσας. Η κάθε υπογλώσσα διαθέτει ένα σύνολο διαφορετικών συνδυασμών από τους κατασκευαστές, που καθορίζει την εκφραστικότητα και την πολυπλοκότητά της. Αυτό έχει ως αποτέλεσμα να χρησιμοποιούνται διαφορετικές μηχανές είτε για να εκμεταλλευτούν την εκφραστικότητα μίας υπογλώσσας, είτε για να εκμεταλλευτούν τις βελτιωτικές διαδικασίες της συλλογιστικής που διαθέτουν οι μηχανές για να ελαττώσουν την πολυπλοκότητα, σε πολυωνυμικό χρόνο.

Μεταξύ των πιο γνωστών μηχανών συλλογιστικής ανάλυσης βρίσκονται οι Pellet, FaCT++, HermiT, KAON2 και RACER οι οποίοι υποστηρίζουν την *SROIQ* εκφραστικότητα της DL που ουσιαστικά είναι η επέκταση της λογικής περιγραφής που βασίζεται στην OWL-DL (*SHOIN*). Στην *SROIQ* υπενθυμίζεται ότι υλοποιούνται κατασκευαστές όπως οι αυτοπεριορισμοί, περιορισμοί ποσοτικοποίησης, ανακλαστικές, μη-ανακλαστικές, ασύμμετρες ιδιότητες αντικειμένων, ξένες ιδιότητες, ιδιότητες αλυσίδα και κλειδιά. Ενδιαφέρον όμως παρουσιάζει ότι τον κατασκευαστή HasKey τον υποστηρίζουν μόνο οι Pellet, FaCT++ και HermiT.

Στον πίνακα που ακολουθεί παρουσιάζονται οι πολυπλοκότητες πέντε κλασικών υπολογιστικών προβλημάτων συλλογισμού σε συνάρτηση με την ταξινομική πολυπλοκότητα, δηλαδή το μέγεθος των αξιωμάτων στην Οντολογία, η πολυπλοκότητα δεδομένων, δηλαδή το συνολικό μέγεθος των ισχυρισμών στην Οντολογία, η πολυπλοκότητα επερωτήσεων, δηλαδή το μέγεθος του ερωτήματος, και η συνδυασμένη πολυπλοκότητα, δηλαδή ο συνδυασμός του μεγέθους των αξιωμάτων, το μέγεθος των ισχυρισμών, και το μέγεθος των ερωτήσεων σε περίπτωση που υπάρχουν.

Πίνακας 5-10. Πολυπλοκότητες σε προβλήματα συλλογισμού

Πηγή: Motic και συν. (2009)

Γλώσσα	Προβλήματα Συλλογισμού	Ταξινομική Πολυπλοκότητα.	Πολυπλοκότητα Δεδομένων	Πολυπλοκότητα Επερωτήσεων	Συνδυασμένη Πολυπλοκότητα
OWL DL	Έλεγχος Συνέπεια, Ικαν/τα, Υπαγωγή, Στιγμιότυπων	NEXPTIME-complete	Αποκρίσιμη, ανοιχτής πολυπλοκότητας (NP-Hard)	Μη εφαρμόσιμη	NEXPTIME-complete
	Απάντηση Συζευγμένων Ερωτημάτων	Ανοιχτή Αποκρισιμότητα	Ανοιχτή Αποκρισιμότητα	Ανοιχτή Αποκρισιμότητα	Ανοιχτή Αποκρισιμότητα
OWL 2 EL	Έλεγχος Συνέπεια, Ικαν/τα, Υπαγωγή, Στιγμιότυπων	PTime-complete	PTime-complete	Μη εφαρμόσιμη	PTime-complete
	Απάντηση Συζευγμένων Ερωτημάτων	PTime-complete	PTime-complete	NP-complete	PSPACE-complete
OWL 2 QL	Έλεγχος Συνέπεια, Ικαν/τα, Υπαγωγή, Στιγμιότυπων	NLogSpace-complete	AC ⁰	Μη εφαρμόσιμη	NLogSpace-complete
	Απάντηση Συζευγμένων Ερωτημάτων	NLogSpace-complete	AC ⁰	NP-complete	NP-complete
OWL 2 RL	Έλεγχος Συνέπεια, Ικαν/τα, Υπαγωγή, Στιγμιότυπων	PTime-complete	PTime-complete	Μη εφαρμόσιμη	PTime-complete
	Απάντηση Συζευγμένων Ερωτημάτων	PTime-complete	PTime-complete	NP-complete	NP-complete

Από τον παραπάνω πίνακα παρατηρείται ότι εκτός του χρόνου απόκρισης, ανάλογα της πολυπλοκότητας, πολλές φορές γίνεται αναφορά στη δυνατότητα *αποκρισιμότητας* (decidability) που καθορίζει ότι είναι δυνατόν ή όχι να παραχθεί μία απάντηση πάνω σε ένα πρόβλημα. Ένα γενικότερο ενδεικτικό παράδειγμα είναι η κατηγορηματική λογική η οποία είναι πάντα αποκρίσιμη, ενώ η περιγραφική λογική δεν είναι πάντα, όπως αυτό συμβαίνει με την OWL FULL.

Όπως παρουσιάστηκαν και παραπάνω, πέντε είναι κυρίως τα προβλήματα συμπερασμού που εξετάζονται από τις μηχανές συλλογιστικής ανάλυσης (Motik, Patel-Schneider & Grau, 2009):

- *Έλεγχος συνέπειας Οντολογίας* (Ontology Consistency): Επιβεβαιώνει ότι τα αξιώματα εισόδου δεν περιέχουν αντιφάσεις. Θεωρείται ότι πρέπει να είναι το πρώτο βήμα πριν από κάθε υπηρεσία του μηχανισμού συλλογιστικής ανάλυσης.
- *Ικανοποιησιμότητα έκφρασης κλάσης* (Class Expression Satisfiability): δοσμένης μίας οντολογίας και μίας κλάσης, ελέγχεται αν υπάρχει μοντέλο στο οποίο η διερμηνευση της κλάσης δεν είναι κενή. Υπάρχει δυνατότητα μία συνεπής Οντολογίας να περιέχει κλάσεις που να μην είναι ικανοποιήσιμες, χωρίς βέβαια να περιέχουν στιγμιότυπα. Σε περίπτωση που δηλωθούν στιγμιότυπα προκαλείται ασυνέπεια.
- *Υπαγωγή έκφρασης κλάσης* (Class Expression Subsumption): Ελέγχει αν μια έννοια είναι υποσύνολο μιας άλλης έννοιας, αν κάτι τέτοιο ισχύει τότε συμπεραίνεται ότι η γνώση είναι σωστή.
- *Έλεγχος στιγμιότυπου* (Instance Checking): Ελέγχεται αν ένα άτομο είναι πράγματι στιγμιότυπο μίας κλάσης.
- *Απάντηση συζευγμένων ερωτημάτων* (Conjunctive Query Answering): Ελέγχει αν σε μία συζευγμένη ερώτηση επιστρέφονται όλες οι πλειάδες των ατόμων που ταιριάζουν στη συγκεκριμένη επερώτηση.

Εκτός από τους παραπάνω ελέγχους, άλλες δύο χρησιμότητες των μηχανών συλλογιστικής ανάλυσης τους καθιστούν απαραίτητους και ωφέλιμους. Η πρώτη είναι η ταξινόμηση (Classification), στην οποία υπολογίζονται οι σχέσεις υποκλάσεων (subClassOf) και ισοδυναμίας (equivalentClass) μεταξύ των κλάσεων που αποτελούν την οντολογία. Η δεύτερη είναι η πραγμάτωση (realization), η οποία βρίσκει την πιο συγκεκριμένη έννοια της οποίας κάποιο άτομο είναι στιγμιότυπο. Για να είναι αποδοτική η πραγμάτωση θα πρέπει να έχει τρέξει πρώτα η διαδικασία της ταξινόμησης. Όλα αυτά γίνονται πάντα λαμβάνοντας υπόψη τις περιγραφές των κλάσεων και την ικανοποίηση των αξιωμάτων. Έτσι πολλές φορές, αν κάποιος φανταστεί την οντολογία ως ένα δέντρο, μετά την χρήση μίας μηχανής συλλογιστικής ανάλυσης, αυτό το δέντρο αλλάζει μορφή ανεβοκατεβάζοντας κάποιες κλάσεις στα επίπεδα του δέντρου αλλά και φτιάχνοντας νέες σχέσεις υπαγωγής. Αυτή ουσιαστικά είναι η κρυμμένη γνώση μέσα από την αναπαράσταση μίας Οντολογίας, που αναδύεται από τις δυνατότητες της εκάστοτε συλλογιστικής μηχανής να την αξιοποιήσει. Συνήθως, τον τελικό χρήστη τον ενδιαφέρει η ανάκτηση (retrieval), που βρίσκει τελικά εκείνα τα άτομα σε μία βάση γνώσης που είναι στιγμιότυπα μίας δοσμένης έννοιας.

5.3 Σημασιολογική Συνάφεια

Μέχρι τώρα εξετάστηκαν δύο πολύ βασικές έννοιες οι Οντολογίες, που χρησιμοποιούνται για την αναπαράσταση μίας σύλληψης και οι Μηχανές Συλλογιστικής Ανάλυσης, που ουσιαστικά διαβάζουν τις Οντολογίες και βοηθάνε το χρήστη να ανακτήσει την απαραίτητη γνώση. Από την μία υπάρχει μια καθορισμένη δομή που δείχνει τις συσχετίσεις μεταξύ των εννοιών και από την άλλη ένα εργαλείο που αξιοποιεί τα χαρακτηριστικά αυτών των σχέσεων. Παρόλα αυτά μέχρι τώρα δεν έχει οριστεί κάποια διαδικασία που να περιγράφει την συνάφεια των εννοιών μεταξύ τους, δηλαδή μία διαδικασία που να παράγει ένα βαθμό συνάφειας μεταξύ των εννοιών, που να μπορεί να καθορίζει το ποσοστό της ομοιότητάς ή ανομοιότητάς τους. Αυτές οι διαδικασίες περιγράφονται μέσω μετρικών ή αλγορίθμων και αξιοποιούν τις σχέσεις των εννοιών μέσα στην Οντολογία. Η γενικότερη κατηγορία αυτών αναφέρεται ως Σημασιολογική Συνάφεια (Semantic Relatedness), βέβαια στην βιβλιογραφία παρουσιάζονται άλλες δύο πιο εξειδικευμένες ορολογίες που πολλές φορές δεν παρουσιάζουν ουσιαστική διαφορά μεταξύ τους. Αυτές είναι η Σημασιολογική Ομοιότητα (Semantic Similarity), που αναφέρεται στο κατά πόσο δύο έννοιες έχουν κοινά χαρακτηριστικά και η Σημασιολογική Απόσταση (Semantic Distance), που αναφέρεται στις αποστάσεις που μπορούν να οριστούν μεταξύ των εννοιών και θεωρείται ότι όσο πιο κοντά είναι δύο έννοιες, τόσο πιο όμοιες είναι ή τόσο υψηλότερη συνάφεια υπάρχει μεταξύ τους (Budanitsky & Hirst, 2006).

Είναι γενικά αποδεκτό ότι δεν υπάρχει μία μοναδική γενική προσέγγιση που να ορίζει την σημασιολογική συνάφεια έτσι ώστε να καλύπτονται όλες οι απαιτήσεις των εφαρμογών. Αντιθέτως, ανάλογα με τις ανάγκες και την δομή της Οντολογίας δημιουργούνται από την αρχή νέες μετρικές και είτε γίνεται συνδυασμός είτε μετεξελίσσονται. Παρόλα αυτά υπάρχουν

κάποια βασικά σημεία που οι μετρικές αυτές θα πρέπει να στηριχτούν ώστε να είναι κατάλληλες προς χρήση. Σύμφωνα με τον Roddick και συν. (2003) είναι πολύ σημαντικό:

- να χρησιμοποιούνται μετρικές ανάλογα με τις ανάγκες ώστε να παράγουν λογικά-χρήσιμα συμπεράσματα. Παραδείγματος χάρη, σε μία χωρική εφαρμογή δε θα μπορούσε να εξαχθεί άμεσο λογικό συμπέρασμα από την άμεση αφαίρεση των αριθμών δύο σπιτιών.
- να παράγονται εύκολα κατανοητά αποτελέσματα, συνήθως προτιμούνται τα σχετικά μέτρα από τα απόλυτα. Δηλαδή ότι η απόσταση A, B είναι ίδια με την Γ, Δ από το να δοθεί ένας απλός αριθμός.
- να συνδυάζονται μέτρα για την πληρέστερη περιγραφή μίας σύγκρισης. Παραδείγματος χάρη, στα χρώματα θα ήταν φρόνιμο να συνδυαστεί η φωτεινότητα τους, η διαβάθμισή του γκρι ώστε να περιγραφεί η σχετική απόσταση μεταξύ δύο χρωμάτων.
- πολλές φορές η διακριτότητα των τιμών που παράγονται δεν μπορεί να είναι γραμμική, αντίθετα απαιτείται η περιγραφή τους μέσα από τα ασαφή σύνολα.
- να λαμβάνεται υπόψη ότι οι αποστάσεις μεταξύ δύο σημείων μπορεί να μην είναι μοναδικές ή ότι ένα μονοπάτι να μην είναι συμμετρικό επομένως να δίνει διαφορετικές αποστάσεις ανάλογα με την αφετηρία.
- τέλος, να λαμβάνονται υπόψη τυχών βάρη- κόστη που μπορούν να υφίστανται για την μεταπήδηση από μία έννοια σε μία άλλη.

Συνήθως μέσα σε μία Οντολογία οι μετρικές χρησιμοποιούν χαρακτηριστικά της δομής της για να ορίσουν κατάλληλα μέτρα. Τα τρία βασικά χαρακτηριστικά είναι το *πλάτος* (width), το *βάθος* (depth) και η *πυκνότητα* (density). Με το πλάτος γίνεται αναφορά στο πλήθος των άμεσων παιδιών μίας έννοιας σε μία Οντολογία. Με το βάθος γίνεται αναφορά στο επίπεδο που βρίσκεται μία έννοια μέσα στο δέντρο με την ρίζα να έχει βάθος ένα. Τέλος με την πυκνότητα γίνεται αναφορά στο ποσοστό ενός υπο-δέντρου με ρίζα μία συγκεκριμένη έννοια, σε σχέση με την Οντολογία. Γίνεται επομένως αντιληπτό ότι τα μέτρα μπορούν να χρησιμοποιήσουν το σύνολο των δομικών στοιχείων (έννοιες, σχέσεις, αξιώματα, στιγμιότυπα) μίας οντολογίας για να επιτύχουν τον στόχο τους.

Για την περιγραφή των πιο διαδεδομένων αλγορίθμων σημασιολογικής συνάφειας θα ακολουθηθεί η κατηγοριοποίηση που δίνεται από τον Saruladha και συν. (2010) και συμβαδίζει με τις περιγραφές των Budanitsky και Hirst (2006). Σύμφωνα με τους Saruladha και συν. (2010) γίνεται ένας χωρισμός μεταξύ τεσσάρων προσεγγίσεων:

- *Μέγεθος Μονοπατιού* (Path Length): το μέτρο ομοιότητας δίνεται από την αξιοποίηση του μονοπατιού με το μικρότερο μήκος, λαμβάνοντας υπόψη ή όχι τα βάρη μεταξύ των εννοιών μέσα σε μία ταξινόμια.
- *Σχετικό Βάθος* (Depth- Relative): χρησιμοποιεί τα μονοπάτια αλλά δίνει ιδιαίτερη βαρύτητα στο βάθος των εννοιών αλλά και σε άλλες σχέσης εκτός της is-a. Εδώ οι προσεγγίσεις που ακολουθούνται είναι η κλιμάκωση σχετικού βάθους, η εννοιολογική ομοιότητα και η κοινωνικοποίηση μήκους μονοπατιού.
- *Βασισμένες στη Συλλογή* (Corpus- based): επιστρατεύει το περιεχόμενο της πληροφορίας από τη θεωρία πληροφορίας για να εμπλουτίσει την εύρεση αποστάσεων και ομοιότητας μεταξύ εννοιών μία οντολογίας ή ταξινόμιας, οι προσεγγίσεις που εξετάζονται είναι του περιεχομένου της πληροφορίας, μιας υβριδικής μεθόδου και ενός γενικού μέτρου ομοιότητας.
- *Πολλαπλά Μονοπάτια* (Multiple- Paths): προς το παρόν οι μέθοδοι που αναφέρθηκαν χρησιμοποιούν το ελάχιστο μονοπάτι και τις σχέσεις is-a, στις προσεγγίσεις των πολλαπλών μονοπατιών λαμβάνονται υπόψη περισσότερα από ένα μονοπάτια μεταξύ των εννοιών αλλά και όλες οι σχέσεις μέσα σε μία οντολογία. Τέτοιες προσεγγίσεις αναφέρονται ως μεσαίες- ισχυρές σχέσεις, γενικευμένη σταθμισμένη συντομότερη διαδρομή, κοινόχρηστοι κόμβοι και ομοιότητα βασισμένη σε σταθμισμένους κοινόχρηστους κόμβους.

Μικρότερο μήκος μονοπατιού (Shortest Path Length). Ο πιο αντιπροσωπευτικός αλγόριθμος είναι αυτός του Rada, στον οποίο το μέτρο απόστασης εξαρτάται από το ελάχιστο πλήθος των ακμών που βρίσκονται μεταξύ των εννοιών

$$dist_{rmbb}(c_i, c_j) = \min_{p \in paths(c_i, c_j)} len_e(p)$$

Κατά την εξέταση της συντομότερης διαδρομής, εξετάζονται μόνο οι σχέσεις is-a μεταξύ των παρακείμενων εννοιών οι οποίες έχουν την ίδια αξία.

Μικρότερο μήκος μονοπατιού με βάρη (Weighted Shortest Path Length). Ένας αλγόριθμος που έχει προταθεί είναι αυτός του Bulskov. Είναι μία γενικότερη προσέγγιση της προηγούμενης μεθόδου, εισάγοντας την έννοια των βαρών πάνω στις ακμές. Οι ακμές που χρησιμοποιούνται είναι αυτές της γενίκευσης και της εξειδίκευσης, οι οποίες μπορούν να πάρουν μία σταθερή τιμή $\sigma \in [0, 1]$ και $\gamma \in [0, 1]$ αντίστοιχα. Θεωρώντας ένα μονοπάτι $P = (p_1, \dots, p_n)$ όπου οι έννοιες σχετίζονται με την σχέση is-a: $p_i \text{ ISA } p_{i+1}$ ή $p_{i+1} \text{ ISA } p_i$, μπορεί να υπολογιστεί ένα σύνολο $s(P)$ με τον αριθμό των σχέσεων εξειδίκευσης και $g(P)$ με τον αριθμό των σχέσεων γενίκευσης για το συγκεκριμένο μονοπάτι P όπως ακολούθως:

$$s(P) = |\{i | p_i \text{ ISA } p_{i+1}\}| \text{ και } g(P) = |\{i | p_{i+1} \text{ ISA } p_i\}|$$

Έτσι αν δοθούν δύο έννοιες x, y και τα μονοπάτια P^1, \dots, P^m που τα συνδέουν, τότε ο βαθμός με τον οποίο το y είναι όμοιο με το x δίνεται από τον τύπο:

$$\text{sim}_{WSP}(x, y) = \max_{j=1, \dots, m} \{\sigma^{s(P^j)} \gamma^{g(P^j)}\}$$

Δηλαδή, επιλέγεται η απόσταση από εκείνο το μονοπάτι με το μέγιστο γινόμενο των σταθερών σ και γ , υψωμένα στους αντίστοιχους αριθμούς των σχέσεων εξειδίκευσης και γενίκευσης

Κλιμάκωση σχετικού βάθους (Depth Relative Scaling). Η προσέγγιση του αλγορίθμου Sussna αξιοποιεί τόσο τις σχέσεις που συνδέουν τις έννοιες όσο και το βάθος στο οποίο βρίσκονται. Η προσέγγιση είναι εμπνευσμένη από μία παρατήρηση που θέλει έννοιες, σε μεγάλο βάθος στην ταξινόμια, να είναι πιο στενά συνδεδεμένες μεταξύ τους από αυτές που είναι πιο ψηλά. Σε αυτή για κάθε σχέση r , ορίζεται ένα βάρος $w(c_i \rightarrow_r c_j)$ από ένα δοσμένο πεδίο $[\min_r, \max_r]$. Αυτό το βάρος υπολογίζεται με την τοπική πυκνότητα που αντιστοιχεί στον αριθμό των σχέσεων του τύπου του r που φεύγουν από την έννοια c_i :

$$w(c_i \rightarrow_r c_j) = \max_r - \frac{\max_r - \min_r}{n_r(c_i)}$$

Ο αλγόριθμος ορίζει την απόσταση μεταξύ δύο γειτονικών εννοιών, όπου για κάθε σχέση ορίζεται και η αντίστροφή της. Έτσι η συνολική ομοιότητα μεταξύ δύο εννοιών, είναι το άθροισμα των τιμών απόστασης μεταξύ ζευγών γειτονικών κόμβων κατά μήκος του ελάχιστου μονοπατιού που συνδέει τις έννοιες αυτές.

$$\text{dist}_{\text{sussna}}(c_i, c_j) = \frac{w(c_i \rightarrow_r c_j) + w(c_j \rightarrow_r c_i)}{2d}$$

Εννοιολογική ομοιότητα (Conceptual Similarity). Ο αλγόριθμος ομοιότητας Wu και Pulmer πρόκειται για ένα μέτρο σημασιολογικής ομοιότητας μεταξύ εννοιών σε μία Οντολογία που περιορίζεται στις σχέσεις is-a. Η μετρική με την οποία υπολογίζεται είναι :

$$\text{sim}_{\text{Wu\&Pulmer}}(c_i, c_j) = \frac{2 * \min_{p \in \text{pths}(m\text{scs}(c_i, c_j), rt)} \text{len}_e(p)}{\min_{p \in \text{pths}(c_i, c_j)} \text{len}_e(p) + 2 * \min_{p \in \text{pths}(m\text{scs}(c_i, c_j), rt)} \text{len}_e(p)}$$

Όπου στον αριθμητή είναι η διπλάσια τιμή της απόστασης από την πιο κοντινή κοινή έννοια των υπό εξέταση εννοιών c_i, c_j , έως την ρίζα και στον παρονομαστή προστίθεται το ελάχιστο μονοπάτι από το c_i στο c_j συν την τιμή του αριθμητή.

Κανονικοποίηση μήκους μονοπατιού (Normalized Path Length). Ο αλγόριθμος Leacock και Chodorow είναι ουσιαστικά ένας μετασχηματισμός της απόστασης Rada σε ένα μέτρο ομοιότητας. Εδώ το ελάχιστο μονοπάτι μεταξύ δύο εννοιών περιορίζεται σε σχέσεις is-a και είναι κανονικοποιημένο, διαιρώντας το με την διπλάσια τιμή του μέγιστου βάθους της ιεραρχίας:

$$\text{sim}_{\text{Leacock\&Chodorow}}(c_i, c_j) = -\log \frac{\min_{p \in \text{pths}(c_i, c_j)} \text{len}_n(p)}{2 * \max_{c \in \text{pths}} (\max_{p \in \text{pths}(c, rt)} \text{len}_n(p))}$$

Περιεχόμενο της Πληροφορίας (Information Content). Ο αλγόριθμος Resnik αναφέρεται στο περιεχόμενο της πληροφορίας, ο οποίος υποστηρίζει ότι από ένα μεγάλο σώμα κειμένου και με τη χρήση στατιστικών εργαλείων μπορεί να διαπιστωθεί το πόσο γενική ή ειδική είναι μία έννοια. Η ιδέα αυτή στηρίζεται στη διαπίστωση ότι όσο περισσότερη πληροφορία

μοιράζονται δύο έννοιες τόσο πιο όμοιες είναι. Και σε αυτή την προσέγγιση οι σχέσεις is-a παίζουν καθοριστικό ρόλο.

Για την εύρεση της ομοιότητας σε δύο έννοιες c_i, c_j , ο αλγόριθμος λαμβάνει τη μέγιστη τιμή του περιεχομένου πληροφορίας του κοινού γονέα. Όπου από την θεωρία της πληροφορίας, το περιεχόμενο της πληροφορίας που μεταφέρει κάθε έννοια ορίζεται ως ο αρνητικός λογάριθμος της πιθανότητας εμφάνισης της έννοιας σε ένα σώμα κειμένου.

$$sim_{Resnik}(c_i, c_j) = \max[-\log P(mscs(c_i, c_j))]$$

Υβριδική μέθοδος (Hybrid Method). Οι Jiang και Conrath πρότειναν μία υβριδική μέθοδο που συνδυάζει το περιεχόμενο πληροφορίας που μεταφέρουν οι δύο εξεταζόμενες έννοιες και ο κοινός τους γονέας, σε αντίθεση με τον Resnik που λάμβανε υπόψη μόνο το περιεχόμενο πληροφορίας του γονέα. Η υβριδικότητα της μεθόδου υφίσταται λόγω των δύο βημάτων που υπάρχουν κατά τον υπολογισμό της, όπου στο πρώτο βρίσκεται η απόσταση των εννοιών:

$$dist_{Jiang\&Conrath}(c_i, c_j) = 2 * \log P(mscs(c_i, c_j)) - (\log P(c_i) + \log P(c_j))$$

Όπου η απόσταση των εννοιών c_i, c_j ορίζεται ως η διαφορά μεταξύ του αθροίσματος του περιεχομένου πληροφορίας των δύο εννοιών και την διπλάσια τιμή του περιεχομένου της πληροφορίας της πιο κοντινής κοινής έννοιας. Ενώ στο δεύτερο βήμα υπολογίζεται η ομοιότητα μεταξύ τους ως:

$$sim_{Jiang\&Conrath}(c_i, c_j) = \frac{1}{dist_{Jiang\&Conrath}(c_i, c_j)}$$

Γενικό μέτρο ομοιότητας (Universal Similarity Measure). Ο αλγόριθμος του Lin είναι μια γενικότερη προσέγγιση από αυτή των Wu και Pulmer, συμπεριλαμβάνοντας το περιεχόμενο πληροφορίας που μεταφέρουν οι δύο εξεταζόμενες έννοιες και ο κοινός τους γονέας. Σε αντίθεση με την προσέγγιση του Resnik και Jiang & Conrath που ενδιαφέρονται κυρίως για τις κοινές πληροφορίες που μοιράζονται δύο έννοιες, εδώ γίνεται αναφορά και στις διαφορές τους. Ο τύπος υπολογισμού της ομοιότητας θυμίζει πολύ αυτό του Jiang & Conrath, που αντί να χρησιμοποιηθεί η αφαίρεση μεταξύ των μέτρων χρησιμοποιείται ο λόγος.

$$sim_{Lin}(c_i, c_j) = \frac{2 * \log P(mscs(c_i, c_j))}{(\log P(c_i) + \log P(c_j))}$$

Μεσαίες- Ισχυρές σχέσεις (Medium- Strong Relations). Οι Hirst και St-Onge προτείνουν μία προσέγγιση κατά την οποία διακρίνονται τέσσερις τύποι σχέσεων μεταξύ δύο εννοιών *πολύ- δυνατή, δυνατή, μέτρια- δυνατή* και *αδύνατη*.

$$rel_{Hist\&St-Onge}(c_i, c_j) = \begin{cases} 3 * C(\text{πολύ} - \text{δυνατή}) & \text{if } C > 2 \\ 2 * C(\text{δυνατή}) & \text{if } C = 2 \\ C - \min_{p \in pths(c_i, c_j)} len_e(p) - K * \min_{p \in pths(c_i, c_j)} trn(p) & \text{if } 1 < C < 2 \\ 0(\text{αδύνατη}) & \text{if } C = 1 \end{cases}$$

όπου C και K σταθερές.

Κυρίως η ερευνά τους κινήθηκε πάνω στο λεξικό WordNet και τις συσχετίσεις μεταξύ των λέξεων, έτσι καθόριζαν τις πολύ-δυνατές συσχετίσεις ως σχέσεις πάνω σε μόνο μία λέξη και την κυριολεκτική επανάληψή της. Δυνατή χαρακτηρίζαν μια σχέση δύο λέξεων που είτε οι δύο λέξεις ανήκαν στο ίδιο σύνολο συνωνύμων - τα οποία συνώνυμα συνδέονται με μία οριζόντια σχέση μεταξύ τους - είτε η μία λέξη είναι σύνθετη και η δεύτερη αποτελεί τμήμα της πρώτης και υπάρχει σύνδεση is-a-kind-of που να συνδέει την πρώτη και τη δεύτερη λέξη. Μέτρια-δυνατη χαρακτηρίζουν μία σχέση όταν μεταξύ των λέξεων υπάρχει ένα ελάχιστο μονοπάτι και δεν υπάρχουν αρκετές αλλαγές στις κατευθύνσεις των ακμών που συναντώνται, όσο πιο μεγάλο το μονοπάτι και όσο περισσότερες αλλαγές κατεύθυνσης γίνονται τόσο μικρότερη συσχέτιση έχουν.

Γενικευμένη σταθμισμένη συντομότερη διαδρομή (Generalized Weighted Shortest Path). Ο Bulskon προτείνει μια γενίκευση στην προσέγγιση του μικρότερου μήκους μονοπατιού με βάρη έτσι ώστε να μη συμπεριλαμβάνει μόνο τις σχέσεις is-a αλλά και τις υπόλοιπες που θα μπορούσαν να υπάρξουν σε μία Οντολογία. Αυτό επιτυγχάνεται με την εισαγωγή παραγόντων ομοιότητας για τις σημασιολογικές σχέσεις. Έτσι υποθέτοντας ότι υπάρχουν k διαφορετικές σημασιολογικές σχέσεις r^1, \dots, r^2 και οι αντίστοιχοι παράγοντες ομοιότητας ρ_1, \dots, ρ_2 τότε η μετρική του μικρότερου μήκους μονοπατιού θα επεκτεινόταν σε:

$$sim_{WSP}(x, y) = \max_{j=1, \dots, m} \{ \sigma^{s(P^j)} \gamma^{g(P^j)} \rho^{r^1(P^j)} \dots \rho^{r^k(P^j)} \}$$

Στον πίνακα που ακολουθεί παρουσιάζονται συνοπτικά κάποια από τα χαρακτηριστικά των μετρικών που συζητήθηκαν. Η στήλη *Χρήση ελάχιστου μονοπατιού root/ mscs* αναφέρεται στους αλγόριθμους που χρησιμοποιούν ελάχιστο μονοπάτι μεταξύ της ρίζας και της κοινής έννοιας μεταξύ των εννοιών που γίνεται η αναζήτηση της ομοιότητάς τους. Η στήλη *Χρήση βάθους έννοιας root/ mscs* αναφέρεται στους αλγόριθμους που χρησιμοποιούν τον αριθμό των απογόνων για κάθε έννοια, η οποία έννοια ανήκει στο μικρότερο μονοπάτι μεταξύ της ρίζας, και της κοινής έννοιας μεταξύ των εννοιών που γίνεται η αναζήτηση της ομοιότητάς τους.

Πίνακας 5-11. Χαρακτηριστικά μετρικών

Κατηγορία	Περιγραφή	Αλγόριθμος	Σηματολογική	Ιδιότητες	Βάρος	Συμμετρίκότητα	Χρήση ελάχιστου μονοπατιού	Χρήση ελάχιστου μονοπατιού root/mscs	Χρήση βάθους έννοιας	Χρήση βάθους έννοιας root/mscs
Μέγεθος Μονοπατιού	Μικρότερο μήκος μονοπατιού	Rada	Απόσταση	is-a		x	x			
	Μικρότερο μήκος μονοπ. με βάρη	Bulskov	Απόσταση	is-a	x		x			
Σχετικό Βάθος	Κλιμάκωση σχετικού βάθους	Sussna	Απόσταση	όλες	x	x	x	x	x	
	Εννοιολογική ομοιότητα	Wu και Pulmer	Ομοιότητα	is-a		x	x	x	x	
	Κανονικοπ μήκους μονοπατιού	Leacock και Chodorow	Ομοιότητα	is-a		x	x			
Βασισμένες στη Συλλογή	Περιεχόμενο της Πληροφορίας	Resnik	Ομοιότητα	is-a		x		x		x
	Υβριδική μέθοδος	Jiang και Conrath	Απόσταση/ Ομοιότητα	is-a		x	x		x	
	Γενικό μέτρο ομοιότητας	Lin	Ομοιότητα	is-a		x	x	x	x	x
Πολλαπλά Μονοπάτια	Μεσαίες-Ισχυρές σχέσεις	Hirst και St-Onge	Συνάφεια	όλες		x	x			
	Γενικευμένη σταθμισμένη συντομότερη διαδρομή	Bulskov	Συνάφεια	όλες	x	x	x			

Μία πάρα πολύ ενδιαφέρουσα και ολοκληρωμένη προσέγγιση σχετικά με την ομοιότητα εννοιών σε μία Οντολογία είναι αυτή του Hotho και συν. (2002). Σε αυτή λαμβάνονται υπόψη τα τρία βασικά χαρακτηριστικά μίας Οντολογίας δηλαδή η ιεραρχία, οι σχέσεις μεταξύ των εννοιών και οι τιμές που λαμβάνουν. Έτσι ορίζει μία μετρική που κανονικοποιεί στο διάστημα [0, 1] το άθροισμα τριών διαφορετικών ομοιοτήτων, αυτά της:

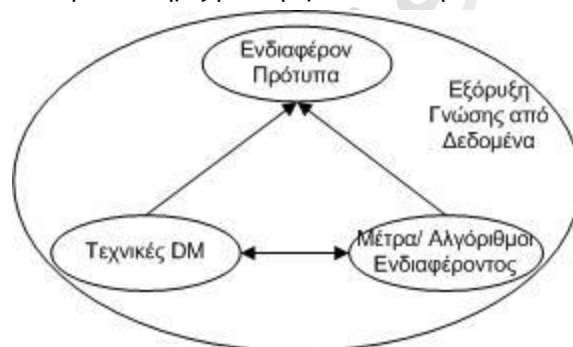
- Ομοιότητας ταξινόμιας (Taxonomy Similarity), που υπολογίζει την ομοιότητα ανάμεσα σε δύο στιγμιότυπα με βάση τις έννοιες που ανήκουν και την θέση τους στην ιεραρχία.
- Ομοιότητα σχέσης (Relation Similarity), που υπολογίζει την ομοιότητα δύο στιγμιότυπων με βάση τις σχέσεις που τα συνδέουν με άλλα αντικείμενα.
- Ομοιότητα τιμών (Attribute Similarity), που υπολογίζει την ομοιότητα μεταξύ δύο περιπτώσεων με βάση τα χαρακτηριστικά τους και τις τιμές που λαμβάνουν αυτά. Εδώ ακόμα πιο εξειδικευμένες μετρικές ορίζονται ώστε να χρησιμοποιηθούν για να μπορέσουν να προσεγγίσουν την ομοιότητα/ απόσταση σε περίπλοκους τύπους τιμών όπως τα αλφαριθμητικά.

Εδώ δεν έγινε περαιτέρω αναφορά των μετρικών που χρησιμοποιούν (αν και έχουν παρουσιαστεί παραπάνω), για να τονιστούν κυρίως τα δομικά χαρακτηριστικά που

χρειάζονται για τον υπολογισμό μιας ολοκληρωμένης σημασιολογικής συνάφειας σε μία Οντολογία, αλλά και να επισημανθεί ότι οι μετρικές που μπορούν να χρησιμοποιηθούν είναι δυνατόν να ποικίλουν ανάλογα με τις απαιτήσεις και τις ανάγκες του εκάστοτε προβλήματος που γίνεται προσπάθεια να επιλυθεί.

6. DM και Οντολογία

Η DM έχει ως στόχο την παραγωγή προτύπων που φανερώνουν στον ειδικό κρυμμένες συσχετίσεις μεταξύ των δεδομένων μέσα σε βάσεις δεδομένων. Βέβαια το τι πρότυπα θα εξαχθούν εξαρτάται από τις απαιτήσεις του ειδικού και τους τύπους των δεδομένων, παραδείγματος χάρη χωρικών, χρονικών. Έτσι ή ειδικές τεχνικές της DM θα πρέπει να ακολουθηθούν (κανόνες συσχετίσεων, κατηγοριοποίηση, συσταδοποίηση, τεχνικές εξόρυξης χρονικής και χωρικής γνώσης), αλλά και ειδικά μέτρα, ή αλγόριθμοι θα πρέπει να οριστούν ώστε να αντανakλούν τα ενδιαφέροντα των ειδικών και να υποστηρίζουν τις τεχνικές της DM. Η ενσωμάτωση του ενδιαφέροντος των ειδικών γίνεται με τις τρεις παρακάτω προσεγγίσεις: αντικειμενικά, υποκειμενικά και σημασιολογικά μέτρα. Αυτό που εξετάζεται στην παρούσα διατριβή είναι μία υποκατηγορία της δεύτερης προσέγγισης που κάνει χρήση των Οντολογιών και στοχεύει κυρίως να καλύψει μέτρα όπως η εξόρυξη ποικιλόμορφων, καινοτόμων, απροσδόκητων ή μη αναμενόμενων προτύπων. Η ενσωμάτωση γίνεται με κατάλληλες μετρικές όπως αυτές που αναφέρθηκαν στην ενότητα *σημασιολογική συνάφεια*, που είτε μπορούν να χρησιμοποιηθούν αυτούσια είτε υποστηρικτικά ή συνδυαστικά με άλλα μέτρα ενδιαφέροντος, όπως είναι η υποστήριξη και η εμπιστοσύνη.



Εικόνα 6-1. Εξόρυξη γνώσης από δεδομένα

Όπως αναφέρθηκε, για όλες τις τεχνικές της DM ορίζονται μετρικές ή αλγόριθμοι πάνω στις οποίες στηρίζεται η απόφαση για την εξόρυξη ενδιαφέροντων προτύπων. Οι Geng και Hamilton (2006) διακρίνουν τρεις εναλλακτικές μεθόδους, που είναι πιθανόν να ακολουθηθούν για να ληφθεί η απόφαση ενδιαφέροντος των προτύπων, και τις προσδιορίζουν ως μεθόδους *αποφασιστικότητας ενδιαφέροντος* (interestingness determination):

- Κατηγοριοποίηση των προτύπων ως ενδιαφέροντα ή μη ενδιαφέροντα
- Καθορισμό μίας σχέσης η οποία ορίζει πως ένα πρότυπο είναι πιο ενδιαφέρον από το άλλο
- Ταξινόμηση των προτύπων

Η χρήση των μετρικών ή των αλγορίθμων μπορεί να χρησιμοποιηθεί πριν το στάδιο της εξόρυξης γνώσης, δηλαδή κατά την KDD στο στάδιο του μετασχηματισμού των δεδομένων ώστε να ελαχιστοποιηθεί ο όγκος τους, να διαχειριστούν ελλιπή δεδομένα ή λανθασμένα δεδομένα, κατά τη διάρκεια της εξόρυξης γνώσης, όπου και εδώ γίνεται προσπάθεια για περιορισμό του χώρου αναζήτησης και έτσι να βελτιωθεί η αποτελεσματικότητα της εξόρυξης, και τέλος μετά το πέρας της εξόρυξης γνώσης που τα πρότυπα μπορούν να ομαδοποιηθούν, να ταξινομηθούν ώστε να είναι κατανοητά στους χρήστες.

Κανόνες Συσχετίσεων. Με την τεχνική των κανόνων συσχετίσεων γίνεται προσπάθεια για εξόρυξη συχνών στοιχειοσυνόλων μέσα από μία βάση δεδομένων. Η έρευνα σχετικά με την χρήση Οντολογιών για την εξόρυξη τέτοιου είδους κανόνων συνήθως σχετίζεται με την ενσωμάτωσή τους στον Αρρίσι αλγόριθμο, παρά στην παραγωγή κανόνων εξ' ολοκλήρου με βάση τις Οντολογίες. Τα συνήθη προβλήματα που τίθενται εδώ είναι:

- Η εξόρυξη μεγάλου συνόλου κανόνων, το οποίο δυσχεραίνει τη διαχείριση τους.
- Μέσα σε αυτό το σύνολο υπάρχουν κανόνες που επαναλαμβάνονται χωρίς να προσφέρουν κάποια επιπλέον πληροφορία.
- Κάποιοι είναι αρκετά γενικοί ή αρκετά ειδικευμένοι για να χρησιμοποιηθούν.
- Πολλοί από αυτούς είναι τετριμμένοι ή αντιπροσωπεύουν μία λανθασμένη γνώση.
- Μη ισχυροί κανόνες, με μικρή υποστήριξη ή εμπιστοσύνη έχουν μεγάλη σημασιολογική αξία αλλά δε μπορούν εύκολα να αναδειχθούν.

Προφανώς, για την επίλυση αυτών των προβλημάτων γίνεται προσπάθεια για συνδυασμό των αντικειμενικών μέτρων, όπως είναι η υποστήριξη και η εμπιστοσύνη, με τα υποκειμενικά μέτρα, που στηρίζονται στις Οντολογίες.

Ο Bellandi και συν. (2007) προτείνουν μία διαδικασία που προηγείται της εξόρυξης κανόνων συσχετίσεων που ουσιαστικά προετοιμάζει τα δεδομένα ώστε να αυξηθεί η ποιότητα των κανόνων κατά την εκτέλεση του Apriori. Αυτό επιτυγχάνεται με το κλάδεμα (pruning) της Οντολογίας ανάλογα με τις έννοιες που αποτελούν τη σύλληψη του χρήστη, απομονώνοντας άχρηστα δεδομένα. Αυτό έχει ως αποτέλεσμα να μειώνονται τα βήματα του αλγορίθμου, ελαχιστοποιώντας την πολυπλοκότητα χρόνου εκτέλεσης του. Επιπλέον, με την αξιοποίηση των περιορισμών που περιλαμβάνουν οι Οντολογίες μπορούν να ξεπεραστούν εμπόδια όπως η έλλειψη δεδομένων μέσα στη βάση δεδομένων.

Η πιο κλασική τεχνική που ακολουθείται είναι η γενίκευση κανόνων, η οποία βασίζεται στην ιεραρχική δομή των Οντολογιών, προσπαθώντας να εξάγουν γενικότερους-ισχυρότερους κανόνες έχοντας μεγαλύτερο βαθμό υποστήριξης (Ferraz και Garcia, 2008). Σε αυτή την λογική οι Xiong και συν. (2010) εισάγουν από την θεωρία της πληροφορίας τον βαθμό της νέας πληροφορίας που μπορεί να παρουσιαστεί από την χρήση της ειδικεύσης ή εξειδίκευση των κανόνων.

Άλλος ένας εναλλακτικός τρόπος είναι να χρησιμοποιηθούν μέτρα όπως αυτά της προηγούμενης ενότητας ώστε να διατηρηθούν στοιχειοσύνολα που διαθέτουν αυξημένη χρησιμότητα, ευκρίνεια, αξιοπιστία (Sebastian και συν., 2009). Εδώ, συνήθως πρώτα εξάγονται οι κανόνες συσχετίσεων και έπειτα κλαδεύονται αυτή που δεν προσφέρουν κάποιο σημασιολογικό ενδιαφέρον.

Στην παρούσα διατριβή υποστηρίζεται ότι με βάση των μετρικών που ορίζονται μέσα στις Οντολογίες, μπορούν να αξιολογηθούν οι κανόνες συσχετίσεων, τόσο κατά τη διάρκεια όσο και με το πέρας του Apriori αλγορίθμου. Επίσης, μπορούν να αντικαταστήσουν τη μετρική της υποστήριξης για να προωθηθούν κατεξοχήν σημασιολογικά ενδιαφέροντες κανόνες, που ανάλογα με τη μετρική μπορεί το ενδιαφέρον να προσδιορίζεται σε μη αναμενόμενη γνώση, καινοτομία αλλά και στη χρησιμότητα.

Κατηγοριοποίηση. Η συγκεκριμένη τεχνική προσπαθεί να κατηγοριοποιήσει νέες εγγραφές σε ένα συγκεκριμένο και διακριτό σύνολο κατηγοριών. Τα βασικά προβλήματα που παρουσιάζονται εδώ είναι:

- Η κατηγοριοποίηση μιας εγγραφής σε μία λανθασμένη κατηγορία, γεγονός το οποίο μπορεί να επιτρέψει να διαιωνιστεί αυτή η λανθασμένη ενέργεια.
- Τα ελλιπή δεδομένα, τα οποία μπορούν να παίξουν καθοριστικό ρόλο στη σωστή κατηγοριοποίηση των νέων εγγραφών.
- Χρήση αποδοτικών αποστάσεων, σε αλγόριθμους όπως οι K- κοντινότεροι γείτονες (k- nearest neighbors- KNN), νευρωνικά δίκτυα.
- Ποια ιεραρχική δομή και ποια χαρακτηριστικά θα πρέπει να χρησιμοποιήσουν οι αλγόριθμοι των δέντρων αποφάσεων.
- Τα δεδομένα εκπαίδευσης δεν είναι πάντα ανεξάρτητα, επομένως τεχνικές όπως η Bayesian κατηγοριοποίηση είναι πιθανό να μην είναι εξίσου αποτελεσματικές όπως άλλες που μπορούν να αξιοποιήσουν την εξάρτηση των δεδομένων.

Ο Wei και συν. (2008), ενσωματώνει τις κατηγορίες μέσα σε μία Οντολογία και ορίζει μια μετρική σχετικά με την συνάφεια μίας νέας εγγραφής με τις αναπαριστάμενες κατηγορίες. Ο Zhang και συν. (2002) εκμεταλλεύονται την ιεραρχική δομή των Οντολογιών για να αναπτύξουν έναν αλγόριθμο δέντρου αποφάσεων, ώστε να επιλέγεται ο καλύτερος εσωτερικός κόμβος μέσα σε αυτά τα δέντρα. Οι Bouza και συν. (2008) βασίζονται στο γεγονός ότι τα στοιχεία εκπαίδευσης δεν είναι ανεξάρτητα μεταξύ τους και ότι οι σχέσεις τους μπορούν να δώσουν πλούσια πληροφορία. Έτσι χρησιμοποιούν Οντολογίες για να μεγιστοποιήσουν το κέρδος πληροφορίας που μπορούν να δώσουν οι εσωτερικοί κόμβοι των δέντρων αποφάσεων. Οι Kharsikar και συν. (2007) εξετάζουν την ενσωμάτωση των

Οντολογιών μέσα στην KNN, κάνοντας χρήση των σημασιολογικών αποστάσεων μέσα στον αλγόριθμο του KNN για να βρεθούν οι κατάλληλες ομοιότητες μιας νέας εγγραφής με μία κατηγορία.

Συσταδοποίηση. Αυτή η τεχνική συσταδοποιεί, δηλαδή ομαδοποιεί ένα σύνολο εγγραφών με βάση κάποιο κριτήριο ομοιότητας. Σε αντίθεση με την κατηγοριοποίηση δεν είναι γνωστές οι διαφορετικές ομάδες που μπορεί να γίνει η ομαδοποίηση. Ένα από τα βασικά της χαρακτηριστικά είναι ότι όλες οι διαδικασίες που την υλοποιούν (ιεραρχικοί αλγόριθμοι, διαμεριστικοί και άλλες προηγμένες τεχνικές), χρησιμοποιούν ως κύριο εργαλείο τους ένα μέτρο ομοιότητας μεταξύ των εγγραφών. Αυτό έρχεται από τον ορισμό του προβλήματος της συσταδοποίησης, που αναζητά μία απεικόνιση που θα μπορεί να αντιστοιχεί μια εγγραφή σε μία συστάδα, έτσι ώστε η νέα εγγραφή να έχει τη μεγαλύτερη ομοιότητα με τις εγγραφές τις συστάδας που πραγματικά ανήκει από ότι με τις εγγραφές μίας άλλης συστάδας. Εδώ γίνεται αντιληπτό ότι οι Οντολογίες μπορούν να παίξουν ένα καθοριστικό ρόλο στην εύρεση αυτής της απεικόνισης, ενσωματώνοντας όλα τα θετικά από τη χρήση των σημασιολογικών αποστάσεων ενός πεδίου (Betat & συν., 2008).

Εκτός της κλασικής χρήσης της μεθόδου συσταδοποίησης, μεγάλο ενδιαφέρον παρουσιάζεται στη βιβλιογραφία σχετικά με την ενσωμάτωση των Οντολογιών στο πεδίο της συσταδοποίησης κειμένων. Όπου Οντολογίες όπως η WordNet χρησιμοποιούνται για να κάνουν προ-επεξεργασία στα διανύσματα που παράγονται από τα κείμενα, ώστε να επιτυγχάνονται καλύτερα αποτελέσματα συσταδοποίησης και να εκμεταλλεύονται τα εννοιολογικά χαρακτηριστικά των κειμένων (Hotho & συν., 2001; Jing & συν., 2006).

Συχνά προβλήματα που επιζητούν λύση στην τεχνικές της συσταδοποίησης είναι:

- Ο χειρισμός ακραίων εγγραφών που μπορούν να επηρεάζουν αρνητικά την ομοιογένεια μίας συστάδας.
- Το μέγεθος των εγγραφών που πρέπει να χρησιμοποιηθούν, καθώς σε μεγάλα σύνολα εγγραφών οι αποστάσεις τείνουν να είναι ίδιες.
- Να υπάρχει η δυνατότητα ερμηνείας των αποτελεσμάτων.

Εξόρυξη χωρικής- χρονικής γνώσης. Περικλείει όλες εκείνες τις τεχνικές που προσπαθούν να εξορύξουν ενδιαφέροντα χωρικά- χρονικά πρότυπα από χώρο-χρονικές βάσεις δεδομένων. Ουσιαστικά είναι ένα υπερσύνολο των υπόλοιπων τεχνικών που συζητήθηκαν καθώς κάτω από την λογική της ύπαρξης ενός χώρο-χρονικού πλέγματος μπορούν να αναζητηθούν κανόνες συσχετίσεων και να πραγματοποιηθούν τεχνικές κατηγοριοποίησης ή συσταδοποίησης (Xu & Huang,). Επομένως, η σχέση τους με τις Οντολογίες δεν μπορεί να μην είναι εφικτές και μη χρήσιμες (Hwang, 2004). Οι ίδιες ανάγκες που απέρρεαν από τις προηγούμενες τεχνικές ισχύουν και εδώ, όπως και η χρήση των Οντολογιών πριν την εκτέλεση μίας χώρο-χρονικής τεχνικής, κατά την διάρκεια εκτέλεσής της αλλά και μετά το πέρας της για την αξιολόγηση των προτύπων. Απλά εδώ θα πρέπει να προσαρμοστούν και οι Οντολογίες έτσι ώστε να μπορούν να αναπαριστούν χωρική- χρονική γνώση (Wang & Li, 2005).

ΜΕΡΟΣ Β – ΑΝΑΛΥΣΗ, ΣΧΕΔΙΑΣΗ & ΥΛΟΠΟΙΗΣΗ

7. Εισαγωγή

Σε αυτό το μέρος της διπλωματικής θα παρουσιαστούν τα βήματα ανάπτυξης ενός συστήματος που ενσωματώνει τις Οντολογίες στην τεχνική Κανόνων Συσχετίσεων της Εξόρυξης Γνώσης από Δεδομένα. Η μέθοδος που ακολουθείται είναι αυτή του Αντικειμενοστραφούς Μοντέλου (Rational Unified Process- RUP) (Sommerville 2001), όπου ο κύκλος ζωής του αποτελείται από τις ακόλουθες τέσσερις διαδικασίες: a) Ανάλυση Απαιτήσεων, b) Ανάλυση και Σχεδίαση, c) Υλοποίηση, d) Έλεγχος. Για τη μοντελοποίηση της συμπεριφοράς του συστήματος χρησιμοποιείται η γλώσσα UML (Unified Modeling Language) (Booch και συν., 2004).

8. Ανάλυση απαιτήσεων

Σε αυτή την ενότητα περιγράφεται, σε φυσική γλώσσα, το σύστημα που θα υλοποιηθεί. Αρχικά θα τεθεί ο στόχος του και θα περιγραφεί αναλυτικά το πρόβλημα. Έπειτα θα εξεταστεί

η παρούσα κατάσταση και στο τέλος θα αναλυθούν οι λειτουργικές και μη-λειτουργικές απαιτήσεις.

8.1 Σκοπός συστήματος

Ο στόχος του συστήματος είναι να ενσωματώσει και να αξιοποιήσει στον αλγόριθμο εύρεσης συχνών στοιχειοσυνόλων Apriori που υλοποιείται από το εργαλείο Weka (Weka 2011), τη γνώση που μπορεί να αποκομισθεί από τις OWL Οντολογίες έτσι ώστε:

1. Να ελαχιστοποιηθούν οι λανθασμένοι σημασιολογικά παραγόμενοι κανόνες.
2. Να αναδειχτούν οι κανόνες με αυξημένο σημασιολογικό ενδιαφέρον.
3. Να επιτευχθεί βελτίωση της λειτουργίας του αλγορίθμου Apriori.

Ο απώτερος στόχος έγκειται στην επέκταση του συστήματος ώστε να υποστηρίζει ομοιόμορφα και άλλες τεχνικές Εξόρυξης Γνώσης από Δεδομένα, όπως είναι η συσταδοποίηση, η κατηγοριοποίηση, οι τεχνικές εξόρυξης χρονικής και χωρικής γνώσης.

8.2 Παρούσα κατάσταση

Η εφαρμογή που θα υλοποιηθεί θα είναι ουσιαστικά ένα υποσύνολο του γενικότερου συστήματος Pattern-Minner. Το τελευταίο ακολουθεί την αρχιτεκτονική του PANDA (Rizzi και συν. 2003) και είναι ένα XML-based PBMS το οποίο είναι ανεξάρτητο από την μηχανή αποθήκευσης των προτύπων αλλά και από τη μηχανή εκτέλεσης εργασιών Data Mining. Στην εικόνα που ακολουθεί (Εικόνα 8-1. Αρχιτεκτονική για PBMS ενισχυμένο με οντολογίες) παρουσιάζεται η αρχιτεκτονική που έχει προταθεί και η οποία συμπεριλαμβάνει τον τρόπο εισαγωγής γνώσης από τις Οντολογίες.

Βέβαια για την υλοποίηση του Pattern-Minner χρησιμοποιήθηκαν συγκεκριμένα εργαλεία. Για την αποθήκευση των προτύπων επιλέχτηκε η XML-based γλώσσα PMML (2011) και για μηχανή Βάσης Προτύπων η Oracle Berkley DB (2011). Ενώ ως μηχανή για την εκτέλεση των Data Mining τεχνικών επιλέχτηκε το Weka. Ο κορμός του Pattern-Minner είναι υλοποιημένος σε Java και η επικοινωνία του συστήματος με το χρήστη γίνεται μέσω μιας ενιαίας και ομοιόμορφης διεπαφής. Ένα παράδειγμα της διεπαφής παρουσιάζεται στην Εικόνα 8-2. Στιγμιότυπο από την εφαρμογή Pattern-Miner ver. 2.0

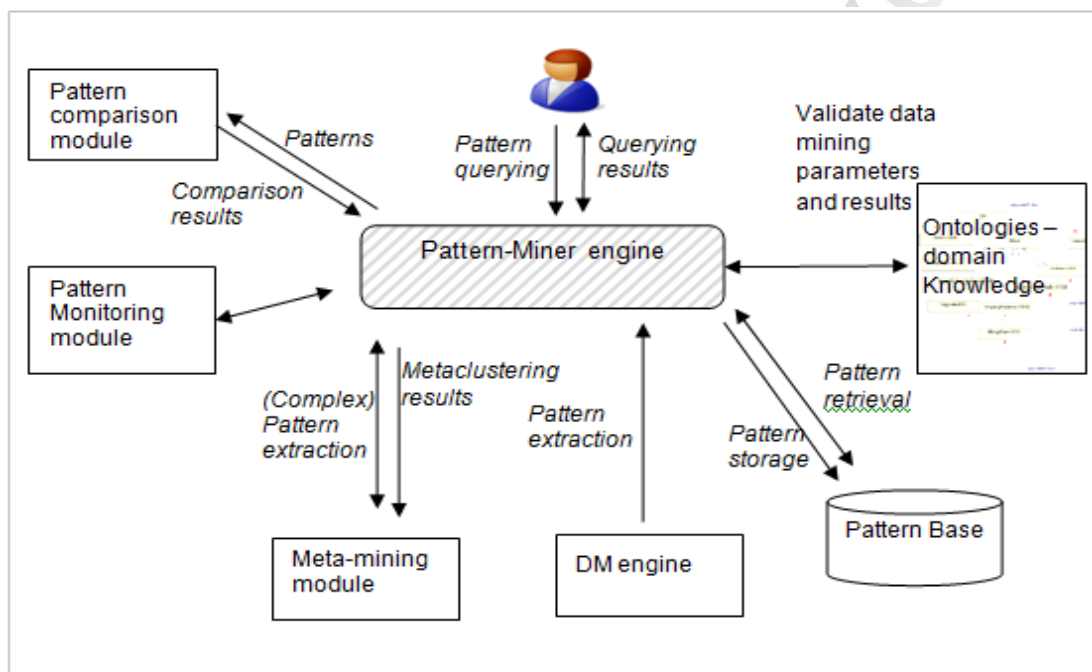
Επίσης, στην Εικόνα 8-2. Στιγμιότυπο από την εφαρμογή Pattern-Miner ver. 2.0, μπορούμε να δούμε τη διεπαφή που έχει υλοποιηθεί για να εισάγονται οι παράμετροι του Apriori αλγορίθμου. Κατά την εκτέλεσή του και αφού έχει επιλεγεί το κουμπί «Execute algorithm and store PMML pattern» καλείται η κλάση Apriori του πακέτου weka.associations του Weka. Η τελευταία επιστρέφει ένα .arff αρχείο, δηλαδή μία φόρμα που ακολουθεί το Weka για να εξάγει τα αποτελέσματα των αλγορίθμων. Στο τέλος καλείται η εσωτερική μέθοδος associationRulesCMD για την παραγωγή και αποθήκευση των PMML αρχείων σε έναν κατάλληλο Container στην Βάση Προτύπων.

Συγκεκριμένα η υλοποίηση του Pattern-Minner αποτελείται από οχτώ πακέτα. Αυτά είναι τα :

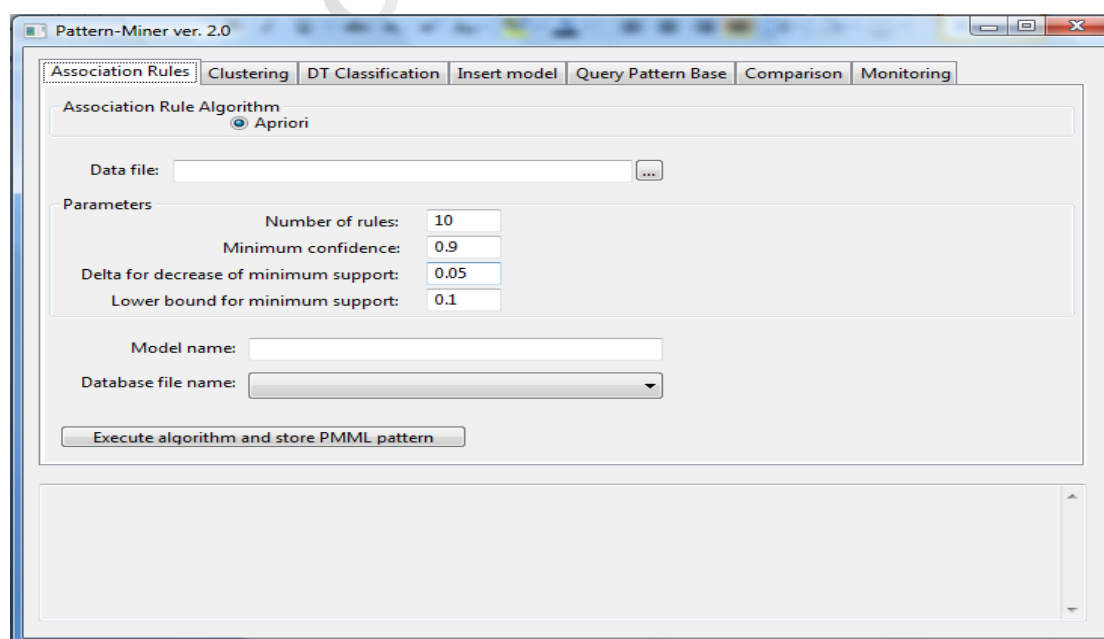
Πίνακας 8-1. Πακέτα υλοποίησης Pattern-Minner

Πακέτο	Περιγραφή
GUIdm	Περιέχει τις κλάσεις για την διεπαφή
associationRules	Περιέχει τις κλάσεις σχετικά με την αποθήκευση και ανάκτηση των προτύπων που παράγονται από τον Apriori
classifiers	Κλάσεις σχετικά με τη συσταδοποίηση
clusterers	Κλάσεις σχετικά με την κατηγοριοποίηση
comparison	Κλάσεις σχετικά με τη σύγκριση μοντέλων προτύπων

Πακέτο	Περιγραφή
core	Κλάσεις που αποτελούν τον βασικό πυρήνα του Pattern-Minner και είναι οι δομές που χρησιμοποιούνται για την διαχείριση των προτύπων.
monitoring	Κλάσεις σχετικά με την προβολή των αποτελεσμάτων από την σύγκριση μοντέλων
storedbxml	Κλάσεις για την αποθήκευση προτύπων



Εικόνα 8-1. Αρχιτεκτονική για PBMS ενισχυμένο με οντολογίες
 Πηγή: Kotsifakos και συν. (2008: 679)



Εικόνα 8-2. Στιγμιότυπο από την εφαρμογή Pattern-Miner ver. 2.0

8.3 Προσδιορισμός συστήματος

Σύμφωνα με τον ορισμό του συστήματος, στο εργαλείο που θα πρέπει να γίνει εστίαση είναι ο αλγόριθμος *Argiori* του *Weka*. Ο συγκεκριμένος αλγόριθμος για να εκτελεστεί θα πρέπει να του εισαχθούν ένα σύνολο παραμέτρων που θα καθορίζουν ποια πρότυπα θα παραχθούν, αλλά φυσικά και το αρχείο με τα δεδομένα με το οποίο θα γίνει η εκπαίδευση του. Το αρχείο αυτό είναι ένα *.arff* αρχείο που ακολουθεί το πρότυπο του *csv* αλλά είναι ειδικευμένο για το *Weka*.

Για την μείωση των παραγόμενων κανόνων συσχετίσεων, αλλά και για την διατήρηση των κανόνων με αυξημένο σημασιολογικά ενδιαφέρον θα γίνει χρήση των *Οντολογιών*. Ωστόσο το σύστημα θα πρέπει να είναι ανεξάρτητο από τη χρήση συγκεκριμένης *Μηχανής Συλλογιστικής Ανάλυσης*, παραδείγματος χάρη *Jena* ή *Pellet*.

Προφανώς, για τη βελτίωση του αλγορίθμου *Argiori* θα χρειαστεί να γίνει μια επέκταση του, ώστε να εισάγεται η πληροφορία από την *Οντολογία* μέσα στην διαδικασία του. Την πληροφορία που θα επιδιώκει να πάρει από την *Οντολογία* θα είναι οι αποστάσεις μεταξύ δύο εννοιών της, ή διαφορετικά θα πρέπει να μπορεί να εξετάζει πόσο όμοιες είναι μεταξύ τους δύο έννοιες που υπάρχουν στην *οντολογία* και είναι πιθανόν να δημιουργήσουν έναν επιθυμητό κανόνα. Φυσικά και εδώ οι μετρικές που θα χρησιμοποιηθούν για να υπολογίζουν την απόσταση ή την ομοιότητα των εννοιών θα πρέπει να είναι ανεξάρτητες τους συστήματος και να είναι στη δυνατότητα του χρήστη να επιλέγεται η επιθυμητή. Επίσης, καθώς ενδιαφερόμαστε για κανόνες με αυξημένο ενδιαφέρον θα πρέπει να δίνεται στο χρήστη η δυνατότητα επιλογής του πεδίου τιμών της μετρικής, που θα αντανακλά την απόσταση ή ομοιότητα των εννοιών για τις οποίες το σύστημα θα προσπαθεί να παραγάγει κανόνες συσχετίσεων.

Για την εύρυθμη λειτουργία του συστήματος θα πρέπει επίσης να δημιουργηθεί μια δομή που θα αντιστοιχεί τα στοιχεία των δεδομένων εκπαίδευσης με τις έννοιες της *Οντολογίας*. Κατά κανόνα τα στοιχεία των δεδομένων εκπαίδευσης αντιστοιχούν σε πεδία μιας βάσης δεδομένων, αλλά αυτά με την σειρά τους δεν αντιστοιχούν σε κάποια έννοια μιας *Οντολογίας*. Βέβαια, για την εκτέλεση της νέας διαδικασίας θα πρέπει να γίνει χαρτογράφηση, (*mapping*) ώστε να δοθεί μια σημασιολογική αναπαράσταση αυτών των δεδομένων.

Τέλος, μετά την εκτέλεση του αλγορίθμου θα πρέπει να γίνεται αποθήκευση των αποτελεσμάτων και να δίνεται η δυνατότητα στο χρήστη να μπορεί να τα ανακτήσει.

8.4 Μη λειτουργικές απαιτήσεις

Ο πρώτος εμφανής περιορισμός που παρουσιάζεται είναι ότι πρέπει να ακολουθηθεί η αρχιτεκτονική του ήδη υλοποιημένου συστήματος. Αυτό είναι φρόνιμο, καθώς θα διατηρήσει τη συνοχή και τη σταθερότητα της προηγούμενης έκδοσης, ενώ θα παραμείνει εύκολα επεκτάσιμη και τροποποιήσιμη. Κατ' επέκταση, θα πρέπει να διατηρηθεί το ίδιο μοτίβο στην εμφάνιση και τη διαχείριση της διεπαφής, ώστε να παρουσιάζεται στο χρήστη ως μία ενιαία εφαρμογή, πράγμα που αυξάνει τη φιλικότητά της.

Η γλώσσα ανάπτυξης είναι εύλογο ότι θα πρέπει να παραμείνει ως έχει. Άλλωστε η *Java* είναι μια γλώσσα που ενδείκνυται για τέτοιου είδους εφαρμογές. Σχετικά με τα πακέτα που υλοποιούν τις εφαρμογές θα πρέπει να τροποποιηθούν στο ελάχιστο ενώ συνίσταται η επέκτασή τους. Έτσι, θα διατηρηθούν οι ήδη υπάρχον δυνατότητες ενώ παράλληλα με την κληρονομικότητα θα μπορέσουν να προστεθούν καινούριες. Χαρακτηριστικό παράδειγμα είναι η επέκταση του αλγορίθμου *Argiori* που θα πρέπει να εκτελεί αβίαστα τις υπάρχουσες λειτουργίες του, αλλά είναι αναγκαίο να επεκταθεί έτσι ώστε να μπουν εμβόλιμα οι πρόσθετες δυνατότητες των *Οντολογιών*.

Άλλος ένα περιορισμός που θα οδηγήσει στην επιλογή των κατάλληλων εργαλείων είναι ότι το σύστημα θα πρέπει να μπορεί να διαχειρίζεται *OWL Οντολογίες*. Είναι ευνόητο ότι θα πρέπει να καλύπτεται όσο το δυνατόν πιο πρόσφατη έκδοση της *OWL* αλλά προφανώς και να μπορεί να εκμεταλλεύεται *Οντολογίες* με αυξημένη εκφραστικότητα.

Τέλος, οι τρόποι ανάγνωσης και αποθήκευσης των μοντέλων και των δεδομένων εκπαίδευσης θα ακολουθούν τις ήδη υπάρχουσες διαδικασίες. Τα δεδομένα εκπαίδευσης θα ακολουθούν το πρότυπο που ακολουθεί το *Weka* ενώ τα αποτελέσματα των τεχνικών του *Data Mining* θα αποθηκεύονται στην μορφή που τα εξάγει το *Weka* (*.arff*), σε καθορισμένο *Container* της βάσης προτύπων σε *PMML* γλώσσα και σε *.xml* μορφή.

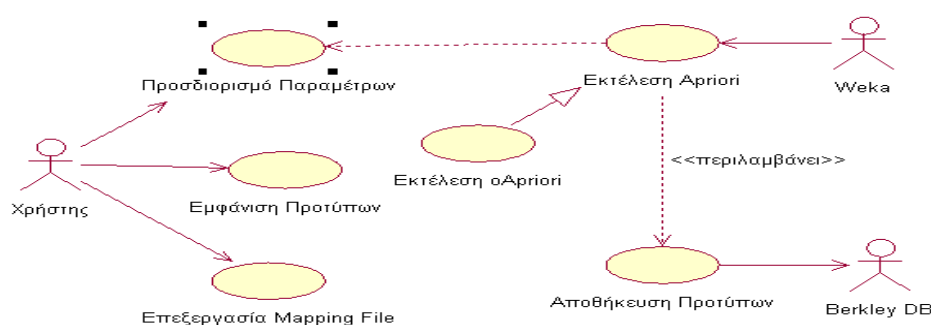
8.5 Λειτουργικές απαιτήσεις

Σύμφωνα με τη μέχρι τώρα ανάλυση ένα σύνολο χαρακτήρων και αλληλεπιδράσεων μπορούν να καταγραφούν μεταξύ του συστήματος και του περιβάλλοντός του. Οι κύριοι χαρακτήρες που αλληλεπιδρούν με το σύστημα είναι ο απλός Χρήστης, το σύστημα Weka και η Berkley DB.

Ο Χρήστης αρχικά θα πρέπει να προσδιορίζει τις παραμέτρους με τους οποίους θα επιθυμεί να εκτελεσθεί ο αλγόριθμος Apriori. Έπειτα θα πρέπει να έχει την δυνατότητα να βλέπει τα παραγόμενα πρότυπα και τέλος να μπορεί να διαχειριστεί τη χαρτογράφηση μεταξύ των εννοιών της οντολογίας και των δεδομένων εκπαίδευσης. Η χαρτογράφηση, ακολουθώντας την λογική του συστήματος, μπορεί να αποθηκεύεται σε ένα εξωτερικό αρχείο με μια προκαθορισμένη δομή και να ανακαλείται όποτε χρειάζεται (Mapping File).

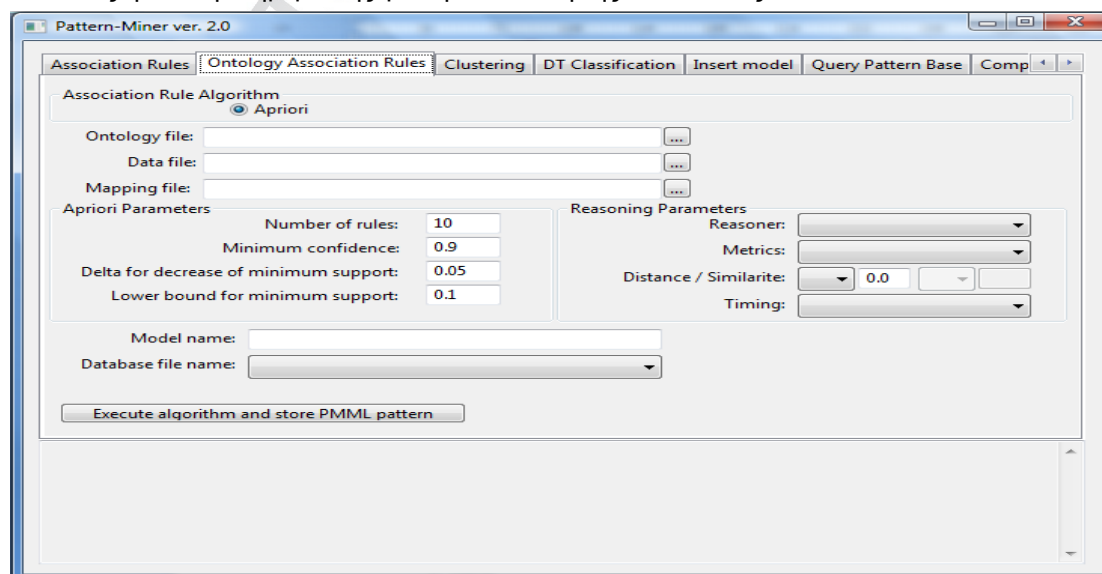
Ο επόμενος χαρακτήρας, ο Weka, θα διαθέτει τον αλγόριθμο Apriori όπου θα περιέχει την διαδικασία παραλαβής των κατάλληλων παραμέτρων, ενώ θα είναι μια γενικευμένη συμπεριφορά της διαδικασίας εκτέλεσης του Apriori σε συνεργασία με την Οντολογία (oApriori).

Τέλος, ο χαρακτήρας Berkley DB θα αναλαμβάνει την αποθήκευση των παραγόμενων προτύπων όπου σαν διαδικασία θα περιλαμβάνεται στην περίπτωση χρήσης της εκτέλεσης του Apriori. Ακολουθεί το διάγραμμα με τις περιπτώσεις χρήσης για περαιτέρω αποσαφήνιση.



Διάγραμμα 8-1. Διάγραμμα Περιπτώσεων Χρήσης

Λαμβάνοντας υπόψη τις μη λειτουργικές απαιτήσεις και τον προσδιορισμό του συστήματος είναι δυνατόν να γίνει μια πρώτη προσέγγιση των παραμέτρων που είναι απαραίτητοι να υπάρχουν για την εύρυθμη λειτουργία του συστήματος αλλά και μια εκτίμηση για την εμφάνιση της διεπαφής. Όπως βλέπουμε στην Εικόνα 8-3. *Διεπαφή συστήματος*, η λειτουργικότητα της διεπαφής παραμένει η ίδια, με το πρώτο μέρος να απαιτεί από τον χρήστη να θέσει τις κατάλληλες παραμέτρους, έπειτα να ακολουθεί το κουμπί εκτέλεσης και στο τέλος η οθόνη ενημέρωσης για την εκτέλεση της διαδικασίας.



Εικόνα 8-3. Διεπαφή συστήματος

Οι αναγκαίοι και επαρκείς παράμετροι που χρειάζεται το σύστημα για να λειτουργήσει μπορούν να ομαδοποιηθούν σε τέσσερις γενικές κατηγορίες. Η πρώτη αφορά τα αρχεία που πρέπει να φορτωθούν στο σύστημα, έπειτα οι παράμετροι του αλγορίθμου Apriori, στη συνέχεια οι παράμετροι για την διαχείριση της οντολογίας και τέλος οι παραμέτροι για την αποθήκευση των αποτελεσμάτων. Αναλυτικότερα έχουμε:

Πίνακας 8-2. Περιγραφή διεπαφής

Ομάδα	Ονομασία	Περιγραφή	Apriori	οApriori
1	Ontology file	Το αρχείο της Οντολογίας		√
	Data file	Τα δεδομένα εκπαίδευσης	√	√
	Mapping file	Το αρχείο με την χαρτογράφηση δεδομένων εκπαίδευσης και εννοιών Οντολογίας		√
2. Apriori Parameters	Number of rules	Αριθμός κανόνων	√	√
	Minimum confidence	Ελάχιστη εμπιστοσύνη	√	√
	Delta for decrease of minimum support	Βήμα μείωσης μέχρι την ελάχιστη υποστήριξη	√	√
	Lower bound for minimum support	Χαμηλότερη δέσμευση για την ελάχιστη υποστήριξη	√	√
3. Reasoner Parameters	Reasoner	Μηχανές συλλογιστικής ανάλυσης		√
	Metrics	Μετρική που υπολογίζει την απόσταση ή ομοιότητα μεταξύ εννοιών Οντολογίας		√
	Distance / Similarite	Τιμή Μετρικής(>= ή = ή <=)		√
	Timing	Πότε ο Apriori κοιτάει την μετρική, κατά την διάρκεια ή στο τέλος του αλγορίθμου		√
4	Model name	Όνομα μοντέλου	√	√
	Database file name	Container της Berkley DB	√	√
5	Execute algorithm and store PMML pattern	Κουμπί έναρξης εκτέλεσης διαδικασίας	√	√
6	Log Screen	Οθόνη μηνυμάτων	√	√

8.6 Επιλογή εργαλείων

Στην τρέχουσα υλοποίηση τα ακόλουθα εργαλεία θα χρησιμοποιηθούν για την ανάπτυξη και τον έλεγχο της:

1. Pattern-Minner v2.0 το οποίο θα είναι το σύστημα που θα επεκταθεί και περιέχει τις κύριες δομές ανάπτυξης.
2. Weka το οποίο περιέχει τους πηγαίους κώδικες για τις τεχνικές του Data Mining
3. Java ως εργαλείο ανάπτυξης
4. Berkley DB για την αποθήκευση των αποτελεσμάτων.
5. OWL για την γλώσσα που θα αναπαριστούνται οι Οντολογίες.
6. Jena ένα ολοκληρωμένο Java πακέτο για την διαχείριση Οντολογιών.
7. Pellet μια από τις πιο δημοφιλείς Μηχανές Συλλογιστικής Ανάλυσης η οποία μπορεί να συνδυαστεί με την Jena.

9. Ανάλυση και Σχεδιασμός

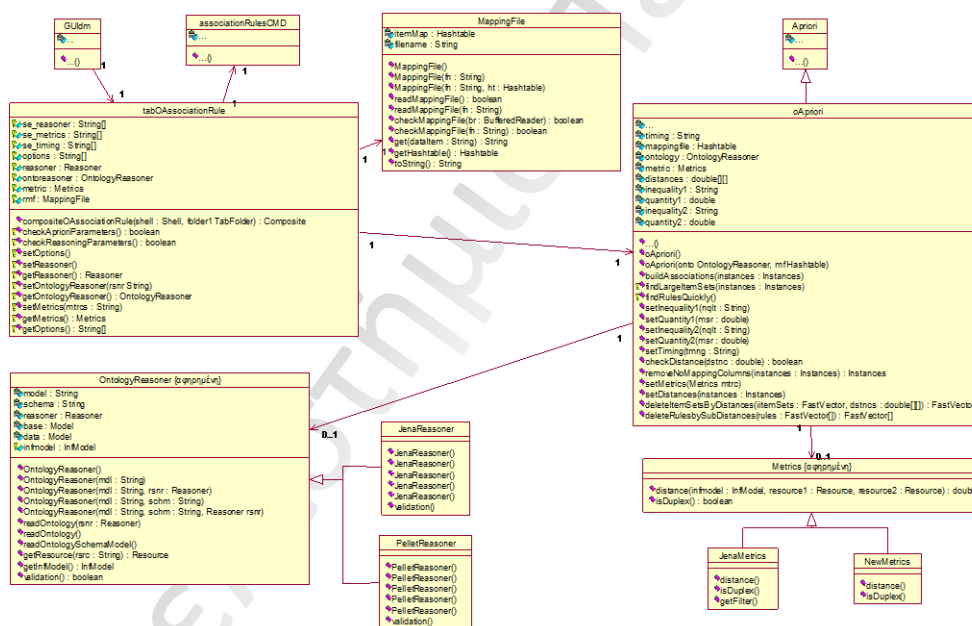
Έχοντας προσδιορίσει τις απαιτήσεις του συστήματος το βήμα που ακολουθεί είναι ο σχεδιασμός του. Στην τρέχουσα ενότητα θα χρησιμοποιηθούν τα Διαγράμματα Τάξεων, Καταστάσεων, Ακολουθίας και Πακέτων για την περαιτέρω ανάλυση του.

9.1 Διάγραμμα Τάξεων

Εξετάζοντας το ήδη υπάρχον σύστημα και τις απαιτήσεις για την επέκτασή του, επτά διαφαίνονται να είναι οι κύριες τάξεις που αποτελούν το διάγραμμα τάξεων.

1. *GUIdm*, το οποίο περιέχει την κύρια δομή της εφαρμογής.
2. *tabOAssociationRule*, αντικαθιστά την *tabAssociationRule*, είναι η τάξη που αλληλεπιδρά με το χρήστη και περιέχει μεθόδους για την εισαγωγή των παραμέτρων στον τροποποιημένο *Apriori*.
3. *MappingFile*, που εξασφαλίζει τις διαδικασίες για την επεξεργασία των επονομαζόμενων *Mapping File*.
4. *oApriori*, που ουσιαστικά επεκτείνει τον *Apriori* για να συγχωνευτούν οι Οντολογίες
5. *OntologyReasoner {αφηρημένη}*, που δίνει την δυνατότητα φόρτωσης και επικοινωνίας με τις Οντολογίες. Καθώς κάθε Μηχανή Συλλογιστικής Ανάλυσης χρησιμοποιεί δικές τις μεθόδους είναι αναγκαίο να γίνει υλοποίηση της λειτουργίας *validation()* για την εκάστοτε μηχανή
6. *Metrics {αφηρημένη}*, που καθορίζει τον τρόπο υπολογισμού των αποστάσεων ή ομοιοτήτων μεταξύ δύο εννοιών μιας Οντολογίας.
7. *associationRulesCMD*, που επιτρέπει την αποθήκευση των αποτελεσμάτων.

Στο σχήμα που ακολουθεί εμφανίζονται οι ιδιότητες και οι λειτουργίες των τάξεων, όπως επίσης και οι σχέσεις των τάξεων μεταξύ τους. Το σύμβολο με τις τρεις τελείες “...” υποδηλώνει ότι υπάρχουν ιδιότητες ή λειτουργίες οι οποίες δεν τροποποιούνται και δεν επισυνάπτονται για εξοικονόμηση χώρου.



Διάγραμμα 9-1. Διάγραμμα Τάξεων

Στους πίνακες που ακολουθούν γίνεται μια συνοπτική περιγραφή της χρησιμότητας των ιδιοτήτων και των λειτουργιών της εκάστοτε τάξης.

Πίνακας 9-1. Τάξη *tabOAssociationRule*

Ιδιότητα ή Λειτουργία	Σχόλια
<i>compositeOAssociationRule(shell : Shell, folder1 TabFolder) : Composite</i>	Δημιουργεί μια ετικέτα στην εφαρμογή που θα αλληλεπιδρά με τον χρήστη.
<i>checkAprioriParameters() : boolean</i>	Ελέγχει για την ύπαρξη των αναγκαίων παραμέτρων του <i>Apriori</i> .
<i>checkReasoningParameters() : boolean</i>	Ελέγχει για την ύπαρξη των αναγκαίων παραμέτρων για την εκτέλεση του <i>Ontology-enhanced Apriori</i> .

Ιδιότητα ή Λειτουργία	Σχόλια
setOptions()	Οργανώνει τις ρυθμίσεις του Apriori
getOptions() : String[]	Επιστρέφει τις επιλεγμένες ρυθμίσεις του Apriori
setReasoner()	Καθορίζει τις Μηχανές Συλλογιστικής Ανάλυσης που θα χρησιμοποιηθεί
getReasoner() : Reasoner	Επιστρέφει τις Μηχανές Συλλογιστικής Ανάλυσης που θα χρησιμοποιηθεί
setOntologyReasoner(rsnr String)	Ορίζει ένα OntologyReasoner αντικείμενο με την κλάση που το υλοποιεί(πχ PelletReasoner)
getOntologyReasoner() : OntologyReasoner	Επιστρέφει το OntologyReasoner αντικείμενο που χρησιμοποιείται
setMetrics(mtrcs : String)	Ορίζει την μετρική που θα χρησιμοποιηθεί
getMetrics() : Metrics	Επιστρέφει το αντικείμενο της μετρικής που θα χρησιμοποιηθεί

Πίνακας 9-2. Τάξη MappingFile

Ιδιότητα ή Λειτουργία	Σχόλια
itemMap : Hashtable	Περιέχει την αντιστοιχία μεταξύ των στοιχείων των δεδομένων εκπαίδευσης με τις έννοιες μιας Οντολογίας
filename : String	Η διεύθυνση του αρχείου
MappingFile()	Κατασκευαστής τάξης
MappingFile(fn : String)	Κατασκευαστής τάξης
MappingFile(fn : String, ht : Hashtable)	Κατασκευαστής τάξης
readMappingFile() : Boolean	Φόρτωση αντιστοιχίας
readMappingFile(fn : String)	Φόρτωση αντιστοιχίας από αρχείο
checkMappingFile(br : BufferedReader) : Boolean	Έλεγχος αν ακολουθούν τα δεδομένα στον BufferedReader μια προκαθορισμένη μορφή για την αντιστοιχία
checkMappingFile(fn : String) : Boolean	Έλεγχος αν ακολουθεί το αρχείο μια προκαθορισμένη μορφή για την αντιστοιχία
get(dataltem : String) : String	Επιστρέφει την ονομασία της έννοιας της Οντολογίας που αντιστοιχεί σε ένα στοιχείο των δεδομένων εκπαίδευσης
getHashtable() : Hashtable	Επιστρέφει την αντιστοιχία
toString() : String	Επιστρέφει την αντιστοιχία σε String μορφή

Πίνακας 9-3. Τάξη OntologyReasoner {αφηρημένη}

Ιδιότητα ή Λειτουργία	Σχόλια
model : String	Το αρχείο με το μοντέλο της Οντολογίας.
schema : String	Το αρχείο με το σχήμα των δεδομένων της Οντολογίας.
reasoner : Reasoner	Οι Μηχανές Συλλογιστικής Ανάλυσης
base : Model	Το αντικείμενο του αρχείου με το μοντέλο της Οντολογίας
data : Model	Το αντικείμενο του αρχείου με το σχήμα των δεδομένων της Οντολογίας

Ιδιότητα ή Λειτουργία	Σχόλια
infmodel : InfModel	Επέκταση των Model για επιπρόσθετο έλεγχο και πρόσβαση στον συμπερασμό της Οντολογίας
OntologyReasoner()	Κατασκευαστής
OntologyReasoner(mdl : String)	Κατασκευαστής, λαμβάνει το αρχείο με το μοντέλο της Οντολογίας
OntologyReasoner(mdl : String, rsnr : Reasoner)	Κατασκευαστής, λαμβάνει το αρχείο με το μοντέλο της Οντολογίας και τον reasoned
OntologyReasoner(mdl : String, schm : String)	Κατασκευαστής, λαμβάνει το αρχείο με το μοντέλο και το αρχείο με το σχήμα των δεδομένων της Οντολογίας
OntologyReasoner(mdl : String, schm : String, Reasoner rsnr)	Κατασκευαστής, λαμβάνει το αρχείο με το μοντέλο της Οντολογίας, το αρχείο με το σχήμα των δεδομένων της Οντολογίας και τον reasoned
readOntology(rsnr : Reasoner)	Φορτώνει ένα μοντέλο Οντολογίας στο InfModel
readOntology()	Φορτώνει ένα μοντέλο Οντολογία στο InfModel
readOntologySchemaModel(rsnr : Reasoner)	Φορτώνει ένα μοντέλο και το σχήμα των δεδομένων μιας Οντολογίας στο InfModel
readOntologySchemaModel()	Φορτώνει ένα μοντέλο και το σχήμα των δεδομένων μιας Οντολογίας στο InfModel
getResource(rsrc : String) : Resource	Επιστρέφει την Πηγή μιας έννοιας της Οντολογίας
getInfModel() : InfModel	Επιστρέφει το InfModel του αντικείμενου της τάξης
validation() : boolean	Ελέγχει αν μια Οντολογία είναι συνεπής

Πίνακας 9-4. Τάξη oApriori

Ιδιότητα ή Λειτουργία	Σχόλια
timing : String	Πότε θα ενσωματώνονται οι λειτουργίες της Οντολογίας. Τρεις επιλογές: 1. κατά τη διάρκεια της εκτέλεσης του oApriori, 2. μετά την εκτέλεση του oApriori, 3. Κατά τη διάρκεια και μετά την εκτέλεση του oApriori
mappingfile : Hashtable	Ένας Hashtable με την αντιστοίχιση των στοιχείων των δεδομένων εκπαίδευσης με τις έννοιες της Οντολογίας
ontology : OntologyReasoner	Ένα αντικείμενο που περιέχει τις Οντολογίες
metric : Metrics	Ένα αντικείμενο που αναφέρεται στη Μετρική που θα χρησιμοποιηθεί
distances : double[][]	Οι αποστάσεις ή ομοιότητες εννοιών τις Οντολογίας
inequality1 : String	Μια ανισότητα για τις επιτρεπόμενες τιμές απόστασης/ ομοιότητας μεταξύ των εννοιών της Οντολογίας (>=, =, <=)
quantity1 : double	Η τιμή της απόστασης/ ομοιότητας μεταξύ των εννοιών της Οντολογίας

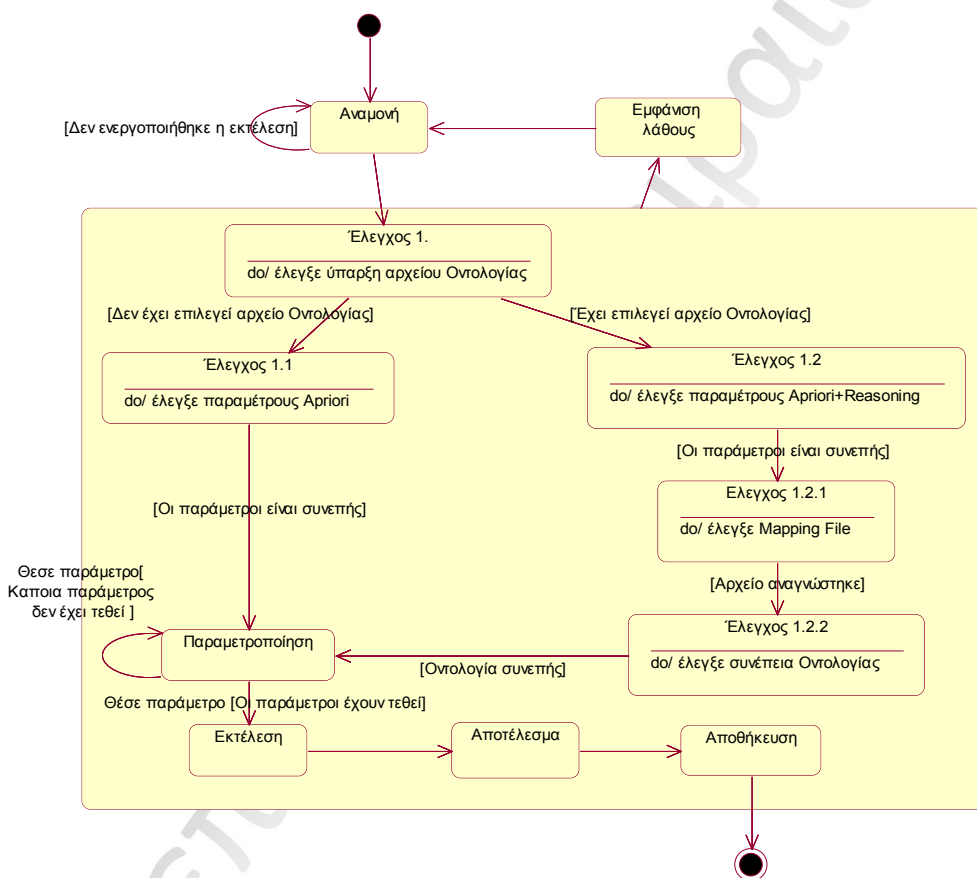
Ιδιότητα ή Λειτουργία	Σχόλια
inequality2 : String	Μια ανισότητα για τις επιτρεπόμενες τιμές απόστασης/ ομοιότητας μεταξύ των εννοιών της Οντολογίας (<=, null). Ενεργή μόνο όταν το inequality1=">="
quantity2 : double	Η τιμή της απόστασης/ ομοιότητας μεταξύ των εννοιών της Οντολογίας. Ενεργή μόνο όταν το inequality1=">="
οApriori()	Κατασκευαστής
οApriori(onto OntologyReasoner, mf Hashtable)	Κατασκευαστής
buildAssociations(instances : Instances)	Υποσκέλιση μεθόδου (override) buildAssociations του Apriori για την ενσωμάτωση των νέων λειτουργιών
findLargelItemSets(instances : Instances)	Υποσκέλιση μεθόδου findLargelItemSets του Apriori για την ενσωμάτωση των νέων λειτουργιών
findRulesQuickly()	Υποσκέλιση μεθόδου findRulesQuickly του Apriori για την ενσωμάτωση των νέων λειτουργιών
setInequality1(nqlt : String)	Ενημερώνει την τιμή της Inequality1
setQuantity1(msr : double)	Ενημερώνει την τιμή της setQuantity1
setInequality2(nqlt : String)	Ενημερώνει την τιμή της Inequality2
setQuantity2(msr : double)	Ενημερώνει την τιμή της setQuantity2
setTiming(tmng : String)	Ενημερώνει την τιμή της Timing
checkDistance(dstnc : double) : Boolean	Ελέγχει αν η απόσταση μεταξύ δύο εννοιών είναι η επιθυμητή
removeNoMappingColumns(instances : Instances) : Instances	Διαγράφει στοιχεία από τα δεδομένα εκπαίδευσης που δεν έχουν αντιστοιχηθεί με έννοιες της Οντολογίας
setMetrics(Metrics mtrc)	Ενημερώνει την τιμή της metric
setDistances(instances : Instances)	Υπολογίζει τις αποστάσεις συγκεκριμένων εννοιών σε καθορισμένη Οντολογία
deleteItemSetsByDistances(iitemSets : FastVector, dstncls : double[][]) : FastVector[]	Διαγράφει από μια συλλογή από ItemSet, τα ItemSet που τα στοιχεία του δεν καλύπτουν τους περιορισμούς των αποστάσεων μεταξύ εννοιών
deleteRulesbySubDistances(rules : FastVector[]) : FastVector[]	Διαγράφει από ένα σύνολο παραγόμενων Κανόνων, τους Κανόνες που τα στοιχεία τους δεν καλύπτουν τους περιορισμούς των αποστάσεων μεταξύ εννοιών

Πίνακας 9-5. Τάξη Metrics {αφηρημένη}

Ιδιότητα ή Λειτουργία	Σχόλια
distance(infmodel : InfModel, resource1 : Resource, resource2 : Resource) : double	Υπολογίζει την απόσταση ή την ομοιότητα μεταξύ δύο πηγών μιας Οντολογίας
isDuplex() : boolean	Επιστρέφει αληθές αν η μετρική είναι αμφίδρομη δηλαδή έχει την ίδια απόσταση για A->B και B->A

9.2 Διαγράμματα Καταστάσεων

Στην εικόνα που ακολουθεί (Διάγραμμα 9-2. Διάγραμμα Καταστάσεων Συστήματος) παρουσιάζονται συνολικά οι καταστάσεις στις οποίες μπορεί να επέλθει το σύστημα. Αρχικά, αναμένει μέχρι να ζητηθεί από το χρήστη να εκτελεσθεί η απαιτούμενη διαδικασία. Έπειτα, εξετάζεται η ύπαρξη του αρχείου της Οντολογίας. Σε περίπτωση που δεν υπάρχει, εισέρχεται στην κατάσταση Έλεγχος 1.1 που εξετάζει την συνέπεια των απαραίτητων παραμέτρων που έχουν εισαχθεί από τον χρήστη. Σε περίπτωση που έχει επιλεγεί αρχείο Οντολογίας, το σύστημα διαδοχικά ελέγχει τις απαραίτητες παραμέτρους που εισάγει ο χρήστης, την συνέπεια του Mapping File και τέλος την συνέπεια της Οντολογίας. Οι καταστάσεις που ακολουθούν πλέον είναι η Παραμετροποίηση της τάξης οΑρriori, η Εκτέλεση του αλγορίθμου like-Αρriori, τα Αποτελέσματα και η Αποθήκευσή τους. Σε περίπτωση που έχει εμφανιστεί κάποιο λάθος ή δεν ακολουθείται η προαναφερθείσα πορεία το σύστημα επέρχεται σε κατάσταση Εμφάνιση λάθους και στη συνέχεια επανέρχεται στην αρχική του κατάσταση.



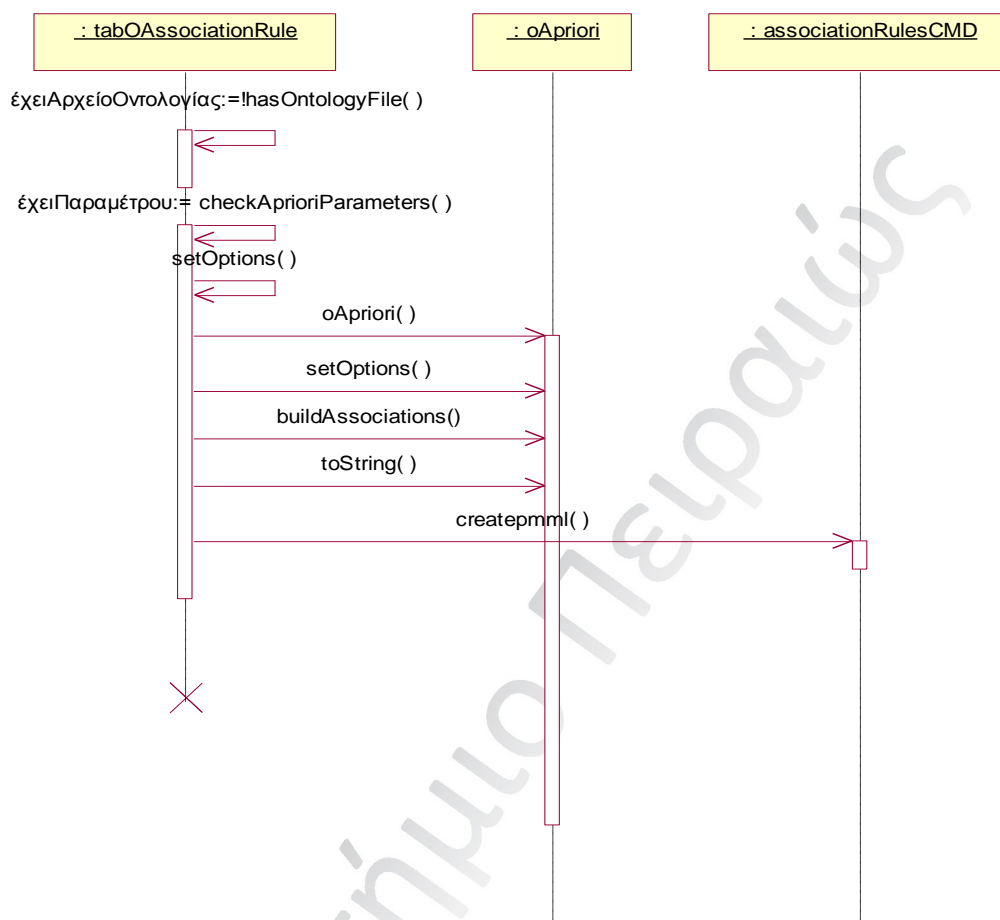
Διάγραμμα 9-2. Διάγραμμα Καταστάσεων Συστήματος

9.3 Διαγράμματα Ακολουθίας

Κατά κανόνα τα Διαγράμματα αλληλεπίδρασης και κατ' επέκταση τα Διαγράμματα Ακολουθίας περιγράφουν τις λειτουργίες μιας και μόνο περίπτωσης χρήσης. Παρόλα αυτά στο παρόν σύστημα καθώς οι περιπτώσεις χρήσης κατά κάποιον τρόπο συσχετίζονται θα εξακολουθήσει μια γενικότερη ανάλυση της συνεργασίας των τάξεων μεταξύ τους. Ουσιαστικά υπάρχουν δύο επιλογές για την περίπτωση χρήσης Εκτέλεση Αρriori. Η πρώτη είναι να επιθυμεί ο χρήστης να τρέξει τον αλγόριθμο χωρίς την ενσωμάτωση Οντολογιών (Διάγραμμα 9-3. Μη ενσωμάτωση Οντολογιών) και η δεύτερη να θέλει να τις ενσωματώσει (Διάγραμμα 9-4. Ενσωμάτωση Οντολογιών).

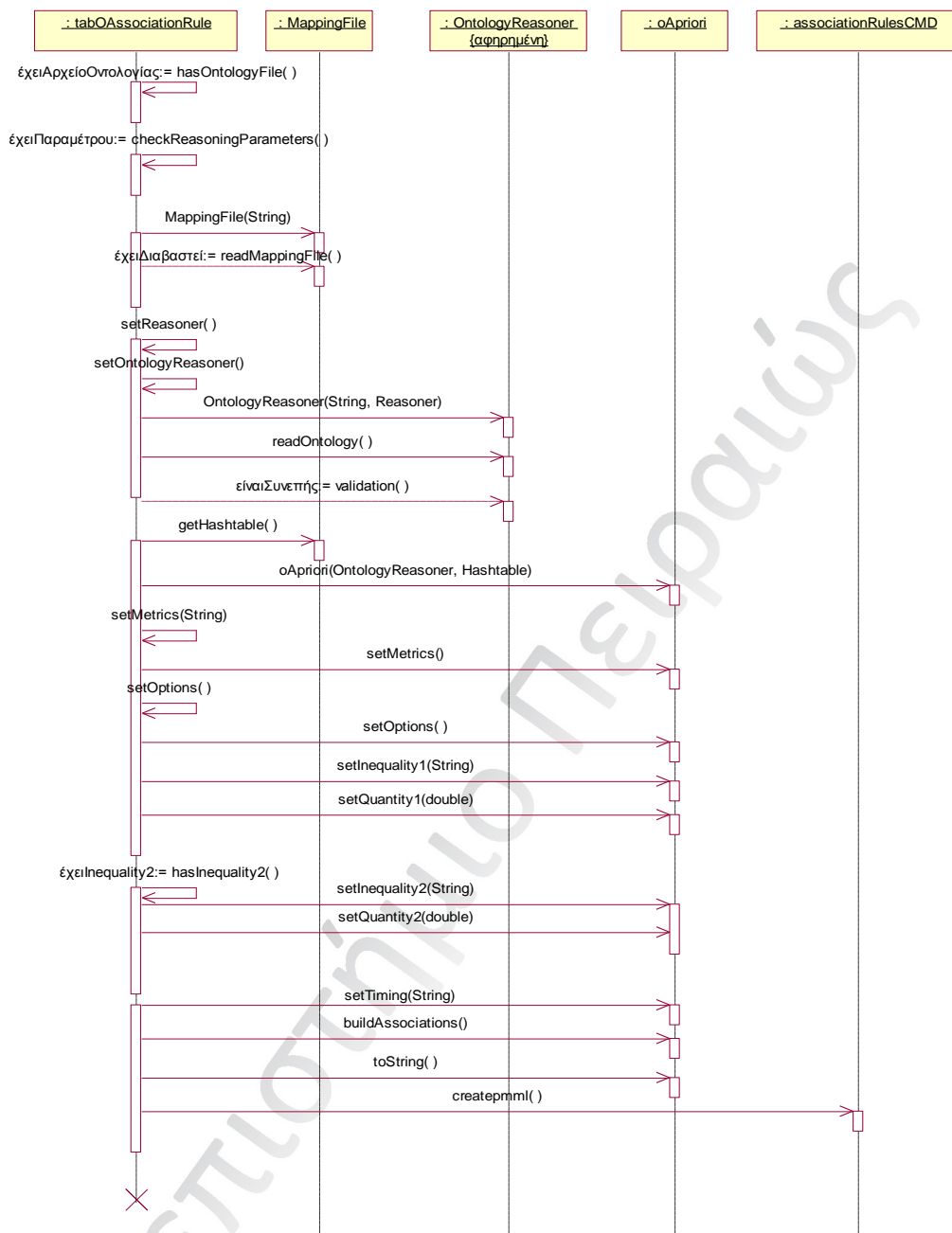
Στην πρώτη περίπτωση γίνεται έλεγχος αν έχει επιλεγεί κάποιο αρχείο Οντολογίας, αν δεν έχει επιλεγεί συνεχίζεται η διαδικασία απρόσκοπτα, διαφορετικά εκτελείται η δεύτερη ακολουθία. Έπειτα, γίνεται έλεγχος αν έχουν συμπληρωθεί όλοι οι παράμετροι του Αρriori.

Στη συνέχεια λαμβάνονται και διατηρούνται οι παράμετροι με μια προκαθορισμένη μορφή που ορίζεται από τον `Apriori`. Μετά ακολουθεί η δημιουργία του αντικείμενου `oApriori`, η ενημέρωση του με τις παραμέτρους και έπειτα ζητείται η εκτέλεση του αλγορίθμου, η μετατροπή των αποτελεσμάτων σε `String` και τέλος η αποθήκευση τους στη βάση.



Διάγραμμα 9-3. Μη ενσωμάτωση Οντολογιών

Στη δεύτερη περίπτωση χρήσης ο ενδιαφερόμενος επιθυμεί να συμμετέχει ενεργά η Οντολογία στην εκτέλεση του `Apriori`. Έτσι αρχικά γίνεται έλεγχος αν υπάρχει κάποιο αρχείο Οντολογίας και αμέσως μετά ελέγχει αν όλοι οι παράμετροι είναι συνεπείς. Έπειτα δημιουργείται ένα αντικείμενο `MappingFile` και εξετάζει αν το κατάλληλο αρχείο που έχει επιλεγεί μπορεί να διαβαστεί και να φορτωθεί στο αντικείμενο. Αν όλα βαίνουν καλώς, αναγνωρίζεται η Μηχανή Συλλογιστικής Ανάλυσης που έχει επιλεγεί από το χρήστη και δημιουργείται ο κατάλληλος απόγονος του `OntologyReasoner`. Στη συνέχεια γίνεται ανάγνωση της Οντολογίας από το αρχείο και ελέγχεται για ασυνέπεια. Σε περίπτωση που είναι συνεπής λαμβάνεται από το `MappingFile` ο `Hashtable`, που περιέχει τις αντιστοιχίες μεταξύ των στοιχείων των δεδομένων εκπαίδευσης και των εννοιών της Οντολογίας. Ακολουθεί η κλήση του κατάλληλου κατασκευαστή της τάξης `oApriori` και συνεχίζεται η επικοινωνία μεταξύ `tabOAssociationRule` και `oApriori` μέχρι να τεθούν όλοι οι απαραίτητοι παράμετροι υπόψη του αλγορίθμου. Στα τελευταία βήματα έχουμε την εκτέλεση του αλγορίθμου, τη μετατροπή του αποτελέσματος σε αλφαριθμητικό και την αποθήκευση του στην βάση.



Διάγραμμα 9-4. Ενσωμάτωση Οντολογιών

9.4 Διαγράμματα Πακέτων

Τα δύο πιο γενικά πακέτα που θα χρησιμοποιηθούν για την υλοποίηση του συστήματος θα είναι αυτά της *Libraries* και του *Source Packages*. Το πρώτο πακέτο θα περιέχει όλες τις βιβλιοθήκες που θα χρειαστεί η εφαρμογή για να εκτελείται ομαλά, ενώ το δεύτερο πακέτο, το οποίο θα εξαρτάται φυσικά από το πρώτο, θα περιέχει τον πηγαίο κώδικα (Διάγραμμα 9-5. Γενικό Διάγραμμα Πακέτων).



Διάγραμμα 9-5. Γενικό Διάγραμμα Πακέτων

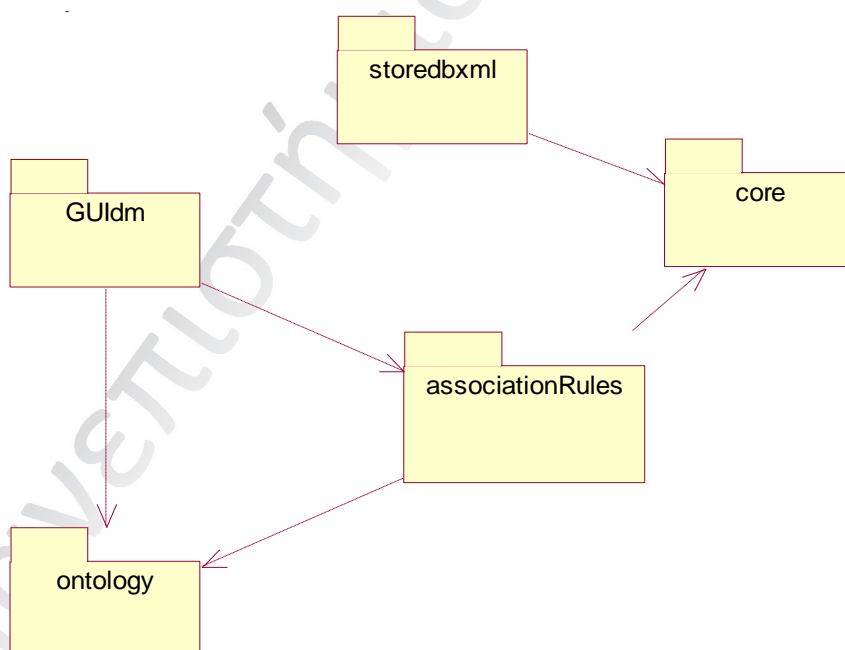
Συγκεκριμένα το πακέτο *Libraries* θα περιέχει:

1. Τις βασικές βιβλιοθήκες που είναι αναγκαίες για να εκτελεστεί η εφαρμογή πριν την επέκταση και είναι οι `db.jar`, `dbxml.jar`, `jdom.jar`, `org-netbeans-api-visual.jar`, `PANDA.jar` και `swt.jar`.
2. Την βιβλιοθήκη του Weka, `weka.jar` η οποία χρησιμοποιείται από την κλάση `οΑpriori` για την επέκταση της μητρικής κλάσης `weka.Apriori`.
3. Τις βιβλιοθήκες του Jena για τη διαχείριση των OWL Οντολογιών
4. Τις βιβλιοθήκες της Pellet για να μπορεί να χρησιμοποιηθεί ο αντίστοιχος Reasoner. Εδώ πρέπει να δοθεί προσοχή καθώς ο συνδυασμός της Jena και της Pellet θα είναι προβληματικός αν στις βιβλιοθήκες συμπεριλαμβάνεται η βιβλιοθήκη της Jena: `slf4j-log4j12-1.5.8.jar`, γι' αυτό συνιστάται η διαγραφή της σε περίπτωση συνύπαρξης της Jena και Pellet.

Αντίστοιχα το *Source Packages* θα περιέχει:

1. Τα πακέτα που ήδη υφίστανται και δε θα τροποποιηθούν, τα οποία είναι `core`, `classifiers`, `clusterers`, `comparison`, `monitoring`.
2. Το `GUIdm` που σχετίζεται κυρίως με την διεπαφή της εφαρμογής και στην οποία θα του προστεθεί άλλη μία κλάση, η `tabOAssociationRule.java` και θα τροποποιηθεί η `GUIdm.java`
3. Το `associationRules`, που σχετίζεται με τις κλάσεις που διαπραγματεύονται τις λειτουργίες των Κανόνων Συσχετίσεων, όπου θα προστεθεί η κλάση `οΑpriori.java`
4. Το νέο πακέτο `ontology`, που θα έχει να διαχειριστεί τις λειτουργίες σχετικά με τις Οντολογίες. Σε αυτό θα υπάρχουν οι βασικές κλάσεις `MappingFile.java`, `Metrics.java`, `OntologyReasoner.java`, όπου οι δύο τελευταίες θα πρέπει να υλοποιούνται από απογόνους καθώς είναι αφηρημένες.

Στην εικόνα που ακολουθεί (Διάγραμμα 9-6. Διάγραμμα Πακέτων *Source Packages*) παρουσιάζονται οι εξαρτήσεις μεταξύ των πακέτων που έχουν άμεση σχέση με το αναπτυσσόμενο σύστημα.



Διάγραμμα 9-6. Διάγραμμα Πακέτων *Source Packages*

10. Υλοποίηση

Στην ενότητα της υλοποίησης θα παρουσιαστούν τα σημαντικότερα κομμάτια του πηγαίου κώδικα κάθε κλάσης που είτε τροποποιήθηκε είτε υλοποιήθηκε από την αρχή. Παράλληλα και όπου χρειάζεται θα γίνεται εκτεταμένη ανάλυση για τον τρόπο που λειτουργούν. Κάθε υποενότητα θα έχει ως τίτλο το όνομα της κλάσης που διαπραγματεύεται.

10.1 GUIdm.java

Η κλάση της GUIdm.java ανήκει στο πακέτο GUIdm το οποίο είναι υπεύθυνο για την διεπαφή και τις βασικές λειτουργίες ολόκληρης της εφαρμογής. Η κλάση προϋπήρχε και προστέθηκε ο παρακάτω κώδικας που υποδηλώνει την ύπαρξη ενός νέου TabItem στην εφαρμογή με την ονομασία *Ontology Association Rules*.

```
TabItem item_011 = new TabItem(folder, SWT.NONE);
    item_011.setText("Ontology Association Rules");
    tabOAssociationRule tabOAR = new tabOAssociationRule();
    Composite c_oar = tabOAR.compositeOAssociationRule(shell, folder);
    c_oar.setLayout(null);
    item_011.setControl(c_oar);
```

Κώδικας 10-1. Επέκταση GUIdm.java

10.2 tabOAssociationRule.java

Η κλάση tabOAssociationRule.java ανήκει στο πακέτο GUIdm. Σχετίζεται με την διεπαφή του συστήματος που υλοποιείται και είναι υπεύθυνο για την παραμετροποίηση του αλγορίθμου Apriori, την κλήση της μεθόδου για την εκτέλεση του προαναφερθέντος αλγορίθμου, την κλήση των μεθόδων για την αποθήκευση των αποτελεσμάτων του αλλά και για τη διαχείριση των λαθών που θα εμφανιστούν. Θεωρητικά οι λειτουργίες του παρουσιάζονται στα διαγράμματα: *Διάγραμμα 9-3. Μη ενσωμάτωση Οντολογιών, Διάγραμμα 9-4. Ενσωμάτωση Οντολογιών.*

Η συγκεκριμένη κλάση μπορεί να χωριστεί νοητά σε τέσσερα κομμάτια. Πρώτα, το κομμάτι που καθορίζει τη μορφή της διεπαφής. Δεύτερον, το κομμάτι που ζητείται να εκτελεσθεί ο αλγόριθμος Apriori χωρίς την ενσωμάτωση των οντολογιών. Τρίτον, από το κομμάτι που τρέχει όταν ζητείται η ενσωμάτωση των οντολογιών στον Apriori. Τέλος το τέταρτο κομμάτι που είναι ανεξάρτητο από τα δύο προηγούμενα που έχει να κάνει με την αποθήκευση των αποτελεσμάτων. Για την εκτέλεση του δεύτερου ή του τρίτου κομματιού ο χρήστης απλά δεν εισάγει ή εισάγει αντίστοιχα κάποιο αρχείο που περιέχει μία OWL Οντολογία.

Θα ακολουθήσει ο πηγαίος κώδικας για το τρίτο και τέταρτο κομμάτι χωρίς τα μπλοκ σύλληψης λαθών για εξοικονόμηση χώρου. Επισημαίνεται ότι η υλοποίηση του δεύτερου κομματιού ακολουθεί την λογική του τρίτου και αυτός είναι ο λόγος που δεν παρουσιάζεται.

```
//Αν έχει επιλεγεί κάποιο αρχείο Οντολογίας
else{
    //Αν όλοι οι παράμετροι είναι συνεπείς
    if(checkReasoningParameters()){
        GUIdm.runinfo.append("Run Apriori With Ontology Intigration\n");
        //Διάβασε το αρχείο MappingFile
        MappingFile rmf = null;
        rmf = new MappingFile(mav_fname.getText());
        if(rmf.readMappingFile()){
            //Αν μπορεί να διαβαστεί το αρχείο MappingFile
            //Θέσε Reasoner
            setReasoner(v_reasoner.getText().toString());
            GUIdm.runinfo.append("Ontology reading...\n");
            //Διάβασε Οντολογία
            setOntologyReasoner(v_reasoner.getText().toString());
            ontoreasoner.readOntology();
            GUIdm.runinfo.append("Ontology readed\n");
```



```

GUIdm.runinfo.append("Ontology Validation...\n");
//Αν η Οντολογία είναι συνεπής
if(ontoreasoner.validation()){
    GUIdm.runinfo.append("Ontology is consistent\n");
    GUIdm.runinfo.append("Execute Apriori... \n");
    //Δημιουργία αντικειμένου οApriori
    oapriori = new οApriori(ontoreasoner, rmf.getHashtable());
    //Ενημέρωση Options Apriori
    setOptions();
    oapriori.setOptions(getOptions());
    //Ενημέρωση Μετρική
    setMetrics(v_metrics.getText());
    oapriori.setMetrics(getMetrics());
    //Ενημέρωση Inequality1, Quantity1, Inequality2, Quantity2
    oapriori.setInequality1(v_inequality1.getText().toString());
    oapriori.setQuantity1(Double.parseDouble(v_quantity1.getText()));
    if( v_inequality2.isEnabled() && !v_inequality2.getText().equals("")){
        oapriori.setInequality2(v_inequality2.getText().toString());
        oapriori.setQuantity2(Double.parseDouble(v_quantity2.getText()));
    }
    //Ενημέρωση timing
    oapriori.setTiming(v_timing.getText().toString());
    Reader reader = null;
    reader = new BufferedReader(new FileReader(v_fname.getText()));
    //Εκτέλεσε τον αλγόριθμο Apriori
    oapriori.buildAssociations(new Instances(reader));
    run=true;
}
else{
    GUIdm.runinfo.append("Ontology is inconsistent\n");
}
}
else{
    GUIdm.runinfo.append("Error Reading Mapping file\n");
}
}
}
}

```

Κώδικας 10-2. Γ' μέρος της tabOAssociationRule.java

```

//Αν δεν έχει παρουσιαστεί κάποιο λάθος συνέχισε
if(run){
    BufferedWriter out = new BufferedWriter(new FileWriter(GUIdm.tempArffFile));
    out.write(oapriori.toString());
    System.out.println(oapriori.toString());
    out.close();
}

```

```

//Αποθήκευσε αποτελέσματα σε μορφή .arff
GUIDm.runinfo.append("Executed \n");
String cmd_copyarff = "copy " + GUIDm.tempArffFile + " " + GUIDm.pathArffFiles + "AR_" +
                    v_modelname.getText() + ".arff";

BufferedWriter outf = new BufferedWriter(new FileWriter("RunWeka_temp.bat"));
outf.write(cmd_copyarff);
outf.close();
Process p = Runtime.getRuntime().exec("RunWeka_temp.bat");
p.waitFor();
//Αποθήκευσε αποτελέσματα στην βάση
GUIDm.runinfo.append("Creating PMML file... \n");
associationRulesCMD demo = new associationRulesCMD();
GUIDm.runinfo.append(demo.createpmml(v_modelname.getText(),
    combo_dbfile.getText()+ GUIDm.dbxmlExtension, v_fname.getText(),
    Integer.parseInt(v_numofrules.getText()), GUIDm.tempArffFile, GUIDm.pathDbxmlFiles));
System.gc();
}
GUIDm.runinfo.append("Finished\n");
GUIDm.runinfo.append("=====\n");

```

Κώδικας 10-3. Δ' μέρος της tabOAssociationRule.java

10.3 οApriori.java

Η κλάση οApriori.java ανήκει στο πακέτο AssociationRules και επεκτείνει τον αλγόριθμο Apriori του Weka. Προφανώς, συμπεριφέρεται όπως ο πρόγονός του αλλά είναι τροποποιημένος έτσι ώστε να ενσωματώνει τις Οντολογίες. Οι μέθοδοι που τροποποιήθηκαν είναι οι buildAssociations, findLargestItemSets, findRulesQuickly.

Στη μέθοδο *buildAssociations* προστέθηκε ο έλεγχος Κώδικας 10-4. Αυτός επιτρέπει, σε περίπτωση που θα ενσωματωθεί στον αλγόριθμο κάποια Οντολογία, να απομακρύνει από τα στοιχεία, που θα αποτελούν τους κανόνες, εκείνα που δεν αντιστοιχούν σε κάποια Οντολογική έννοια. Δηλαδή, κοιτάζει αν τα στοιχεία εκπαίδευσης έχουν χαρτογραφηθεί (Κώδικας 10-5. Μέθοδος οApriori.removeNoMappingColumns).

```

if(ontology!= null){
    instances = removeNoMappingColumns(instances);
    setDistances(instances);
}

```

Κώδικας 10-4. Προσθήκη στη μέθοδο οApriori.buildAssociations

```

protected Instances removeNoMappingColumns(Instances instances) throws Exception {
    int numInstances = instances.numInstances();
    StringBuffer deleteString = new StringBuffer();
    boolean first = true;
    int maxCount = 0;
    for (int i=0;i<instances.numAttributes();i++)
        if (!mappingfile.containsKey(instances.attribute(i).name()))
            if (first) {
                deleteString.append((i+1));
                first = false;
            } else

```

```

        deleteString.append(", "+(i+1));
    if (deleteString.toString().length() > 0) {
        Remove af = new Remove();
        af.setAttributeIndices(deleteString.toString());
        af.setInvertSelection(false);
        af.setInputFormat(instances);
        Instances newInst = Filter.useFilter(instances, af);
        return newInst;
    }
    return instances;
}

```

Κώδικας 10-5. Μέθοδος οApriori.removeNoMappingColumns

Στην μέθοδο *findLargestItemSets* προστέθηκε ο έλεγχος Κώδικας 10-6. Ο σκοπός της μεθόδου είναι να βρίσκει όλα τα στοιχειοσύνολα (ItemSets) από τα οποία θα παραχθούν στη συνέχεια οι κανόνες. Ο έλεγχος που προστέθηκε ενεργοποιείται αν ο χρήστης έχει επιλέξει στην Οντολογία να επεμβαίνει πριν την παραγωγή κανόνων και καλεί τη μέθοδο *deleteltemSetsByDistances*. Η τελευταία μέθοδος ελέγχει σε κάθε στοιχειοσύνολο αν όλες οι σημασιολογικές αποστάσεις μεταξύ των στοιχείων καλύπτουν τη συνθήκη ομοιότητας που έχει επιλέξει ο χρήστης (μέθοδος *checkDistance*). Αν ένα στοιχειοσύνολο δεν ακολουθεί τον περιορισμό, διαγράφεται.

Για παράδειγμα, έστω ότι έχουμε το στοιχειοσύνολο {A, B, C} και ο χρήστης επιθυμεί οι σημασιολογικές τους αποστάσεις να είναι μεγαλύτερες ή ίσες του 3. Επίσης, υποθέτουμε ότι οι ακόλουθοι δύο πίνακες περιέχουν τις σημασιολογικές αποστάσεις των στοιχείων για δύο διαφορετικές περιπτώσεις.

Παράδειγμα 10-1. Αποδοχή Στοιχειοσυνόλου

	A	B	C
A	0	4	5
B	6	0	8
C	7	8	0

	A	B	C
A	0	2	5
B	6	0	8
C	7	8	0

i) Δεύτερη περίπτωση

i) Πρώτη περίπτωση

Τότε θα γινόταν έλεγχος, όλες οι δυνατές αποστάσεις, δηλαδή $distance(A, B)$, $distance(B, A)$, $distance(A, C)$, $distance(C, A)$, $distance(B, C)$, $distance(C, B)$ να είναι μεγαλύτερες ή ίσες του 3. Επομένως, στην πρώτη περίπτωση το στοιχειοσύνολο δε θα διαγραφόταν ενώ στη δεύτερη περίπτωση θα διαγραφόταν καθώς η σημασιολογική απόσταση $distance(A, B)=2$ δε θα κάλυπτε τον περιορισμό.

```

if(timing.equals("During") || timing.equals("Over_all")){
    kSets = deleteltemSetsByDistances(kSets,distances);
}

```

Κώδικας 10-6. Προσθήκη στην μέθοδο οApriori.findLargestItemSets

```

public FastVector deleteltemSetsByDistances(FastVector itemSets, double [][] dstncs) {
    FastVector newVector = new FastVector(itemSets.size());
    boolean step=true;
    for (int i = 0; i < itemSets.size(); i++) {
        AprioriItemSet current = (AprioriItemSet)itemSets.elementAt(i);
        for(int i1 = 0; i1<current.items().length && step == true; i1++){

```

```

if(current.itemAt(i1) != -1)
    for(int i2 = i1+1; i2<current.items().length && step == true; i2++)
        if(current.itemAt(i2) != -1)
            if(!checkDistance(distances[i1][i2]) && !checkDistance(distances[i2][i1]))
                step = false;
if(step == true)
    newVector.addElement(current);
else
    step = true;
}
return newVector;
} //deleteItemSetsByDistances

```

Κώδικας 10-7. Μέθοδος οApriori.deleteItemSetsByDistances

```

public boolean checkDistance (double dstnc){
    if(inequality1.equals("="))
        return ((dstnc == quantity1) ? true:false);
    if(inequality1.equals("<="))
        return ((dstnc >= 0.0 && dstnc <= quantity1) ? true:false);
    if(inequality1.equals(">=") && inequality2.equals(""))
        return ((dstnc >= quantity1) ? true:false);
    if(inequality1.equals(">=") && inequality2.equals("<="))
        return (((dstnc >= quantity1)&&(dstnc <= quantity2)) ? true:false);
    return false;
}

```

Κώδικας 10-8. Μέθοδος οApriori.checkDistance

Στην μέθοδο *findRulesQuickly* προστέθηκε ο έλεγχος Κώδικας 10-9. Ο σκοπός της μεθόδου είναι να βρίσκει όλους τους επιτρεπτούς παραγόμενους κανόνες από ένα σύνολο στοιχειοσυνόλων. Ο έλεγχος που προστέθηκε ενεργοποιείται αν ο χρήστης έχει επιλέξει στην Οντολογία να επεμβαίνει μετά την παραγωγή κανόνων και καλεί την μέθοδο *deleteRulesBySubDistances*. Οι τελευταία ελέγχει σε κάθε κανόνα αν όλες οι σημασιολογικές αποστάσεις που δημιουργούνται από τα στοιχεία του αριστερού μέρους του κανόνα προς τα στοιχεία του δεξιού μέρους του κανόνα καλύπτουν τη συνθήκη ομοιότητας που έχει επιλέξει ο χρήστης. Αν ένας κανόνας δεν ακολουθεί τον περιορισμό, διαγράφεται.

Για παράδειγμα, έστω ότι έχουμε τον κανόνα $\{A, B \Rightarrow C, D\}$ και ο χρήστης επιθυμεί οι σημασιολογικές τους αποστάσεις να είναι μεγαλύτερες ή ίσες του 3. Επίσης, υποθέτουμε ότι οι ακόλουθοι δύο πίνακες περιέχουν τις σημασιολογικές αποστάσεις των στοιχείων για δύο διαφορετικές περιπτώσεις.

Τότε θα γίνονταν έλεγχος, για τις αποστάσεις, δηλαδή $distance(A, C)$, $distance(A, D)$, $distance(B, C)$, $distance(B, D)$ να είναι μεγαλύτερες ή ίσες του 3. Επομένως, στην πρώτη περίπτωση ο κανόνας δεν θα διαγραφόταν ενώ στην δεύτερη περίπτωση θα διαγραφόταν καθώς η σημασιολογική απόσταση $distance(A, C)=2$ δε θα κάλυπτε τον περιορισμό.

Παράδειγμα 10-2. Αποδοχή Κανόνα

	A	B	C	D
A	0	4	5	7
B	6	0	8	8
C	7	8	0	6
D	7	4	9	0

i) Πρώτη περίπτωση

	A	B	C	D
A	0	4	2	7
B	6	0	8	8
C	7	8	0	6
D	7	4	9	0

ii) Δεύτερη περίπτωση

```

if(timing.equals("Ending") || timing.equals("Over_all")){
    rules = deleteRulesBySubDistances(rules);
}

```

Κώδικας 10-9. Προσθήκη στην μέθοδο οApriori.findRulesQuickly

```

public FastVector[] deleteRulesBySubDistances(FastVector[] rules) {
    boolean step = true;
    for (int i = rules[0].size()-1; i >= 0; i--) {
        AprioriItemSet current1 = (AprioriItemSet)rules[0].elementAt(i);
        for(int i1 = 0; i1<current1.items().length && step == true; i1++)
            if(current1.itemAt(i1) != -1){
                AprioriItemSet current2 = (AprioriItemSet)rules[1].elementAt(i);
                for(int i2 = 0; i2<current2.items().length && step == true; i2++)
                    if(current2.itemAt(i2) != -1)
                        if(!checkDistance(distances[i1][i2]))
                            step = false;
            }
        if(step == false){
            rules[0].removeElementAt(i);
            rules[1].removeElementAt(i);
            rules[2].removeElementAt(i);
            step = true;
        }
    }
    return rules;
}

```

Κώδικας 10-10. Μέθοδος οApriori.deleteRulesBySubDistances

10.4 MappingFile.java

Η κλάση MappingFile.java ανήκει στο πακέτο ontology. Σκοπός της είναι να παρέχει μεθόδους για τη διαχείριση αρχείων που περιέχουν τις αντιστοιχίες των στοιχείων των δεδομένων εκπαίδευσης με τις κατάλληλες έννοιες μιας Οντολογίας.

Αρχικά, ορίστηκε ένας νέος τύπος αρχείων με την κατάληξη .map όπου θα αποθηκεύονται μαζί με τις Οντολογίες προς χρήση στο τοπικό φάκελο C:\patternMiner\ontology_files. Τα συγκεκριμένα αρχεία θα περιέχουν σε κάθε γραμμή ένα όνομα από τα δεδομένα εκπαίδευσης ακολουθούμενο από το χαρακτήρα tab και το όνομα της έννοιας που αντιστοιχεί στην Οντολογία. Ένα στοιχείο από τα δεδομένα εκπαίδευσης θα μπορεί να αντιστοιχηθεί με μόνο μία έννοια Οντολογίας, ενώ μία έννοια Οντολογίας θα μπορεί να είναι αντιστοιχισμένη με πολλά στοιχεία.



Διάγραμμα 10-1. Έννοιες – Συσχετίσεις Mapping File

Ακολουθεί ένα παράδειγμα ενός Mapping File αρχείου, όπου οι τιμές {crit1, attacktype1, gname} είναι τα ονόματα των στοιχείων των δεδομένων εκπαίδευσης και οι τιμές {http://sa.aktivespace.org/ontologies/aktivesa#Goal, http://sa.aktivespace.org/ontologies/aktivesa#Phenomenon, http://sa.aktivespace.org/ontologies/aktivesa#TerroristOrganization } αντιστοιχούν στην ονομασία εννοιών της Οντολογίας AKTiveSAOntology.owl (2011).

Παράδειγμα 10-3. Mapping File

```
crit1 http://sa.aktivespace.org/ontologies/aktivesa#Goal
attacktype1 http://sa.aktivespace.org/ontologies/aktivesa#Phenomenon
gname http://sa.aktivespace.org/ontologies/aktivesa#TerroristOrganization
```

Ενδιαφέρουσες μέθοδοι που έχουν υλοποιηθεί εδώ είναι οι ακόλουθες:

readMappingFile(fn : String)	Φόρτωση αντιστοιχίας από αρχείο	Κώδικας 10-11. Μέθοδος <i>MappingFile.readMappingFile</i>
checkMappingFile(fn : String) : Boolean	Έλεγχος αν ακολουθεί το αρχείο μια προκαθορισμένη μορφή για την αντιστοιχία	Κώδικας 10-12. Μέθοδος <i>MappingFile.checkMappingFile</i>
get(dataItem : String) : String	Επιστρέφει την ονομασία της έννοιας της Οντολογίας που αντιστοιχεί σε ένα στοιχείο των δεδομένων εκπαίδευσης	Κώδικας 10-13. Μέθοδος <i>MappingFile.get</i>
getHashtable() : Hashtable	Επιστρέφει την αντιστοιχία	Κώδικας 10-14. Μέθοδος <i>MappingFile.getHashtable</i>
toString() : String	Επιστρέφει την αντιστοιχία σε String μορφή	Κώδικας 10-15. Μέθοδος <i>MappingFile.toString</i>

```
public boolean readMappingFile() throws Exception{
    FileInputStream fstream = new FileInputStream(filename);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    if(checkMappingFile(filename)){
        String strLine;
        String split[];
        while ((strLine = br.readLine()) != null) {
            split = strLine.split("\t");
            itemMap.put(split[0],split[1]); } }
    else{
        in.close();
        return false; }
    in.close();
    return true;
} //readMappingFile
```

Κώδικας 10-11. Μέθοδος MappingFile.readMappingFile

```
public boolean checkMappingFile(String fn) throws Exception{
    filename = fn;
    FileInputStream fstream = new FileInputStream(filename);
    DataInputStream in = new DataInputStream(fstream);
    BufferedReader br = new BufferedReader(new InputStreamReader(in));
    boolean r =checkMappingFile(br);
    in.close();
    return r;
} //checkMappingFile
```

Κώδικας 10-12. Μέθοδος MappingFile.checkMappingFile

```
public String get(String dataItem){
    return (String) itemMap.get(dataItem);
}
```

Κώδικας 10-13. Μέθοδος MappingFile.get

```
public Hashtable getHashtable(){
    return itemMap;
}
```

Κώδικας 10-14. Μέθοδος MappingFile.getHashtable

```
public String toString(){
    StringBuffer text = new StringBuffer();
    Set set = itemMap.keySet();
    Iterator itr = set.iterator();
    while(itr.hasNext()){
        String str = (String) itr.next();
        text.append(str+"\t" + itemMap.get(str)+"\n");
    }
    return text.toString();
}
```

Κώδικας 10-15. Μέθοδος MappingFile.toString

10.5 OntologyReasoner.java

Η κλάση OntologyReasoner.java ανήκει στο πακέτο ontology. Σκοπός της είναι να διαβάζει αρχεία που περιέχουν OWL οντολογίες και να επιστρέφει χρήσιμες πληροφορίες. Ενδιαφέρουσες μέθοδοι που έχουν υλοποιηθεί εδώ είναι:

Πίνακας 10-1. Μέθοδοι της OntologyReasoner.java

readOntology()	Φορτώνει ένα μοντέλο Οντολογία σε ένα InfModel	Κώδικας 10-16
readOntologySchemaModel()	Φορτώνει ένα μοντέλο και το σχήμα των δεδομένων μιας Οντολογίας σε ένα InfModel	Κώδικας 10-17
getResource(rsrc : String) : Resource	Επιστρέφει την Πηγή μιας έννοιας της Οντολογίας	Κώδικας 10-18
getInfModel() : InfModel	Επιστρέφει το InfModel του αντικειμένου της τάξης	Κώδικας 10-19
validation() : boolean {αφηρημένη}	Ελέγχει αν μια Οντολογία είναι συνεπής. Υλοποιείται από της κλάσεις που την επεκτείνουν. Όπως η PelletReasoner, JenaReasoner	Κώδικας 10-20
<pre>public void readOntology()throws JenaException{ base = ModelFactory.createDefaultModel(); base.read(model); infmodel = ModelFactory.createInfModel(reasoner, base); }</pre>		

Κώδικας 10-16. Μέθοδος OntologyReasoner.readOntology

```
public void readOntologySchemaModel()throws JenaException{
    base = FileManager.get().loadModel(model);
    data = FileManager.get().loadModel(schema);
    reasoner = reasoner.bindSchema(base);
    reasoner.setDerivationLogging(true);
    infmodel = ModelFactory.createInfModel(reasoner, data);
}
```

Κώδικας 10-17. Μέθοδος `OntologyReasoner.readOntologySchemaModel`

```
public Resource getResource(String rsrc){
    return infmodel.getResource(rsrc);
}
```

Κώδικας 10-18. Μέθοδος `OntologyReasoner.getResource`

```
public InfModel getInfModel(){
    return infmodel;
}
```

Κώδικας 10-19. Μέθοδος `OntologyReasoner.getInfModel`

```
abstract public boolean validation();
```

Κώδικας 10-20. Μέθοδος `OntologyReasoner.validation`

10.6 JenaReasoner.java

Η κλάση `JenaReasoner.java` του πακέτου `ontology` επεκτείνει την κλάση `ontology.OntologyReasoner` υλοποιώντας την αφηρημένη μέθοδο `validation()`. Η συγκεκριμένη τάξη χρησιμοποιεί τον `Reasoner` αλλά και τις μεθόδους της `Jena` για να εξετάσει αν μια οντολογία είναι συνεπής.

```
public boolean validation() {
    ValidityReport validity = infmodel.validate();
    if (validity.isValid()) {
        StringBuffer text = new StringBuffer();
        for (Iterator i = validity.getReports(); i.hasNext(); ) {
            ValidityReport.Report report = (ValidityReport.Report)i.next();
            text.append(report+"\n");
        }
        System.out.println(text);
        return false;
    }
    return true;
} //validation
```

Κώδικας 10-21. Μέθοδος `JenaReasoner.validation`

10.7 PelletReasoner.java

Η κλάση `PelletReasoner.java` του πακέτου `ontology` επεκτείνει την κλάση `ontology.OntologyReasoner` υλοποιώντας την αφηρημένη μέθοδο `validation()`. Η συγκεκριμένη τάξη χρησιμοποιεί τον `Reasoner` αλλά και τις μεθόδους της `Pellet` για να εξετάσει αν μια οντολογία είναι συνεπής.


```

public boolean validation() {
    PelletInfGraph pellet = (PelletInfGraph) infmodel.getGraph();
    // check for inconsistency
    // create an inferencing model using Pellet reasoner
    if( !pellet.isConsistent() ) {
        //System.out.println("\nInConsistency Explanation: ");
        // create an inferencing model using Pellet reasoner
        Model explanation = pellet.explainInconsistency();
        // print the explanation
        explanation.write( System.out );
        return false;
    }
    return true;
}
}

```

Κώδικας 10-22. Μέθοδος PelletReasoner.validation

10.8 Metrics

Η αφηρημένη κλάση Metrics του πακέτου ontology δίνει τις αναγκαίες μεθόδους που πρέπει να υλοποιούν οι απόγονοί της για τον υπολογισμό αποστάσεων μέσα σε μια Οντολογία. Η χρήση της κυρίως είναι να προσφέρει ένα πρότυπο για την αποφυγή ασυνεπειών μέσα στον κώδικα.

```

abstract public double distance(InfModel infmodel, Resource resource1, Resource resource2);

```

Κώδικας 10-23. Μέθοδος Metrics.distance

```

abstract public boolean isDuplex();

```

Κώδικας 10-24. Μέθοδος Metrics.isDuplex

10.9 JenaMetrics.java

Η κλάση JenaMetrics.java είναι ένα παράδειγμα υλοποίησης μιας επέκτασης της Metrics για τον υπολογισμό αποστάσεων. Εδώ χρησιμοποιείται η μέθοδος της Jena findShortestPath η οποία επιστρέφει την απόσταση που έχουν δύο πόροι (Resource) μέσα σε μία Οντολογία. Η findShortestPath έχει την τάση να κινείται στο ιεραρχούμενο γράφο μιας οντολογίας από κάτω προς τα πάνω με αποτέλεσμα πολλές φορές να δίνει τιμές προς μόνο μία κατεύθυνση A->B ενώ για την κατεύθυνση B->A να επιστρέφει τιμή -1, δηλαδή ότι δεν υπάρχει μονοπάτι. Η μέθοδος getFilter χρησιμοποιείται για να ενημερώσει την findShortestPath ποιες θα είναι οι σχέσεις τις οποίες θα χρησιμοποιεί για να πάει από έναν πόρο σε έναν άλλο.

```

@Override
public double distance(InfModel infmodel, Resource resource1, Resource resource2) {
    shortestPath = OntTools.findShortestPath(infmodel, resource1, resource2, new OntTools.PredicatesFilter(getFilter()));
    double length = -1;
    if(shortestPath!=null)
        length = shortestPath.size();
    return length;
}

```

Κώδικας 10-25. Μέθοδος JenaMetrics.distance

```

public static Collection<Property> getFilter(){
    Collection<Property> m_preds = new HashSet<Property>();
    m_preds.add(OWL.onProperty);
    m_preds.add(OWL.equivalentClass);
    m_preds.add(RDFS.domain);
    m_preds.add(RDFS.range);
    m_preds.add(RDFS.label);
    m_preds.add(RDFS.subClassOf);
    m_preds.add(RDFS.subPropertyOf);
    return m_preds;
} //getFilter

```

Κώδικας 10-26 Μεθοδος JenaMetrics.getFilter

```

@Override
public boolean isDuplex() {
    return false;
}

```

Κώδικας 10-27 Μεθοδος JenaMetrics.isDuplex

10.10 Σύνοψη

Στην ενότητα αυτή εκτός από την παρουσίαση μέρους του πηγαίου κώδικα της εφαρμογής παρουσιάστηκαν έξι πολύ σημαντικές λειτουργίες του συστήματος. Σε αυτή την υποενότητα είναι σκόπιμο να επισημανθούν και να αριθμηθούν.

1. Απομάκρυνση στοιχείων από τα δεδομένα εκπαίδευσης που δεν είναι αντιστοιχισμένα με έννοιες της Οντολογίας. Αυτό παρέχει τη δυνατότητα στο χρήστη να τρέχει με τα ίδια δεδομένα εκπαίδευσης διαφορετικά σενάρια ανάλογα με τις σχέσεις που τους ορίζει μέσω μιας Οντολογίας. Επίσης, βοηθάει τον αλγόριθμο Apriori να απομακρύνει από πολύ νωρίς άχρηστα δεδομένα.
2. Απομάκρυνση στοιχείοσυνόλων κατά την φάση της δημιουργίας τους. Ο αλγόριθμος πρώτα δημιουργεί ένα σύνολο στοιχείοσυνόλων και σε μια δεύτερη φάση παράγει τους κανόνες μέσα από τα συγκεκριμένα στοιχείοσύνολα. Θεωρώντας, ότι η συγκεκριμένη φάση εκτελείται κατά την διάρκεια του Apriori έχει προταθεί μια σαφώς ορισμένη διαδικασία για την περαιτέρω μείωση των στοιχείοσυνόλων με βάση των σημασιολογικών αποστάσεων, με αποτέλεσμα να βοηθάται ο αλγόριθμος στο να βρεθούν πιο εύκολα χρήσιμοι κανόνες. Ο ορισμός της διαδικασίας δίνεται στην υποενότητα oApriori.java στην υλοποίηση της μεθόδου findLargeltemSets (Παράδειγμα 10-1. Αποδοχή Στοιχείοσυνόλου). Η βασική ιδέα είναι να μη διατηρούνται στοιχείοσύνολα που τα στοιχεία που τα αποτελούν δεν έχουν μεταξύ τους μια σημασιολογική σύνδεση.
3. Απομάκρυνση Κανόνων μετά το τέλος της εκτέλεσης του αλγορίθμου Apriori. Προφανώς, η συγκεκριμένη λειτουργία εκτελείται μετά το πέρας της διαδικασίας και δε στοχεύει τόσο στο να κάνει τον αλγόριθμο πιο γρήγορο αλλά να αναδείξει κανόνες που καλύπτουν τους σημασιολογικούς περιορισμούς που εισάγει ο χρήστης. Ο ορισμός της διαδικασίας δίνεται στην υποενότητα oApriori.java στην υλοποίηση της μεθόδου findRulesQuickly (Παράδειγμα 10-2. Αποδοχή Κανόνα). Εδώ η βασική ιδέα είναι να μην διατηρούνται κανόνες που τα στοιχεία που τα αποτελούν δεν έχουν μεταξύ τους μια σημασιολογική σύνδεση. Η διαφορά με το προηγούμενο κλάδεμα είναι ότι η συγκεκριμένη διαδικασία ενδιαφέρεται για την κατεύθυνση του κανόνα, δηλαδή θα πρέπει να υπάρχει μια σημασιολογική σύνδεση μεταξύ του αριστερού μέρους του κανόνα προς το δεξιό, ενώ το αντίθετο δεν το απασχολεί. Βέβαια κάτι τέτοιο είναι χρήσιμο για μετρικές που δίνουν άλλες τιμές για A->B και άλλες για B->A, όπως είναι η JenaMetrics.java

4. Διαχείριση συγκεκριμένων αρχείων Mapping File, όπου ο ορισμός τους και η δομή τους παρουσιάζεται στην υποενότητα MappingFile.java (Παράδειγμα 10-3. Mapping File).
5. Δυνατότητα εύκολης ενσωμάτωσης και χρήσης μίας Μηχανής Συλλογιστικής Ανάλυσης με τον μόνο περιορισμό να είναι συμβατός με την Jena. Εδώ γίνεται χρήση της κλάσης OntologyReasoner.
6. Δυνατότητα ενσωμάτωσης και χρήσης διαφορετικών Μετρικών για τον ορισμό αποστάσεων ή ομοιότητας μεταξύ εννοιών μιας OWL Οντολογίας υλοποιώντας την μέθοδο distance της κλάσης Metrics.

11. Έλεγχος (Εφαρμογή – Αποτίμηση)

Σε αυτή την ενότητα θα εξετασθεί κατά πόσο το σύστημα ικανοποιεί τις απαιτήσεις ως προς την βελτίωση των παραγόμενων κανόνων. Ο σκοπός ουσιαστικά είναι να παραχθούν κανόνες που να καλύπτουν τους σημασιολογικούς περιορισμούς που θέτει ο χρήστης. Οι σημασιολογικοί περιορισμοί έχουν να κάνουν με το πόσο θα πρέπει να είναι η απόσταση ή η ομοιότητα μεταξύ των στοιχείων εκπαίδευσης κατά την διάρκεια εκτέλεσης του Apriori ή μετά το πέρας της διαδικασίας.

Για το λόγο αυτό θα εκτελεστούν κάποια αντιπροσωπευτικά σενάρια βασιζόμενα σε αντικειμενικά δεδομένα. Αρχικά, τα δεδομένα εκπαίδευσης του αλγορίθμου θα προέρχονται από την Global Terrorism Database (2011) από την οποία θα χρησιμοποιηθούν μόνο τα πεδία {crit1, attacktype1, gname}. Κάθε μία εγγραφή αυτών των τριών πεδίων αναφέρεται σε μία τρομοκρατική επίθεση που έχει πραγματοποιηθεί. Συγκεκριμένα το crit1 παίρνει τιμές 0 ή 1 όπου η τιμή 1 εκφράζει ότι η τρομοκρατική επίθεση πραγματοποιήθηκε για την επίτευξη ενός πολιτικού, οικονομικού, θρησκευτικού, ή κοινωνικού στόχου. Το πεδίο attacktype1 παίρνει τιμές από 1 έως 9 και εκφράζει κωδικοποιημένα τον τύπο της επίθεσης. Τέλος το πεδίο gname περιέχει το όνομα του κύριου δράστη του συμβάντος, μετά από επεξεργασία το πεδίο αντί να παίρνει αλφαριθμητικές τιμές έχει ως πεδίο τιμών το φυσικό σύνολο από 1 έως 1508 ή την τιμή -9 σε περίπτωση που ο δράστης δεν αναφέρεται. Με βάση τα παραπάνω δημιουργείται το .arff αρχείο το οποίο αποτελεί είσοδο για τον αλγόριθμο. Μέρος του αρχείου παρουσιάζεται στην επόμενη εικόνα

```
@relation terrorism

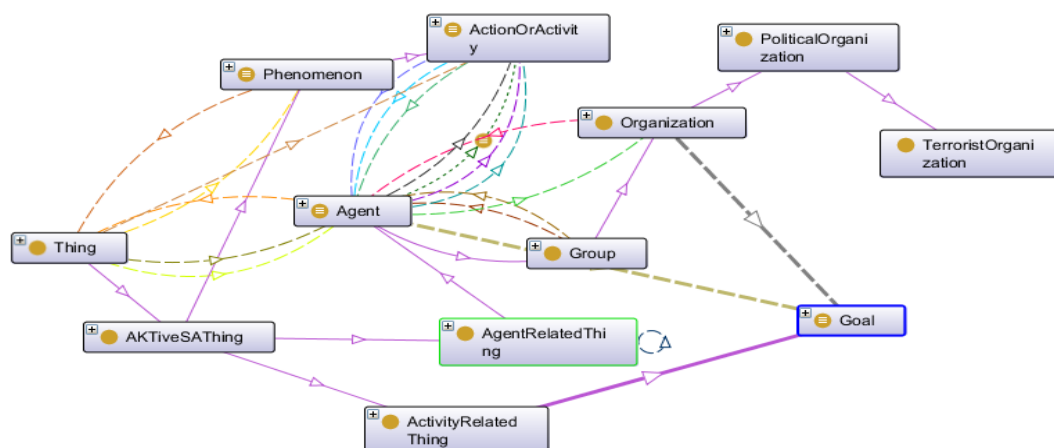
@attribute crit1 {0,1}
@attribute attacktype1 {1,2,3,4,5,6,7,8,9}
@attribute gname {1,2,3,...,1508}

@data
"1","7","1452"
"1","3","994"
"1","1","532"
"1","2","994"
"1","1","1170"
"1","1","1170"
```

Εικόνα 11-1. Μέρος αρχείου δεδομένων εκπαίδευσης Apriori

Για την σημασιολογική απεικόνιση των δεδομένων χρησιμοποιείται η OWL Οντολογία AKTiveSAOntology (2011). Μέρος της οποίας παρουσιάζεται στην εικόνα που ακολουθεί. Η εικόνα λήφθηκε μέσω του εργαλείου Protégé. Σύμφωνα με την οντολογία και την μετρική JenaMetrics που έχει υλοποιηθεί οι σημασιολογικές αποστάσεις μεταξύ των πόρων είναι:

- απόσταση(TerroristOrganization-> Goal)=3
- απόσταση(Goal -> TerroristOrganization)=-1, δηλαδή δεν υπάρχει σύνδεση
- απόσταση(TerroristOrganization-> Phenomeno)=4
- απόσταση(Phenomeno-> TerroristOrganization)=-1
- απόσταση(Phenomeno->Goal)=-1
- απόσταση(Goal-> Phenomeno)=-1



Εικόνα 11-2. Μέρος OWL Οντολογίας AKTiveSAOntology

Βάσει της Οντολογίας και των πεδίων της βάσης δημιουργείται το Mapping File με την αντιστοίχησή τους (Παράδειγμα 10-3. Mapping File). Όπου το πεδίο crit1 αντιστοιχίζεται με τον πόρο Goal, το πεδίο attacktype1 με τον πόρο Phenomenon και τέλος το πεδίο gname με τον πόρο TerroristOrganization

11.1 Πρώτο Σενάριο

Στο πρώτο σενάριο θα εκτελεσθεί ο Αρriori Αλγόριθμος χωρίς την ενσωμάτωση της Οντολογίας ώστε τα αποτελέσματά του να αποτελούν σημείο ορισμού για τους ελέγχους που θα πραγματοποιηθούν. Θα γίνουν δύο δοκιμές, η πρώτη με τιμή εμπιστοσύνης ίση με 0.9 και η δεύτερη με τιμή εμπιστοσύνης ίση με 0.4 ενώ η υποστήριξη θα μένει ίδια και για τις δύο με δυνατότητα μείωσης ως την τιμή 0.1 ώστε να παραχθούν όσο το δυνατόν περισσότεροι κανόνες. Στους δύο πίνακες που ακολουθούν παρουσιάζονται τα αποτελέσματα του αλγορίθμου.

Πίνακας 11-1. Αποτελέσματα Αρriori με min Conf: 0.9

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.9, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1
Αποτελέσματα	<ol style="list-style-type: none"> 1. attacktype1=1 5380 ==> crit1=1 5347 conf:(0.99) 2. attacktype1=2 11475 ==> crit1=1 11158 conf:(0.97) 3. attacktype1=3 18174 ==> crit1=1 17607 conf:(0.97) 4. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 5. attacktype1=2 gname=1452 5152 ==> crit1=1 4953 conf:(0.96) 6. attacktype1=3 gname=1452 10044 ==> crit1=1 9595 conf:(0.96)

Πίνακας 11-2. Αποτελέσματα Αρriori με min Conf: 0.3

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1
Αποτελέσματα	<ol style="list-style-type: none"> 1. attacktype1=1 5380 ==> crit1=1 5347 conf:(0.99) 2. attacktype1=2 11475 ==> crit1=1 11158 conf:(0.97) 3. attacktype1=3 18174 ==> crit1=1 17607 conf:(0.97) 4. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 5. attacktype1=2 gname=1452 5152 ==> crit1=1 4953 conf:(0.96) 6. attacktype1=3 gname=1452 10044 ==> crit1=1 9595 conf:(0.96) 7. attacktype1=3 18174 ==> gname=1452 10044 conf:(0.55) 8. crit1=1 attacktype1=3 17607 ==> gname=1452 9595 conf:(0.54)

9. attacktype1=3 18174 ==> crit1=1 gname=1452 9595 conf:(0.53)
10. crit1=1 41706 ==> gname=1452 20627 conf:(0.49)
11. gname=1452 21384 ==> attacktype1=3 10044 conf:(0.47)
12. crit1=1 gname=1452 20627 ==> attacktype1=3 9595 conf:(0.47)
13. attacktype1=2 11475 ==> gname=1452 5152 conf:(0.45)
14. gname=1452 21384 ==> crit1=1 attacktype1=3 9595 conf:(0.45)
15. crit1=1 attacktype1=2 11158 ==> gname=1452 4953 conf:(0.44)
16. attacktype1=2 11475 ==> crit1=1 gname=1452 4953 conf:(0.43)
17. crit1=1 41706 ==> attacktype1=3 17607 conf:(0.42)

11.2 Δεύτερο Σενάριο

Στο συγκεκριμένο σενάριο θα ακολουθήσουμε τις παραμέτρους της πρώτης περίπτωσης του πρώτου σεναρίου αλλά θα εισάγουμε και την κατάλληλη Οντολογία ώστε να εκτελεστούν οι νέες λειτουργίες του αλγορίθμου. Η διαδικασία θα εκτελεσθεί τρεις φορές με τις ίδιες παραμέτρους αλλά θα αλλάζει μόνο το πότε θα γίνεται χρήση της Οντολογίας, δηλαδή θα αλλάζει μόνο η παράμετρος Timing. Έτσι έχουμε:

Πίνακας 11-3. Δεύτερο Σενάριο Περίπτωση α'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.9, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: During
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96)

Πίνακας 11-4. Δεύτερο Σενάριο Περίπτωση β'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.9, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Ending
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96)

Πίνακας 11-5. Δεύτερο Σενάριο Περίπτωση γ'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.9, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Over_all
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96)

Και για τις τρεις περιπτώσεις οι κανόνες από έξι μειώνονται σε έναν δηλαδή έχουμε μια μείωση της τάξης του 83%.

11.3 Τρίτο Σενάριο

Στο συγκεκριμένο σενάριο θα ακολουθήσουμε τις παραμέτρους της δεύτερης περίπτωσης του πρώτου σεναρίου αλλά θα εισάγουμε και την κατάλληλη Οντολογία ώστε να εκτελεστούν οι νέες λειτουργίες του αλγορίθμου. Η διαδικασία θα εκτελεσθεί τρεις φορές με τις ίδιες παραμέτρους αλλά θα αλλάζει μόνο το πότε θα γίνεται χρήση της Οντολογίας, δηλαδή θα αλλάζει μόνο η παράμετρος Timing. Έτσι έχουμε:

Πίνακας 11-6. Τρίτο Σενάριο Περίπτωση α'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: During
------------	---

Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 2. attacktype1=3 18174 ==> gname=1452 10044 conf:(0.55) 3. crit1=1 41706 ==> gname=1452 20627 conf:(0.49) 4. gname=1452 21384 ==> attacktype1=3 10044 conf:(0.47) 5. attacktype1=2 11475 ==> gname=1452 5152 conf:(0.45)
--------------	--

Πίνακας 11-7. Τρίτο Σενάριο Περίπτωση β'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Ending
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 2. gname=1452 21384 ==> attacktype1=3 10044 conf:(0.47) 3. gname=1452 21384 ==> crit1=1 attacktype1=3 9595 conf:(0.45)

Πίνακας 11-8. Τρίτο Σενάριο Περίπτωση γ'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Over_all
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 2. gname=1452 21384 ==> attacktype1=3 10044 conf:(0.47)

Εδώ από τους δεκαεπτά κανόνες που παρήγαγε ο αρχικός αλγόριθμος τώρα πέφτουμε στους πέντε, τρεις, δύο για την πρώτη, δεύτερη και τρίτη περίπτωση αντίστοιχα. Δηλαδή, έχουμε μείωση κανόνων της τάξης από 70 έως 88%.

11.4 Τέταρτο Σενάριο

Στο συγκεκριμένο σενάριο θα ακολουθήσουμε τις παραμέτρους της δεύτερης περίπτωσης του πρώτου σεναρίου αλλά θα εισάγουμε και την κατάλληλη Οντολογία ώστε να εκτελεστούν οι νέες λειτουργίες του αλγορίθμου. Η διαδικασία θα εκτελεσθεί τρεις φορές με τις ίδιες παραμέτρους αλλά θα αλλάζει μόνο το πότε θα γίνεται χρήση της Οντολογίας, δηλαδή θα αλλάζει μόνο η παράμετρος Timing. Η διαφορά του τρίτου και τέταρτου σεναρίου έγκειται στις τιμές που λαμβάνει η παράμετρος Distance/ Similarite Έτσι έχουμε:

Πίνακας 11-9. Τέταρτο Σενάριο, Περίπτωση α'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0 && <=4, Timing: During
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96) 2. crit1=1 41706 ==> gname=1452 20627 conf:(0.49)

Πίνακας 11-10. Τέταρτο Σενάριο, Περίπτωση β'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Ending
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96)

Πίνακας 11-11. Τέταρτο Σενάριο, Περίπτωση γ'

Παράμετροι	Αριθμός κανόνων: 1000, Ελάχιστη Εμπιστοσύνη: 0.3, Δελτα: 0.05, Ελάχιστη Υποστήριξη: 0.1 Reasoner: Pellet, Μετρική: JenaMetric, Απόσταση: >=0, Timing: Over_all
Αποτελέσματα	1. gname=1452 21384 ==> crit1=1 20627 conf:(0.96)

Εδώ από τους δεκαεπτά κανόνες που παρήγαγε ο αρχικός αλγόριθμος τώρα πέφτουμε στους δύο, έναν, έναν για την πρώτη, δεύτερη και τρίτη περίπτωση αντίστοιχα. Δηλαδή, έχουμε μείωση κανόνων της τάξης από 88 έως 94%.

11.5 Σύνοψη

Όπως πάντα σε μια επέκταση ενός συστήματος υπάρχουν θετικά αλλά και αρνητικά αποτελέσματα. Στην διαπραγματευόμενη περίπτωση δύο είναι τα κύρια μειονεκτήματα. Το πρώτο είναι ότι σε σχέση με τον πρωτότυπο αλγόριθμο του Weka Arriori αυξάνεται η χρήση της μνήμης, καθώς η νέα έκδοσή χρειάζεται έναν δυσδιάστατο πίνακα μεγέθους όσα είναι τα χαρτογραφημένα στοιχεία εκπαίδευσης στο τετράγωνο, το οποίο θα διατηρεί τις σημασιολογικές τους αποστάσεις. Δεν συμπεριλαμβάνεται ο χρόνος κατασκευής του καθώς μπορεί να κατασκευαστεί μία φορά, να αποθηκευτεί και έπειτα να χρησιμοποιείται χωρίς κανένα επιπλέον κόστος. Το δεύτερο μειονέκτημα που παρατηρείται είναι ότι έχουν προστεθεί εξ' ορισμού άλλοι δύο έλεγχοι, ο πρώτος για την μείωση των στοιχειοσυνόλων και ο δεύτερος για την μείωση των παραγόμενων κανόνων που κατά βάση θα μπορούσε κάποιος να υποστηρίξει ότι σε συγκεκριμένες περιπτώσεις αυξάνουν το χρόνο εκτέλεσης του αλγορίθμου.

Στον αντίποδα έχουμε το πλεονέκτημα ότι από πολύ νωρίς ο αλγόριθμος αποδεδεσμεύεται από μη χρήσιμα στοιχειοσύνολα. Αυτό έχει ως αποτέλεσμα ο αλγόριθμος να κάνει λιγότερους κύκλους για την εύρεση των συνόλων των στοιχειοσυνόλων. Για παράδειγμα όπως βλέπουμε στον ακόλουθο πίνακα που παρουσιάζει συνοπτικά τα αποτελέσματα των προηγούμενων σεναρίων, το *Σενάριο β', Περίπτωση α'* έχει τρία στοιχειοσύνολα λιγότερα στον L(2) ενώ τελειώνει σε ένα κύκλο λιγότερο από το *Σενάριο α', Περίπτωση α'*. Το δεύτερο πλεονέκτημα είναι η μείωση των παραγόμενων κανόνων έως και 94% από αυτούς του αρχικού αλγορίθμου, για παράδειγμα στο *Σενάριο δ', Περίπτωση γ'*. Φυσικά, το τρίτο αλλά όχι το τελευταίο πλεονέκτημα είναι ότι οι παραγόμενοι κανόνες αποκτούν μια σημασιολογική δυναμική που είναι απαραίτητη για την αξιολόγηση τους από τον εκάστοτε ειδικό.

Πίνακας 11-12. Σύγκριση Σεναρίων

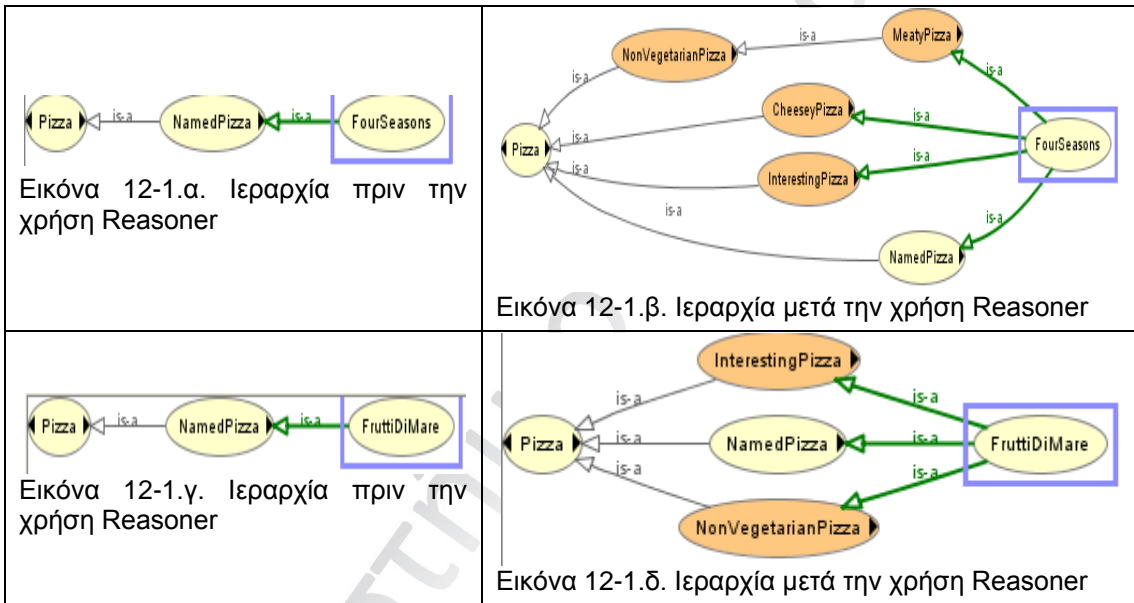
	L(1)	L(2)	L(3)	Κανόνες
Σενάριο α' Περίπτωση α'	5	6	2	6
Σενάριο α' Περίπτωση β'	5	6	2	17
Σενάριο β' Περίπτωση α'	5	3	-	1
Σενάριο β' Περίπτωση β'	5	6	2	1
Σενάριο β' Περίπτωση γ'	5	3	-	1
Σενάριο γ' Περίπτωση α'	5	3	-	5
Σενάριο γ' Περίπτωση β'	5	6	2	3
Σενάριο γ' Περίπτωση γ'	5	3	-	2
Σενάριο δ' Περίπτωση α'	5	1	-	2
Σενάριο δ' Περίπτωση β'	5	6	2	1
Σενάριο δ' Περίπτωση γ'	5	1	-	1

12. Επεκτάσεις

Στην ενότητα της υλοποίησης παρουσιάστηκε ο πηγαίος κώδικας που θα μπορούσε να χαρακτηριστεί ως ο κορμός του συστήματος. Σε αυτή τη νέα ενότητα θα διαπραγματευθούν κάποια επιπρόσθετα κομμάτια που επεκτείνουν τις δυνατότητες του.

12.1 Μετρική TaxonomyMetrics.java

Η TaxonomyMetrics.java υλοποιεί την κλάση Metrics και παρέχει στο σύστημα άλλη μία μετρική για τον υπολογισμό της απόστασης/ ομοιότητας μεταξύ δύο εννοιών μίας Οντολογίας. Στόχος της είναι να εκμεταλλευτεί τη νέα ιεραρχική δομή που παράγει ο Reasoner, ο οποίος διαβάξει την Οντολογία. Στην επόμενη εικόνα παρουσιάζεται ένα παράδειγμα από την Οντολογία *pizza* του Protégé. Στις εικόνες α, γ εμφανίζεται η ιεραρχία όπως δίνεται από την Οντολογία ενώ στις εικόνες β, δ εμφανίζονται οι νέες ιεραρχίες που παράχθηκαν από τον Pellet Reasoner μετά τον συμπερασμό. Αυτό πραγματοποιήθηκε γιατί στην δεύτερη περίπτωση ο Reasoner εκμεταλλεύτηκε τις ιδιότητες που συνδέουν τις έννοιες για να συμπεράνει τις πραγματικές σχέσεις που τις διέτουν.



Εικόνα 12-1. Ιεραρχίες πριν και μετά την χρήση Reasoner

Έτσι η TaxonomyMetrics προσπαθεί να εκμεταλλευτεί τη νέα ιεραρχία που παράγει η εκάστοτε Μηχανή Συλλογιστικής Ανάλυσης και κατ' επέκταση τις ιδιότητες που διέτουν τις έννοιες της Οντολογίας. Έτσι αν γίνει η θεώρηση ότι οι πρόγονοι μιας έννοιας C_i εμφανίζονται στο σύνολο:

$$Pa(C_i, H^C) = \{C_j \in c \mid H^C(C_i, C_j)\} \text{ όπου } H^C \text{ δηλώνει ότι η } C_i \text{ είναι υπο-έννοια της } C_j$$

Τότε μπορεί να υπολογιστεί η ομοιότητα μεταξύ δύο εννοιών από την σχέση:

$$TaxonomyMetrics = \frac{|(Pa(C_i, H^C) \cap Pa(C_j, H^C))|}{|(Pa(C_i, H^C) \cup Pa(C_j, H^C))|}$$

Δηλαδή, το πηλίκο των κοινών απογόνων προς το σύνολο όλων των εννοιών που σχετίζονται ιεραρχικά με τις έννοιες (C_i, C_j) .

Ο πηγαίος κώδικας που την υλοποιεί αποτελείται από πέντε μεθόδους:

Πίνακας 12-1. Μέθοδοι της TaxonomyMetrics.java

distance(infmodel : InfModel, resource1 : Resource, resource2 : Resource) : double	Υλοποιεί την αφηρημένη μέθοδο distance της Metrics.java	Κώδικας 12-1
isDuplex() : boolean	Υλοποιεί την αφηρημένη μέθοδο distance της Metrics.java	Κώδικας 12-2
taxonomySimilarity(infmodel :	Υπολογίζει τον τύπο ομοιότητας αφού	Κώδικας

Model, resource1 : Resource, resource2: Resource) : double	καλέσει τις μεθόδους, getHierarchySimilarity, getHierarchyDeference	12-3
getHierarchySimilarity(subClassSet1 : ArrayList, subClassSet2: ArrayList): int	Υπολογίζει τους κοινούς προγόνους των δύο εννοιών, οι πρόγονοι δίνονται σε δύο ArrayList	Κώδικας 12-4
getHierarchyDeference(subClassSet1 : ArrayList, subClassSet2: ArrayList): int	Υπολογίζει τους διαφορετικούς πρόγονους των δύο εννοιών, οι πρόγονοι δίνονται σε δύο ArrayList	Κώδικας 12-5

Συγκεκριμένα οι μέθοδοι είναι:

```
@Override
public double distance(InfModel infmodel, Resource resource1, Resource resource2) {
    double dstnc=taxonomySimilarity(infmodel, resource1, resource2);
    return dstnc;
}
```

Κώδικας 12-1. Μέθοδος TaxonomyMetrics.distance()

```
@Override
public boolean isDuplex() {
    return true;
}
```

Κώδικας 12-2. Μέθοδος TaxonomyMetrics.isDuplex()

```
public double taxonomySimilarity(Model infmodel, Resource resource1, Resource resource2){
    StmtIterator subClassSet1 = infmodel.listStatements(resource1,
        RDFS.subClassOf,(Resource) null);
    ArrayList subClass1= (ArrayList) subClassSet1.toList();
    StmtIterator subClassSet2 = infmodel.listStatements(resource2,
        RDFS.subClassOf,(Resource) null);
    ArrayList subClass2= (ArrayList) subClassSet2.toList();
    int similar=getHierarchySimilarity(subClass1, subClass2);
    int difference1=getHierarchyDeference(subClass1, subClass2);
    int difference2=getHierarchyDeference(subClass2, subClass1);
    double sim= (double)similar/(double)(similar+difference1+difference2);
    return sim;
}
```

Κώδικας 12-3. Μέθοδος TaxonomyMetrics.taxonomySimilarity()

```
public int getHierarchySimilarity(ArrayList subClassSet1, ArrayList subClassSet2){
    int similarity=0;
    Iterator itr1 = subClassSet1.iterator();
    while(itr1.hasNext()) {
        Statement stmt1 = (Statement) itr1.next();
        Iterator itr2 = subClassSet2.iterator();
        if(stmt1.getObject().isURIResource())
            while(itr2.hasNext()) {
                Statement stmt2 = (Statement) itr2.next();
                if(stmt2.getObject().isURIResource())
```

```

        if(stmt1.getObject().equals(stmt2.getObject())){
            similarity++;
            break;
        }
    }
}
return similarity;
}

```

Κώδικας 12-4. Μέθοδος `TaxonomyMetrics.getHierarchySimilarity()`

```

public int getHierarchyDeference(ArrayList subClassSet1, ArrayList subClassSet2){
    int deference=0;
    boolean step =true;
    Iterator itr1 = subClassSet1.iterator();
    while(itr1.hasNext()) {
        Statement stmt1 = (Statement) itr1.next();
        Iterator itr2 = subClassSet2.iterator();
        if(stmt1.getObject().isURIResource())
            while(itr2.hasNext() && step) {
                Statement stmt2 = (Statement) itr2.next();
                if(stmt2.getObject().isURIResource())
                    if(stmt1.getObject().equals(stmt2.getObject()))
                        step=false;
            }
        else
            step=false;
        if(step)
            deference++;
        else
            step=true;
    }
    return deference;
}

```

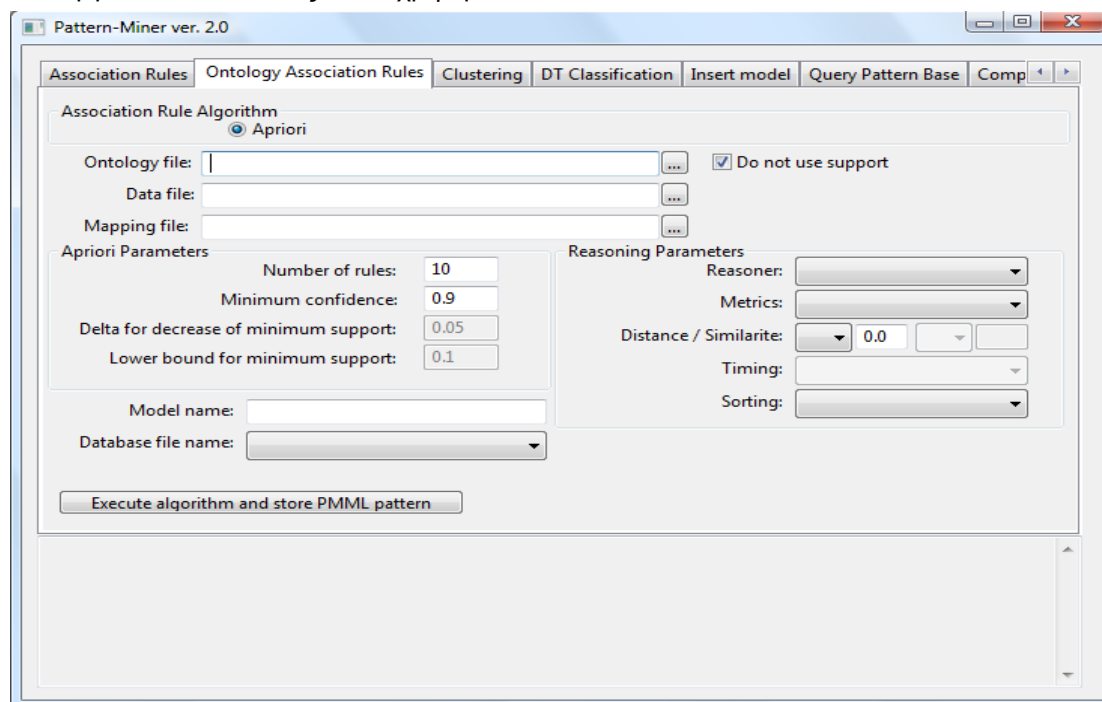
Κώδικας 12-5. Μέθοδος `TaxonomyMetrics.getHierarchyDeference()`

12.2 OAprioriWithoutSupport.java

Με αυτή την κλάση γίνεται προσπάθεια στο να δοθεί μεγαλύτερη ευθύνη στις αποστάσεις, που σχηματίζουν οι έννοιες μίας Οντολογίας, μέσα στον αλγόριθμο Apriori. Το διαφορετικό που παρουσιάζεται εδώ είναι ότι αντί ο αλγόριθμος να διαγράφει τα μη επιθυμητά στοιχειοσύνολα με βάση την υποστήριξη (support), τα απορρίπτει μόνο βασιζόμενος πάνω στις επιθυμητές αποστάσεις που δημιουργούνται στην Οντολογία. Επομένως, τα στοιχεία που εξετάζει ο νέος αλγόριθμος για την παραγωγή κανόνων είναι αρχικά οι αποστάσεις των εννοιών που αποτελούν τα στοιχειοσύνολα και έπειτα την εμπιστοσύνη (confidence) των κανόνων. Καθώς δεν δίνεται σημασία στην μετρική της υποστήριξης θα είναι απαραίτητος ο ορισμός ενός νέου τρόπου διάταξης των αποτελεσμάτων.

Η νέα διαδικασία θα εκτελείται μέσα από την διεπαφή *Ontology Association Rules*, η οποία τροποποιείται κατάλληλα για να μπορούν να ορισθούν οι παράμετροι της *oAprioriWithoutSupport*. Όπως παρουσιάζεται στην επόμενη εικόνα δύο βασικά πεδία προστίθενται. Το *Do not use support* με το οποίο όταν θα είναι τσεκαρισμένο θα ζητείται να τρέξει η νέα διαδικασία ενώ το *Sorting* θα καθορίζει τον τρόπο που θα γίνεται η διάταξη των

αποτελεσμάτων. Στην περίπτωση που θα επιλέγεται να τρέξει η νέα διαδικασία τα πεδία *Delta for decrease of minimum support*, *Lower bound for minimum support* και το *Timing* θα απενεργοποιούνται καθώς δε θα χρησιμεύουν πια.



Εικόνα 12-2. Νέα διεπαφή συστήματος

Η κλάση *oAprioriWithoutSupport* επεκτείνει την *Apriori* ενώ διατηρεί κάποια χαρακτηριστικά ίδια με την *oApriori*. Εδώ γίνεται παρουσίαση των μεθόδων που διαφοροποιούνται σε σχέση με την κλάση *oApriori*.

Στη βασική μέθοδο *buildAssociations* πλέον δε γίνεται πια αναφορά στο *support*, ενώ η διάταξη των κανόνων μπορεί να γίνει με δύο διαφορετικούς τρόπους. Αρχικά για κάθε κανόνα βρίσκεται μια μέση εννοιολογική απόσταση από τη μέθοδο *meanRuleNoDublexDistance*, η οποία προσθέτει τις αποστάσεις που σχηματίζουν μέσα στην οντολογία οι έννοιες του αριστερού μέρους ενός κανόνα με τις έννοιες του δεξιού μέρους του κανόνα δια τις συνολικές αποστάσεις. Έπειτα, ανάλογα με την επιλογή του χρήστη θα γίνει η διάταξη είτε αρχίζοντας από τις μεγαλύτερες αποστάσεις είτε από τις μικρότερες, ενώ στο τέλος γίνεται και μια δεύτερη διάταξη με βάση το *confidence*.

Παράδειγμα 12-1. Μέση σημασιολογική απόσταση κανόνα

Έστω ότι έχουμε τον κανόνα $A, B \Rightarrow C, D$ η μέση εννοιολογική απόσταση του κανόνα ορίζεται ως
$$\frac{\text{απόσταση}(A,C) + \text{απόσταση}(A,D) + \text{απόσταση}(B,C) + \text{απόσταση}(B,D)}{4}$$

Στη μέθοδο *findLargestItemSets* η μόνη αλλαγή που πραγματοποιείται είναι ότι δεν εκτελείται πια η μέθοδος που διέγραφε τα στοιχειοσύνολα ανάλογα με το επιθυμητό support ενώ η μέθοδος *deleteItemSetsByDistances* εκτελείται πάντα.

Η μέθοδος *findRulesQuickly* υλοποιείται όπως η ομώνυμη της προγονικής κλάσης Apriori, απλά εισάγεται ακόμα μια μέθοδος η *deleteRulesWithNaNConf* η οποία διαγράφει κανόνες όπου το confidence παίρνει μη επιτρεπτές τιμές NaN λόγω της ύπαρξης διαιρετέου στον τύπο του confidence με τιμή μηδέν. Αυτό γίνεται διότι δεν έχουν διαγραφεί στοιχειοσύνολα με support ίσο του μηδενός.

```
public FastVector[] deleteRulesWithNaNConf(FastVector[] rules) {
    boolean step = true;
    for (int i = rules[0].size()-1; i >= 0; i--) {
        if(Double.isNaN(((Double)rules[2].elementAt(i)).doubleValue())){
            rules[0].removeElementAt(i);
            rules[1].removeElementAt(i);
            rules[2].removeElementAt(i);
        }
    }
    return rules;
}
```

Κώδικας 12-8. Μέθοδος οAprioriWithoutSupport.deleteRulesWithNaNConf()

Τα αποτελέσματα που παράγονται από την εκτέλεση της παραπάνω διαδικασίας δείχνουν ότι ο αριθμός τους είναι πολύ μεγαλύτερος από αυτά των προηγούμενων διαδικασιών (Apriori, οApriori), παρόλα αυτά παρατηρείται να διατηρούν μια δυνατότερη και πιο πλούσια σημασιολογική ερμηνεία.

ΜΕΡΟΣ Γ – ΠΕΡΙΠΤΩΣΗ ΧΡΗΣΗΣ

13. Εισαγωγή

Οι επιστήμονες του χώρου ασχολήθηκαν από πολύ νωρίς με το πεδίο έρευνας των Οντολογιών και της Εξόρυξης Γνώσης από Δεδομένα, αλλά η εμπλοκή τους με τον τομέα της Τρομοκρατίας δεν είχε το ίδιο μέγεθος ενασχόλησης. Τα δεδομένα αυτά άλλαξαν μετά την 11^η Σεπτεμβρίου 2001, όπου και η ημερομηνία αυτή θεωρείται το σημείο μηδέν για την αλλαγή του τρόπου σκέψης στο ευρύτερο πεδίο της Ασφάλειας. Μετέπειτα, εμφανίζεται μία αυξημένη ενεργητικότητα στην έρευνα του πεδίου της τρομοκρατίας συνδυαστικά με άλλους τομείς. Ένα ευρέως γνωστό πρόγραμμα που εκτελέστηκε ήταν το Terrorism Information Awareness (TIA-αργότερα μετονομάστηκε σε Total Information Awareness) κάτω από την αιγίδα του Department of Defense Advanced Research Projects Agency (DARPA). Σε αυτή την ενότητα γίνεται μία προσπάθεια συνύπαρξης των τριών πεδίων, συγκεκριμένα θα εξεταστεί η ενσωμάτωση μιας Οντολογίας στην τεχνική Εξόρυξης Κανόνων Συσχέτισης της DM στο Πεδίο Γνώσης της Τρομοκρατίας.

Γενικά για τον όρο *Τρομοκρατία* δεν υπάρχει ένας καθολικά αποδεκτός ορισμός. Αυτή είναι και μια ένδειξη για την περιπλοκότητα του πεδίου, που μπορεί να εξετασθεί από διαφορετικές οπτικές γωνίες με διαφορετικά επιχειρήματα και στόχους. Σύμφωνα με την εγκυκλοπαίδεια Britanica (2011) η τρομοκρατία ορίζεται ως «η συστηματική χρήση βίας για να δημιουργήσει ένα γενικό κλίμα φόβου σε έναν πληθυσμό και να υλοποιήσει έτσι ένα συγκεκριμένο πολιτικό στόχο. Η τρομοκρατία έχει ασκηθεί από πολιτικές οργανώσεις με δεξιές και αριστερές πεποιθήσεις, από εθνικιστικές και θρησκευτικές ομάδες, από επαναστάτες, και ακόμα και από κρατικά θεσμικά όργανα όπως οι στρατοί, οι μυστικές υπηρεσίες, και η αστυνομία».

Ο Thuraisingham (2004) και η DeRosa (2004) δίνουν πραγματικά το στίγμα της DM ως εργαλείο κατά της τρομοκρατίας. Την αντιλαμβάνονται ως μια διαδικασία που θέτει ερωτήματα και εξάγει χρήσιμα πρότυπα ή τάσεις που συχνά είναι άγνωστες μέσα σε ένα σύνολο δεδομένων, τα οποία μπορούν να χρησιμοποιηθούν για πρόβλεψη συμπεριφοράς, αξιολόγηση κινδύνου, καθορισμού συσχετίσεων και άλλων ειδών ανάλυσης. Ο πρώτος

ερευνητής πραγματεύεται τις διάφορες τεχνικές προκλήσεις της DM για την εφαρμογή της στο πεδίο της τρομοκρατίας, και τις διάφορες απειλές και λύσεις που δίνει. Στο ίδιο μοτίβο η DeRosa αναφέρεται στην χρησιμότητα της DM και την αναγκαιότητα για κατανόηση των τεχνικών της. Ενδιαφέρον παρουσιάζει η κοινή τους αντίληψη ότι παρόλο που η DM μπορεί να θεωρηθεί ένα πολύ ισχυρό εργαλείο στα χέρια των αναλυτών και των ερευνητών, κάθε άλλο από μια πλήρης λύση κατά του προβλήματος ασφαλείας είναι. Πάντα πρέπει να θεωρείται ως ενισχυτικό και συμπληρωματικό εργαλείο πληροφόρησης που θα υποστηρίζει τα συμπεράσματα και τις αποφάσεις που παίρνουν οι ειδικοί. Τέλος, σε αυτό που καταλήγουν είναι ότι μέσω της DM υφίσταται ευκολότερη και πιο ισχυρή ανάλυση δεδομένων, τα οποία τα κάνουν πιο χρήσιμα και ελκυστικά για τις κυβερνήσεις. Βέβαια, τα αποτελέσματα του DM πρέπει να εξάγονται συνετά και μετά από αυστηρό έλεγχο, ενώ θα πρέπει να δίνεται πάντα ιδιαίτερο βάρος στην προστασία των προσωπικών δεδομένων.

Τα θετικά για τη χρήση της DM μέσα στο πεδίο της αντιτρομοκρατίας είναι ουσιαστικά οι ιδιαίτερες λειτουργίες που προσφέρει. Ένα από τα βασικά πλεονεκτήματά της είναι ότι παρουσιάζει τον ακριβή προσδιορισμό των αποτελεσμάτων βάσει φορμαλιστικών διαδικασιών. Αυτό την κάνει αξιόπιστη και επαναχρησιμοποιήσιμη, ενώ δίνει την δυνατότητα τροποποίησης της σε περίπτωση εξαγωγής λανθασμένων προτύπων. Πλέον ως ενισχυτικό, αξιόπιστο εργαλείο επιτρέπει στους αναλυτές να ασχολούνται περισσότερο με την ερμηνεία των αποτελεσμάτων, χωρίς να χάνουν υπολογιστικό χρόνο για την εύρεσή τους. Πολλές φορές κάτι τέτοιο είναι καθοριστικό, όπως για περιπτώσεις που οι αποφάσεις πρέπει να παρθούν σε πραγματικό χρόνο. Επιπλέον, προσφέρει διαδικασίες που έχουν αξία πρώτης γραμμής, όπως είναι η αναζήτηση σύνδεσης (link analysis), διαπίστωση ανωμαλιών (anomaly detection), ομοιότητας μεταξύ προτύπων, απομυθοποίηση προτύπων που θεωρούνταν επικίνδυνα ενώ στην ουσία δεν ήταν, την εύρεση νέων μεταβλητών που μπορούν να βοηθήσουν στην ανάλυση των τρομοκρατικών γεγονότων. Επίσης προσφέρει διαδικασίες αιχμής όπως την εξόρυξη γνώσης από πολυμεσικά δεδομένα, γράφους, τα πρότυπα και τον ιστό. Κατά τη διάρκεια του ψυχρού πολέμου ισχυρές διαδικασίες και μέσα εξελίχθηκαν για τη φύλαξη των πληροφοριών των εκάστοτε κρατών, επομένως σήμερα είναι πολύ σημαντικό να υπάρχουν εργαλεία που να βοηθούν στο συμπερασμό νέας γνώσης από ελάχιστα δυνατά δεδομένα. Παράδειγμα χρήσης της DM θα μπορούσε να είναι το χτύπημα της 11ης Σεπτεμβρίου. Στο χτύπημα αυτό υπήρχαν ενδείξεις, οι οποίες αν είχαν αναγνωρισθεί, συγκριθεί και αναλυθεί θα ήταν δυνατή η ελαχιστοποίηση των επιπτώσεων.

Καθώς η συλλογή δεδομένων για τρομοκρατικές ενέργειες είχε ξεκινήσει από πολύ νωρίς, η συσχέτιση τους με τις Οντολογίες ακολούθησε δύο κατευθύνσεις. Δηλαδή, γίνονταν προσπάθεια για ανακάλυψη των Οντολογιών μέσα από τα δεδομένα αλλά και η κατανόηση των υπάρχοντων δεδομένων μέσα από Οντολογίες που δημιουργούνταν με τη βοήθεια αναλυτών (Gruenwald και συν., 2003; Inyaeu και συν., 2009; Chmielewski και συν., 2009). Οι Gruenwald και συν. επεκτείνουν την βάση γνώσης της Oracle Text για να βελτιώσουν την αναζήτηση, ανάκτηση και κατηγοριοποίηση εγγράφων, τόσο μέσα σε μία βάση αλλά και από έγγραφα που προέρχονται από το Σημαιολογικό Ιστό, ο οποίος είναι και η κύρια πηγή για την ενημέρωση νέων τρομοκρατικών ομάδων και ενεργειών παγκοσμίως. Οι Inyaeu και συν. χρησιμοποιούν και αυτοί τις Οντολογίες, για εξαγωγή κατάλληλων δεδομένων από ειδησεογραφικά έγγραφα μέσα από το διαδίκτυο και για την ενημέρωση ενός Συστήματος Υποστήριξης Αποφάσεων. Τα οποία προαναφερθέντα δεδομένα ασχολούνται με την τρομοκρατία. Οι Chmielewski και συν. υποστηρίζουν ότι η ανάλυση των δεδομένων σχετικά με τη τρομοκρατία μπορεί να παράγει ενδιαφέροντα συμπεράσματα και χρησιμοποιούν τα εργαλεία του Σημαιολογικού Ιστού για να επιτύχουν επεκτασιμότητα και ευελιξία στην αναπαράσταση της γνώσης. Συγκεκριμένα συνδέουν τα δεδομένα μιας βάσης δεδομένων περί τρομοκρατίας με κατάλληλα ορισμένη οντολογία για να υποστηρίξουν ένα εργαλείο υποστήριξης κρίσιμων αποφάσεων, μέσω της γλώσσας SPARQL.

Ο Mannes και Golbeck (2005) παρουσιάζουν ένα σύνολο επιχειρημάτων για τη χρήση του Σημαιολογικού Ιστού και κατ' επέκταση των οντολογιών στο Πεδίο της Τρομοκρατίας. Θεωρούν ότι μπορούν να χρησιμεύσουν ως πολύτιμα εργαλεία για τη συλλογή, οργάνωση και διάδοση πληροφοριών. Ο στόχος τους είναι να δημιουργήσουν, τόσο μια πλήρη οντολογία όσο και μία συνεπή βάση δεδομένων, τα οποία θα συνδέονται μεταξύ τους. Απώτερος στόχος τους είναι να υπάρξει ένα εργαλείο πρόληψης αλλά και κατανόησης ενός τρομοκρατικού χτυπήματος, το οποίο θα μπορέσει μελλοντικά να επεκταθεί για την έρευνα άλλων εγκληματικών ενεργειών. Όσον αφορά τον λόγο για τον οποίο επιλέγουν τη χρήση οντολογιών, επισημαίνουν ότι τους δίνεται η δυνατότητα να εκφράσουν περίπλοκες και μερικές φορές αντιφατικές πληροφορίες, με προσιτό και διαισθητικό τρόπο.

Τα επιχειρήματα που προτάσσουν είναι η αδυναμία των κλασικών βάσεων δεδομένων να ανακτούν μικρή σε όγκο αλλά περίπλοκη ή έξυπνη πληροφορία. Αυτή θα μπορούσε να είναι αναζήτηση ενός ατόμου με βάση το παρατσούκλι του ή με παραπλήσια γραφή όσο και η σύνδεση άσχετων φαινομενικά ατόμων με μια τρομοκρατική ενέργεια τα οποία όμως έχουν θέσεις κλειδιά. Καθώς ο σημασιολογικός ιστός προσδίδει νόημα στα δεδομένα είναι εξαιρετικά χρήσιμος για την παρακολούθηση ενεργειών μέσα στο χώρο και το χρόνο. Όπως σχέσεις μεταξύ ατόμων, πλάνα συνάντησης, συνεχής εκπαίδευση συγκεκριμένων προσώπων. Αυτό θα έδινε έναν επιπλέον ισχυρό μηχανισμό για την κατανόηση των τρομοκρατικών δικτύων (Terrorist Networks). Επίσης, μπορεί να φανεί χρήσιμο για την αποκωδικοποίηση των συνομιλιών και των κειμένων επικοινωνίας, οι οποίες συνηθίζεται να περιέχουν πληθώρα κωδικοποιημένων στοιχείων. Άλλη μία σημαντική δυνατότητα που δίνει είναι η ανάλυση οικονομικών συναλλαγών, όπως η μεταφορά χρημάτων μεταξύ συγκεκριμένων λογαριασμών αλλά και η αγορά υπηρεσιών ή υλικών. Για το τελευταίο και κυρίως για την αγορά όπλων, η ανάλυση μπορεί να χρησιμεύσει για την εξαγωγή συμπερασμάτων από τον τύπο των όπλων που έχουν αγοραστεί αλλά και από την ιστορία τους, δηλαδή από πού προήρθαν ή μέσω ποιων μεταφορικών μέσων- οδών διοχετεύθηκαν κτλ. Επιπλέον, μπορεί να χρησιμοποιηθεί για ανάλυση περίπλοκων ή ιδιαίτερων δεδομένων όπως είναι ο ήχος και η εικόνα, τα οποία πολλές φορές περιέχουν πιο ουσιαστική πληροφορία από ένα κείμενο. Επίσης μπορεί να χρησιμοποιηθεί για την αυθεντικοποίησή τους. Τέλος, αναφέρουν τη χρησιμότητα του Σημασιολογικού Ιστού για την διαλειτουργικότητα των εφαρμογών μεταξύ υπηρεσιών και την ευχέρεια που τους δίνεται για γρήγορη, κατανοητή και συμβατή μεταφορά πληροφοριών.

Μία πάρα πολύ ενδιαφέρουσα προσπάθεια για αναπαράσταση της γνώσης στο πεδίο της τρομοκρατίας ολοκληρώθηκε το 2008 με το πρόγραμμα Terrorist Knowledge Base (TKB) (Lenat & Deaton, 2008). Με την επίβλεψη των Αμερικάνικων αρχών και τις ήδη εξειδικευμένες γνώσεις της εταιρείας Cyc, σχεδιάστηκε και υλοποιήθηκε μια Βάση Γνώσης που παρέχει τη δυνατότητα στους αναλυτές να αναζητήσουν πληροφορία για τρομοκρατικές ομάδες, μέλη ή και γεγονότα. Βέβαια, από τα πιο σημαντικά της χαρακτηριστικά είναι ότι η TKB εφαρμόζει εξελιγμένη συλλογιστική χρησιμοποιώντας το πεδίο γνώσης, υποστηρίζεται από εξωτερικά αποθηκευμένα δεδομένα, παρουσιάζει την υπάρχουσα γνώση σύμφωνα με το τι ο αναλυτής θεωρεί ενδιαφέρον και έχει τη δυνατότητα να απαντάει σε ερωτήσεις δίνοντας επεξηγήσεις.

Από τις ανάγκες που προσπαθούν να καλύψουν οι συνδυασμοί της τρομοκρατίας με την DM και της τρομοκρατίας με τις Οντολογίες, εύκολα μπορεί να αντιληφθεί κανείς τις ιδιαιτερότητες και τα θετικά από μια συνένωση μεταξύ των τριών τομέων. Πολλοί στόχοι είναι κοινοί και επικαλύπτονται όπως η δυνατότητα να προσφέρουν ευκολίες στον ειδικό και να εξάγουν ενδιαφέρουσα γνώση. Από την άλλη υπάρχουν πολλές λειτουργίες που φαίνεται να αλληλοσυμπληρώνονται, όπως η δημιουργία ισχυρότερων διαδικασιών ελέγχου των δεδομένων αλλά και των αποτελεσμάτων των τεχνικών DM, ο ανασχεδιασμός στο τι πραγματικά είναι ενδιαφέρουσα γνώση, δημιουργία αξιόπιστων συλλογών δεδομένων εκπαίδευσης, ανακάλυψη νέων εννοιών που θα έπρεπε να εισαχθούν στις Οντολογίες και είναι οδηγούμενες από δεδομένα (Data Driven).

Η σκοπός της ενότητας είναι αρχικά να παρουσιαστούν επιβεβαιωμένες ή μη τάσεις πάνω στα δεδομένα Τρομοκρατικών Βάσεων Δεδομένων που έχουν προταθεί από αναλυτές του πεδίου. Αυτές οι τάσεις θα περικλείουν την περίπτωση χρήσης και θα είναι το βασικό σενάριο για το σημασιολογικό έλεγχο της εφαρμογής που έχει υλοποιηθεί. Έπειτα θα παρουσιαστούν τα δεδομένα εκπαίδευσης, τα οποία είναι ένα υποσύνολο της Global Terrorism Database (GTD) (2011), μία από τις πιο γνωστές και ενημερωμένες βάσεις πάνω στην τρομοκρατία. Θα ακολουθήσει η παρουσίαση της Οντολογίας, που αποτελείται από την συνένωση των εννοιών της Umbel, υποσυνόλου της OpenCyc, οι οποίες έννοιες αντιστοιχίζονται με τα δεδομένα της περίπτωσης χρήσης και τις μεταβλητές των Δεδομένων Εκπαίδευσης. Η επόμενη υποενότητα παρουσιάζει τις σημασιολογικές αποστάσεις, που δημιουργούνται μεταξύ των εννοιών μέσω της κλάσης TaxonomyMetrics, και ορίζει την έννοια του σημασιολογικού πλαισίου. Τέλος, παρουσιάζονται κάποια γενικά στατιστικά στοιχεία των δεδομένων της βάσης και μετά εξετάζονται οι υποθέσεις της μελέτης περίπτωσης. Μέσα από την απόδειξη ή μη των υποθέσεων θα αποδειχθεί η σημασιολογική χρησιμότητα αλλά και εγκυρότητα της επέκτασης που έγινε στον αλγόριθμο Arriori, με την ενσωμάτωση των Οντολογιών.

14. Σενάριο χρήσης

Το σενάριο χρήσης βασίζεται στην δημοσίευση του Chasdi (1997) που προσπαθεί να ξεφύγει από την απλή έρευνα του τι έγινε και πού, επιζητώντας μια πιο αυστηρή επιστημονική προσέγγιση για την κατανόηση των τρομοκρατικών ενεργειών. Ο απώτερος στόχος του είναι η απόδειξη ότι οι τρομοκρατικές οργανώσεις λαμβάνουν αποφάσεις με ορθολογικό τρόπο και δε βασίζονται σε τυχαία ή ευκαιριακά γεγονότα. Στην προσέγγισή του κάνει χρήση εμπειρικών δεδομένων για να εξετάσει αν οι μεταβλητές, όπως η επιλογή ενός στόχου ή οι τρομοκρατικές ενέργειες συνδέονται με διάφορους τύπους τρομοκρατικών οργανώσεων. Ουσιαστικά εξετάζει την ύπαρξη στατιστικής σύνδεσης μεταξύ μεταβλητών της τρομοκρατίας που περιλαμβάνονται σε συγκεκριμένες θεωρητικές υποθέσεις.

Συνοπτικά, οι μεταβλητές που εξετάζονται είναι η πολιτική ιδεολογία της εκάστοτε τρομοκρατικής οργάνωσης, η γεωγραφική τοποθεσία του γεγονότος, ο τύπος του στόχου, ο αριθμός των θυμάτων, ο αριθμός των τραυματιών, οι υλικές ζημιές και η αφορμή της διενέργειας μιας τρομοκρατικής επίθεσης από ένα πολιτικό γεγονός. Το πλαίσιο της ανάλυσης περιλαμβάνει μια συζήτηση σχετικά με τις γενικές τάσεις στη συμπεριφορά και την επίδραση της εκάστοτε μεταβλητής.

Οι υποθέσεις που εξετάζει είναι τάσεις μεταξύ μεταβλητών που έχουν παρατηρηθεί από αναλυτές των τρομοκρατικών γεγονότων. Τέτοιες υποθέσεις ή αρχές είναι πολύ σημαντικές καθώς επιτρέπουν την διάκριση των τρομοκρατικών ομάδων. Κατ' επέκταση, μεγάλο ενδιαφέρον παρουσιάζει η αναζήτηση νέων μεταβλητών για την ταξινόμηση τρομοκρατικών ομάδων, εκτός του παραδοσιακού τρόπου, ανάλογα με την τοποθεσία του συμβάντος και το είδος της τρομοκρατικής δραστηριότητας. Οι υπό εξέταση υποθέσεις είναι:

- Υπόθεση 1: Ανάλογα την πολιτική ιδεολογία της τρομοκρατικής οργάνωσης υπάρχει μια τάση προς την επιλογή συγκεκριμένου στόχου. Εν μέρει την επιβεβαιώνει.
- Υπόθεση 2: Ανάλογα την πολιτική ιδεολογία της τρομοκρατικής οργάνωσης υπάρχει μια τάση προς τον αριθμό των νεκρών που αποφέρει μια τρομοκρατική ενέργεια. Την οποία επιβεβαιώνει.
- Υπόθεση 3: Ανάλογα την πολιτική ιδεολογία της τρομοκρατικής οργάνωσης υπάρχει μια τάση προς τον αριθμό των τραυματιών που αποφέρει μια τρομοκρατική ενέργεια. Την οποία επιβεβαιώνει.
- Υπόθεση 4: Ανάλογα την πολιτική ιδεολογία της τρομοκρατικής οργάνωσης υπάρχει μια τάση προς τις υλικές ζημιές που αποφέρει μια τρομοκρατική ενέργεια. Την οποία επιβεβαιώνει.
- Υπόθεση 5: Ανάλογα την περιοχή που διενεργείται μια τρομοκρατική ενέργεια υπάρχει μια τάση προς την επιλογή συγκεκριμένου είδους στόχου. Την οποία επιβεβαιώνει.
- Υπόθεση 6: Υπάρχει συσχέτιση μεταξύ των πολιτικών γεγονότων και την επιλογή του είδους του στόχου. Την οποία δεν επιβεβαιώνει. Η μεταβλητή «πολιτικό γεγονός» μπορεί να είναι στρατιωτικές επιθέσεις, όπως ο Πόλεμος των Έξι Ημερών, οι θρησκευτικές εκδηλώσεις, όπως το Ραμαζάνι, οι επισκέψεις από μεγάλες πολιτικές προσωπικότητες, οι διπλωματικές πρωτοβουλίες, οι κοσμικές εκδηλώσεις, οι αντιτρομοκρατικές δραστηριότητες.

Στη συγκεκριμένη δημοσίευση οι υποθέσεις που αναλύονται είναι πιο ειδικευμένες και αναφέρονταν σε συγκεκριμένους τύπους τρομοκρατικών ομάδων που έδρασαν κατά τα έτη 1968 έως 1993 στην Μέση Ανατολή. Στο παρόν όμως σενάριο έχει γίνει μια γενίκευση των υποθέσεων για να συμβαδίσει με τους νέους στόχους.

Σκοπός της παρούσας περίπτωσης χρήσης είναι να εξετάσει τις υποθέσεις πάνω σε μία περί τρομοκρατίας Βάση Δεδομένων και να προσπαθήσει να εξάγει τις προηγούμενες τάσεις χρησιμοποιώντας ως βοηθητικό εργαλείο μια Οντολογία που σχετίζεται με τον παρών τομέα. Στην περίπτωση εξαγωγής των ίδιων συμπερασμάτων, ενδιαφέρον θα παρουσιαστεί για την εύρεση νέων τάσεων ή μεταβλητών μέσα από το εργαλείο που έχει υλοποιηθεί. Στις επόμενες δύο ενότητες θα αναπτυχθούν τα κατάλληλα δεδομένα εκπαίδευσης και η Οντολογία πεδίου που θα χρησιμοποιηθούν.

Γίνεται αντιληπτό ότι ο στόχος της περίπτωσης χρήσης δεν είναι η παραγωγή στατιστικών αποτελεσμάτων που θα απαντούν σε ερωτήσεις τύπου, ποιο είναι το ποσοστό των τρομοκρατικών ενεργειών από συγκεκριμένους τύπους οργανώσεων ή το ποσοστό των θανάτων ή τραυματιών στο σύνολο των επιθέσεων. Αντίθετα, επιζητούνται οι τάσεις και οι

συσχετίσεις που είναι κρυμμένες μεταξύ των εννοιών και των δεδομένων μέσα στο πεδίο γνώσης της τρομοκρατίας.

15. Δεδομένα εκπαίδευσης

Τα δεδομένα εκπαίδευσης είναι ένα υποσύνολο της βάσης δεδομένων GTD (2011). Η Global Terrorism Database είναι μια βάση δεδομένων ελεύθερης χρήσης που συμπεριλαμβάνει πληροφορίες σχετικά με τις τρομοκρατικές ενέργειες που έχουν διενεργηθεί σε όλο τον κόσμο από το 1970 έως το 2008. Η διαφορά της με αντίστοιχες βάσεις δεδομένων έγκειται στο ότι συμπεριλαμβάνει στοιχεία τόσο από εγχώρια όσο και διακρατικά και διεθνή τρομοκρατικά συμβάντα που σημειώθηκαν κατά τη διάρκεια αυτής της περιόδου και τώρα περιλαμβάνει περισσότερα από 87.000 περιπτώσεις. Για κάθε περιστατικό, η GTD καταγράφει πληροφορίες για την ημερομηνία, τον τόπο του συμβάντος, τα όπλα που χρησιμοποιήθηκαν, τον τύπο του στόχου, τον αριθμό των θυμάτων και όταν είναι αναγνωρίσιμα, τα στοιχεία της ομάδας ή των ατόμων που είναι υπεύθυνα.

Γενικά περιλαμβάνει από 45 έως 120 μεταβλητές για την περιγραφή των τρομοκρατικών επιθέσεων, οι οποίες για να εισαχθούν στη βάση θα πρέπει να καλύπτουν συγκεκριμένα κριτήρια. Οι εγγραφές ελέγχονται από μία συμβουλευτική επιτροπή 12 εμπειρογνομόνων πάνω στην έρευνα της τρομοκρατίας. Η πηγή ενημέρωσης για τις τρομοκρατικές επιθέσεις είναι ένα σύνολο συγκεκριμένων ειδησεογραφικών πρακτορείων που έχουν μια φήμη ελάχιστης συστηματικής προκατάληψης και ιδίως εκείνα που δεν είναι συνδεδεμένα με κάποια ομάδα συμφερόντων ή με κυβερνήσεις. Παραδείγματα, αξιόπιστων πηγών ειδήσεων περιλαμβάνουν το Associated Press, Reuters, Agence-France Presse, Deutsche Presse Agentur, κτλ. Συνολικά, και μόνο για την περίοδο 1998 έως 2008, πάνω από 3,5 εκατομμύρια άρθρα ειδήσεων και 25 χιλιάδες πηγές ειδήσεων εξετάστηκαν για τη συλλογή δεδομένων.

Το υποσύνολο των δεδομένων που θα χρησιμοποιηθεί αναφέρεται σε 21 δημοφιλείς τρομοκρατικές οργανώσεις που έδρασαν την περίοδο 1991 έως 2008. Οι οργανώσεις δεν προέρχονται από μια συγκεκριμένη περιοχή ενώ είναι ομοιόμορφα ομαδοποιημένες σε τρεις γενικές ιδεολογικές κατηγορίες. Εκτός των μεταβλητών της υπάρχουσας ΒΔ εισήχθηκαν επιπλέον τρεις μεταβλητές που καθορίζουν την πολιτική ιδεολογία των οργανώσεων, την περιοχή και χώρα «καταγωγής» τους. Οι τιμές που παίρνουν οι νέες μεταβλητές συμβαδίζουν με τις πληροφορίες που δίνονται από την GTD για το προφίλ των τρομοκρατικών οργανώσεων. Στον ακόλουθο πίνακα εμφανίζονται οι οργανώσεις με την πολιτική ιδεολογία που συνδέονται.

Πίνακας 15-1. Πολιτική ιδεολογία και οργανώσεις

Ιδεολογία	Ονομασία Οργάνωσης	Χώρα Καταγωγής	Περιοχή Καταγ.
1. Σοσιαλιστές/ Κομμουνιστές	Red Army Faction (RAF)	Γερμανία	Δυτική Ευρώπη
	Red Brigades	Ιταλία	
	Revolutionary People's Struggle (ELA)	Ελλάδα	
	Popular Front for the Liberation of Palestine (PFLP)	Παλαιστίνη- (Δυτική Όχθη και Λωρίδα της Γάζας)	Μέση Ανατολή και Βόρεια Αφρική
	Shining Path (SL)	Περού	Νότια Αμερική
2. Εθνικό Απελευθερωτικό	Basque Fatherland and Freedom (ETA)	Ισπανία- (Βάσκοι)	Δυτική Ευρώπη
	Irish Republican Army (IRA)	Ιρλανδία	Μέση Ανατολή και Βόρεια Αφρική
	Palestine Liberation Organization (PLO)	Παλαιστίνη- (Δυτική Όχθη και Λωρίδα της Γάζας)	
	Kurdistan Workers' Party (PKK)	Κούρδοι	
	Liberation Tigers of Tamil Eelam (LTTE)	Σρι Λάνκα	Νότια Ασία

Ιδεολογία	Ονομασία Οργάνωσης	Χώρα Καταγωγής	Περιοχή Καταγ.
3. Θρησκευτικό-Πολιτικές	Aum Shinri Kyo	Ιαπωνία	Ανατολική Ασία
	Ku Klux Klan	Ηνωμένες Πολιτείες	Βόρεια Αμερική
	Al-Qa`ida in Iraq	Ιράκ	Μέση Ανατολή και Βόρεια Αφρική
	Al-Qa`ida in Levant and Egypt	Αίγυπτος	
	Al-Qa`ida in the Arabian Peninsula (AQAP)	Σαουδική Αραβία	
	Al-Qa`ida in the Lands of the Islamic Maghreb (AQLIM)	Αλγερία	
	Hamas (Islamic Resistance Movement)	Παλαιστίνη- (Δυτική Όχθη και Λωρίδα της Γάζας)	
	Hizballah	Λίβανο	
	Islamic Jihad (Ideological Grouping)	Παλαιστίνη- (Δυτική Όχθη και Λωρίδα της Γάζας)	
	Al-Qa`ida	Διεθνή	Νότια Ασία
	Abu Sayyaf Group (ASG)	Φιλιππίνες	Νοτιοανατολική Ασία

Στον πίνακα που ακολουθεί παρουσιάζονται οι μεταβλητές του GTD που θα αποτελούν τα δεδομένα εκπαίδευσης. Στην πρώτη στήλη καταγράφεται η ονομασία της μεταβλητής, στη δεύτερη στήλη ο τύπος των τιμών που παίρνει από τη μητρική βάση. Υπάρχουν τρεις κυρίως τύποι *αριθμητική*, *κατηγορηματική* και *κείμενο*. Η τρίτη στήλη αναφέρεται στις τιμές που παίρνει η εκάστοτε μεταβλητή. Στην επόμενη στήλη περιγράφεται λεκτικά τι αντιπροσωπεύει η κάθε μεταβλητή. Για τις μεταβλητές *country*, *region*, *targettype1* και *natly1* ο λεκτικός τύπος των τιμών παρουσιάζεται σε επόμενη ενότητα. Η στήλη *Νέο Πεδίο Τιμών* αναφέρεται στο πεδίο τιμών που καθορίζεται να έχουν οι μεταβλητές στα δεδομένα εκπαίδευσης. Τέλος η στήλη Σχόλια δίνει κάποιες περαιτέρω διευκρινίσεις.

Πίνακας 15-2. Χρήσιμες μεταβλητές από GTD

Μεταβλητή	Τύπος	Τιμές	Επεξήγηση	Πεδίο Τιμών	Σχόλια
<i>imonth</i>	Αριθμ-ός	Μήνας ή 0	Μήνας συμβάντος	0..12	
<i>country</i>	Κατηγορηματική	4 έως 1003	Προσδιορίζει τη χώρα ή τη θέση όπου το γεγονός διαπράχθηκε.	4..1003	
<i>region</i>	Κατηγορηματική	1 έως 13	Προσδιορίζει την περιοχή στην οποία το γεγονός διαπράχθηκε. Οι χώρες διαιρούνται σε 13 περιφέρειες.	1..13	
<i>attacktype1</i>	Κατηγορηματική	1 έως 9	Το πεδίο καταγράφει την γενική μέθοδο επίθεσης και συχνά απεικονίζει την ευρεία κατηγορία τακτικής που χρησιμοποιήθηκε. 1 = Assassination 2 = Armed Assault 3 = Bombing/Explosion 4 = Hijacking 5 = Hostage Taking (Barricade Incident) 6 = Hostage Taking (Kidnapping) 7 = Facility/Infrastructure Attack 8 = Unarmed Assault 9 = Unknown	1..9	
<i>targettype1</i>	Κατηγορηματική	1 έως 22	Το συγκεκριμένο πεδίο αντιπροσωπεύει τον γενικό τύπο του στόχου, σύμφωνα με τις 22 εφιστάμενες κατηγορίες	1..22	
<i>natly1</i>	Κατηγορηματική	4 έως 1003	Αντιπροσωπεύει την υπηκοότητα του στόχου, και δεν είναι απαραίτητως η ίδια με τη χώρα στην οποία το γεγονός εμφανίστηκε, αν και στις περισσότερες περιπτώσεις είναι. Για μια αεροπορική πειρατεία, καταγράφεται η υπηκοότητα του αεροπλάνου και όχι αυτή των επιβατών. Οι κώδικες υπηκοότητας είναι οι ίδιοι με του πεδίου <i>country</i>	4..1003, -9	
<i>gname</i>	Κείμενο	Όνομ. οργ.	Στο συγκεκριμένο πεδίο γίνεται αναφορά στο όνομα της ομάδας	1..1508, -9	

Μεταβλητή	Τύπος	Τιμές	Επεξήγηση	Πεδίο Τιμών	Σχόλια
<i>weaptype1</i>	Κατηγορηματική	1 έως 13	Στο πεδίο γίνεται αναφορά στον τύπο του όπλου ανάλογα με 13 κατηγορίες που έχουν οριστεί.	1..13	
<i>nkill</i>	Αριθμός	Θετ. Πραγμ. ή -99	Αποθηκεύεται ο συνολικός αριθμός επιβεβαιωμένων θανάτων κατά το γεγονός. Ο αριθμός περιλαμβάνει τα θύματα αλλά και τους επιτιθεμένους που πέθαναν ως άμεσο αποτέλεσμα του γεγονότος. Σε περίπτωση που ο αριθμός είναι άγνωστος εισάγεται η τιμή -99.	I, WP, DP, UNKNOW N	I= Κανένας, WP= Λίγοι- [1, 12], DP= Πολύ [13, 1381]
<i>nwound</i>	Αριθμός	Θετ. Πραγμ. ή -99	Επιβεβαιωμένος αριθμός μη θανατηφόρων τραυματισμών	I, WP, DP, UNKNOW N	I= Κανένας, WP= Λίγοι- [1, 55], DP= Πολύ [55, 5500]
<i>propextent</i>	Κατηγορηματική	1 ή 2 ή 3 ή 4 ή -9	Αν το <i>property</i> = 1 τότε μια από τις κατηγορίες περιγράφουν την έκταση της ζημιάς. 1 = Catastrophic (likely > \$1 billion) 2 = Major (likely > \$1 million but < \$1 billion) 3 = Minor (likely < \$1 million) 4 = Unknown	0, 1, 2, 3, 4	'0' όταν το <i>property</i> =0 δηλαδή όταν θεωρείται ότι δεν υπάρχει ζημιά.

Στο I Παράρτημα παρουσιάζονται τα *scripts* για την δημιουργία των δεδομένων εκπαίδευσης. Όπως γίνεται αντιληπτό από τον παραπάνω πίνακα τα ακατέργαστα δεδομένα της GTD πρέπει να επεξεργαστούν κατάλληλα ώστε να ακολουθούν τον επιτρεπτό τύπο δεδομένων εισαγωγής στην εφαρμογή (.arff) αλλά και να μπορούν να εξαγουν ενδιαφέρουσα πληροφορία.

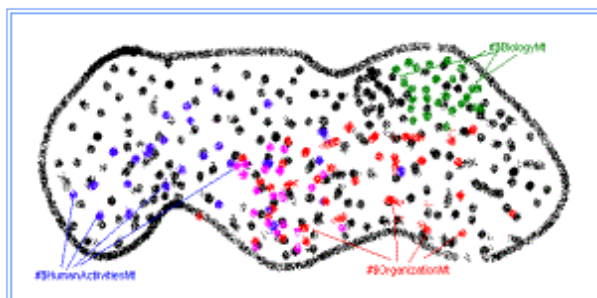
Επίσης, στα πρωταρχικά δεδομένα της GTD, οπότε και στα δεδομένα εκπαίδευσης δεν υπάρχει μεταβλητή που να ενημερώνεται με τα πολιτικά γεγονότα που συσχετίζονται με μία τρομοκρατική ενέργεια. Αυτό καθιστά αδύνατο τον έλεγχο της έκτης υπόθεσης, που θέλει τα πολιτικά γεγονότα να συσχετίζονται με το είδος του στόχου που θα επιλεγεί.

16. Οντολογία πεδίου

Για την επιλογή της Οντολογίας Πεδίου έγινε μια γενικότερη έρευνα για το ποια θα ήταν η κατάλληλη, η οποία θα έδινε τη δυνατότητα για πλήρη αντιστοίχιση των μεταβλητών των δεδομένων εκπαίδευσης με τις έννοιες μιας OWL Οντολογίας. Οι οντολογίες που εξετάστηκαν ήταν η COSMO, SUMO, AKTiveSAOntology, Mindswap, OpenCyc, Umbel. Από αυτές η COSMO κατηγοριοποιείται στις οντολογίες υψηλού επιπέδου με αποτέλεσμα να περιέχει κυρίως βασικές έννοιες, οι οποίες είναι αναγκαίες και ικανές για την κατασκευή πιο συγκεκριμένων εννοιών, όπως είναι η τρομοκρατία. Σε αυτή, η κλάση *TerroristAgent* είναι η μόνη που συσχετίζεται με το πεδίο της τρομοκρατίας, ενώ ο τύπος *Terror* είναι μια πολύ γενική έννοια του συναισθήματος του τρόμου και όχι μόνο το συγκεκριμένο γεγονός που προκαλούν οι τρομοκράτες. Επίσης, η SUMO ανήκει στις οντολογίες υψηλού επιπέδου αλλά παρέχει την Mid-Level Ontology, οντολογία που όπως προδιαθέτει η ονομασία της είναι μεσαίου επιπέδου, δηλαδή πιο αναλυτική από την SUMO και περιέχει όλες τις προσδιοριστικές έννοιες της τρομοκρατίας. Παρόλα αυτά είναι γραμμένη στην γλώσσα KIF, ενώ το σύστημα που υλοποιήθηκε αναγνωρίζει μόνο OWL οντολογίες. Βέβαια, υπάρχει μια OWL μεταφρασμένη οντολογία, η οποία εκτός ότι αντιστοιχεί σε παλιότερη έκδοση της Mid-Level Ontology δεν είναι τόσο ισχυρή στην διαδικασία του συμπερασμού όσο η πρωτότυπη. Η Mid-Level Ontology δεν περιλαμβάνει τις έννοιες του πεδίου της τρομοκρατίας. Οι AKTiveSAOntology, Mindswap είναι οντολογίες ειδικευμένες στο πεδίο της τρομοκρατίας, όμως είναι εννοιολογικά συμπυκνωμένες και δεν μπορεί να γίνει πλήρης αντιστοίχιση όλων των μεταβλητών. Οι OpenCyc, Umbel ήταν αυτές που κάλυπταν όλες τις απαιτήσεις και γι' αυτό εξετάζονται ξεχωριστά.

Το OpenCyc αποτελεί την ελεύθερη έκδοση της τεχνολογίας Cyc και φιλοδοξεί να αποτελέσει τη μεγαλύτερη και πιο πλήρη βάση γενικής γνώσης, υποστηριζόμενη από μία ισχυρή μηχανή συμπερασμού. Το πιο δυνατό της σημείο είναι η χρήση Μικροθεωριών (Mt), οι οποίες διαμερίζουν τη βάση γνώσης. Αν υποθέσουμε ότι η Cyc Βάση Γνώσης είναι μια μεγάλη συλλογή ισχυρισμών, τότε οι Μικροθεωρίες είναι ένα υποσύνολο των προηγούμενων ισχυρισμών, οι οποίες την ταυτοποιούν. Οι ισχυρισμοί επομένως ομαδοποιούνται σε μια Μικροθεωρία με βάση τις κοινές υποθέσεις, κοινά θέματα, κοινές πηγές ή άλλα

χαρακτηριστικά. Με αυτό τον τρόπο η OpenCyc καταφέρνει να δομήσει κατά κάποιο τρόπο τη γνώση και να την εξειδικεύσει στα ανάλογα πεδία εφαρμογής. Παραδείγματα τέτοιων Μικροθεωριών είναι οι `#$OrganizationMt`, `#$HumanActivitiesMt`.



Εικόνα 16-1. Μικροθεωρίες στην OpenCyc

Πηγή: CYCORP, http://www.cyc.com/doc/tut/ppoint/Microtheories_files/v3_document.htm

Το αρνητικό με την OpenCyc είναι ότι το αρχείο που διατίθεται σε OWL μορφή περιέχει χιλιάδες έννοιες με αποτέλεσμα να είναι δύσκολη η διαχείρισή του και η απομόνωση των χρήσιμων στοιχείων. Ενώ το εργαλείο OpenCyc KB Browser που δίνει στον χρήστη τη δυνατότητα πλοήγησης μεταξύ των εννοιών και εξαγωγής συμπερασμάτων, δε διαθέτει κάποια λειτουργία εξαγωγής τους σε οποιαδήποτε μορφή. Για να αποφευχθεί αυτό το πρόβλημα γίνεται χρήση της Umbel (2011), ενός Upper Mapping and Binding Exchange Layer, το οποίο προσδιορίζεται κυρίως για να βοηθήσει την σημασιολογική διαλειτουργικότητα στο διαδίκτυο. Ουσιαστικά είναι ένα αυστηρό υποσύνολο των εννοιών και των σχέσεων της OpenCyc, από το οποίο μπορεί να αποσπασθεί η περιγραφή της εκάστοτε έννοιας σε OWL μορφή. Η διαδικασία που πραγματοποιήθηκε για την δημιουργία του πεδίου γνώσης της τρομοκρατίας ήταν να βρεθούν όλες οι απαραίτητες περιγραφές εννοιών σε OWL μορφή από το Umbel και να «συραφτούν» από το εργαλείο Protégé.

Συνολικά ανακτήθηκαν και χρησιμοποιήθηκαν για τη δημιουργία της οντολογίας 112 έννοιες. Το κριτήριο για την επιλογή των εννοιών ήταν να βρεθούν οι ελάχιστες δυνατές έννοιες που να μπορέσουν να προσδιορίσουν επαρκώς τις έννοιες με τις οποίες θα γίνονταν αντιστοίχιση οι μεταβλητές των δεδομένων εκπαίδευσης. Ουσιαστικά, η τεχνική που ακολουθήθηκε ήταν να βρίσκεται η έννοια με την οποία μια μεταβλητή θα αντιστοιχούσε και έπειτα να γινόταν αναζήτηση των γονέων της μέχρι να βρεθεί η κλάση Thing, που είναι και η γενικότερη τάξη της OpenCyc. Στον πίνακα που ακολουθεί παρουσιάζονται μερικές από τις έννοιες που αποτελούν την Οντολογία. Από τις 112 έννοιες μόνο σε δύο δεν ήταν δυνατή η ανάκτηση του ορισμού τους, αυτές είναι οι http://umbel.org/umbel/rc/Artifact_Agentive και <http://umbel.org/umbel/rc/Dying>. Από αυτές, η δεύτερη έννοια για την οποία δεν έχουν οριστεί σωστά οι απόγονοί της. Αποτέλεσμα αυτού είναι να δημιουργηθούν κάποια αναμενόμενα προβλήματα στα αποτελέσματα των σημασιολογικών αποστάσεων μέσα στην οντολογία, τα οποία θα αναλυθούν αργότερα.

Πίνακας 16-1. Υποσύνολο εννοιών που αποτελούν την Οντολογία Τρομοκρατίας

http://umbel.org/umbel/rc/Thing	http://umbel.org/umbel#Activities
http://umbel.org/umbel#SuperType	http://umbel.org/umbel/rc/Politics
http://umbel.org/umbel#TopicsCategories	http://umbel.org/umbel/rc/PoliticalEvent
http://umbel.org/umbel#AbstractLevel	http://umbel.org/umbel#Products
http://umbel.org/umbel/rc/ObjectType	http://umbel.org/umbel/rc/Weapon
http://umbel.org/umbel/rc/Place	http://umbel.org/umbel/rc/ParamilitaryOrganization
http://umbel.org/umbel/rc/GeographicalPlace	http://umbel.org/umbel/rc/PoliticalOrganization
http://umbel.org/umbel/rc/GeographicalRegion	http://umbel.org/umbel/rc/TerroristAgent
http://umbel.org/umbel/rc/TemporalThing	http://umbel.org/umbel/rc/TerroristGroup
http://umbel.org/umbel/rc/TemporallyExtendedThing	http://umbel.org/umbel/rc/AttackOnTangible
http://umbel.org/umbel/rc/TemporalObjectType	http://umbel.org/umbel/rc/TerroristAct
http://umbel.org/umbel/rc/TemporallyContinuousThing	http://umbel.org/umbel/rc/TerroristAttack
http://umbel.org/umbel/rc/CalendarMonth	http://umbel.org/umbel/rc/PopulatedPlace

http://umbel.org/umbel#Events	http://umbel.org/umbel#Geopolitical
http://umbel.org/umbel/rc/Action	http://umbel.org/umbel/rc/Country
http://umbel.org/umbel/rc/ActionOnObject	http://umbel.org/umbel/rc/Dying
http://umbel.org/umbel/rc/AttackOnObject	http://umbel.org/umbel/rc/DeathOfTheVictim

17. Συσχέτιση δεδομένων εκπαίδευσης & εννοιών Οντολογίας

Με βάση την περιγραφή των εννοιών που ορίζεται από το OpenCyc επιλέχθηκε η ακόλουθη αντιστοιχία. Στην πρώτη στήλη περιγράφεται λεκτικά η ονομασία της μεταβλητής, η επόμενη ενημερώνεται με την ονομασία της μεταβλητής, η τρίτη στήλη περιέχει της ονομασία της έννοιας στην Umbel και η τελευταία στήλη περιέχει την ονομασία της έννοιας που αντιστοιχίστηκε στην OpenCyc.

Πίνακας 17-1. Αντιστοιχία μεταβλητών δεδομένων εκπαίδευσης & εννοιών Οντολογίας

Περιγραφή	Μεταβλητή	Ονομασία Umbel	Ονομασία OpenCyc
Μήνας	imonth	http://umbel.org/umbel/rc/CalendarMonth	Mx4rvJAKZwpEbGdrcN5Y29ycA
Χώρα (γεγονότος, τρομοκρατών, στόχου)	country, terroristcountry natlty1	http://umbel.org/umbel/rc/Country	Mx4rvVileZwpEbGdrcN5Y29ycA
Περιοχή (γεγονότος, τρομοκρατών)	region, terroristregion	http://umbel.org/umbel/rc/GeographicalRegion	Mx4rvViACZwpEbGdrcN5Y29ycA
Τύπος επίθεσης	attcktype1	http://umbel.org/umbel/rc/TerroristAttack	Mx4rwFYGEpwpEbGdrcN5Y29ycA
Είδος Στόχου	targtype1	http://umbel.org/umbel/rc/AttackOnObject	Mx4rvbfzAZwpEbGdrcN5Y29ycA
Οργάνωση	gname	http://umbel.org/umbel/rc/TerroristGroup	Mx4rv7MhJwpEbGdrcN5Y29ycA
Όπλα	weaptype	http://umbel.org/umbel/sc/Weapon	Mx4rvVj8V5wpEbGdrcN5Y29ycA
Θύμα	nkill	http://umbel.org/umbel/rc/DeathOfTheVictim	Mx4rEIFC4JYjEdaAAACgycbQqA
Τραυματίες	nwound	http://umbel.org/umbel/rc/Injury_Object	Mx4rwP1IHpwpEbGdrcN5Y29ycA
Ζημιές	propextent	http://umbel.org/umbel/rc/AttackOnObject	Mx4rvVifsJwpEbGdrcN5Y29ycA
Πολιτική Ιδεολογία	ideology	http://umbel.org/umbel/sc/PoliticalBeliefSystem	Mx4rvVjUHpwpEbGdrcN5Y29ycA
Πολιτικά Γεγονότα	-	http://umbel.org/umbel/sc/PoliticalEvent	Mx4rwxOT5wpEbGdrcN5Y29ycA

Σύμφωνα με τον παραπάνω πίνακα δημιουργείται το αρχείο αντιστοιχίσης (.map) των μεταβλητών με τις έννοιες της οντολογίας που θα διαβάζει η εφαρμογή.

Πίνακας 17-2. Mapping File Οντολογία τρομοκρατίας

imonth	http://umbel.org/umbel/rc/CalendarMonth
country	http://umbel.org/umbel/rc/Country
region	http://umbel.org/umbel/rc/GeographicalRegion
attacktype1	http://umbel.org/umbel/rc/TerroristAttack
targtype1	http://umbel.org/umbel/rc/AttackOnObject
natlty1	http://umbel.org/umbel/rc/Country
gname	http://umbel.org/umbel/rc/TerroristGroup
weaptype1	http://umbel.org/umbel/rc/Weapon
nkill	http://umbel.org/umbel/rc/DeathOfTheVictim
nwound	http://umbel.org/umbel/rc/Injury_Object
propextent	http://umbel.org/umbel/rc/AttackOnObject
terroristcountry	http://umbel.org/umbel/rc/Country
terroristregion	http://umbel.org/umbel/rc/GeographicalRegion
ideology	http://umbel.org/umbel/rc/PoliticalBeliefSystem

18. Σημασιολογικές αποστάσεις

Σε αυτή την ενότητα θα εξεταστούν οι σηματολογικές αποστάσεις που προκύπτουν μεταξύ των εννοιών της οντολογίας με βάση την μετρική TaxonomyMetrics, η οποία θα χρησιμοποιηθεί και για την εφαρμογή της περίπτωσης χρήσης. Υπενθυμίζεται ότι η αναφερόμενη μετρική είναι αμφίδρομη και ότι βασίζεται στους απογόνους των εξεταζόμενων εννοιών για να υπολογίσει την ομοιότητά τους. Ο όρος αμφίδρομος υποδηλώνει ότι δε δίνει διαφορετικά αποτελέσματα για τις αποστάσεις A->B και B->A. Όσον αφορά το βαθμό ομοιότητας, αυξάνεται καθώς αυξάνεται και ο αριθμός των κοινών απογόνων μεταξύ των εννοιών. Το πεδίο τιμών του κυμαίνεται από 0 έως 1, όπου όσο πιο κοντά στο μηδέν είναι η τιμή απόστασης τόσο πιο ανόμοιες είναι οι δύο έννοιες. Στον πίνακα που ακολουθεί εμφανίζονται οι σηματολογικές αποστάσεις των ενδιαφερόμενων εννοιών, στρογγυλοποιημένες στο τρίτο δεκαδικό ψηφίο.

Πίνακας 18-1. Σηματολογικές αποστάσεις εννοιών Οντολογίας

	A	B	C	D	E	F	G	H	I	J	K
A	1.0	0.294	0.163	0.157	0.523	0.421	0.15	0.214	0.103	0.146	0.176
B		1.0	0.162	0.2	0.25	0.333	0.175	0.193	0.125	0.178	0.3
C			1.0	0.428	0.206	0.224	0.425	0.357	0.055	0.325	0.21
D				1.0	0.219	0.29	0.175	0.541	0.125	0.65	0.368
E					1.0	0.714	0.116	0.239	0.085	0.204	0.21
F						1.0	0.14	0.27	0.12	0.264	0.285
G							1.0	0.173	0.06	0.162	0.162
H								1.0	0.086	0.481	0.32
I									1.0	0.105	0.166
J										1.0	0.318
K											1.0

A	PoliticalEvent
B	CalendarMonth
C	Country
D	GeographicalRegion

E	TerroristAttack
F	AttackOnObject
G	TerroristGroup

H	Weapon
I	DeathOfTheVictim
J	Injury_Object
K	PoliticalBeliefSystem

Από τον πίνακα παρατηρείται ότι στις θέσεις χαμηλής ομοιότητας βρίσκεται η έννοια DeathOfTheVictim, γεγονός το οποίο οφείλεται στη μη δυνατότητα ορισμού σωστά των απογόνων του. Αυτό δείχνει για ακόμα μια φορά πόσο σημαντική είναι η διαδικασία δημιουργίας της οντολογίας και πόσο εύκολο είναι να παρουσιαστούν σηματολογικά λάθη από μικρές παραβλέψεις.

Άλλη μία παρατήρηση που μπορεί να γίνει είναι ότι όλες οι έννοιες συνδέονται μεταξύ τους με μια θετική τιμή μεγαλύτερη του μηδενός, που υποδηλώνει ότι δεν υπάρχουν μεταβλητές στα δεδομένα εκπαίδευσης που δεν συσχετίζονται μεταξύ τους. Επομένως, τα δεδομένα εκπαίδευσης έχουν ένα αυξημένο βαθμό σηματολογικής συνέπειας.

Τέλος, ενδιαφέρον παρουσιάζει η μεταβλητή CalendarMonth, όπου έχει αντιστοιχηθεί με την μεταβλητή που διατηρεί το μήνα πραγματοποίησης της επίθεσης. Το περίεργο είναι η αυξημένη ομοιότητάς της με την πολιτική ιδεολογία (PoliticalBeliefSystem) και τον τύπο του στόχου ή τις ζημιές που προκαλεί μια επίθεση (AttackOnObject). Παρόλα αυτά εξετάζοντας διεξοδικά την Οντολογία αυτές οι συνδέσεις είναι αιτιολογημένες καθώς οι έννοιες με τις οποίες συνδέεται έχουν μια αφηρημένη σύνδεση με την έννοια χρόνος γενικότερα. Αυτό επισημαίνει στους χρήστες μιας οντολογίας που προσπαθούν να την αντιστοιχήσουν με τις μεταβλητές μίας βάσης ότι πρέπει να είναι πολύ προσεκτικοί με τις πολλαπλές ερμηνείες που πιθανόν να μπορούν να αποδοθούν σε μία έννοια. Η λύση εδώ είναι να αναζητηθούν πιο εξειδικευμένες και καλά ορισμένες έννοιες.

Σε αυτό το σημείο θα γίνει προσπάθεια για τον ορισμό μιας νέας έννοιας η οποία θα χρησιμοποιηθεί για την απόδειξη των υποθέσεων της περίπτωσης χρήσης και έχει να κάνει με τη δυνατότητα που προσφέρει η εφαρμογή που υλοποιήθηκε για την αναζήτηση στοχευόμενων κανόνων. Αν γίνει διάταξη των σημασιολογικών αποστάσεων του παραπάνω πίνακα από το μικρότερο στο μεγαλύτερο τότε στις πρώτες θέσεις θα εμφανίζονται έννοιες που σημασιολογικά είναι μακριά, ενώ στις τελευταίες θέσεις έννοιες που έχουν μεγάλη σημασιολογική ομοιότητα. Αναμένεται επομένως στο ενδιάμεσο αυτής της διάταξης να υπάρχουν συνδέσεις εννοιών που σημασιολογικά να παρουσιάζουν μεγάλο ενδιαφέρον για έναν ειδικό καθώς δεν είναι εντελώς ξένες μεταξύ τους αλλά ούτε και έχουν μια εύκολα αντιληπτή σύνδεση. Παραδείγματος χάρη, οι υποθέσεις που εξετάζονται από τον Chasdi και επιβεβαιώνεται θα μπορούσαν να αναπαρασταθούν από τις αποστάσεις 31. *AttackOnObject - PoliticalBeliefSystem*, 32. *GeographicalRegion - AttackOnObject*, 34. *Injury_Object - PoliticalBeliefSystem* οι οποίες πράγματι βρίσκονται στο μέσο της διάταξης, που συνάδει με την υπόθεση ότι στο μέσο των διατεταγμένων σημασιολογικών αποστάσεων κρύβονται οι πραγματικά ενδιαφέρουσες πληροφορίες.

Όπως παρατηρείται όμως μεταξύ των αποστάσεων 31 και 34, από τον παρακάτω πίνακα, μπορεί να δημιουργηθεί ένα άνω και κάτω φράγμα στις αποστάσεις που πιθανόν να ενδιαφέρουν τον ειδικό. Αυτό το άνω και κάτω φράγμα το ορίζουμε Σημασιολογικό Πλαίσιο, δηλαδή ένα πλαίσιο ή ένα συνεχές διάστημα τιμών το οποίο είναι σημαντικό για τον ειδικό.

Πίνακας 18-2. Αύξουσα διάταξη αποστάσεων εννοιών

	Μεταβλητή 1	Μεταβλητή 2	Ομ.
1	Country	DeathOfTheVictim	0.055
2	TerroristGroup	DeathOfTheVictim	0.06
3	TerroristAttack	DeathOfTheVictim	0.085
4	Weapon	DeathOfTheVictim	0.086
5	DeathOfTheVictim	Injury_Object	0.105
6	TerroristAttack	TerroristGroup	0.116
7	AttackOnObject	DeathOfTheVictim	0.12
8	CalendarMonth	DeathOfTheVictim	0.125
9	GeographicalRegion	DeathOfTheVictim	0.125
10	AttackOnObject	TerroristGroup	0.14
11	CalendarMonth	Country	0.162
12	TerroristGroup	Injury_Object	0.162
13	TerroristGroup	PoliticalBeliefSystem	0.162
14	<i>DeathOfTheVictim</i>	<i>PoliticalBeliefSystem</i>	<i>0.166</i>
15	TerroristGroup	Weapon	0.173
16	CalendarMonth	TerroristGroup	0.175
17	GeographicalRegion	TerroristGroup	0.175
18	CalendarMonth	Injury_Object	0.178
19	CalendarMonth	Weapon	0.193
20	CalendarMonth	GeographicalRegion	0.2
21	TerroristAttack	Injury_Object	0.204
22	Country	TerroristAttack	0.206
23	<i>Country</i>	<i>PoliticalBeliefSystem</i>	<i>0.21</i>
24	TerroristAttack	PoliticalBeliefSystem	0.21
25	GeographicalRegion	TerroristAttack	0.219
26	Country	AttackOnObject	0.224
27	TerroristAttack	Weapon	0.239
28	CalendarMonth	TerroristAttack	0.25
29	AttackOnObject	Injury_Object	0.264
30	AttackOnObject	Weapon	0.27
31	<i>AttackOnObject</i>	<i>PoliticalBeliefSystem</i>	<i>0.285</i>
32	<i>GeographicalRegion</i>	<i>AttackOnObject</i>	<i>0.29</i>
33	CalendarMonth	PoliticalBeliefSystem	0.3
34	<i>Injury_Object</i>	<i>PoliticalBeliefSystem</i>	<i>0.318</i>
35	Weapon	PoliticalBeliefSystem	0.32
36	Country	Injury_Object	0.325
37	CalendarMonth	AttackOnObject	0.333
38	Country	Weapon	0.357
39	GeographicalRegion	PoliticalBeliefSystem	0.368
40	Country	TerroristGroup	0.425
41	Country	GeographicalRegion	0.428
42	Weapon	Injury_Object	0.481
43	GeographicalRegion	Weapon	0.541
44	GeographicalRegion	Injury_Object	0.65
45	TerroristAttack	AttackOnObject	0.714
46	CalendarMonth	CalendarMonth	1.0
47	Country	Country	1.0
48	GeographicalRegion	GeographicalRegion	1.0
49	TerroristAttack	TerroristAttack	1.0
50	AttackOnObject	AttackOnObject	1.0
51	TerroristGroup	TerroristGroup	1.0
52	Weapon	Weapon	1.0
53	DeathOfTheVictim	DeathOfTheVictim	1.0
54	Injury_Object	Injury_Object	1.0
55	PoliticalBeliefSystem	PoliticalBeliefSystem	1.0

19. Ανάλυση αποτελεσμάτων

Στην ενότητα που ακολουθεί δύο είναι οι βασικοί άξονες που επιδιώκονται να επιτευχθούν. Ο πρώτος προσπαθεί να αναδείξει τις ιδιαιτερότητες των δεδομένων εκπαίδευσης και ο δεύτερος να αξιοποιήσει την υλοποίηση της επέκτασης του Arjori αλγορίθμου. Αρχικά θα εξεταστούν γενικά στατιστικά στοιχεία που ενημερώνουν κυρίως τον αναγνώστη για την ποιότητα και την ποικιλία των δεδομένων. Έπειτα, έχοντας ως γνώμονα τις υποθέσεις της περίπτωσης χρήσης θα γίνει προσπάθεια απόδειξης τους, όπου θα συμβάλει στην επικύρωση και την ανάδειξη της χρησιμότητας της υλοποίησης. Μέσα από τις υποενότητες *περιπτώσεις ανάλυσης* γίνεται επιπλέον προσπάθεια για την ανάδειξη των διαφορετικών λειτουργιών της υλοποίησης. Εδώ σε αντίθεση με το κεφάλαιο της σχεδίασης δίνεται μεγαλύτερο βάρος στην αξιολόγηση της εφαρμογής με βάση την σημασιολογική ορθότητα των κανόνων που παράγει.

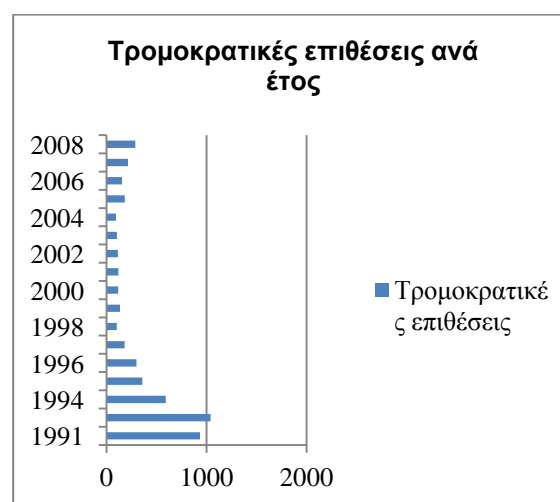
19.1 Στατιστικά στοιχεία

Σε αυτή την υποενότητα θα παρουσιαστούν στατιστικά στοιχεία για τις μεταβλητές των δεδομένων εκπαίδευσης για να υπάρχει μια ολοκληρωμένη εικόνα των δεδομένων πάνω στα οποία θα εκτελεστούν οι διαδικασίες της Εξόρυξης Γνώσης από Δεδομένα.

Τρομοκρατικές επιθέσεις ανά έτος και ανά μήνα. Όπως φαίνεται στα παρακάτω διαγράμματα από το 1991 έως το 1994 υπήρξε μια έξαρση των τρομοκρατικών επιθέσεων με αποκορύφωμα το 1992 που καταγράφηκαν 1040 επιθέσεις. Τα επόμενα χρόνια οι οργανώσεις που εξετάζονται μειώνουν σημαντικά τις επιθέσεις που οργανώνουν, παρόλα αυτά τα τελευταία χρόνια και συγκεκριμένα από το 2005 παρουσιάζεται μια συνεχής αύξηση. Το διάγραμμα με τις τρομοκρατικές επιθέσεις ανά μήνα δε παρουσιάζει κάποιο σημαντικό ενδιαφέρον εκτός του ότι υπενθυμίζει ότι οι οργανώσεις δεν έχουν κάποιο μήνα που προτιμούν να επιτίθενται, ενώ περίπου τέσσερεις (4) είναι ο μέσος αριθμός τρομοκρατικών επιθέσεων το μήνα.

Πίνακας 19-1. Τρομοκρατικές επιθέσεις ανά έτος

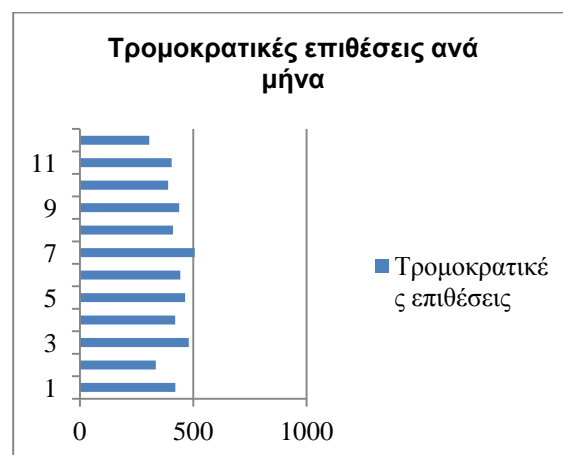
Έτος	Τρομοκρατικές επιθέσεις	%
1991	933	19%
1992	1040	21%
1994	591	12%
1995	358	7%
1996	298	6%
1997	180	4%
1998	102	2%
1999	134	3%
2000	116	2%
2001	118	2%
2002	113	2%
2003	104	2%
2004	93	2%
2005	181	4%
2006	155	3%
2007	213	4%
2008	288	6%



Διάγραμμα 19-1. Τρομοκρατικές επιθέσεις ανά έτος.

Πίνακας 19-2. Τρομοκρατικές επιθέσεις ανά μήνα

Μήνας	Τρομοκρατικές επιθέσεις	%
1	421	8%
2	335	7%
3	480	10%
4	420	8%
5	464	9%
6	443	9%
7	506	10%
8	410	8%
9	438	9%
10	389	8%
11	405	8%
12	306	6%

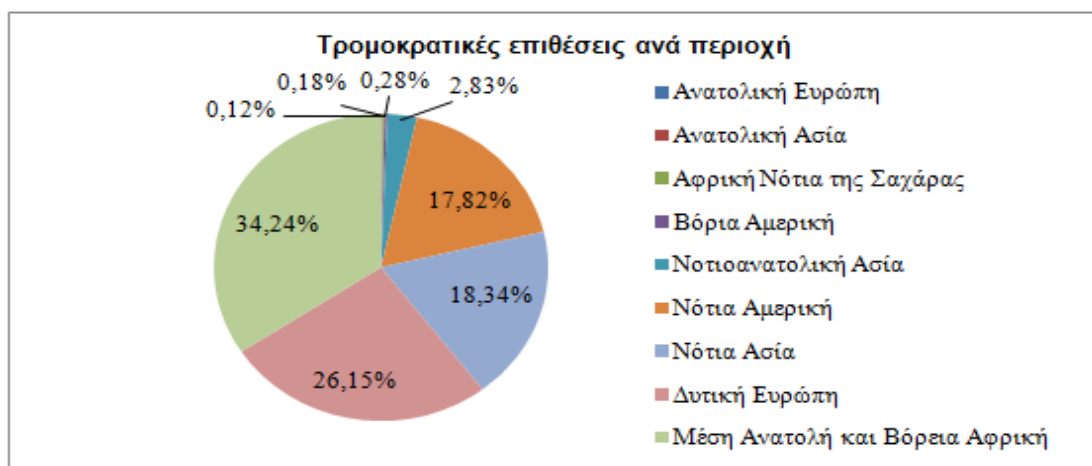
**Διάγραμμα 19-2. Τρομοκρατικές επιθέσεις ανά μήνα**

Τρομοκρατικές επιθέσεις ανά περιοχή και ανά χώρα. Όπως και φημιολογείται η περιοχή της Μέσης ανατολής και της Βόρειας Αφρικής βρίσκεται στην πρώτη θέση με τις περισσότερες τρομοκρατικές επιθέσεις, όπου σε αυτές οι πέντε πιο ενεργές χώρες είναι η Τουρκία, Ισραήλ, Δυτική Όχθη και Λωρίδα της Γάζας, Λίβανο και το Ιράκ. Ακολουθούν οι περιοχές της Δυτική Ευρώπη, Νότιας Ασίας και Νότια Αμερική με ποσοστά 26,15%, 18,34% και 17,82% αντίστοιχα. Ενδιαφέρον παρουσιάζει η Ελλάδα που βρίσκεται στην 13^η θέση μεταξύ των 53 χωρών που εξετάζονται, με 32 τρομοκρατικές επιθέσεις ή σε ποσοστό 0,64%.

Πίνακας 19-3. Τρομοκρατικές επιθέσεις ανά περιοχή, χώρα

Κωδικός Περιοχής	Ονομασία Περιοχής	Τρομοκρατικές Επιθέσεις ανά περιοχή	%	Κωδικός Χώρας	Ονομασία Χώρας	Τρομοκρατικές Επιθέσεις ανά χώρα	%
9	Ανατολική Ευρώπη	2	0,04%	28	Βοσνία-Ερζεγοβίνη	1	0,02%
				166	Ρουμανία	1	0,02%
4	Ανατολική Ασία	6	0,12%	101	Ιαπωνία	6	0,12%
11	Αφρική Νότια της Σαχάρας	9	0,18%	99	Ακτή Ελεφαντοστού	1	0,02%
				128	Μαυριτανία	1	0,02%
				146	Νιγηρία	1	0,02%
				203	Τανζανία	1	0,02%
				182	Σομαλία	2	0,04%
				104	Κένυα	3	0,06%
1	Βόρια Αμερική	14	0,28%	38	Καναδάς	1	0,02%
				130	Μεξικό	2	0,04%
				217	Ηνωμένες Πολιτείες	11	0,22%
5	Νοτιοανατολική Ασία	142	2,83%	121	Μαλαισία	1	0,02%
				160	Φιλιππίνες	141	2,81%
3	Νότια Αμερική	894	17,82%	26	Βολιβία	1	0,02%
				30	Μπρουνέι	1	0,02%
				11	Αργεντινή	2	0,04%
				159	Περού	890	17,74%

Κωδικός Περιοχής	Ονομασία Περιοχής	Τρομοκρατικές Επιθέσεις ανά περιοχή	%	Κωδικός Χώρας	Ονομασία Χώρας	Τρομοκρατικές Επιθέσεις ανά χώρα	%
6	Νότια Ασία	920	18,34%	92	Ινδία	15	0,30%
				153	Πακιστάν	15	0,30%
				4	Αφγανιστάν	20	0,40%
				186	Σρι Λάνκα	870	17,34%
8	Δυτική Ευρώπη	1312	26,15%	55	Δανία	1	0,02%
				199	Ελβετία	1	0,02%
				151	Νορβηγία	1	0,02%
				68	Φιλανδία	1	0,02%
				21	Βέλγιο	6	0,12%
				142	Ολλανδία	6	0,12%
				15	Αυστρία	7	0,14%
				96	Ιρλανδία	8	0,16%
				69	Γαλλία	17	0,34%
				98	Ιταλία	21	0,42%
				78	Ελλάδα	32	0,64%
				216	Μεγάλη Βρετανία	121	2,41%
				75	Γερμανία	132	2,63%
				233	Βόρια Ιρλανδία	470	9,37%
10	Μέση Ανατολή και Βόρεια Αφρική	1718	34,24%	422	Διεθνή	1	0,02%
				106	Κουβέιτ	1	0,02%
				136	Μαρόκο	1	0,02%
				200	Συρία	1	0,02%
				60	Αίγυπτος	2	0,04%
				208	Τυνησία	2	0,04%
				228	Υεμένη	2	0,04%
				102	Ιορδανία	3	0,06%
				94	Ιράν	4	0,08%
				173	Σαουδική Αραβία	14	0,28%
				6	Αλγερία	32	0,64%
				95	Ιράκ	171	3,41%
				110	Λίβανο	198	3,95%
				155	Δυτική Όχθη και Λωρίδα της Γάζας	231	4,60%
97	Ισραήλ	285	5,68%				
209	Τουρκία	770	15,35%				

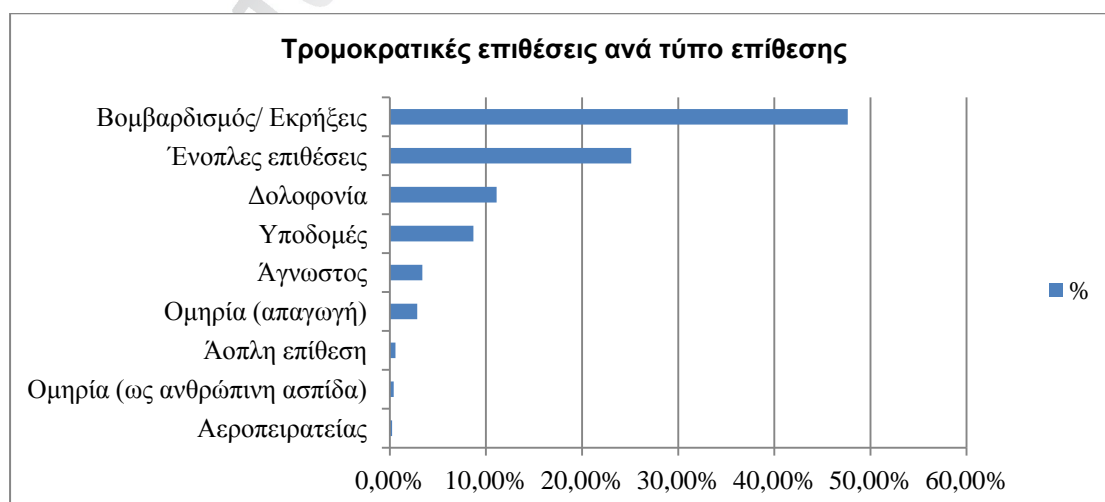


Διάγραμμα 19-3. Τρομοκρατικές επιθέσεις ανά περιοχή

Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης. Στη μελέτη καταγράφονται εννιά τύποι τρομοκρατικών επιθέσεων, μεταξύ των οποίων και μια γενική κατηγορία που θεωρεί τον τύπο της επίθεσης άγνωστο και αντιστοιχεί μόνο στο 3,39% των επιθέσεων. Οι τέσσερις πιο δημοφιλείς μέθοδοι επίθεσης είναι οι βομβαρδισμοί ή εκρήξεις με ποσοστό 47,64%, οι ένοπλες επιθέσεις 25,11%, οι δολοφονίες με 11,10% και οι επιθέσεις σε υποδομές με ποσοστό 8,69%.

Πίνακας 19-4. Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης

Κωδικός	Τύπος επίθεσης	Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης	%
4	Αεροπειρατεία	12	0,24%
5	Ομηρία (ως ανθρώπινη ασπίδα)	20	0,40%
8	Άοπλη επίθεση	29	0,58%
6	Ομηρία (απαγωγή)	143	2,85%
9	Άγνωστος	170	3,39%
7	Υποδομές	436	8,69%
1	Δολοφονία	557	11,10%
2	Ένοπλες επιθέσεις	1260	25,11%
3	Βομβαρδισμός/ Εκρήξεις	2390	47,64%



Διάγραμμα 19-4. Τρομοκρατικές επιθέσεις ανά τύπο επίθεσης

Τρομοκρατικές επιθέσεις ανά τύπο στόχου. Οι στόχοι των τρομοκρατικών οργανώσεων εμφανίζονται να είναι εικοσι-τριώνδιαφορετικών τύπων. Μεταξύ αυτών οι επτά πιο δημοφιλείς στόχοι, οι οποίοι και αποτελούν το 87,30% των επιθέσεων, είναι οι ιδιώτες και ιδιωτικές περιουσίες με ποσοστό 20,73%, ο στρατός με 20,59%, οι επιχειρήσεις με 16,24%, η αστυνομία με 13,43%, οι κυβερνητικοί στόχοι με 7,63%, οι στόχοι μεταφορών εκτός τις αεροπορίας με 5,5% και τέλος οι στόχοι κοινής ωφέλειας με ποσοστό 3,17%.

Πίνακας 19-5. Τρομοκρατικές επιθέσεις ανά τύπο στόχου

Κωδικός	Τύπος στόχου	Τρομοκρατικές επιθέσεις ανά τύπο στόχου	%
14	Ιδιώτες & ιδιωτική περιουσία	1040	20,73%
4	Στρατός	1033	20,59%
1	Επιχειρήσεις	815	16,24%
3	Αστυνομία	674	13,43%
2	Κυβερνητικοί (γενικά)	383	7,63%
19	Μεταφορών (εκτός αεροπορίας)	276	5,50%
21	Κοινής ωφέλειας	159	3,17%
13	Άλλοι	103	2,05%
17	Τρομοκράτες	103	2,05%
7	Κυβερνητικοί (διπλωματικοί)	80	1,59%
15	Θρησκευτικές προσωπικότητες/ ιδρύματα	65	1,30%
20	Άγνωστοι	48	0,96%
8	Εκπαιδευτικά ιδρύματα	47	0,94%
10	Δημοσιογράφοι & ΜΜΕ	41	0,82%
6	Αεροδρόμια & αερογραμμές	31	0,62%
18	Τουρίστες	29	0,58%
12	Μη κυβερνητικές οργανώσεις	25	0,50%
22	Βίαια πολιτικά κόμματα	25	0,50%
11	Ναυτιλιακοί (λιμένες και ναυτιλιακές εγκαταστάσεις)	18	0,36%
16	Τηλεπικοινωνίες	17	0,34%
9	Εγκαταστάσεις τροφίμων ή νερού	5	0,10%

Τρομοκρατικές επιθέσεις ανά εθνικότητα στόχου. Η εθνικότητα του στόχου εξετάζεται καθώς μια τρομοκρατική οργάνωση δεν είναι απαραίτητο ότι θα επιτίθεται μόνο σε ομοεθνείς, ούτε μόνο στο έδαφος από το οποίο προέρχεται. Γι' αυτό και παρατηρούνται διαφορές μεταξύ του επόμενου πίνακα με τον πίνακα που συσχετίζει τις επιθέσεις με τις χώρες που διαπράττονται. Παρόλα αυτά η Τουρκία, το Περού, η Σρι Λάνκα, η Ισπανία εξακολουθούν να είναι ψηλά και στους δύο πίνακες. Βέβαια, πλέον στις εθνικότητες που παρουσιάζεται μεγάλος αριθμός επιθέσεων συγκαταλέγεται το Ισραήλ, η Μεγάλη Βρετανία και το Λίβανο, ενώ λιγότερες επιθέσεις σε εθνικότητες όπως της Βόρειας Ιρλανδίας, της Δυτικής Όχθης και της Λωρίδας της Γάζας διαπράττονται σε σχέση με τις επιθέσεις που γίνονται σε αυτά τα κράτη.

Πίνακας 19-6. Τρομοκρατικές επιθέσεις ανά εθνικότητα στόχου

Κωδικός Χώρας	Όνομασία Χώρας	Τρομοκρατικές επιθέσεις ανά εθνικότητα	%	Κωδικός Χώρας	Όνομασία Χώρας	Τρομοκρατικές επιθέσεις ανά εθνικότητα	%
209	Τουρκία	890	17,74%	14	Αυστραλία	2	0,04%
159	Περου	860	17,14%	21	Βέλγιο	2	0,04%
186	Σρι Λάνκα	840	16,74%	26	Βολιβία	2	0,04%
97	Ισραήλ	527	10,50%	93	Ινδονησία	2	0,04%
185	Ισπανία	483	9,63%	184	Νότια Κορέα	2	0,04%
233	Βόρια Ιρλανδία	323	6,44%	208	Τυνησία	2	0,04%
216	Μεγάλη Βρετανία	264	5,26%	11	Αργεντινή	1	0,02%
95	Ιράκ	134	2,67%	15	Αυστρία	1	0,02%
160	Φιλιππίνες	118	2,35%	18	Μπαχρέιν	1	0,02%
110	Λίβανο	111	2,21%	28	Βοσνία-Ερζεγοβίνη	1	0,02%
-9	Άγνωστο	71	1,42%	30	Μπρουνέι	1	0,02%
155	Δυτική Όχθη και Λωρίδα της Γάζας	64	1,28%	32	Βουλγαρία	1	0,02%
217	Ηνωμένες Πολιτείες	49	0,98%	68	Φινλανδία	1	0,02%
422	Διεθνή	46	0,92%	106	Κουβέιτ	1	0,02%
6	Αλγερία	26	0,52%	111	Λεσότο	1	0,02%
69	Γαλλία	23	0,46%	128	Μαυριτανία	1	0,02%
75	Γερμανία	19	0,38%	130	Μεξικό	1	0,02%
4	Αφγανιστάν	18	0,36%	144	Νέα Ζηλανδία	1	0,02%
78	Ελλάδα	17	0,34%	146	Νιγηρία	1	0,02%
92	Ινδία	16	0,32%	161	Πολωνία	1	0,02%
101	Ιαπωνία	13	0,26%	166	Ρουμανία	1	0,02%
153	Πακιστάν	12	0,24%	179	Δημοκρατία της Σλοβακίας	1	0,02%
98	Ιταλία	10	0,20%	182	Σομαλία	1	0,02%
38	Καναδάς	7	0,14%	195	Σουδάν	1	0,02%
44	Κίνα	7	0,14%	199	Ελβετία	1	0,02%
96	Ιρλανδία	7	0,14%	200	Συρία	1	0,02%
173	Σαουδική Αραβία	7	0,14%	201	Ταϊβάν	1	0,02%
60	Αίγυπτος	4	0,08%	205	Ταϊλάνδη	1	0,02%
142	Ολλανδία	4	0,08%	215	Ηνωμένα Αραβικά Εμιράτα	1	0,02%
94	Ιράν	3	0,06%	228	Υεμένη	1	0,02%
102	Ιορδανία	3	0,06%	296	Κούρδοι	1	0,02%
167	Ρωσία	3	0,06%	359	Σοβιετική Ένωση	1	0,02%

Τρομοκρατικές επιθέσεις ανά Ιδεολογία και οργάνωση. Οι οργανώσεις που υπόκεινται στον Εθνικό Απελευθερωτικό Αγώνα εμφανίζονται να έχουν το μεγαλύτερο ποσοστό δράσης με 58,66% από τις οποίες η οργάνωση Kurdistan Workers' Party (PKK) έχει διαπράξει τις περισσότερες ενέργειες (18,70% στον σύνολο των τρομοκρατικών επιθέσεων). Ακολουθούν οι Θρησκευτικές- Πολιτικές οργανώσεις με 21,75% με ενεργητικότερη την Hamas (Islamic Resistance Movement) με 5,44%. Τέλος, οι Σοσιαλιστικές/ Κομμουνιστικές λαμβάνουν το ποσοστό των 19,59% με πιο ενεργή την Shining Path (SL) με 17,80%.

Πίνακας 19-7. Τρομοκρατικές επιθέσεις ανά Ιδεολογία, Οργάνωση

Κωδικός Ιδεολογίας	Ιδεολογία	Τρομοκρ. Επιθ. ανά Ιδεολογία	%	Κωδικός Οργάνωσης	Ονομασία Οργάνωσης	Τρομοκρ. Επιθ. ανά Οργάνωση	%
1	Σοσιαλιστές/ Κομμουνιστές	983	19,59%	1147	Red Army Faction (RAF)	3	0,06%
				1148	Red Brigades	5	0,10%
				1182	Revolutionary People's Struggle (ELA)	30	0,60%
				1096	Popular Front for the Liberation of Palestine (PFLP)	52	1,04%
				1266	Shining Path (SL)	893	17,80%
3	Θρησκευτικό- Πολιτικές	1091	21,75%	65	Al-Qa`ida in Levant and Egypt	1	0,02%
				66	Al-Qa`ida in the Arabian Peninsula (AQAP)	5	0,10%
				758	Ku Klux Klan	7	0,14%
				204	Aum Shinri Kyo	14	0,28%
				67	Al-Qa`ida in the Lands of the Islamic Maghreb (AQLIM)	34	0,68%
				63	Al-Qa`ida	84	1,67%
				14	Abu Sayyaf Group (ASG)	142	2,83%
				64	Al-Qa`ida in Iraq	142	2,83%
				632	Islamic Jihad (Ideological Grouping)	161	3,21%
				567	Hizballah	228	4,54%
549	Hamas (Islamic Resistance Movement)	273	5,44%				
2	Εθνικό Απελευθερωτικό	2943	58,66%	1034	Palestine Liberation Organization (PLO)	19	0,38%
				238	Basque Fatherland and Freedom (ETA)	512	10,21%
				617	Irish Republican Army (IRA)	599	11,94%
				806	Liberation Tigers of Tamil Eelam (LTTE)	875	17,44%
				778	Kurdistan Workers' Party (PKK)	938	18,70%

Τρομοκρατικές επιθέσεις ανά είδος όπλου. Δέκα κύριοι τύποι όπλων εμφανίζονται να χρησιμοποιούνται για τις τρομοκρατικές επιθέσεις. Μέσα σε αυτούς είναι και η γενική κατηγορία στην οποία δεν είναι γνωστό ο τύπος του όπλου και αποτελεί το 6,12% των

επιθέσεων. Πάντως η πιο διαδεδομένη χρήση όπλου είναι οι εκρηκτικές ύλες με ποσοστό 50,03%, ενώ ακολουθούν τα πυροβόλα όπλα με 32,89% και ο εμπρησμός με 8,85%. Οι υπόλοιποι τύποι όπως οχήματα, χημικά, μάχη σώμα με σώμα (melee), δολιοφθορά εξοπλισμού και ψεύτικα όπλα αποτελούν συνολικά ένα μικρό ποσοστό του 2,11%

Πίνακας 19-8. Τρομοκρατικές επιθέσεις ανά είδος όπλου

Κωδικός όπλου	Ονομασία όπλου	Τρομοκρατικές επιθέσεις ανά είδος όπλου	%
6	Εκρηκτικές ύλες/ Βόμβες/ Δυναμίτες	2510	50,03%
5	Πυροβόλα όπλα	1650	32,89%
8	Εμπρηστικός	444	8,85%
13	Άγνωστο	307	6,12%
9	Όχημα	85	1,69%
2	Χημικά	12	0,24%
12	Άλλα	4	0,08%
10	Σώμα με σώμα	2	0,04%
11	Δολιοφθορά εξοπλισμού	2	0,04%
7	Ψεύτικα όπλα	1	0,02%

Τρομοκρατικές επιθέσεις ανά αριθμό νεκρούς. Στις τέσσερις κατηγορίες που χωρίζονται οι τρομοκρατικές επιθέσεις με βάση τον αριθμό των νεκρών παρατηρείται ότι στο μεγαλύτερο ποσοστό, αυτό του 48,30%, εμφανίζονται να μην υπάρχουν νεκροί. Το αμέσως επόμενο μεγάλο ποσοστό διατηρούν οι επιθέσεις που έχουν ως αντίκτυπο έναν έως δώδεκα νεκρούς (43,37%). Από δεκατρείς νεκρούς και πάνω το ποσοστό των επιθέσεων μειώνεται αισθητά και πλησιάζει το 6,16%. Τέλος, σε ένα ποσοστό 2,17% των επιθέσεων δεν έχει διευκρινιστεί εάν υπάρχουν νεκροί.

Πίνακας 19-9. Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών

Αριθμός νεκρών	Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών	%
Άγνωστος	109	2,17%
0	2423	48,30%
1 έως 12	2176	43,37%
13 έως 1381	309	6,16%

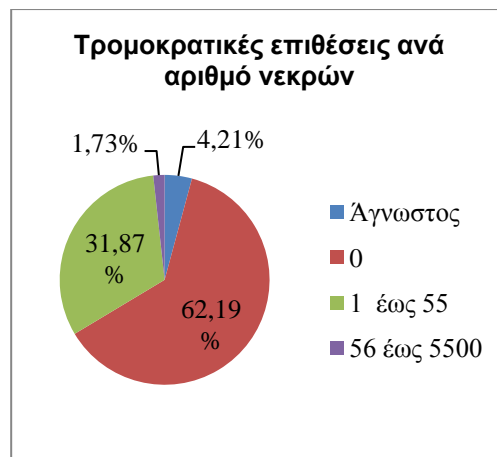


Διάγραμμα 19-5. Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών

Τρομοκρατικές επιθέσεις ανά αριθμό τραυματιών. Από τις επιθέσεις που πραγματοποιήθηκαν στο 62,19% δεν υπήρξαν τραυματίες ενώ από τις 31,87% τραυματίστηκαν από ένας έως πενήντα-πέντε άνθρωποι. Πάνω από πενήντα-έξη άτομα υπήρχαν μόνο στο 1,73% των περιπτώσεων. Για το υπόλοιπο ποσοστό 4,21% ο αριθμός των τραυματιών παραμένει άγνωστος.

Πίνακας 19-10. Τρομοκρατικές επιθέσεις ανά αριθμό τραυματιών

Αριθμός τραυματιών	Τρομοκρατικές επιθέσεις ανά αριθμό νεκρών	%
Άγνωστος	211	4,21%
0	3120	62,19%
1 έως 55	1599	31,87%
56 έως 5500	87	1,73%

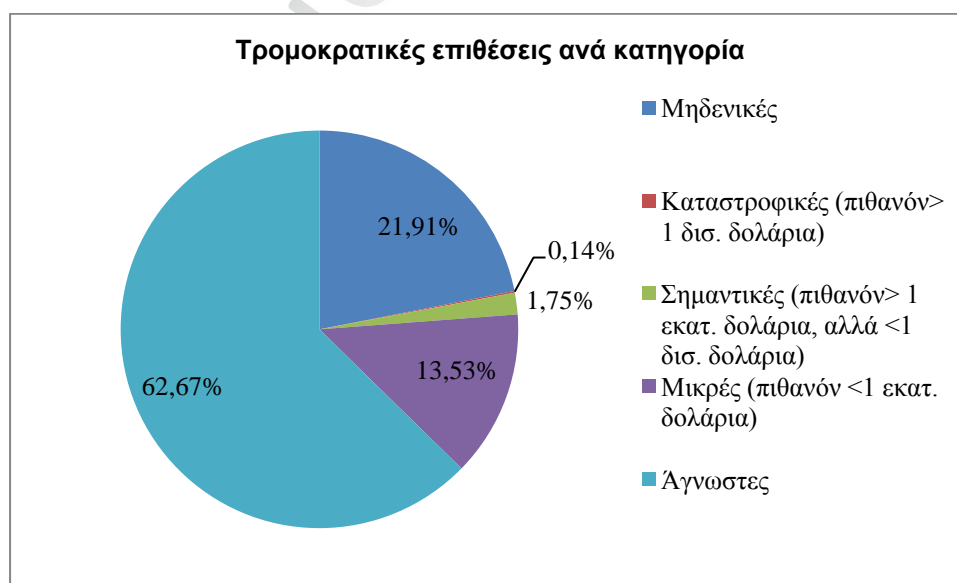


Διάγραμμα 19-6. Τρομοκρατικές επιθέσεις ανά αριθμό τραυματιών

Τρομοκρατικές επιθέσεις ανά μέγεθος υλικών ζημιών. Για το μεγαλύτερο μέρος των τρομοκρατικών επιθέσεων είναι άγνωστες οι υλικές ζημιές αποτιμημένες σε δολάρια (62,67%). Για τις υπόλοιπες επιθέσεις, στο 21,91% δεν παρουσιάζονται ζημιές, στο 13,53% υπολογίζεται ότι υπάρχουν μικρές θεωρητικά ζημιές, που κοστολογούνται σε λιγότερο από 1 εκατομμύριο δολάρια, στο 1,75% εμφανίζονται σημαντικές ζημιές που ανέρχονται σε λιγότερο από 1 δισεκατομμύριο δολάρια αλλά περισσότερες από 1 εκατομμύριο δολάρια και τέλος στο 0,14% υπολογίζονται οι καταστροφικές ζημιές που ξεπερνούν το 1 δισεκατομμύριο δολάρια.

Πίνακας 19-11. Τρομοκρατικές επιθέσεις ανά μέγεθος υλικών ζημιών

Κωδικός	Κατηγορίες ζημιών	Τρομοκρατικές επιθέσεις ανά κατηγορία	%
0	Μηδενικές	1099	21,91%
1	Καταστροφικές (πιθανόν > 1 δισ. δολάρια)	7	0,14%
2	Σημαντικές (πιθανόν > 1 εκατ. δολάρια, αλλά <1 δισ. δολάρια)	88	1,75%
3	Μικρές (πιθανόν <1 εκατ. δολάρια)	679	13,53%
4	Άγνωστες	3144	62,67%



Διάγραμμα 19-7. Τρομοκρατικές επιθέσεις ανά μέγεθος υλικών ζημιών

19.2 Πρώτη περίπτωση ανάλυσης

Στην πρώτη περίπτωση ανάλυσης θα εξεταστούν τα αποτελέσματα που παράγονται από την κλασική διαδικασία παραγωγής Κανόνων Συσχετίσεων μέσω του Apriori. Αυτά θα θεωρηθούν η βάση σύγκρισης για το κατά πόσο υπάρχει βελτίωση των αποτελεσμάτων από τις νέες διαδικασίες που ενσωματώνουν τις Οντολογίες. Η διαφορά με το προηγούμενο κεφάλαιο της υλοποίησης είναι ότι τώρα μεγαλύτερο ενδιαφέρον θα παρουσιαστεί στην σημασιολογική βελτίωση των αποτελεσμάτων και όχι μόνο στη μείωση ή γρήγορη παραγωγή των κανόνων.

Για τη δημιουργία της πρώτης περίπτωσης ανάλυσης αρκεί να εισαχθούν τα δεδομένα εκπαίδευσης που παρουσιάστηκαν στην προηγούμενη ενότητα και να εκτελεσθεί ο Apriori μέσω του Pattern Miner εισάγοντας τις ακόλουθες τιμές (10000, 0.9, 0.05, 0.1) στις μεταβλητές (Number of rules, Minimum confidence, Delta for decrease of minimum support, Lower bound for minimum support). Ένα υποσύνολο των αποτελεσμάτων παρουσιάζεται στον ακόλουθο πίνακα.

Πίνακας 19-12. Αποτελέσματα Apriori με min Conf: 0.9

1. gname=778 938 ==> terroristcountry=296 terroristregion=10 ideology=2 938	conf:(1)
2. terroristcountry=296 938 ==> gname=778 terroristregion=10 ideology=2 938	conf:(1)
...	
34. gname=1266 893 ==> terroristcountry=159 terroristregion=3 ideology=1 893	conf:(1)
35. terroristcountry=159 893 ==> gname=1266 terroristregion=3 ideology=1 893	conf:(1)
...	
77. region=3 gname=1266 892 ==> terroristcountry=159 terroristregion=3 ideology=1 892	conf:(1)
78. region=3 terroristcountry=159 892 ==> gname=1266 terroristregion=3 ideology=1 892	conf:(1)
...	
127. country=159 890 ==> region=3 gname=1266 terroristcountry=159 terroristregion=3 ideology=1 890	conf:(1)
128. country=159 region=3 890 ==> gname=1266 terroristcountry=159 terroristregion=3 ideology=1 890	conf:(1)
...	
188. country=159 890 ==> region=3 gname=1266 terroristregion=3 ideology=1 890	conf:(1)
...	
338. natly1=209 gname=778 886 ==> terroristcountry=296 terroristregion=10 ideology=2 886	conf:(1)
339. natly1=209 terroristcountry=296 886 ==> gname=778 terroristregion=10 ideology=2 886	conf:(1)
...	
1886. gname=1266 nbound=l 729 ==> terroristcountry=159 terroristregion=3 ideology=1 729	conf:(1)
1887. nbound=l terroristcountry=159 729 ==> gname=1266 terroristregion=3 ideology=1 729	conf:(1)
...	
2220. gname=778 propextent=4 703 ==> terroristcountry=296 terroristregion=10 ideology=2 703	conf:(1)
2221. propextent=4 terroristcountry=296 703 ==> gname=778 terroristregion=10 ideology=2 703	conf:(1)
...	
4620. gname=806 nkill=WP 523 ==> terroristcountry=186 terroristregion=6 ideology=2 523	conf:(1)
4621. nkill=WP terroristcountry=186 523 ==> gname=806 terroristregion=6 ideology=2 523	conf:(1)

Στις πρώτες εκατό θέσεις οι μεταβλητές που εμφανίζονται στους παραγόμενους κανόνες είναι οι gname, terroristcountry, terroristregion και ideology, οι οποίες έχουν άμεση σχέση μεταξύ τους, με αποτέλεσμα οι συνδυασμοί τους να μην προσφέρουν καμία χρήσιμη πληροφορία στο χρήστη που τις εξετάζει. Έπειτα εισάγονται οι μεταβλητές country και natly1 όπου θα μπορούσαν να προσφέρουν ενδιαφέρουσες πληροφορίες αλλά επειδή οι προηγούμενες τέσσερις μεταβλητές που αναφέρονται στις τρομοκρατικές οργανώσεις είναι πολύ ισχυρές, συγχωνεύονται μεταξύ τους με αποτέλεσμα να δημιουργούνται δυο νέοι κανόνες αλλά με πάρα πολύ μικρή σημασιολογική αξία. Το προαναφερόμενο μοτίβο επαναλαμβάνεται μεταξύ των κανόνων 400 έως 1880, όπου η μόνη διαφορά είναι ότι γίνεται αναφορά των μεταβλητών σε διαφορετικές τρομοκρατικές οργανώσεις.

Στη συνέχεια εισάγεται η μεταβλητή nbound που αναπαράγεται μέχρι τον 2600^ο κανόνα όπου εισάγεται η μεταβλητή propextent, έπειτα ακολουθεί η μεταβλητή nkill γύρω στον 4600^ο κανόνα ενώ συνεχίζει και πέρα από τον 10000^ο. Δυστυχώς, και σε αυτούς τους κανόνες οι μεταβλητές gname, terroristcountry, terroristregion και ideology διαδραματίζουν σημαντικό ρόλο με αποτέλεσμα είτε να συνεχίζεται η παραγωγή ανώφελων κανόνων είτε να μη μπορούν να διακριθούν εύκολα οι πραγματικά ενδιαφέροντες κανόνες.

Επομένως τα συμπεράσματα που μπορούν γρήγορα να εξαχθούν για ακόμη μια φορά για την κλασική διαδικασία του Apriori είναι ότι:

1. Παράγεται μεγάλος όγκος δεδομένων.
2. Παράγεται μεγάλος όγκος τριμμένων δεδομένων.

3. Ένας μεγάλος αριθμός κανόνων απλά επαναλαμβάνεται άσκοπα.
4. Υπάρχει αρμονία στην παραγωγή αποτελεσμάτων, καθώς σε κάθε νέα μεταβλητή που εισάγεται παλιότεροι κανόνες ανακυκλώνονται. Βέβαια, στην πραγματικότητα αυτή δεν είναι μια παρατήρηση που θα βοηθούσε έναν ειδικό, όπου θα χανόταν στην πληθώρα των κανόνων.
5. Δεν μπορεί να παράγει άμεσα στοχευόμενους κανόνες. Παραδείγματος χάρη, είναι πολύ σημαντικό για την περίπτωση χρήσης που εξετάζεται να αναδειχθούν σχέσεις μεταξύ της πολιτικής ιδεολογίας των οργανώσεων με άλλες μεταβλητές. Στους παραγόμενους κανόνες που γίνεται έλεγχος η μεταβλητή ideology δεν εμφανίζεται καθόλου. Σημειώνεται, ότι η μη εμφάνιση της δεν αποδεικνύει και τη μη ύπαρξη σχέσεων της με άλλες μεταβλητές, καθώς μπορεί η παρουσία της να εμπίπτει σε μικρότερο ποσοστό εμπιστοσύνης (confidence).
6. Είναι ανακουφιστικό ότι δε παράγονται στα συγκεκριμένα δεδομένα μη αληθείς κανόνες. Αυτό όμως δεν είναι κάτι που έχει να κάνει με τον αλγόριθμο του Apriori αλλά με το ότι τα δεδομένα εκπαίδευσης έχουν προέλθει από μια συλλογή δεδομένων που οι μεταβλητές τους έχουν μια άμεση σχέση μεταξύ τους, όπως αποδεικνύεται και από την οντολογία στην προηγούμενη ενότητα.

19.3 Δεύτερη περίπτωση ανάλυσης

Στη δεύτερη περίπτωση ανάλυσης εξετάζεται ένα απλό σενάριο, κυρίως για λόγους πληρότητας. Σε αυτή την περίπτωση θα θεωρηθεί ότι ο χρήστης ενδιαφέρεται να εισάγει στην διαδικασία γνώσης από την Οντολογία και να εκμεταλλευτεί όλες τις σχέσεις που υπάρχουν μεταξύ των μεταβλητών που ορίζονται μέσα σε αυτή. Επομένως δε θα γίνει διάκριση μεταξύ των τιμών των αποστάσεων, αρκεί να μην έχουν τιμή -1 που υποδηλώνει ότι δεν υπάρχει σύνδεση μεταξύ δύο εννοιών και κατ' επέκταση δύο μεταβλητών. Για να τρέξει το συγκεκριμένο σενάριο θα πρέπει στη διεπαφή Ontology Association Rules του Pattern Miner να εισαχθούν οι τιμές (10000, 0.9, 0.05, 0.1, Pellet via Jena, TaxonomyMetrics, >=, 0.0, During) στις μεταβλητές (Number of rules, Minimum confidence, Delta for decrease of minimum support, Lower bound for minimum support, Reasoner, Metrics, Distance/ Similarity 1, Distance/ Similarity 2, Timing).

Τα αποτελέσματα όπως ήταν αναμενόμενο δε διαφέρουν από την πρώτη περίπτωση ανάλυσης. Παρόλο που εκ πρώτης όψεως φαίνεται αρνητικό, καθόλα δεν είναι. Αυτή η διαπίστωση μπορεί να χρησιμοποιηθεί για να αποδειχθεί ότι τόσο οι μεταβλητές της Βάσης Δεδομένων όσο και η επιλεγμένη Οντολογία συμβαδίζουν στα σωστά πρότυπα. Θεωρώντας ότι οι κατασκευαστές της Βάσης Δεδομένων λάμβαναν υπόψη μόνο τις κατάλληλες μεταβλητές που μπορούν να περιγράψουν επαρκώς και ορθώς τις τρομοκρατικές επιθέσεις, αυτό συνεπάγεται ότι έχει επιλεγεί η κατάλληλη Οντολογία, καθώς υπήρξε συνεπής αντιστοίχιση των μεταβλητών της ΒΔ με τις έννοιες της Οντολογίας. Αντίστροφα, θεωρώντας ότι οι κατασκευαστές της Οντολογίας χρησιμοποίησαν τις κατάλληλες έννοιες για να περιγράψουν επαρκώς το πεδίο γνώσης, τότε καθώς οι μεταβλητές της ΒΔ μπόρεσαν να αντιστοιχηθούν με τις έννοιες υποδηλώνει ότι η ΒΔ περιέχει τις κατάλληλες μεταβλητές για την αποθήκευση των δεδομένων του πεδίου γνώσης.

Όπως παρατηρείται από τον πίνακα με τις αποστάσεις των εννοιών, όλες οι έννοιες συνδέονται με μία θετική τιμή μικρή ή μεγάλη, που υποδηλώνει ότι μεταξύ τους υπάρχει μια μικρή ή μεγάλη ομοιότητα. Επομένως, από την στιγμή που επιλέχτηκε να αξιοποιηθούν όλες οι αποστάσεις χωρίς εξαίρεση ήταν αναμενόμενο να μη διαγραφεί κανένας κανόνας που κάλυπτε το κατώφλι του support και του confidence. Από αυτό μπορεί να διαπιστωθεί ότι παρόλο που οι κανόνες που παράχθηκαν δεν είναι ενδιαφέροντες, έχουν όμως σημασιολογική υπόσταση.

Από τα δύο προηγούμενα σενάρια γίνεται αντιληπτό ότι για να μπορέσουν να ελεγχθούν οι υποθέσεις της περίπτωσης χρήσης, αρκεί να γίνει εκμετάλλευση των δυνατοτήτων που δίνονται από την διεπαφή Ontology Association Rules για τον προσδιορισμό στοχευόμενων δεδομένων, αλλά και χρήση της δυνατότητας παραγωγής κανόνων χωρίς να δίνεται σημασία στο support παρά μόνο στις αποστάσεις που ορίζονται μέσα στην Οντολογία. Για τον προσδιορισμό στοχευόμενων δεδομένων αρκεί να προσδιορίζεται κάθε φορά ένα σημασιολογικό πλαίσιο μέσα στο οποίο θα πρέπει να κινούνται οι κανόνες. Το σημασιολογικό πλαίσιο θα ορίζεται με ένα πάνω και κάτω φράγμα

στις δυνατές τιμές που μπορούν να λάβουν οι αποστάσεις. Στις επόμενες περιπτώσεις ανάλυσης παίρνει οστά η παραπάνω λογική.

19.4 Τρίτη περίπτωση ανάλυσης

Εξετάζοντας τις σημασιολογικές αποστάσεις παρατηρείται ότι μπορεί να οριστεί ένα σημασιολογικό πλαίσιο μέσα στο οποίο είναι δυνατόν να υποστηριχθούν οι περισσότερες υποθέσεις. Το σημασιολογικό πλαίσιο είναι το διάστημα [0.28, 0.319] το οποίο περιλαμβάνει τις ακόλουθες τέσσερις σημασιολογικές αποστάσεις:

- AttackOnObject- PoliticalBeliefSystem, η οποία συνδέει την πολιτική ιδεολογία μιας οργάνωσης με το είδος του στόχου και την πολιτική ιδεολογία μιας οργάνωσης με τις υλικές ζημιές που προκαλούνται από μια επίθεσή της. Η διπλή έννοια οφείλεται στη διπλή σημασιολογία που έχει αποδοθεί στην έννοια AttackOnObject.
- GeographicalRegion- AttackOnObject, η οποία συνδέει την περιοχή διεξαγωγής της επίθεσης με το είδος του στόχου ή τις υλικές ζημιές.
- CalendarMonth- PoliticalBeliefSystem, η οποία συνδέει τον μήνα διεξαγωγής μιας επίθεσης με την πολιτική ιδεολογία. Εδώ βέβαια υφίσταται, όπως έχει συζητηθεί μια σημασιολογική σύγχυση καθώς για τα δεδομένα εκπαίδευσης η έννοια CalendarMonth σηματοδοτεί τον χρόνο πραγματοποίησης του γεγονότος ενώ για την οντολογία η σύνδεση του CalendarMonth- PoliticalBeliefSystem αναδεικνύει τη χρονική διάρκεια ύπαρξης της έννοιας PoliticalBeliefSystem.
- Injury_Object- PoliticalBeliefSystem, η οποία συνδέει την πολιτική ιδεολογία μιας οργάνωσης με τον αριθμό των τραυματιών από την διενέργεια μιας επίθεσης.

Έχοντας ορίσει το σημασιολογικό πλαίσιο στο οποίο είναι επιθυμητό να κυμαίνονται οι παραγόμενοι κανόνες, γίνεται εισαγωγή των τιμών (10000, 0.9, 0.05, 0.1, Pellet via Jena, TaxonomyMetrics, >=, 0.28, <=, 0.319, During) στις μεταβλητές (Number of rules, Minimum confidence, Delta for decrease of minimum support, Lower bound for minimum support, Reasoner, Metrics, Distance/ Similarity 1, Distance/ Similarity 2, Distance/ Similarity 3, Distance/ Similarity 4, Timing) και εκτελείται ο αλγόριθμος του οποίου τα αποτελέσματα φαίνονται παρακάτω.

Πίνακας 19-13. Αποτελέσματα Apriori με ενσωμάτωση Οντολογίας

1. ideology=1 983 ==> nwound=I 791 conf:(0.8)
2. targtype1=1 815 ==> ideology=2 554 conf:(0.68)
3. ideology=2 2943 ==> propextent=4 1995 conf:(0.68)
4. region=6 920 ==> propextent=4 612 conf:(0.67)
5. terroristregion=6 959 ==> propextent=4 625 conf:(0.65)
6. terroristregion=10 1853 ==> propextent=4 1206 conf:(0.65)
7. targtype1=4 1033 ==> ideology=2 672 conf:(0.65)
8. region=8 1312 ==> propextent=4 849 conf:(0.65)
9. ideology=2 2943 ==> nwound=I 1900 conf:(0.65)
10. propextent=4 3144 ==> ideology=2 1995 conf:(0.63)
11. region=10 1718 ==> propextent=4 1080 conf:(0.63)
12. terroristregion=8 1149 ==> propextent=4 711 conf:(0.62)
13. nwound=I 3120 ==> ideology=2 1900 conf:(0.61)
14. terroristregion=3 893 ==> propextent=4 542 conf:(0.61)
15. region=3 894 ==> propextent=4 542 conf:(0.61)
16. ideology=1 983 ==> propextent=4 595 conf:(0.61)
17. nwound=WP 1599 ==> ideology=2 920 conf:(0.58)
18. propextent=0 1099 ==> ideology=2 574 conf:(0.52)
19. ideology=3 1091 ==> propextent=4 554 conf:(0.51)
20. ideology=3 1091 ==> nwound=WP 504 conf:(0.46)

Ομαδοποιώντας τα αποτελέσματα σύμφωνα με τις υποθέσεις, για τις οποίες γίνεται προσπάθεια ελέγχου, παρουσιάζονται οι ακόλουθες κατηγορίες.

Πολιτική ιδεολογία => Είδος στόχου. Η οποία αναφέρεται στην πρώτη υπόθεση και μπορεί να επιβεβαιωθεί εν μέρει από τους δύο κανόνες που ακολουθούν.

- targtype1=1 815 ==> ideology=2 554 conf:(0.68)
- targtype1=4 1033 ==> ideology=2 672 conf:(0.65)

Πράγματι, επιβεβαιώνεται μια σύνδεση μεταξύ της πολιτικής ιδεολογίας και το είδος στόχου, αλλά δεν επιβεβαιώνεται και η κατεύθυνση δηλαδή ότι έχοντας ως γνώμονα την πολιτική ιδεολογία μιας οργάνωσης θα υπήρχε μία τάση για συγκεκριμένου είδους στόχους. Παρόλο που σημασιολογικά δεν υπάρχει πρόβλημα, αφού συνειδητά επιλέχτηκε μια μετρική υπολογισμού αποστάσεων που να μην ενδιαφέρεται για την κατεύθυνση των συσχετίσεων μεταξύ των εννοιών, ο τύπος υπολογισμού του confidence δεν υποστηρίζει την ύπαρξη των επιθυμητών κανόνων. Υπενθυμίζεται ότι η εμπιστοσύνη του $AR(a) X \Rightarrow Y$ είναι η εξαρτημένη πιθανότητα $P(X \cup Y|X) = P(X \cup Y)/P(X)$. Εντούτοις, ενδιαφέρον παρουσιάζει ότι αν το είδος του στόχου είναι *επιχειρήσεις* ή *κυβερνητικοί (γενικά)* τότε με εμπιστοσύνη 0.68 και 0.65 αντίστοιχα συνδέονται με τρομοκρατικές ομάδες που η ιδεολογία τους είναι η *Εθνικό-Απελευθερωτικοί*.

Πολιτική ιδεολογία => Αριθμός τραυματιών. Η οποία αναφέρεται στην τρίτη υπόθεση και μπορεί να επιβεβαιωθεί από τους κανόνες που ακολουθούν.

- 1. ideology=1 983 ==> nbound=1 791 conf:(0.8)
- 9. ideology=2 2943 ==> nbound=1 1900 conf:(0.65)
- 20. ideology=3 1091 ==> nbound=WP 504 conf:(0.46)
- 13. nbound=1 3120 ==> ideology=2 1900 conf:(0.61)
- 17. nbound=WP 1599 ==> ideology=2 920 conf:(0.58)

Σύμφωνα με τους κανόνες υπάρχει μια αναμφισβήτητη σύνδεση μεταξύ της πολιτικής ιδεολογίας και των αριθμό των τραυματιών. Συγκεκριμένα οι πολιτικές ιδεολογίες *Σοσιαλιστές/Κομμουνιστές* και *Εθνικό-Απελευθερωτικό* έχουν την τάση για μηδενικό αριθμό τραυματιών ενώ η ιδεολογία *Θρησκευτικό- Πολιτικές* έχουν την τάση η ενέργειές τους να προκαλούν μικρό αριθμό τραυματιών από 1 έως 55.

Πολιτική ιδεολογία => Υλικές ζημιές. Η οποία αναφέρεται στην τέταρτη υπόθεση και μπορεί να επιβεβαιωθεί από τους κανόνες που ακολουθούν.

- 3. ideology=2 2943 ==> propextent=4 1995 conf:(0.68)
- 16. ideology=1 983 ==> propextent=4 595 conf:(0.61)
- 19. ideology=3 1091 ==> propextent=4 554 conf:(0.51)
- 18. propextent=0 1099 ==> ideology=2 574 conf:(0.52)
- 10. propextent=4 3144 ==> ideology=2 1995 conf:(0.63)

Σύμφωνα με τους κανόνες παρουσιάζεται μια τάση σύνδεσης της πολιτικής ιδεολογίας με τις υλικές ζημιές, όμως δεν δίνεται πλούσια σημασιολογική ερμηνεία, εκτός του ότι και για τις τρεις ιδεολογίες οι υλικές ζημιές που προκαλούν είναι άγνωστες. Βέβαια υπάρχει μια αντίστροφη σύνδεση μεταξύ πολιτικής ιδεολογίας και υλικών ζημιών που αναφέρει ότι σε περίπτωση μικρών υλικών ζημιών (πιθανόν < 1 εκατ. δολάρια) υπάρχει υποστήριξη 11,44% και εμπιστοσύνη 52% να διαπράχθηκαν από την ιδεολογία *Εθνικό-Απελευθερωτικό*.

Περιοχή => Επιλογή είδους στόχου. Η οποία δεν επιβεβαιώνεται

Άλλοι κανόνες. Υπάρχει ένα μικρό σύνολο κανόνων που παράγονται μέσα στο σημασιολογικό πλαίσιο και οι οποίοι μπορούν να δώσουν επιπλέον γνώση.

- 4. region=6 920 ==> propextent=4 612 conf:(0.67)
- 8. region=8 1312 ==> propextent=4 849 conf:(0.65)
- 11. region=10 1718 ==> propextent=4 1080 conf:(0.63)
- 15. region=3 894 ==> propextent=4 542 conf:(0.61)
- 5. terroristregion=6 959 ==> propextent=4 625 conf:(0.65)
- 6. terroristregion=10 1853 ==> propextent=4 1206 conf:(0.65)
- 12. terroristregion=8 1149 ==> propextent=4 711 conf:(0.62)
- 14. terroristregion=3 893 ==> propextent=4 542 conf:(0.61)

Η ερμηνεία των κανόνων είναι ότι για τέσσερις συγκεκριμένες περιοχές, που συμπίπτουν και ως περιοχές δράσης τρομοκρατικών οργανώσεων αλλά είναι και τόπος «καταγωγής» των οργανώσεων, υπάρχει μια τάση να είναι άγνωστο το ποσό των υλικών ζημιών.

Τέλος, ενδιαφέρον παρουσιάζει η μη ύπαρξη κανόνων σχετικά με το μήνα και την πολιτική ιδεολογία που ουσιαστικά δεν υποστηρίζονται από τα δεδομένα εκπαίδευσης και αποφεύγεται η σύγχυση της ερμηνείας τους λόγω του διπτού σημασιολογικού χαρακτήρα που θα μπορούσαν να λάβουν.

Συμπεράσματα. Σε αντίθεση με τα συμπεράσματα από την πρώτη περίπτωση ανάλυσης έχουμε:

1. Παράγεται μικρός όγκος κανόνων, στην συγκεκριμένη περίπτωση μόνο είκοσι κανόνες.
2. Παράγονται ενδιαφέροντες κανόνες. Το 60% των παραγόμενων κανόνων χρησιμοποιήθηκαν για την απόδειξη των υποθέσεων, ενώ το υπόλοιπο 40% ήταν μη αναμενόμενοι κανόνες που παρουσίαζαν αυξημένο ενδιαφέρον.
3. Δυνατότητα παραγωγής στοχευόμενων κανόνων. Όλοι οι κανόνες είχαν μεταβλητές που οι μεταξύ τους αποστάσεις ήταν μέσα στο διάστημα του σημασιολογικού πλαισίου που είχε οριστεί.
4. Η παραγωγή κανόνων μόνο με δύο μεταβλητές έχει να κάνει με την επιλογή να τρέξει η διαδικασία ελέγχου των σημασιολογικών αποστάσεων κατά την διάρκεια της εκτέλεσης του αλγορίθμου. Υπενθυμίζεται ότι όταν δημιουργούνται τα στοιχειοσύνολα γίνεται έλεγχος μεταξύ όλων των στοιχείων για το αν καλύπτουν τα επιθυμητά όρια σημασιολογικής απόστασης. Προφανώς, καθώς το σημασιολογικό πλαίσιο ήταν μικρό δεν έδωσε τη δυνατότητα για παραγωγή μεγάλων στοιχειοσυνόλων. Υπάρχει και η επιλογή να γίνει ο έλεγχος στο τέλος του αλγορίθμου ο οποίος δίνει τα ίδια αποτελέσματα συν κάποιους επιπλέον κανόνες με περισσότερες από 2 μεταβλητές, παρόλα αυτά δεν εμφανίζεται κάποια απόκλιση των συμπερασμάτων του, από το τρέχον σενάριο και γι αυτό το λόγο δεν παρουσιάζεται διεξοδικά.

19.5 Τέταρτη περίπτωση ανάλυσης

Σε αυτή την περίπτωση ανάλυσης θα εξετασθεί η υπόθεση δύο όπου αναφέρεται στην συσχέτιση της πολιτικής ιδεολογίας με τον αριθμό των νεκρών σε μια τρομοκρατική ενέργεια. Επιλέχθηκε, να παρουσιαστεί μόνη της ως περίπτωση καθώς για να συμπεριλαμβάνονταν οι κανόνες που διέπουν τις έννοιες *DeathOfTheVictim* και *PoliticalBeliefSystem* θα έπρεπε το σημασιολογικό πλαίσιο να ήταν αρκετά μεγάλο, με αποτέλεσμα να παραγόταν μεγαλύτερος αριθμός κανόνων, όπου θα δυσκόλευε τον απώτερο στόχο της ανάλυσης, να απόδειξει την ορθότητα των υποθέσεων.

Για την εκτέλεση του σεναρίου αρκεί να εισαχθούν οι τιμές (1000, 0.4, 0.05, 0.1, Pellet via Jena, TaxonomyMetrics, >=, 0.163, <=, 0.17, During) στις μεταβλητές (Number of rules, Minimum confidence, Delta for decrease of minimum support, Lower bound for minimum support, Reasoner, Metrics, Distance/ Similarity 1, Distance/ Similarity 2, Timing). Η εκτέλεση του αλγορίθμου θα αποφέρει τους κανόνες που ακολουθούν.

Πίνακας 19-14. Αποτελέσματα *Apriori* με ενσωμάτωση *Οντολογίας* για υπόθεση 2

- | | |
|--------------------------------------|-------------|
| 1. nkill=I 2423 ==> ideology=2 1468 | conf:(0.61) |
| 2. nkill=WP 2176 ==> ideology=2 1264 | conf:(0.58) |
| 3. ideology=1 983 ==> nkill=I 540 | conf:(0.55) |
| 4. ideology=2 2943 ==> nkill=I 1468 | conf:(0.5) |
| 5. ideology=3 1091 ==> nkill=WP 517 | conf:(0.47) |
| 6. ideology=2 2943 ==> nkill=WP 1264 | conf:(0.43) |

Για ακόμα μια φορά γίνονται εύκολα αντιληπτές οι δυνατότητες που μας δίνει η ενσωμάτωση των *Οντολογιών* μέσα στην διαδικασία του *AR* και συγκεκριμένα στον αλγόριθμο *Apriori*. Λίγοι κανόνες, στο σύνολο έξι! Απόλυτα στοχευόμενη, χωρίς περιττή πληροφορία. Όσο για την υπόθεση τα αποτελέσματα την υποστηρίζουν πλήρως, δηλαδή ότι ανάλογα την πολιτική ιδεολογία της τρομοκρατικής οργάνωσης υπάρχει μια τάση προς τον αριθμό των νεκρών που αποφέρει μια τρομοκρατική ενέργεια. Συγκεκριμένα η οργανώσεις της ιδεολογίας *Σοσιαλιστές/ Κομμουνιστές* έχουν την τάση να μην επιφέρουν νεκρούς σε αντίθεση με αυτές της ιδεολογίας *Θρησκευτικό- Πολιτικές* όπου επιφέρουν μικρό σύνολο νεκρών από 1 έως 12. Οι οργανώσεις της *Εθνικό- Απελευθερωτικής* ιδεολογίας έχουν την τάση με υποστήριξη 29% και εμπιστοσύνη 50% να μην έχουν κανένα νεκρό και με υποστήριξη 25% και εμπιστοσύνη 43% να αποφέρουν λίγους θανάτους σε 1 έως 12 άτομα.

19.6 Πέμπτη περίπτωση ανάλυσης

Στη τρίτη περίπτωση ανάλυσης ελέγχθηκε, μεταξύ άλλων, η υπόθεση πέντε, που υποστηρίζει ότι ανάλογα την περιοχή που διενεργείται μια τρομοκρατική επίθεση υπάρχει μια τάση προς

την επιλογή συγκεκριμένου στόχου. Στην τρέχουσα περίπτωση θα εξετασθεί αν υπάρχει κάποια τάση σύνδεσης μεταξύ των χωρών εκτέλεσης της τρομοκρατικής ενέργειας με τους στόχους που επιλέγονται.

Οι Οντολογικές έννοιες που χρειάζονται είναι οι *Country* και *AttackOnObject* οι οποίες έχουν μια σημασιολογική απόσταση της τάξης του 0.224. Επομένως, για την εξέταση της υπόθεσης αρκεί να οριστεί το σημασιολογικό πλαίσιο ως [0.223, 0.23]. Έτσι εισάγονται οι τιμές (1000, 0.4, 0.05, 0.1, Pellet via Jena, TaxonomyMetrics, >=, 0.223, >=, 0.23, During) στις μεταβλητές (Number of rules, Minimum confidence, Delta for decrease of minimum support, Lower bound for minimum support, Reasoner, Metrics, Distance/ Similarity 1, Distance/ Similarity 2, Distance/ Similarity 3, Distance/ Similarity 4, Timing). Τα αποτελέσματα από την εκτέλεση του αλγορίθμου εμφανίζονται στον επόμενο πίνακα.

Πίνακας 19-15. Αποτελέσματα Apriori με ενσωμάτωση Οντολογίας για υπόθεση 5

1. natlty1=209 890 ==> propextent=4 670 conf:(0.75)
2. terroristcountry=296 938 ==> propextent=4 703 conf:(0.75)
3. country=209 770 ==> propextent=4 556 conf:(0.72)
4. natlty1=186 840 ==> propextent=4 574 conf:(0.68)
5. terroristcountry=186 875 ==> propextent=4 595 conf:(0.68)
6. country=186 870 ==> propextent=4 591 conf:(0.68)
7. natlty1=159 860 ==> propextent=4 523 conf:(0.61)
8. terroristcountry=159 893 ==> propextent=4 542 conf:(0.61)
9. country=159 890 ==> propextent=4 539 conf:(0.61)

Όπως γίνεται αντιληπτό η υπόθεση δεν μπορεί να επιβεβαιωθεί. Τα αποτελέσματα δείχνουν μία τάση μεταξύ των εθνικοτήτων που έχουν πληγή από τρομοκρατική ενέργεια, των χωρών που έχουν πραγματοποιηθεί τέτοιες επιθέσεις και των χωρών «καταγωγής» των οργανώσεων με τις υλικές ζημιές και συγκεκριμένα ότι δεν είναι γνωστό το μέγεθος τους. Υπενθυμίζεται ότι η έννοια *AttackOnObject* έχει διττή σημασία λόγω της αντιστοίχισης του με τις τιμές *targetype1* και *propextent*.

19.7 Έκτη περίπτωση ανάλυσης

Σε αυτή την τελευταία περίπτωση ανάλυσης θα επανεξεταστεί η πέμπτη υπόθεση, η οποία είναι και η μοναδική που δεν έχει επιβεβαιωθεί. Αυτή τη φορά όμως θα χρησιμοποιηθεί η δεύτερη σημαντική επέκταση που έχει υλοποιηθεί στον αλγόριθμο. Δηλαδή, θα αντικατασταθεί το κατώφλι του support με τον έλεγχο του σημασιολογικού πλαισίου. Αυτή η λειτουργία θα επιτρέψει σημασιολογικά ορθούς κανόνες να αναδυθούν, ενώ ανάλογα με το πόσο ισχυροί είναι θα επιτρέπεται η όχι η εμφάνισή τους στον χρήστη.

Οι μεταβλητές που θα πρέπει να αρχικοποιηθούν είναι οι (Do not use support, Number of rules, Minimum confidence, Reasoner, Metrics, Distance/ Similarite 1, Distance/ Similarite 2, Distance/ Similarite 3, Distance/ Similarite 4, Sorting) με τις μεταβλητές (True, 1000, 0.4, , Pellet via Jena, TaxonomyMetrics, >=, 0.223, >=, 0.23, max). Ως αποτελέσματα παράγονται 257 κανόνες όπου ένα μέρος αυτών παρουσιάζονται παρακάτω.

Πίνακας 19-16. Αποτελέσματα με ενσωμάτωση Οντολογίας χωρίς support για υπόθεση 5

1. terroristcountry=60 1 ==> propextent=2 1 conf:(1)
2. terroristcountry=98 5 ==> propextent=0 5 conf:(1)
3. natlty1=359 1 ==> propextent=0 1 conf:(1)
4. natlty1=296 1 ==> propextent=4 1 conf:(1)
...
38. terroristcountry=60 1 ==> targetype1=14 1 conf:(1)
39. natlty1=201 1 ==> targetype1=14 1 conf:(1)
...
59. country=422 1 ==> propextent=2 1 conf:(1)
60. country=422 1 ==> targetype1=11 1 conf:(1)
...
146. targetype1=21 159 ==> terroristcountry=159 99 conf:(0.62)
147. targetype1=21 159 ==> natlty1=159 99 conf:(0.62)
148. targetype1=21 159 ==> country=159 99 conf:(0.62)

...
 162. propextent=1 7 ==> terroristcountry=422 4 conf:(0.57)
 163. propextent=1 7 ==> natlty1=217 4 conf:(0.57)
 ...

Από τους κανόνες που παράγονται όχι μόνο επιβεβαιώνεται η υπόθεση αλλά αναπαράγονται πλούσιες πληροφορίες. Στον πίνακα που ακολουθεί παρουσιάζονται τα συμπεράσματα που μπορούν να εξαχθούν από τους κανόνες τύπου *country => targtype1*. Αυτό ενθαρρύνει τη χρήση της συγκεκριμένης διαδικασίας παρόλο που επιβαρύνει περισσότερο τους πόρους του υπολογιστικού συστήματος κατά την εκτέλεσή του από ότι ο απλός Αρriori.

Πίνακας 19-17. Σχέση χωρών με τύπο στόχου

Χώρα	Τύπος επίθεσης	Χώρα	Τύπος επίθεσης	
Μεξικό	Επιχειρήσεις	Αργεντινή	Κυβερνητικοί (διπλωματικοί)	
Αυστρία		Νορβηγία		
Βέλγιο		Σομαλία		
Μεγάλη Βρετανία		Τανζανία		
Γερμανία		Φιλανδία		
Ιταλία		Ιράν		
Σομαλία		Κυβερνητικοί (γενικά)		Μαρόκο
Ελβετία	Καναδάς			
Τυνησία	Αστυνομία	Διεθνή	Ναυτιλιακοί (λιμένες και ναυτιλιακές εγκαταστάσεις)	
Αίγυπτος		Νιγηρία	Μη κυβερνητικές οργανώσεις	
Κουβέιτ	Στρατός	Ιορδανία	Ιδιώτες & ιδιωτική περιουσία	
Λίβανο		Κένυα		
Μαυριτανία		Σαουδική Αραβία		
Ολλανδία		Ηνωμένες Πολιτείες		
Πακιστάν		Αίγυπτος		
Συρία		Ιρλανδία		
Τυνησία		Ισραήλ		
Υεμένη		Αργεντινή		Θρησκευτικές προσωπικότητες/ ιδρύματα
Μπρουνέι		Τυνησία		
Ινδία		Βοσνία-Ερζεγοβίνη		
Ακτή Ελεφαντοστού		Αεροδρόμια & αερογραμμές		Μαλαισία
Δανία	Μεξικό		Μεταφορών (εκτός αεροπορίας)	
Ρουμανία	Άγνωστοι	Βολιβία		

Εκτός από τη ζητούμενη σχέση ενδιαφέρον εμφανίζουν άλλοι δύο τύποι κανόνων που παράγονται. Αυτοί είναι οι *natlty1 => targtype1*, *terroristcountry => targtype1*, οι οποίοι υποδηλώνουν μία τάση συσχέτισης μεταξύ της εθνικότητας των στόχων και της «καταγωγής» των τρομοκρατικών οργανώσεων με τους τύπους των στόχων που επιλέγονται. Οι πίνακες που ακολουθούν καταγράφουν τις συσχετίσεις που εξαγονται από τους κανόνες.

Πίνακας 19-18. Σχέση εθνικότητας στόχου με τύπο στόχου

Χώρα	Τύπος επίθεσης	Χώρα	Τύπος επίθεσης	
Ολλανδία	Επιχειρήσεις	Κίνα	Ναυτιλιακοί (λιμένες και ναυτιλιακές εγκαταστάσεις)	
Νότια Κορέα		Ινδονησία		
Ταϊλάνδη		Νέα Ζηλανδία	Μη κυβερνητικές οργανώσεις	
Βουλγαρία		Νιγηρία		
Σοβιετική Ένωση		Κυβερνητικοί (γενικά)	Αυστραλία	Ιδιώτες & ιδιωτική περιουσία
Καναδάς			Ιορδανία	
Σομαλία	Νότια Κορέα			
Ελβετία	Ταϊβάν			
Ελλάδα	Διεθνή			
Μαυριτανία	Στρατός		Φιλανδία	
Πακιστάν		Ινδονησία		
Συρία		Ιρλανδία		
Τυνησία		Ισραήλ		
Μπρουνέι		Αργεντινή	Θρησκευτικές προσωπικότητες/ ιδρύματα	
Αυστραλία		Πολωνία		
Κουβέιτ	Τυνησία			
Μεξικό	Κυβερνητικοί (διπλωματικοί)	Βοσνία-Ερζεγοβίνη	Τρομοκράτες	
Αυστρία		Λίβανο		
Ρωσία		Λεσότο		
Δημοκρατία της Σλοβακίας		Βέλγιο	Τουρίστες	
Μπαχρέιν		Μεταφορών (εκτός αεροπορίας)	Ρουμανία	
Σουδάν			Ηνωμένα Αραβικά Εμιράτα	
Βέλγιο			Βολιβία	
Βολιβία			Κούρδοι	Βίαια πολιτικά κόμματα
Αίγυπτος			Υεμένη	Αστυνομία

Πίνακας 19-19. Σχέση «καταγωγής» οργάνωσης με τύπο στόχου

Χώρα	Τύπος επίθεσης
Ιταλία	Κυβερνητικοί (γενικά)
Λίβανο	Στρατός
Δυτική Όχθη και Λωρίδα της Γάζας	Ιδιώτες & ιδιωτική περιουσία
Σαουδική Αραβία	
Αίγυπτος	
Ηνωμένες Πολιτείες	Θρησκευτικές προσωπικότητες/ ιδρύματα

20. Σύνοψη

Σε αυτή την ενότητα εξετάστηκε η ενσωμάτωση των Οντολογιών στην εξαγωγή Κανόνων Συσχετίσεων και συγκεκριμένα στον Αργιορί αλγόριθμο, από την πλευρά της αξίας των σημασιολογικών πλέον αποτελεσμάτων. Η βασική ιδιότητα που δανείστηκε από τις Οντολογίες είναι οι σημασιολογικές αποστάσεις που μπορούν να ορισθούν μεταξύ εννοιών και δείχνουν πόσο όμοιες είναι. Με αφορμή αυτή την ιδιότητα ορίστηκε μια νέα έννοια, το σημασιολογικό πλαίσιο, το οποίο είναι ένα σύνολο ζευγών εννοιών που η σημασιολογική τους απόσταση περιορίζεται μέσα σε ένα άνω και κάτω όριο. Η ενσωμάτωση των Οντολογιών γίνεται με τρεις διαφορετικούς τρόπους, δηλαδή κατά τη διάρκεια της εκτέλεσης του αλγορίθμου, στο τέλος του αλγορίθμου που έχουν παραχθεί οι κανόνες και τέλος κατά τη διάρκεια του αλγορίθμου αλλά αυτή την φορά ο αλγόριθμος αφηφά το *support* και ελέγχει μόνο την ισχύ του περιορισμού των σημασιολογικών αποστάσεων. Ο έλεγχος της μεθόδου που προτείνεται γίνεται μέσω πραγματικών απαιτήσεων και δεδομένων στο Πεδίο της Τρομοκρατίας.

Μέσω της περίπτωσης χρήσης παρουσιάζονται κάποιες αρχές- υποθέσεις που έχουν επιβεβαιωθεί από αναλυτές του Πεδίου της Τρομοκρατίας, ορίζεται και η διαδικασία ελέγχου του συστήματος, το οποίο θα πρέπει να εξαγάγει παρόμοιες ή ίδιες τάσεις. Από τις έξι υποθέσεις, αυτή που αναφερόταν στη συσχέτιση των Πολιτικών Γεγονότων με το είδος του Στόχου που επιλέγεται, δεν ήταν δυνατόν να ελεγχθεί καθώς δεν υπήρχαν δεδομένα για την έννοια Πολιτικό Γεγονός. Βέβαια, ενδιαφέρον παρουσιάζει ότι παρόλο που ο Chasdi δεν βρίσκει κάποια συσχέτιση, στην Οντολογία που χρησιμοποιείται οι δύο διαπραγματευόμενες έννοιες φαινόταν να έχουν υψηλή ομοιότητα.

Από τις υπόλοιπες πέντε υποθέσεις, στην υπόθεση 2, όπου αναφέρεται στη συσχέτιση της Πολιτικής Ιδεολογίας με τον Αριθμό των Νεκρών υπήρξε μία μικρή αστοχία, καθώς σε μία βασική έννοια, που κληρονομεί (<http://umbel.org/umbel/rc/Dying>) η αντιστοιχούμενη έννοια με τον Αριθμό των Νεκρών (*DeathOfTheVictim*), δεν βρέθηκε ο ακριβής ορισμός της. Αυτό είχε ως αποτέλεσμα να φέρνει τα ζεύγη με τη μεταβλητή *nkill* σε πολύ χαμηλή κατάταξη στην ταξινόμηση των σημασιολογικών αποστάσεων, οπότε τα ενδιαφερόμενα ζεύγη ήταν εκτός από το αρχικό σημασιολογικό πλαίσιο που είχε οριστεί. Βέβαια, μέσω της διαδικασίας επιβεβαιώθηκε ότι υπάρχει μία τάση που αποδεικνύει την συσχέτιση τους.

Οι υποθέσεις 1, 3, 4, τις οποίες υποστήριξε και ο Chasdi, όπου συσχετιζόταν η έννοια Πολιτική Ιδεολογία με την Επιλογή Συγκεκριμένου Στόχου, των Αριθμό των Νεκρών και τις Υλικές Ζημιές επιβεβαιώνεται με την παραγωγή συγκεκριμένων κανόνων.

Ιδιαίτερο ενδιαφέρον παρουσιάζεται για την υπόθεση 5, όπου υπενθυμίζεται ότι ο Chasdi την επιβεβαιώνει. Με την απλή ενσωμάτωση των Οντολογιών στον αλγόριθμο Αργιορί είτε κατά την διάρκεια εκτέλεσης του αλγορίθμου είτε στο τέλος του δεν δίνονται τα κατάλληλα αποτελέσματα, ώστε να επιβεβαιωθεί μια τέτοια τάση. Παρόλα αυτά αν ο αναλυτής επιλέξει να αφηφήσει το *support* και να δώσει έμφαση μόνο στο σημασιολογικό πλαίσιο, αυτή η τάση πράγματι επιβεβαιώνεται. Η ιδέα που κρύβεται πίσω από αυτό είναι ότι αντί να παράγονται ισχυρά στοιχειοσύνολα με βάση το *support*, θεωρείται ισχυρό ένα στοιχειοσύνολο που βρίσκεται μέσα σε ένα σημασιολογικό πλαίσιο, ενώ η παραγωγή ενός κανόνα εξακολουθεί να εξαρτάται από το βαθμό της εμπιστοσύνης του μέσα στα δεδομένα.

Άλλες δύο ενδιαφέρουσες παρατηρήσεις μπορούν να γίνουν για την υπόθεση 5. Αρχικά, εξετάζοντας το συγκεκριμένο σημασιολογικό πλαίσιο, νέοι κανόνες παρουσιάζονται που πιθανόν να είναι πραγματικά ενδιαφέροντες όσο και οι κανόνες που γίνεται προσπάθεια να επιβεβαιωθούν. Η δεύτερη παρατήρηση έχει να κάνει με την ιεραρχία των εννοιών. Όπως έχει επισημανθεί στη θεωρία υπάρχουν ειδικοί αλγόριθμοι που εξετάζουν την γενίκευση ή ειδίκευση των κανόνων με βάση την ιεραρχία των εννοιών στην Οντολογία. Κάτι τέτοιο μπορεί να γίνει και στην περίπτωση των σημασιολογικών πλαισίων. Για παράδειγμα η έννοια *GeographicalRegion* είναι γενικότερη από την έννοια *Country*, η πρώτη αναφέρεται σε μία περιοχή όπως Δυτική Ευρώπη, ενώ η δεύτερη σε μία χώρα όπως η Ελλάδα. Εξετάζοντας την σημασιολογική τους απόσταση σε σχέση με την έννοια *AttackOnObject* παρατηρείται ότι το πρώτο ζεύγος (*GeographicalRegion*, *AttackOnObject*) έχει σημασιολογική απόσταση 0.29 ενώ το δεύτερο (*Country*, *AttackOnObject*) 0,224 η οποία διαφορά εξηγείται λόγω της ιεραρχικής τους δομής. Επομένως, θα μπορούσαν να αναπαραχθούν γενικότεροι κανόνες αν το σημασιολογικό πλαίσιο περικλείει την πρώτη σημασιολογική απόσταση.

Από αυτή την ενότητα μπορούν να εξαχθούν και κάποιες επιπλέον γενικές παρατηρήσεις. Όπως το πόσο σημαντικό είναι να χρησιμοποιούνται καλά κατασκευασμένες Οντολογίες, που καλύπτουν τις ανάγκες της εκάστοτε περίπτωσης χρήσης. Προσοχή, θα πρέπει να δίνεται στο πως «απλώνονται» οι έννοιες μέσα στο χώρο της Οντολογίας και κατ' επέκταση στο να είναι τόσο αυστηρές- αναλυτικές όσο πρέπει. Επίσης, η ύπαρξη καλά ορισμένων συσχετίσεων δίνει τη δυνατότητα για πλούσιο συμπερασμό σε συνδυασμό με κατάλληλες Μηχανές Συλλογιστικής Ανάλυσης. Στην εφαρμογή, μεγάλο βάρος δίνεται στις σημασιολογικές αποστάσεις, οπότε οι Μετρικές που θα χρησιμοποιηθούν για τον υπολογισμό τους είναι μείζον σημασίας και πρέπει να καλύπτουν τόσο τις ανάγκες του σεναρίου χρήσης αλλά και τις ιδιαιτερότητες της Οντολογίας. Για παράδειγμα, η Μετρική TaxonomyMetrics που βασίζεται στους κοινούς απογόνους των εννοιών που εξετάζει δε θα πρέπει να χρησιμοποιείται για Οντολογίες που δεν έχουν πλούσια πληροφορία για τους απογόνους των εννοιών. Τέλος, πολύ σημαντικό κομμάτι για την εξαγωγή πλούσιων σημασιολογικά κανόνων είναι η χρήση κατάλληλων δεδομένων εκπαίδευσης και προφανώς να γίνει προσεχτική αντιστοίχιση των μεταβλητών τους με τις έννοιες της Οντολογίας. Όσο πιο σημασιολογικά συνεπή είναι τα δεδομένα μιας μεταβλητής, με την έννοια με την οποία γίνεται η αντιστοίχιση, τόσο πιο πλούσια και σημασιολογικά ορθά θα είναι τα αποτελέσματα. Έτσι, ανάλογα το σενάριο και τους στόχους που ορίζονται κάθε φορά, θα πρέπει να επιλέγεται η κατάλληλη Οντολογία, η Μηχανή Συλλογιστικής Ανάλυσης, η Μετρική, το Σημασιολογικό Πλαίσιο και το Αρχείο Αντιστοίχισης, επομένως διαφορετικός συνδυασμός μεταξύ αυτών λύνει διαφορετικά προβλήματα.

Σύμφωνα με τα αποτελέσματα της περίπτωσης χρήσης αποδείχθηκε η χρησιμότητα και η σημασιολογική εγκυρότητα των αποτελεσμάτων και για τις δύο λογικές εκτέλεσης της εφαρμογής. Με τον ορισμό του σημασιολογικού πλαισίου, μπορεί κάποιος να εξετάσει την συσχέτιση στοχευόμενων εννοιών, να εξαγει κανόνες μέσα σε ένα καθορισμένο εννοιολογικό πλαίσιο και να ελέγξει την γενικότητα των κανόνων. Μια επέκταση που φαίνεται δυνατή και ενδιαφέρον είναι να μπορούν να συσχετισθούν διάφορα σημασιολογικά πλαίσια μεταξύ τους. Τέλος, είναι σαφές ότι η εισαγωγή ίδιων τιμών στις παραμέτρους (*Ontology File, Mapping File, Reasoner, Metrics, Distance/ Similarity Timing*) του αλγορίθμου, δε θα επιλύει αναγκαστικά διαφορετικούς τύπους προβλημάτων.

Στον πίνακα που ακολουθεί παρουσιάζονται συνοπτικά οι διαφορές των τριών μεθόδων Apriori, oApriori, oAprioriWithoutSupport που έχουν συζητηθεί στην τρέχουσα και την προηγούμενη ενότητα.

Πίνακας 20-1. Απόδοση επεκτάσεων Apriori

Κατηγορία	Χαρακτηριστικά	Apriori	oApriori	oAprioriWithout Support
Απόδοση	Χρήση μνήμης		+πίνακα σημασιολογικών αποστάσεων	+πίνακα σημασιολογικών αποστάσεων
	Κόστος κατασκευής πίνακα αποστάσεων		1 φορά	1 φορά
	Κόστος ελέγχου	support/ confidence	support/ confidence/ σημασιολογικός έλεγχος στοιχειοσυνόλων/ σημασιολογικός έλεγχος κανόνων	confidence/ σημασιολογικός έλεγχος στοιχειοσυνόλων/ σημασιολογικός έλεγχος κανόνων
	Στοιχειοσύνολο		Μείωση στοιχειοσυνόλων	Μεγάλο σύνολο στοιχειοσυνόλων
	Κύκλοι εκτέλεσης		Μείωση	1
	Κανόνες	Πολύ	Λίγοι	Μέτρια
Σημασιολογική	Πλούσια σημ. ερμηνία	Χαμηλή	Υψηλή	Υψηλή
	Κατανόηση αποτελ.	Χαμηλή	Υψηλή	Υψηλή
	Στοχευόμενοι κανόνες	Όχι	Ναι	Ναι

Κατηγορία	Χαρακτηριστικά	Apriori	οApriori	οAprioriWithout Support
Σηματολογική	Ανάδειξη κανόνων με μικρό support/ confidence	Χαμηλή	Μέτρια	Υψηλή
	Τετριμμένη γνώση	Υψηλή (ισχυροί κανόνες μικρή σηματολογ.)	Μέτρια	Χαμηλή
	Λανθασμένη	Υψηλή	Χαμηλή	Χαμηλή
	Επανάληψη κανόνων	Υψηλή	Χαμηλή	Μέτρια
	Ενδιαφέρον κανόνες	Χαμηλή	Υψηλή/ ευελιξία μετρικών	Υψηλή/ ευελιξία μετρικών

21. Μελλοντική μελέτη και βελτιώσεις

Σε αυτή την ενότητα θα παρουσιαστούν περιληπτικά οι μελλοντικές μελέτες και βελτιώσεις που μπορούν να πραγματοποιηθούν ως επεκτάσεις της παρούσας διατριβής. Οι βελτιώσεις ακολουθούν τη δομή της διπλωματικής και είναι χωρισμένες σε τρία κομμάτια, τις επεκτάσεις θεωρίας, επεκτάσεις εφαρμογής και επεκτάσεις πεδίου. Έτσι για το εκάστοτε κομμάτι μπορεί να πραγματοποιηθούν οι παρακάτω βελτιώσεις:

Επεκτάσεις Θεωρίας:

- Εκτεταμένη μελέτη σε άλλες τεχνικές DM εκτός των κανόνων συσχετίσεων και κυρίως στις τεχνικές χωρικής- χρονικής γνώσης.
- Στην εξέταση Οντολογιών χωρικής και χρονικής γνώσης.
- Επιπρόσθετα πιο περίπλοκα μέτρα που συνδυάζουν συγκεκριμένες διαδικασίες DM και τις Οντολογίες.

Επεκτάσεις Εφαρμογής:

- Να συμπεριληφθούν νέες τεχνικές DM.
- Εισαγωγή διεπαφών που να βοηθούν τον χρήστη να διαχειρίζεται τα Mapping File και να μπορεί να βλέπει τα αποτελέσματα των σηματολογικών αποστάσεων για συγκεκριμένο σύνολο εννοιών, Οντολογίας, μετρικής και μηχανής συλλογιστικής ανάλυσης.
- Η εφαρμογή να εκμεταλλεύεται πλήρως τις βάσεις γνώσης και τα δεδομένα που περικλείει, καθώς προς το παρόν εξετάζει μόνο τη δομή των Οντολογιών.
- Επιβεβαίωση υπόθεσης σηματολογικού πλαισίου.
- Να είναι εφικτή η εισαγωγή πολλαπλών μετρικών σηματολογικών αποστάσεων παραδείγματος χάρι χώρου και χρόνου για χωρικό- χρονικές Οντολογίες.
- Να εισάγονται και να λαμβάνονται υπόψη περισσότερα σηματολογικά πλαίσια.
- Δημιουργία νέων μετρικών.

Επεκτάσεις Πεδίου:

- Εξέταση πιο περίπλοκων αρχών ή υποθέσεων με την βοήθεια ειδικών.
- Επέκταση δεδομένων εκπαίδευσης, περισσότερες μεταβλητές και δεδομένα.
- Επέκταση Οντολογίας που να αναπαριστά καλύτερα το πεδίο εφαρμογής.

22. Επίλογος

Στην παρούσα διπλωματική παρουσιάστηκαν οι διαφορετικές προσεγγίσεις σχετικά με την αναπαράσταση της γνώσης ενώ δόθηκε μεγαλύτερο βάρος στην προσέγγιση των Οντολογιών και συγκεκριμένα στην γλώσσα αναπαράστασης OWL. Επίσης, παρουσιάστηκαν άλλα δύο συνοδευτικά εργαλεία, αυτά των Μηχανών Συλλογιστικής Ανάλυσης και των Μετρικών αποστάσεων και ομοιότητας που είναι αναπόσπαστα κομμάτια με τις Οντολογίες. Έπειτα, εξετάστηκε η ανάγκη της ενσωμάτωσης των Οντολογιών στις τεχνικές της Εξόρυξης

Γνώσης και υλοποιήθηκε η ενσωμάτωσή τους στην τεχνική των Κανόνων Συσχετίσεων και συγκεκριμένα στον αλγόριθμο Arriogi. Στον αλγόριθμο αυτό έγινε ενσωμάτωση για την αξιολόγηση των παραγόμενων προτύπων κατά την διάρκεια της εκτέλεσής του και μετά το πέρας του. Ενδιαφέρον παρουσίασε και η προσπάθεια για εμπλουτισμό του αλγορίθμου Arriogi με μεγαλύτερη σημασιολογία απομονώνοντας και μη χρησιμοποιώντας τη μετρική της εμπιστοσύνης, αλλά θέτοντας στην θέση του σημασιολογικές μετρικές ορισμένες πάνω στις Οντολογίες. Τέλος, έγινε σημασιολογικός έλεγχος και έλεγχος απόδοσης των επεκτάσεων του Arriogi πάνω σε μία πραγματική περίπτωση χρήσης, αυτής της τρομοκρατίας. Εδώ εξετάστηκε η αναγκαιότητα της αξιοποίησης στο συγκεκριμένο πεδίο γνώσης, των Οντολογιών και της Εξόρυξης Γνώσης και αναδείχθηκαν τα θετικά που μπορούν να παραχθούν. Ενώ παράλληλα αποδείχθηκαν έξι βασικές αρχές που είχαν οριστεί από ειδικούς της τρομοκρατίας και υποστηρίζονται από απλές στατιστικές προσεγγίσεις.

Αφού αποδείχθηκε επιτυχώς η χρησιμότητα της ενσωμάτωσης των Οντολογιών στην τεχνική των Κανόνων Συσχετίσεων παρατηρήθηκαν τέσσερα σημεία στα πρέπει κάποιος να εντρυφήσει για την επιτυχημένη εφαρμογή αυτής της προσέγγισης. Η πρώτη είναι η αναζήτηση και χρησιμοποίηση κατάλληλων Οντολογιών που αναπαριστούν πλήρως ένα πεδίο γνώσης ή κατ' ελάχιστον ανταποκρίνονται στις απαιτήσεις των ειδικών. Εδώ μεγάλη προσοχή πρέπει να δίνεται στον συνδυασμό της εκφραστικότητα της Οντολογίας και της πολυπλοκότητάς της, καθώς έχουν μια γραμμική συσχέτιση. Για παράδειγμα, η αυξημένη εκφραστικότητα συνεπάγεται μεγάλη πολυπλοκότητα και μπορεί να οδηγήσει σε αδυναμία απόκρισης. Το δεύτερο σημείο είναι η χρήση κατάλληλων Μηχανών Συλλογιστικής Ανάλυσης που θα είναι σε θέση να εκμεταλλευτούν την εκφραστικότητα των Οντολογιών, αλλά και να βελτιώσουν την πολυπλοκότητα του ελέγχου και την αναζήτησης γνώσης μέσα στις Οντολογίες. Το τρίτο σημείο έχει να κάνει με τις μετρικές και την σημαντικότητά τους καθώς είναι αυτές που ουσιαστικά μπορούν να ερμηνεύσουν κατά μία έννοια την γνώση μέσα στις Οντολογίες. Όσο καλύτερες μετρικές εφαρμόζονται, τόσο καλύτερα συμπεράσματα μπορούν να εξαχθούν από τις συσχετίσεις των εννοιών. Τέλος, το τέταρτο σημείο είναι η χρήση εκείνων των τεχνικών της Εξόρυξης γνώσης που αντανακλούν τους στόχους των ειδικών. Φυσικά, εδώ άλλα δύο σημεία πρέπει κάποιος να εξετάζει με προσοχή. Πρώτα, το σημείο στο οποίο θα ενσωματωθεί η γνώση των Οντολογιών και έπειτα την αναγκαιότητα συνεργασίας με άλλες αντικειμενικές, υποκειμενικές ή σημασιολογικές μετρικές για να επιτευχθεί η εξαγωγή προτύπων με αυξημένο ενδιαφέρον.

Γλωσσάρι

ΣΥΜΠΤΥΞΗ ΑΓΓΛΙΚΑ	ΑΓΓΛΙΚΑ	ΣΥΜΠΤΥΞΗ ΕΛΛΗΝΙΚΑ	ΕΛΛΗΝΙΚΑ
-	Abduction	-	Απαγωγή
-	Agent	-	Πράκτορας
-	Alignment Ontology	-	Ευθυγράμμιση Οντολογιών
-	Anomaly Detection	-	Διαπίστωση Ανωμαλιών
-	Applicability	-	Δυνατότητα Εφαρμογής
-	Application Ontology	-	Οντολογία Εφαρμογής
AI	Artificial Intelligent	TN	Τεχνητής Νοημοσύνης
ABox	Assertional Box	-	Σώμα Ισχυρισμών
AR	Association Rules	-	Κανόνες Συσχέτισης
-	Attributes	-	Ιδιότητες
-	Axiom	-	Αξίωμα
-	Backward chaining	-	Προς τα Πίσω Σύνδεση
BFO	Basic Formal Ontology	-	-
-	Cardinality Restriction	-	Περιορισμός Ποσοτικοποίησης
-	Catalog	-	Κατάλογος
CCO	Cell-Cycle Ontology	-	-
-	Clarity	-	Σαφήνεια
-	Class	-	Κλάσεις
-	Class Expression Satisfiability	-	Ικανοποιησιμότητα Έκφρασης Κλάσης
-	Class Expression Subsumption	-	Υπαγωγή Έκφρασης Κλάσης
-	Classification	-	Κατηγοριοποίηση, Ταξινόμηση
-	Clausal Form of Logic	-	Διαζευκτική Μορφή της Λογικής
-	Clustering	-	Συσταδοποίηση
-	Cognitive Knowledge	-	Γνωσιακή Επιστήμη
-	Coherence	-	Συνοχή
CWM	Common Warehouse Model	-	-
-	Commonsense knowledge	-	Γνώση Κοινής Λογικής
-	Concept	-	Έννοια
-	Conceptual Dependency	-	Εννοιολογική Εξάρτηση
-	Conciseness	-	Περιεκτικότητα
-	Confidence	-	Εμπιστοσύνη
-	Conjunctive Query Answering	-	Απάντηση Συζευγμένων Ερωτημάτων
CICQ	Consortium on Discovering Knowledge with Inductive Queries	-	-
-	Controlled Vocabulary	-	Ελεγχόμενο Λεξιλόγιο
-	Data Driven	-	Οδηγούμενα από Δεδομένα
-	Data Intelligence	-	Νοημοσύνη από τα δεδομένα
DM	Data Mining	-	Εξόρυξη Γνώση από Δεδομένα
-	Data- Model	-	Μοντέλο Δεδομένων
DBMS	Database Management System	-	Συστημάτων Διαχείρισης Δεδομένων
-	Decideability	-	Αποκρισιμότητα
-	Decision Support System	-	Συστημάτων Υποστήριξης Αποφάσεων
-	Declarative knowledge	-	Δηλωτική Γνώση
-	Deduction	-	Παραγωγή

DARPA	Defense Advanced Research Projects Agency	-	-
-	Density	-	Πυκνότητα
-	Depth	-	Βάθος
DL	Description Logic	ΠΛ	Περιγραφική Λογική
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering	-	-
-	Diversity	-	Ποικιλομορφία
-	Domain	-	Πεδίο Ορισμού
-	Domain Intelligence	-	Νοημοσύνη από το Πεδίο
-	Domain Knowledge	-	Πεδίο Γνώσης
-	Domain Specific Ontology	-	Οντολογίες Εξειδικευμένες σε ένα Πεδίο
e-BMO	e-Business Model Ontology	-	-
-	Enumeration	-	Απαρίθμηση
-	Events	-	Γεγονότα
∃	Existential Quantifier	∃	Υπαρξιακός Ποσοδείκτης
-	Expert systems	-	Ευφυή συστήματα
-	Expressiveness	-	Εκφραστικότητα
-	Extendibility	-	Επεκτασιμότητα
XML	Extensible Markup Language	-	-
XMLs	Extensible Markup Language schema	-	-
-	Facets	-	Εκφάνσεις
-	Facts	-	Γεγονότα
-	Formal Taxonomy	-	Τυπική Ταξινόμια
-	Forward Chaining	-	Πρόσθια Σύνδεση
-	Frame	-	Πλαίσιο
F-Logic	Frame Logic	-	Λογική Πλαισίων
-	Function	-	Συνάρτηση
-	Functional Property	-	Συναρτησιακή Ιδιότητα
-	Fuzzy Logic	-	Ασαφής Λογική
-	General Logical Constraints	-	Περιορισμοί Γενικής Λογικής
-	Generality	-	Γενικότητα
GTD	Global Terrorism Database	-	-
-	Glossary	-	Γλωσσάρι
-	Heuristic knowledge	-	Ευρετική Γνώση
-	Highly Informal	-	Υψηλά Άτυπη
-	Human Knowledge	-	Ανθρώπινης Γνώσης
HTML	HyperText Markup Language	-	-
-	Individual Axioms	-	Αξιώματα Ατόμων
-	Induction	-	Επαγωγή
-	Inductive Databases Approach	-	Βάσεις Επαγωγικής Προσέγγισης
-	Inexact and uncertain knowledge	-	Ανακριβής και Αβέβαιη Γνώση
-	Inference Mechanisms	-	Μηχανισμούς Εξαγωγής Συμπερασμάτων
-	Inference mechanisms	-	Μηχανισμοί εξαγωγής συμπερασμάτων
-	Inferenceability	-	Ικανότητα συμπερασμού
-	Informal Taxonomy	-	Άτυπη Ταξινόμια
IR	Information Retrieval	-	Ανάκτηση Πληροφοριών
-	Instance	-	Στιγμιότυπο
-	Instance Checking	-	Έλεγχος Στιγμιότυπου
-	Interestingness Determination	-	Αποφασιστικότητα

			Ενδιαφέροντος
-	Interestingness Measures	-	Μέτρα Ενδιαφέροντος
IRI	Internationalized Resource Identifiers		Διεθνοποιημένοι Προσδιοριστές Πόρων
-	Interoperability	-	Διαλειτουργικότητα
-	Intersection	-	Τομή
-	Inverse Property	-	Ιδιότητα Αντιστροφής
-	Item Set	-	Στοιχειοσύνολο
JDMAPI	Java Data Mining API	-	-
KB	Knowledge Base	-	Βάση Γνώσης
-	Knowledge Based Systems	-	Βασισμένα στη Γνώση Συστήματα
KDD	Knowledge Discovery of Data	-	Ανακάλυψη Γνώσης σε Βάσεις Δεδομένα
KIF	Knowledge Interchange Format	-	-
KM	Knowledge Management	-	Διαχείριση Γνώσης
KR	Knowledge Representation	-	Αναπαράσταση της Γνώσης
KR	Knowledge Representation	ΑΓ	Αναπαράσταση Γνώσης
KSE	Knowledge Sharing Effort	-	-
-	Link Analysis	-	Ανάλυση Συνδέσεων
-	Logic	-	Λογική
-	Mapping	-	Χαρτογράφηση
-	Markup Language	-	Γλώσσα Σήμανση
-	Merging Ontology	-	Συγχώνευση Οντολογιών
-	Meta- Ontology	-	Μετα- Οντολογία
Mt	Microtheories	-	Μικροθεωρίες
-	Middle Level Ontology	-	Οντολογία Μεσαίου Επιπέδου
MILO	Mid-Level Ontology	-	Οντολογία Μεσαίου Επιπέδου της SUMO
-	Multiple Inheritance	-	Πολλαπλή Κληρονομικότητα
-	Network and web Intelligence	-	Νοημοσύνη από τα Δίκτυα
-	Neural Network	-	Νευρωνικά Δίκτυα
-	No Symbolic Methods	-	Μη συμβολικές Μέθοδοι
-	Novelty	-	Καινοτομία
-	Object	-	Αντικείμενο
-	Objective Measures	-	Αντικειμενικά Μέτρα
-	Objects	-	Αντικείμενα
-	Ontological Engineering	-	Οντολογικής Μηχανικής
-	Ontological Knowledge	-	Οντολογική Γνώση
-	Ontology	-	Οντολογία
-	Ontology Consistency	-	Έλεγχος Συνέπειας Οντολογίας
-	Ontology Evaluation	-	Αξιολόγηση Οντολογιών
-	Ontology Evolution	-	Εξέλιξη Οντολογιών
-	Ontology Reengineering	-	Ανασχεδιασμό Οντολογιών
-	Ontology Versioning	-	Υποστήριξη Πολλαπλών Εκδόσεων Οντολογιών
-	Organizational Intelligence	-	Οργανωτική νοημοσύνη
-	Override	-	Υποσκέλιση Μεθόδου
OWL	Web Ontology Language		
-	Parts	-	Μέρη
-	Pattern	-	Πρότυπο
PBMS	Pattern Based Management Systems	-	Συστήματα Διαχείρισης Βάσεων Προτύπων
-	Pattern Management	-	Διαχείριση Προτύπων
PANDA	Patterns for Next-generation	-	-

	Database Systems		
-	Peculiarity	-	Ιδιαιτερότητα
-	Predicate Logic	-	Κατηγορηματική Λογική
PMML	Predictive Model Markup Language	-	-
-	Probability Networks	-	Δίκτυα Πιθανοτήτων
-	Procedural knowledge	-	Διαδικαστική Γνώση
-	Procedure	-	Διαδικασίες
-	Properties	-	Ιδιότητες
-	Property Restrictions	-	Περιορισμοί ιδιοτήτων
-	Propositional Logic	-	Προτασιακή Λογική
-	Pruning	-	Κλάδεμα
-	Punning	-	Τεχνική της Παρονομασίας
-	Query Language	-	Γλώσσα Ερωτήσεων
-	Range	-	Πεδίο Τιμών
RUP	Rational Unified Process	-	Αντικειμενοστραφές Μοντέλο
-	Realization	-	Πραγμάτωση
-	Reasoner	-	Μηχανές Συλλογιστικής Ανάλυσης
-	Reasoning	-	Συλλογιστική
-	Relation	-	Σχέση
-	Reliability	-	Αξιοπιστία
-	Resource	-	Πόρος
RDF	Resource Description Framework	-	-
RDFs	Resource Description Framework schema	-	-
-	Restriction	-	Περιορισμός
-	Retrieval	-	Ανάκτηση
-	Reusability	-	Επαναχρησιμοποίηση
-	Rigorously Formal	-	Αυστηρά Τυπική
-	Role	-	Ρόλοι
-	Role Restriction	-	Περιορισμός Ρόλου
-	Rules	-	Κανόνες
-	Scripts	-	Σενάρια
-	Semantic Distance	-	Σημασιολογική Απόσταση
-	Semantic Networks	-	Σημασιολογικά Δίκτυα
-	Semantic Relatedness	-	Σημασιολογική Συνάφεια
-	Semantic Similarity	-	Σημασιολογική Ομοιότητα
-	Semantic Web	-	Σημασιολογικός Ιστός
SDW	Semantic Web Document	-	Αρχεία του Σημασιολογικού Ιστού
-	Semantics-based Measures	-	Μέτρα Βασισμένα σε Σημασιολογία
-	Semi- Formal	-	Ημι- Τυπική
-	Semi- Informal	-	Ημι- Άτυπη
-	Semi-Structured Approach	-	Ημι-Δομημένη Προσέγγιση
-	Slots	-	Σχισμές
-	Social Intelligence	-	κοινωνική νοημοσύνη
-	Spatial Data Mining)	-	Εξόρυξη Χωρικής Γνώσης
-	String	-	Αλφαριθμητικό
-	Structural Knowledge	-	Δομημένη Γνώση
-	Subjective Measures	-	Υποκειμενικά Μετρά
-	Subsumption	-	Υπαγωγή
SUMO	Suggested Upper Merged Ontology	-	-
-	Support	-	Υποστήριξη

-	Surprisingness	-	Απροσδόκητο
-	Symmetric Property	-	Συμμετρική Ιδιότητα
-	Tableau	-	Ταμπλό
-	Task Ontology	-	Οντολογίες Διεργασιών
-	Taxonomies	-	Ταξινόμια
-	Temporal Data Mining)	-	Εξόρυξης Χρονικής Γνώσης
-	Temporal Logic	-	Χρονική Λογική
-	Temporal Representation	-	Αναπαράσταση του Χρόνου
TBox	Terminological Box	-	Σώμα Ορολογίας
TIA	Terrorism Information Awareness or Total Information Awareness	-	-
TKB	Terrorist Knowledge Base	-	Βάση Γνώση για την Τρομοκρατία
-	Terrorist Network	-	Τρομοκρατικό Δίκτυο
-	Thesaurus	-	Θησαυρός Γνώσης
-	Time Interval Logic	-	Λογική Χρονικών Διαστημάτων
-	Top-Level Ontology	-	Οντολογιών Υψηλού Επιπέδου
-	Transitive Property	-	Ιδιότητα Μεταβατικότητας
-	Uncertainty and Fuzzy Knowledge	-	Αβέβαιης και Ασαφής Γνώσης
-	Unexpectedness	-	Μη αναμενόμενο
UML	Unified Modeling Language	-	-
URL	Uniform/ Universal Resource Locator	-	Ενιαίος/ Καθολικός Εντοπιστής Πόρου
-	Union	-	Ένωση
▼	Universal Quantifier	▼	Καθολικός Ποσοδείκτης
URI	Universal Resource Locator	-	Παγκόσμιο Προσδιοριστή
-	Upper Mapping and Binding Exchange Layer	-	-
-	Upper Ontology	-	Οντολογιών Υψηλού Επιπέδου
-	Utility	-	Χρησιμότητα
-	Value	-	Τιμές
-	Value Restrictions	-	Περιορισμοί Τιμών
-	Visibility	-	Ορατότητα
-	Width	-	Πλάτος

Βιβλιογραφία

- Agrawal R, Imielinski T, Swami A (1993) Mining Association Rules Between Sets of Items in Large Databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216, Washington, D.C.
- Agrawal R, Srikant R (1994) Fast Algorithms for Mining Association Rules in Large Databases. *Proceedings of the Twentieth International Conference on Very Large Databases*, pp. 487-499, Santiago, Chile
- AKTiveSAOntology.owl, Ontology (2011) Available: <http://www.zaltys.net/ontology/AKTiveSAOntology.owl> Accessed 5 February 2011
- ANSI/NISO Z39.19 – 2003 *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*, Niso Press, Bethesda, Maryland, U.S.A.
- Antoniou G, Harmelen F, (2004) *A Semantic Web Primer*. The MIT Press, Cambridge, MA, USA
- Batet M, Valls A, Gibert K (2008) Improving classical clustering with ontologies, *In Proc. of IASC08*, Japan
- Bellandi A, Furlotti B, Grossi V, Romei A (2007) Pushing Constraints in Association Rules Mining: An Ontology-based Approach, *In Proceedings of the International Conference WWW/Internet 2007*, Volume 1, pp. 179-186, Vila Real, Portugal
- Berkeley DB XML release 2.2.13 <http://www.oracle.com/technology/software/products/berkeley-db/xml/index.html> Accessed 4 April 2011
- Berners-Lee T, Hendler J, Lassila O (2001) The Semantic Web. *Scientific American*.
- BFO, Basic Formal Ontology (BFO), Upper Level Ontology (2011) <http://www.ifomis.org/bfo> Accessed 5 February 2011
- Booch G, Rumbaugh J, Jacobson I (2004) *The Unified Modeling Language User Guide*, Addison – Wesley
- Bouza A, Reif G, Bernstein A, Gall H (2008) Semtree: Ontology-based decision tree algorithm for recommender systems, *in Proceedings of the Seventh International Semantic Web Conference (ISWC2008)*
- Brachman R (1992) What is knowledge representation, and where is it going?, *Lecture Notes in Computer Science*, Volume 653/1992, pp. 187-203, doi: 10.1007/3-540-56320-2
- Britanica, online encyclopedia (2011) <http://www.britannica.com/EBchecked/topic/588371/-terrorism> Accessed 17 May 2011
- Budanitsky A, Hirst G (2006) Evaluating wordnet-based measures of semantic distance, *Computational Linguistics* 32(1), pp. 13–47
- Cai H, Fu W, Cheng H, Kwong W (1998) Mining association rules with weighted items. *In Proceedings of the International Database Engineering and Applications Symposium (IDEAS '98)*. Cardiff, UK. 68–77
- Cao L, Yu P S, Zhang C, Zhao Y (2010) Ubiquitous Intelligence. In: *Domain Driven Data Mining*, Springer DOI 10.1007/978-1-4419-5737-5
- Carter L, Hamilton J, Cercone N (1997) Share-Based measures for itemsets. *In Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery (PKDD '97)*. Trondheim, Norway. 14–24.
- Catania A, Maddalena A (2006) Pattern Management: Practice and Challenges. *Processing and Managing Complex Data for Decision Support*. Idea Group Publishing, pp. 280–317
- CCO, Cell-Cycle Ontology (2011) <http://www.semantic-systems-biology.org/cco/> Accessed 5 February 2011
- Chandrasekaran B, Josephson J, Benjamins R (1999) What Are Ontologies, and Why Do We Need Them? *IEEE Intelligent Systems*, v.14 n.1, p.20-26, doi>10.1109/5254.747902
- Chasdi R (1997) Middle East Terrorism 1968-1993: An Empirical Analysis of Terrorist Group-Type Behavior, *Journal of Conflict Studies*, Volume 17, issue 2
- Chen M S, Han J, Yu P (1996) Data Mining: An Overview from a Database Perspective, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883

- Chen X, Zhou X, Scherl R, Geller J (2003) Using an interest ontology for improved support in rule mining. In *DaWaK 2003*. pp. 320-329.
- Chmielewski M, Gałka A, Jarema P, Krasowski K, Kosiński A (2009) Semantic Knowledge Representation in Terrorist Threat Analysis for Crisis Management Systems, In: *Nguyen N T* (ed.) et al., *ICCCI 2009, LNCS (LNAI)*, vol. 5796, pp. 460-471, Springer, Heidelberg
- CINQ, Consortium on Discovering Knowledge with Inductive Queries (2011) project homepage. <http://www.cinq-project.org> Accessed 5 February 2011
- Corcho O, Fernández-López M, Gómez-Pérez A (2003) Methodologies, tools and languages for building ontologies: where is their meeting point? *Data & Knowledge Engineering*, 46(1):41–64
- Corcho O, Fernandez-Lopez M, Gómez-Pérez A, Vicente O (2002). WebODE: An integrated workbench for ontology representation, reasoning and exchange. In *Lecture notes on artificial intelligence: Vol. 2473. 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW'02)* (pp. 138-153). Springer-Verlag.
- Corcho O, Fernández-López, M, Gómez-Pérez A (2007) *Ontological Engineering: What are Ontologies and How Can We Build Them?* Semantic Web Services. Premier Reference Source, Nueva York, pp. 44-70.
- CWM, Common Warehouse Model (2011) project homepage. <http://www.omg.org/cwm> Accessed 5 February 2011
- Davis R , Shrobe H, Szolovits P (1993) What is knowledge Representation?, *AI Magazine* 14(1), σσ. 17-33
- Dean M, Schreider G (2004) OWL Web Ontology Language Reference. *W3C Recommendation. World Wide Web Consortium (W3C)* Available: <http://www.w3.org/TR/owl-ref/>
- DeRosa M (2004) Data Mining and Data Analysis for. Counter Terrorism. *Center for Strategic and. International Studies, CSIS Press*
- Devedzic V (2002) Understanding Ontological Engineering. *Communications of the ACM*, 45(4):136-144
- Ding L, Kolari P, Ding Z, Avancha S (2006) Using Ontologies in the Semantic Web: A Survey. In R Sharman, R Kishore, & R Ramesh (Eds.), *Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems* (pp. 79–114), Berlin: Springer
- Ding L, Kolari P, Ding Z, Avancha S, Finin T, Joshi A (2004) Swoogle: A Search and Metadata Engine for the Semantic Web. *Thirteenth ACM Conference on Information and Knowledge Management*
- DOLCE, Descriptive Ontology fro Linguistic and Cognitive Engineering, Upper Level Ontology (2011), <http://www.loa-cnr.it/DOLCE.html> Accessed 5 February 2011
- Dragan G, Dragan D, Vladan D (2009) Knowledge Representation, *Model Driven Engineering and Ontology Development*, Berlin Heidelberg: Springer-Verlag, Part 1, pp. 3-43
- Durkin J (1994). Expert systems: design and development, London, Prentice Hall International, In: Dragan G, Dragan D, Vladan D (2009) Knowledge Representation, Model Driven Engineering and Ontology Development, Berlin Heidelberg: Springer-Verlag, Part 1, pp. 3-43
- Euzenat J, Shvaiko P (2007) Ontology matching. *Springer-Verlag*, 978-3-540-49611-3.
- FaCT++ Documentation Last modified: 2005, <http://owl.man.ac.uk/documentation.shtml> Accessed 5 February 2011
- Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery, *AI Magazine* 17(3) 37-54.
- Fensel D, et al. (2001) OIL: an ontology infrastructure for the semantic web. *IEEE Intelligent Systems & Their Applications*, vol. 16, no. 2, pp. 38–45
- Ferraz N, Garcia B (2008) Ontology in Association Rules Pre-Processing And Post-Processing, In *Proceedings of IADIS European Conf. Data Mining*, pp. 87-91
- Floyd C, Ukena S (2005) On Designing Situated Ontologies for Knowledge Sharing in Communities of Practice, *Workshop on Philosophical Foundations of Information Systems Engineering, Proceedings of the CAISE.05 Workshops*, Vol. 2

- Freitas, A.A. (1999) On rule interestingness measures, *Knowledge-Based Systems*, Vol. 12. Number 5-6. pp. 309-315, October 1999. Elsevier
- Fridman-Noy N, Hafner C D (1997) The State of the Art in Ontology Design: A Survey and Comparative Review. *AI Magazine* 18(3): 53—74
- García C, Ferraz I, Vivacqua S (2009) From data to knowledge mining, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* 23 (4), pp. 427-441
- García J, García-Peñalvo F, Therón R (2010) A Survey on Ontology Metrics, *Knowledge Management, Information Systems, E-Learning, and Sustainability Research Communications in Computer and Information Science*, vol. 111, pp.22-27, DOI: 10.1007/978-3-642-16318-0_4
- Genesereth M (1991) Knowledge Interchange Format. Principles of Knowledge Representation and Reasoning, *Proceedings of the Second International Conference*, Cambridge, MA, pages 599-600. Morgan Kaufmann
- Geng L, Hamilton H (2006) Interestingness measures for data mining: A survey, *ACM Computing Surveys*, 38(3):9
- Global Terrorism Database (2011), START, Available: <http://www.start.umd.edu/start/> Accessed 5 February 2011
- Golbreich C, Wallace K E (2009) OWL 2 Web Ontology Language New Features and Rationale, W3C Recommendation, available: http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/#Features_.26_Rationale Accessed 5 June 2011
- Gómez-Pérez A (2001) Evaluation of ontologies, *International Journal of Intelligent Systems* 16 (3) 10:1–10: 36.
- Gómez-Pérez A, Fernandez-Lopez M, Corcho O (2005) Ontological Engineering, (Eds) Springer, ISBN: 1852335513,p.403
- Gómez-Pérez A, Fernández-López M, Vicente A (1996) Towards a Method to Conceptualize Domain Ontologies, *ECAI96 Workshop on Ontological Engineering*, Budapest, pp. 41–51.
- Gómez-Pérez A, Rojas M (1999) Ontological reengineering and reuse, in: D. Fensel, R. Studer (Eds.), *11th European Workshop on Knowledge Acquisition, Modeling and Management (EKAW_99)*, *Lecture Notes in Artificial Intelligence*, vol. 1621, Springer, Berlin, pp. 139–156.
- Gómez-Pérez A (1999) Ontological engineering: A state of the art, *Expert Update* 2.3
- Gruber T (1993) Toward principles for the design of ontologies used for knowledge sharing, *Padua workshop on Formal Ontology*
- Gruenwald L, McNutt G, Mercier A (2003) Using an Ontology to Improve Search in a Terrorism Database System, in *Proc. DEXA Workshops* pp.753-757
- Gruenwald L, McNutt G, Mercier A (2003) Using an Ontology to Improve Search in a Terrorism Database System. *dexa, 14th International Workshop on Database and Expert Systems Applications (DEXA'03)*, pp.753
- Guarino N (1998) Formal Ontology and Information Systems, In N. Guarino (ed.). *Formal Ontology in Information Systems. Proceedings of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.
- Guarino N, Giaretta P (1995) Ontologies and knowledge bases: Towards a terminological clarification. In Mars, N., editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995*, pages 25-32, Amsterdam, NL. IOS Press.
- Harmelen F, Patel-Schneider P, Horrocks I (2001) Reference Description of the DAML+OIL (March 2001) Ontology Markup Language, Technical report. Available: <http://www.daml.org/2001/03/reference.html>
- Heflin J, Hendler J (2000) Dynamic Ontologies on the Web. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, AAAI/MIT Press, Menlo Park, CA, pp. 443-449
- Hermit (2011) Available: <http://hermit-reasoner.com/> Accessed 5 February 2011
- Hitzler P, Krötzsch M, Parsia B, Patel-Schneider F P, Rudolph S (2009) OWL 2 Web Ontology Language Primer, W3C Recommendation, available: http://www.w3.org/TR/2009/REC-owl2-primer-20091027/#What_is_OWL_2.3F Accessed 5 June 2011

- Hjørland B, Albrechtsen H (1995) Toward A New Horizon in Information Science: Domain Analysis. *Journal of the American Society for Information Science*, 46(6), p.p. 400-425.
- HorrIDGE M, Drummond N, Jupp S, Moulton G, Stevens R (2009) A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.2. Technical report, The University Of Manchester
- HorrIDGE M, Patel-Schneider F P (2009) OWL 2 Web Ontology Language Manchester Syntax, *W3C Recommendation*, available: <http://www.w3.org/TR/2009/NOTE-owl2-manchester-syntax-20091027/> Accessed 5 June 2011
- Hotho A, Maedche A, Staab S (2001) Ontology-based Text Document Clustering. *Künstliche Intelligenz* 16(4), pp. 48-54
- Hotho A, Maedche A, Staab S, Zacharias V (2002) On knowledgeable unsupervised text mining. In *Proceedings of the DaimlerChrysler Workshop on Text Mining*, Ulm, Springer.
- Huang J, Dou D, He J, Dang J, Hayes P (2010) Ontology-Based Knowledge Discovery and Sharing in Bioinformatics and Medical Informatics: A Brief Survey. *Proc. 7th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD-10)*, Yantai, China
- Hwang S (2004), Using Formal Ontology for Integrated Spatial Data Mining, In: *Laganà A et al (ed) Computational Science and Its Applications – ICCSA2004, LNCS Vol. 3044*, Springer-Verlag, pp. 1026-1035
- IEEE P1600.1, Standard Upper Ontology Working Group (SUO WG) Home Page <http://suo.ieee.org/index.html> Accessed 5 February 2011
- Inyaem U, Meesad P, Haruechaiyasak C, Tran D, (2009) Ontology-Based Terrorism Event Extraction. 2009 *First IEEE International Conference on Information Science and Engineering*, Nanjing, Jiangsu, China, 18-20/12/09, 912-915
- JDMAPI, Java Data Mining API (2011) project homepage. <http://www.jcp.org/jsr/detail/73.prt> Accessed 5 February 2011
- Jena Documentation, Last modified: 2010, <http://openjena.org/documentation.html> Accessed 5 February 2011
- Jing L, Zhou L, Ng M, Huang J (2006) Ontology-based Distance Measure for Text Clustering, *SIAM Text Mining 2006 Workshop*
- Kaon2 Documentation, Last modified: 2011, <http://kaon2.semanticweb.org/#documentation> Accessed 5 February 2011
- Kharsikar S, Mugler D, Sheffer D, Moore F, Duan Z (2007) A Weighted k-Nearest Neighbor Method for Gene Ontology Based Protein Function Prediction. *IMSCCS 07 Proceedings, IEEE Computer Society, USA. 2: 25-31*
- Kifer M, Lausen G, Wu J (1995) Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, vol. 42, no. 4, pp. 741–843.
- Klein M, Fensel D (2001) Ontology versioning on the Semantic Web, *First International Semantic Web Workshop(SWWS01)*, Stanford
- KNIME, DM tool (2011) <http://www.knime.org/> Accessed 5 February 2011
- Knowledge Interchange Format (KIF) Documentation <http://www.ksl.stanford.edu/knowledge-sharing/kif/> Accessed 5 February 2011
- Kotsifakos E, Marketos G, Theodoridis Y (2007) A framework for integrating ontologies and pattern-bases. In Nigro, H.O., Cisaró, S.G. and Xodo, D. (eds), *Data Mining with Ontologies: Implementations, Findings, and Frameworks*. Idea Group Inc., Hershey
- Kotsifakos E, Ntoutsis I, Theodoridis Y (2005) Database Support for Data Mining Patterns. *10th Panhellenic Conference on Informatics (PCI'2005)*, Volos, Greece. *Advances in Informatics - Springer-Verlag LNCS #3746*, 2005
- Kotsifakos E, Ntoutsis I, Vrahoritís Y, Theodoridis Y (2008) Monitoring Patterns through an Integrated Management and Mining Tool (Demo), *Proc. of Machine Learning and Knowledge Discovery in Databases, European Conference (ECML/PKDD'08)*, Antwerp, Belgium
- Kotsifakos E, Ntoutsis I, Vrahoritís Y, Theodoridis Y (2008) PATTERN-MINER: Integrated Management and Mining over Data Mining Models (Demo), *Proc. of 14th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'08)*, Las Vegas, USA

- Kuo Y, Lonie A, Sonenberg L, Paizis K (2007) Domain ontology driven data mining - a medical case study. *Proc. International Conference on Knowledge Discovery and Data Mining, Proc. 2007 International Workshop on Domain Driven Data Mining*, San Jose, California, USA., pp.11-17, doi:10.1145/1288552.1288554.
- Lassila O, Swick R (1999) Resource description framework (RDF) model and syntax specification, *W3C Recommendation*
- Lenat D, Deaton C (2008) TERRORISM KNOWLEDGE BASE (TKB), Cycorp Inc, Technical report. Available: <http://www.dtic.mil>
- Lenca P, Meyer P, Vaillant B, Lallich S (2004) A multicriteria decision aid for interestingness measure selection. Tech. Rep. LUSI-TR-2004-01-EN. LUSI Department, GET/ENST, Bretagne, France.
- Liu B, Hsu W, Chen S (1997) Using general impressions to analyze discovered classification rules. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97)*, Newport Beach, CA. 31–36.
- Liu B, Hsu W, Mun L, Lee H (1999) Finding interesting patterns using user expectations, *IEEE Trans. Knowl. Data Eng.* 11, 6, 817–832.
- Loom Documentation Last modified: 2009 <http://www.isi.edu/isd/LOOM/documentation/> Accessed 5 February 2011
- Lu S, Hu H, Li F (2001) Mining weighted association rules, *Intell. Data Anal.* 5, 3, 211–225
- Mannes A, Golbeck J (2005) Building a Terrorism Ontology, *Proceedings of the ISWC Workshop on Ontology Patterns for the Semantic Web*
- Mizoguchi R, Vanwelkenhuysen J, Ikeda M (1995) Task Ontology for reuse of problem solving knowledge. In *N.J.I. Mars Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, IOS Press, pp.46-57
- Motik B, Grau C B, Horrocks I, Wu Z, Fokoue A, Lutz C (2009) OWL 2 Web Ontology Language Profiles, *W3C Recommendation*, available: http://www.w3.org/TR/2009/REC-owl2-profiles-20091027/#Computational_Properties Accessed 5 June 2011
- Motik B, Patel-Schneider F P, Grau C B (2009) OWL 2 Web Ontology Language Direct Semantics, *W3C Recommendation*, available: http://www.w3.org/TR/2009/REC-owl2-direct-semantics-20091027/#Inference_Problems Accessed 5 June 2011
- Niles I, Pease A (2001) Toward a Standard Upper Ontology, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*.
- Noy N, McGuinness D (2001) Ontology Development 101: A Guide to Creating Your First Ontology, Knowledge Systems, AI Laboratory, Stanford University (KSL-01-05)
- Ntoutsis I, Theodoridis Y. (2007) Pattern Management. UNIPI-ISL-TR-2007-01. Technical Report Series, Information Systems Lab., University of Piraeus, Piraeus, Hellas
- Ontolingua Server, Ontology management tool (2011) <http://www.ksl.stanford.edu/software/ontolingua/> Accessed 5 February 2011
- OntoStudio, Ontology management tool (2011) <http://www.ontoprise.de> Accessed 5 February 2011
- Orange, DM tool (2011) <http://orange.biolab.si/> Accessed 5 February 2011
- Osterwalder A (2002) An e-Business Model Ontology for Modeling e-Business. Bled, Slovenia
- OWLIM Documentation Last modified: 2011 http://www.ontotext.com/owlim/owlim_documentation.html Accessed 5 February 2011
- Padmanabhan B, Tuzhilin A (2000) Small is beautiful: Discovering the minimal set of unexpected patterns. In *Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (KDD 2000)*, Boston, MA. 54–63.
- PANDA, Patterns for Next-generation Database Systems (2011) project homepage. <http://dke.cti.gr/panda> Accessed 5 February 2011
- Pellet Documentation, Last modified 2011: <http://clarkparsia.com/pellet/docs/> Accessed 5 February 2011
- Piatetsky-Shapiro G (1991) Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and W. J. Frawley, Eds. MIT Press, Cambridge, MA, 229–248.

- PMML, Predictive Model Markup Language (2011) project homepage. <http://www.dmg.org/pmml-v4-0-1.html> Accessed 5 February 2011
- PowerLoom Documentation, Last modified: 2006 <http://www.isi.edu/isd/LOOM/PowerLoom/documentation/documentation.html> Accessed 5 February 2011
- Profiles in Terror, ontology in terrorism (2011) <http://www.mindswap.org/2003/owl/swint/terrorism> Accessed 5 February 2011
- Protégé 2000 Ontology Editor and Knowledge Acquisition System (2011) <http://protege.stanford.edu/> Accessed 5 February 2011
- RacerPro (2010) <http://www.racer-systems.com/products/racerpro/index.phtml> Accessed 5 February 2011
- RapidMiner, DM tool (2011) <http://rapid-i.com/> Accessed 5 February 2011
- RDF, Resource Description Framework, W3.org (2011) <http://www.w3.org/RDF/> Retrieved 14-01-2011 Accessed 5 February 2011
- RDFs, Resource Description Framework Schema (RDFs), W3.org (2011) <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/> Accessed 5 February 2011
- Rich E, Knight K (1991) Artificial Intelligence, McGraw-Hill, In: Βλαχαβάς Ι, Κεφαλάς Π, Βασιλειάδης Ν, Ρεφανίδης Ι, Κόκκορας Φ, Σακελλαρίου Η (2002) *Τεχνητή Νοημοσύνη, Θεσσαλονίκη: Εκδόσεις Γαρταγάνη*, 87-126
- Rizzi S, Bertino E, Catania B, Golfarelli M, Halkidi M, Terrovitis M, Vassiliadis P, Vazirgiannis, M, & Vrahnos E (2003) Towards a logical model for patterns. *Proceedings of ER'03 conference*
- Roddick F, Hornsby K, Vries D (2003) A unifying semantic distance model for determining the similarity of attribute values, *Proc. 26th Australasian Computer Science Conf.*, pp. 111–118.
- Russell, S., & Norvig, P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall
- Sahar S (1999) Interestingness via what is not interesting. In *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD-99)*, San Diego, CA. 332–336.
- Saruladha K, Aghila G, Raj S (2010) A Survey of Semantic Similarity Methods for Ontology Based Information Retrieval, *2010 Second International Conference on Machine Learning and Computing*, pp.297-301
- Schaffert S, Gruber A, Westenthaler R (2005) A Semantic Wiki for Collaborative Knowledge Formation, In: *Semantics 2005*, Vienna, Austria
- SchemaWeb, list from RDF schemas expressed in languages RDFS, OWL and DAML+OIL schema (2011) <http://www.schemaweb.info/> Accessed 5 February 2011
- Sebastian Y, Brian C S Loh, Patrick H H Then (2009) Towards Natural Language Interface Framework for Automated Medical Analysis, *Proceedings of 2nd IEEE International Conference on Computer Science and Information Technology (IEEE ICCSIT 2009)*, vol.3, pp.5-9.
- Semy S, Pulvermacher M, Obrst L (2004) Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains: An Evaluation. MITRE Technical Report, MTR 04B0000063
- Shen D, Zhang Z, Yang Q (2002) Objective-Oriented utility-based association mining. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM '02)*, Maebashi City, Japan. 426–433.
- Silberschatz A, Tuzhilin A (1995) On subjective measures of interestingness in knowledge discovery. In *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining (KDD-95)*, Montreal, Canada. 275–281.
- Silberschatz A, Tuzhilin A (1996) What makes patterns interesting in knowledge discovery systems, *IEEE Trans. Knowl. Data Eng.* 8, 6, 970–974.
- Sirin E, Parsia B, Grau C B, Kalyanpur A, Katz K (2007) Pellet: A practical OWL-DL reasoned, *Journal of Web Semantics*, 5(2)
- Smith B (2003) Ontology Ch. in: Luciano Floridi (ed.) *The Blackwell Guide to Philosophy of Computing and Information*, pp. 155–166. Malden: Blackwell

- Sommerville I (2006) *Software Engineering: (Update)*, 8/E, Addison-Wesley.
- SQL/MM part 6 Data Mining (2006) INTERNATIONAL STANDARD, ISO/IEC 13249-6:2006
- Srikant R, Agrawal R (1997) Mining Generalized Association Rules, *Future Generation Computer Systems*, 13(2-3)
- Staab S, Schnurr H, Studer R, Sure Y (2001) Knowledge processes and ontologies, *IEEE Intelligent Systems* 16 (1) 26–34.
- Stojanovic L, Motik B (2002) Ontology evolution with ontology, *EKAW02 Workshop on Evaluation of Ontologybased Tools (EON2002)*, *CEUR Workshop Proceedings*, Siguenza, vol. 62, pp. 53–62.
- SUMO, Suggested Upper Merged Ontology, Upper Level Ontology (2011) <http://www.ontologyportal.org/> Accessed 5 February 2011
- Swoogle, ontology search machine (2011) <http://swoogle.umbc.edu/> Accessed 5 February 2011
- Tan P, Kumar V, Srivastava J (2002) Selecting the right interestingness measure for association patterns. In *Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining (KDD 2002)*, Edmonton, Canada. 32–41.
- Theodoridis Y, Vazirgiannis M, et al. (2003) A manifesto for pattern bases. PANDA Technical Report TR-2003-03, 2003. Available at <http://dke.cti.gr/panda>.
- Thompson H (2010) What Is a URI and Why Does It Matter?, *Ariadne*, Issue 65
- Thuraisingham B (2004) Data mining for counterterrorism. In Kargupta H, Joshi A, Sivakumar K, Yesha Y, editors, *Data Mining: Next Generation Challenges and Future Directions*, pages 157,183, AAAI/MIT Press
- Travel Ontology, ontology in travel industry (2011), <http://swatproject.org/travelOntology.asp> Accessed 5 February 2011
- Umbel, Vocabulary and Reference Concept Ontology (2011) <http://umbel.org/> Accessed 5 February 2011
- Uschold M, Grüninger M (1996), Ontologies: Principles, Methods, and Applications, *Knowledge Engineering Review*, 1:96-137
- Van Heijst G, Schreiber A Th, Wielinga B J (1997) Using explicit ontologies in KBS development, *International Journal of Human-Computer Studies*, 45, pp.183-292
- Velásquez J, Jain L (2010) Advanced Techniques in Web Intelligence. *Springer-Verlag*, Berlin Heidelberg, SCI 311, pp. 191–224
- Wang K, Zhou S, HAN J (2002) Profit mining: From patterns to actions. In *Proceedings of the 8th Conference on Extending Database Technology (EDBT 2002)*, Prague, Czech Republic. 70–87.
- Wang, H. and Li, L. (2005) Classification based on geographical information ontology, in *Proceedings of International Symposium on Spatio-temporal Modeling, Spatial Reasoning, Analysis, Data Mining and Data Fusion*, 27 – 29 August, Beijing, China, pp.27-29.
- WebOnto, Ontology management tool (2011) <http://projects.kmi.open.ac.uk/webonto/> Accessed 5 February 2011
- Wei G, Wu G, Gu Y, Ling Y (2008) An ontology based approach for Chinese web texts classification, *Information Technology Journal*, vol. 5, issue 5, pp. 796-801
- Weka Documentation (2011) <http://www.cs.waikato.ac.nz/ml/weka/> Accessed 5 February 2011
- Wordnet, a large lexical database of English (2011) <http://wordnet.princeton.edu/> Accessed 24 May 2011
- Xiong X, Fan L, Lei Z (2010) Ontology-based Association Rule Quality Evaluation Using Information Theory, *International Conference on Computational and Information Sciences*, pp. 170-173
- XML Extensible Markup Language 1.1 (Second Edition), W3.org (2011) <http://www.w3.org/TR/xml11/#sec-xml11> Accessed 5 February 2011
- Xu W, Huang H (2006) Research and Application of Spatio-temporal Data Mining Based on Ontology, *ICICIC* (2) pp.535-538

Yang Q, Wu X (2006) 10 Challenging Problems in Data Mining Research, *International Journal of Information Technology & Decision Making*, Vol. 5, No. 4 (2006) 597--604 World Scientific Publishing Company, 2006

Yao Y, Chen H, Yang D (2006) A measurement-theoretic foundation of rule interestingness evaluation, In *Foundations and Novel Approaches in Data Mining*, T. Y. Lin et al., Eds. Springer-Verlag, Berlin, 41–59.

Zhang J, Silvescu A, Honavar V (2002) Ontology-Driven Induction of Decision Trees at Multiple Levels of Abstraction, Technical Report ISU-CS-TR 02-13, Computer Science, Iowa State University

Zhou X, Geller J (2007) *Raising, to Enhance Rule Mining in Web Marketing with the Use of an Ontology*, Data Mining with Ontologies: Implementations, Findings and Frameworks, pp. 18-36, Idea Group Reference

Βλαχαβάς Ι, Κεφαλάς Π, Βασιλειάδης Ν, Ρεφανίδης Ι, Κόκκορας Φ, Σακελλαρίου Η (2002) *Τεχνητή Νοημοσύνη*. Θεσσαλονίκη: Εκδόσεις Γαρταγάνη, 87-126

ΠΑΡΑΡΤΗΜΑΤΑ

A Παράρτημα - Σύμβολα εκφραστικότητας DL & Πολυπλοκότητα

Η εκφραστικότητα των γλωσσών που ακολουθούν την περιγραφική λογική ποικίλει ανάλογα με τις λειτουργίες που τους έχουν προσδοθεί. Στον παρακάτω πίνακα παρουσιάζονται τα σύμβολα που χρησιμοποιούνται για τον λεκτικό προσδιορισμό αυτής της εκφραστικότητας.

Πίνακας A-1. Σύμβολα εκφραστικότητας DL

Συμβολισμός	Εκφραστικότητα
\mathcal{F}	Λειτουργικοί ρόλοι
\mathcal{E}	Πλήρης υπαρξιακή ποσοτικοποίηση
\mathcal{U}	Ένωση εννοιών
\mathcal{C}	Σύνθετη άρνηση
\mathcal{S}	Σύντμηση \mathcal{ALC} και μεταβατικούς ρόλους
\mathcal{H}	Ιεραρχία ατομικών ρόλων
\mathcal{R}	Αξιώματα υπαγωγής σύνθετων ρόλων, ανακλαστικοί ρόλοι και μη ανακλαστικοί ρόλοι, ασύμβατοι ρόλοι
\mathcal{O}	Απαριθμητικά
\mathcal{I}	Αντίστροφοι ρόλοι
\mathcal{N}	Περιορισμοί αριθμών
\mathcal{Q}	Ποιοτικοί περιορισμοί αριθμών
(\mathcal{D})	Ρόλοι τύπου δεδομένων, τιμές δεδομένων, τύποι δεδομένων
\mathcal{AL}	Ατομική άρνηση, τομή εννοιών, περιορισμός τιμής, περιορισμένη υπαρξιακή ποσοτικοποίηση
\mathcal{FL}	\mathcal{AL} χωρίς ατομική άρνηση
\mathcal{FL}_o	\mathcal{FL} χωρίς περιορισμένη υπαρξιακή ποσοτικοποίηση
\mathcal{EL}	Διατομή και πλήρης υπαρξιακός περιορισμός

Παρακάτω παρουσιάζονται οι όροι για τις πολυπλοκότητες που εμφανίζονται στις υπογλώσσες της OWL.

- *Ανοιχτή Αποκρισιμότητα* (Decidability Open): σημαίνει ότι δεν είναι γνωστό αν το πρόβλημα συλλογισμού μπορεί να καταλήξει σε κάποια απόφαση.
- *Αποκρίσιμη ανοιχτής πολυπλοκότητας* (Decidable, but complexity open): σημαίνει ότι είναι γνωστό για το πρόβλημα συλλογισμού ότι μπορεί να καταλήξει σε απόφαση, αλλά δεν είναι ακριβής η υπολογιστική του πολυπλοκότητα. Με την ονομασία NP-Hard συμβολίζεται ότι το πρόβλημα είναι τουλάχιστον τόσο δύσκολο όσο ένα NP πρόβλημα.
- X-complete: το X συμβολίζει την κατηγορία πολυπλοκότητας, ακολουθούν κατηγορίες από τις πιο πολύπλοκες προς τις απλούστερες.
 - *NEXPTIME-complete*: είναι η κλάση προβλημάτων που επιλύονται από μη-ντετερμινιστικούς αλγόριθμους σε χρόνο $O(2^{p(n)})$ για πολυώνυμο $p(n)$ με απεριόριστο χώρο.

- *PSPACE- complete*: είναι η κλάση προβλημάτων που επιλύονται από ντετερμινιστικό αλγόριθμο, χρόνο που είναι το πολύ πολυωνυμικός στο μέγεθος των εισροών.
- *NP- complete*: είναι η κλάση προβλημάτων που επιλύονται από μη-ντετερμινιστικούς αλγόριθμους που χρησιμοποιούν χρόνο που είναι το πολύ πολυωνυμικός στο μέγεθος των εισροών.
- *PTIME- complete*: είναι η κλάση των προβλημάτων που επιλύονται από ντετερμινιστικό αλγόριθμο χρησιμοποιώντας χρόνο ο οποίος είναι το πολύ πολυωνυμικός με το μέγεθος των εισροών.
- *LogSpace- complete*: είναι η κλάση προβλημάτων που επιλύονται από ντετερμινιστικό αλγόριθμο που χρησιμοποιεί χώρο ο οποίος είναι το πολύ λογαριθμικός με το μέγεθος των εισροών. Το *NLogSpace* είναι η μη-ντετερμινιστική έκδοση αυτής της κατηγορίας.
- AC^0 : είναι η κλάση προβλημάτων στην οικογένεια της κυκλωματικής πολυπλοκότητα συνεχούς βάθους και πολυωνυμικού μεγέθους, η οποία μπορεί να παραχθεί από μία ντετερμινιστική μηχανή Turing σε λογαριθμικό χρόνο.

B Παράρτημα – Διαφορές μεταξύ OWL 1 Lite- DL- Full

Στον πίνακα που ακολουθεί παρουσιάζονται οι κατασκευαστές που χρησιμοποιούνται στην OWL 1. Ενδεικτικό είναι ότι η κάθε μία υπογλώσσα είναι υποσύνολο της επόμενης με την OWL Full να είναι η πιο εκφραστική αλλά και η πιο πολύπλοκη με αποτέλεσμα να υστερεί σε απόδοση κατά την εκτέλεση των συλλογιστικών διαδικασιών. Η OWL Lite φαίνεται να είναι ο αδύναμος κρίκος λόγω της μικρής της εκφραστικότητας που δεν συμβαδίζει με τις απαιτήσεις των επιστημονικών και εμπορικών εφαρμογών, γι αυτό και στην OWL 2 αντικαταστάθηκε πλήρως.

Πίνακας B-2. Διαφορές μεταξύ OWL 1 Lite- DL- Full

Discription	Definition	OWL Lite	OWL DL	OWL Full
RDF Schema Features:	Class (Thing, Nothing)	X	X	X
	rdfs:subClassOf	X	X	X
	rdf:Property	X	X	X
	rdfs:subPropertyOf	X	X	X
	rdfs:domain	X	X	X
	rdfs:range	X	X	X
	Individual	X	X	X
Property Restrictions:	Restriction	X	X	X
	onProperty	X	X	X
	allValuesFrom	X	X	X
	someValuesFrom	X	X	X
Class Intersection:	intersectionOf	X	X	X
Datatypes	xsd datatypes	X	X	X
(In)Equality:	equivalentClass	X	X	X
	equivalentProperty	X	X	X
	sameAs	X	X	X
	differentFrom	X	X	X
	AllDifferent	X	X	X
	distinctMembers	X	X	X
Restricted	minCardinality	(only 0 or 1)	X	X

Cardinality:	maxCardinality	(only 0 or 1)	X	X
	cardinality	(only 0 or 1)	X	X
Versioning:	versionInfo	X	X	X
	priorVersion	X	X	X
	backwardCompatibleWith	X	X	X
	incompatibleWith	X	X	X
	DeprecatedClass	X	X	X
	DeprecatedProperty	X	X	X
Property Characteristics:	ObjectProperty	X	X	X
	DatatypeProperty	X	X	X
	inverseOf	X	X	X
	TransitiveProperty	X	X	X
	SymmetricProperty	X	X	X
	FunctionalProperty	X	X	X
	InverseFunctionalProperty	X	X	X
Header Information:	Ontology	X	X	X
	imports	X	X	X
Annotation Properties:	rdfs:label	X	X	X
	rdfs:comment	X	X	X
	rdfs:seeAlso	X	X	X
	rdfs:isDefinedBy	X	X	X
	AnnotationProperty	X	X	X
	OntologyProperty	X	X	X
Class Axioms:	one of		X	X
	dataRange		X	X
	disjointWith		X	X
	equivalentClass		applied to class expressions	X
	rdfs:subClassOf		applied to class expressions	X
Arbitrary Cardinality:	minCardinality		X	X
	maxCardinality		X	X
	cardinality		X	X
Boolean Combinations of Class Expressions:	unionOf		X	X
	complementOf		X	X
	intersectionOf		X	X
Filler Information:	hasValue		X	X
Meta-Classes				X

Γ Παράρτημα – Διαφορές μεταξύ OWL 2 EL- QL- RL

Στον πρώτο πίνακα που ακολουθεί συνοψίζονται οι διαφορές μεταξύ των τριών υπογλωσσών της OWL 2 σχετικά με τους περιορισμούς και τα αξιώματα που τα χαρακτηρίζουν. Στην πρώτη στήλη γίνεται μία τυπική περιγραφή, στη δεύτερη στήλη παρατίθενται οι ονομασίες των αντίστοιχων περιορισμών/ αξιωμάτων και στις άλλες τρεις στήλες με το σύμβολο X επισημαίνεται η παρουσία ή όχι του χαρακτηριστικού. Στις στήλες OWL- QL και OWL- RL εμφανίζονται άλλες δύο εγγραφές, *Subclass/ Superclass Expressions*, αυτό γίνεται γιατί η OWL 2 επιτρέπει να οριστούν περιορισμοί ανάλογα το τύπο των κλάσεων, όπως σε υποτάξεις ή υπερτάξεις. Στον δεύτερο πίνακα παρουσιάζονται οι τύποι δεδομένων που κάθε υπογλώσσα μπορεί να χρησιμοποιεί.

Πίνακας Γ-3. Χαρακτηριστικά OWL 2 EL- QL- RL

Description	Definition	OWL EL	OWL QL	OWL RL
Existential quantification to a class expression	ObjectSomeValuesFrom	X	Subclass/ Superclass Expressions	Subclass Expressions
Existential quantification to a data range	DataSomeValuesFrom	X	Subclass/ Superclass Expressions	Subclass Expressions
Existential quantification to an individual	ObjectHasValue	X		Subclass/ Superclass Expressions
Existential quantification to a literal	DataHasValue	X		Subclass/ Superclass Expressions
Self-restriction	ObjectHasSelf	X		X
Enumerations involving a <i>single</i> individual	ObjectOneOf	X		Subclass Expressions
Enumerations involving a <i>single</i> literal	DataOneOf	X		X
Intersection of classes	ObjectIntersectionOf	X	Superclass Expressions	Subclass/ Superclass Expressions
Intersection of data ranges	DataIntersectionOf	X		X
Class inclusion	SubClassOf	X	X	X
Class equivalence	EquivalentClasses	X	X	X
Class disjointness	DisjointClasses	X	X	X
Object property inclusion	SubObjectPropertyOf	X	X	X
Object property inclusion with or without property chains, and data property inclusion	SubDataPropertyOf	X	X	X
Property equivalence	EquivalentObjectProperties	X	X	X
	EquivalentDataProperties	X	X	X
Transitive object	TransitiveObjectProperty	X		X

properties				
Reflexive object properties	ReflexiveObjectProperty	X	X	
Domain restrictions	ObjectPropertyDomain	X	X	X
	DataPropertyDomain	X	X	X
Range restrictions	ObjectPropertyRange	X	X	X
	DataPropertyRange	X	X	X
Assertions	SameIndividual	X		X
	DifferentIndividuals	X	X	X
	ClassAssertion	X	X	X
	ObjectPropertyAssertion	X	X	X
	DataPropertyAssertion	X	X	X
	NegativeObjectPropertyAssertion	X		X
NegativeDataPropertyAssertion	X		X	
Functional data properties	FunctionalDataProperty	X		X
Keys	HasKey	X		X
Universal quantification to a class expression	ObjectAllValuesFrom			Superclass Expressions
Universal quantification to a data range	DataAllValuesFrom			Superclass Expressions
Cardinality restrictions	ObjectMaxCardinality			Superclass Expressions
	ObjectMinCardinality			X
	ObjectExactCardinality			X
	DataMaxCardinality			Superclass Expressions
	DataMinCardinality			X
	DataExactCardinality			X
Disjunction	ObjectUnionOf			Subclass Expressions
	DisjointUnion			
	DataUnionOf			X
Class negation	ObjectComplementOf		Superclass Expressions	Superclass Expressions
Enumerations involving more than one individual	ObjectOneOf			X
	DataOneOf			X
Disjoint properties	DisjointObjectProperties		X	X
	DisjointDataProperties		X	X
Irreflexive object properties	IrreflexiveObjectProperty		X	X
Inverse object properties	InverseObjectProperties			X
Unctional and	FunctionalObjectProperty			X

inverse-functional object properties	InverseFunctionalObjectProperty			X
Symmetric object properties	SymmetricObjectProperty		X	X
Asymmetric object properties	AsymmetricObjectProperty		X	X

Πίνακας Γ-4. Τύποι δεδομένων OWL 2 EL- QL- RL

Datatypes	OWL EL	OWL QL	OWL RL
<i>rdf:PlainLiteral</i>	X	X	X
<i>rdf:XMLLiteral</i>	X	X	X
<i>rdfs:Literal</i>	X	X	X
<i>owl:real</i>	X	X	
<i>owl:rational</i>	X	X	
<i>xsd:decimal</i>	X	X	X
<i>xsd:integer</i>	X	X	X
<i>xsd:nonNegativeInteger</i>	X	X	X
<i>xsd:string</i>	X	X	X
<i>xsd:normalizedString</i>	X	X	X
<i>xsd:token</i>	X	X	X
<i>xsd:Name</i>	X	X	X
<i>xsd:NCName</i>	X	X	X
<i>xsd:NMTOKEN</i>	X	X	X
<i>xsd:hexBinary</i>	X	X	X
<i>xsd:base64Binary</i>	X	X	X
<i>xsd:anyURI</i>	X	X	X
<i>xsd:dateTime</i>	X	X	X
<i>xsd:dateTimeStamp</i>	X	X	X
<i>xsd:long</i>			X
<i>xsd:language</i>			X
<i>xsd:boolean</i>			X
<i>xsd:byte</i>			X
<i>xsd:float</i>			X
<i>xsd:int</i>			X
<i>xsd:double</i>			X
<i>xsd:negativeInteger</i>			X
<i>xsd:nonPositiveInteger</i>			X
<i>xsd:positiveInteger</i>			X
<i>xsd:short</i>			X
<i>xsd:unsignedByte</i>			X
<i>xsd:unsignedInt</i>			X
<i>xsd:unsignedLong</i>			X
<i>xsd:unsignedShort</i>			X

Δ Παράρτημα – Βάσεις Δεδομένων Τρομοκρατίας

- ❖ American Terrorism Database
 - ♦ Δικτυακός Ιστότοπος: <http://www.icpsr.umich.edu/cocoon/ICPSR/STUDY/04639.xml>
 - ♦ Πηγή: Brent L. Smith, Kelly R. Damphousse
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις στην Αμερική
 - ♦ Περίοδος: 1968- 2002
 - ♦ Επιπλέον στοιχεία: Περίπου 7,306 επιθέσεις, χωρίζεται σε πέντε μέρη που διαπραγματεύονται διαφορετικές μεταβλητές
 - ♦ Λήψη Δεδομένων: Ναι
- ❖ Aviation Safety Database
 - ♦ Δικτυακός Ιστότοπος: <http://aviation-safety.net/database/>
 - ♦ Πηγή: Aviation Safety Database
 - ♦ Δεδομένα: Σχετικά με ασφάλεια πτήσεων
 - ♦ Περίοδος: 1943-2009
 - ♦ Επιπλέον στοιχεία: πάνω από 12,200 περιπτώσεις αεροπορικών πτήσεων
 - ♦ Λήψη Δεδομένων: Όχι
- ❖ Global Terrorism Database (GTD)
 - ♦ Δικτυακός Ιστότοπος: <http://www.start.umd.edu/start/>
 - ♦ Πηγή: National Consortium for the Study of Terrorism and Responses to Terrorism (START) at the University of Maryland; funded by the U.S. Department of Homeland
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις από όλο τον κόσμο
 - ♦ Περίοδος: 1997- 2007
 - ♦ Επιπλέον στοιχεία: Πάνω από 87.000 επιθέσεις με πάνω από 45 μεταβλητές που χαρακτηρίζουν την κάθε επίθεση.
 - ♦ Λήψη Δεδομένων: Ναι
- ❖ How Terrorist Groups End
 - ♦ Πηγή: RAND Corporation
 - ♦ Δικτυακός Ιστότοπος: http://www.rand.org/pubs/monographs/2008/RAND_MG741-1.pdf
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις από την Al-Qaida και άλλες οργανώσεις
 - ♦ Περίοδος: 1994-2007
 - ♦ Επιπλέον στοιχεία: Δίνονται δύο πίνακες με τρομοκρατικές επιθέσεις της Al-Qaida και άλλων οργανώσεων στο πρώτο παράρτημα του αρχείου.
 - ♦ Λήψη Δεδομένων: Ναι, σε μορφή pdf
- ❖ International Terrorism: Attributes of Terrorist Events (ITERATE)
 - ♦ Δικτυακός Ιστότοπος: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7947>
 - ♦ Πηγή: Edward F. Mickolus
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις από όλο τον κόσμο
 - ♦ Περίοδος: 1968-1977
 - ♦ Επιπλέον στοιχεία: 3,329 επιθέσεις, 14 μεταβλητές
 - ♦ Λήψη Δεδομένων: Ναι
- ❖ Terrorism in Western Europe: Events Data
 - ♦ Δικτυακός Ιστότοπος: <http://www.nsd.uib.no/macrodataloguide/set.html?id=39&sub=1>
 - ♦ Πηγή: The MacroDataGuide- An International Social Science Resource
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις στην Ευρώπη
 - ♦ Περίοδος: 1950- 2004
 - ♦ Επιπλέον στοιχεία: Τρομοκρατικές επιθέσεις σε 18 Ευρωπαϊκές χώρες, αποτελείται από 52 μεταβλητές
 - ♦ Λήψη Δεδομένων: Ναι
- ❖ Worldwide Incidents Tracking System (WITS)
 - ♦ Δικτυακός Ιστότοπος: <http://wits.nctc.gov/>
 - ♦ Πηγή: U.S. National Counterterrorism Center (U.S. government)
 - ♦ Δεδομένα: Τρομοκρατικές επιθέσεις από όλο τον κόσμο

- ♦ Περίοδος: 2004-2007
- ♦ Επιπλέον στοιχεία: Δυνατότητα επιλογής δεδομένων λήψης σε διάφορες μορφές
- ♦ Λήψη Δεδομένων: Ναι

Ε Παράρτημα – Οντολογίες Τρομοκρατίας

- ❖ AKTiveSAOntology
 - ♦ Δικτυακός Ιστότοπος: <http://www.zaltys.net/ontology/AKTiveSAOntology.owl>
 - ♦ Πηγή: Owens A.
 - ♦ Σχόλια: Μια αρκετά εκτενής και ετερογενής Οντολογία
- ❖ COSMO
 - ♦ Δικτυακός Ιστότοπος: <http://www.micra.com/COSMO/COSMO.owl>
 - ♦ Πηγή: COSMO working group
 - ♦ Σχόλια: Οντολογία υψηλού επιπέδου, η μόνη κλάση που σχετίζεται με το πεδίο της τρομοκρατίας είναι η *TerroristAgent* την οποία δανείζεται από την OpenCYC, ενώ ο τύπος *Terror* είναι μια πολύ γενική έννοια για την αίσθηση του τρόμου.
- ❖ Mindswap
 - ♦ Δικτυακός Ιστότοπος: <http://www.mindswap.org/2003/owl/swint/terrorism>
 - ♦ Πηγή: Alford R.
 - ♦ Σχόλια: Οντολογία πάνω στο πεδίο της τρομοκρατίας από το εργαστήριο mindswap του Πανεπιστημίου του Μέριλαντ.
- ❖ OpenCYC
 - ♦ Δικτυακός Ιστότοπος: <http://www.opencyc.org/downloads>
 - ♦ Πηγή: Bergman M., Giasson F.
 - ♦ Σχόλια: Περιλαμβάνει ένα αυστηρό υποσύνολο των εννοιών και των σχέσεων της OpenCyc.
- ❖ Umbel
 - ♦ Δικτυακός Ιστότοπος: <http://vocab.umbel.org/>
 - ♦ Πηγή: Cycorp
 - ♦ Σχόλια: Από τις μεγαλύτερες και πληρέστερες βάσεις γενικής γνώσης.
- ❖ SUMO
 - ♦ Δικτυακός Ιστότοπος: <http://sigmakee.cvs.sourceforge.net/viewvc/sigmakee/KBs/Mid-level-ontology.kif>
 - ♦ Πηγή: Niles I., Pease A.
 - ♦ Σχόλια: Γνωστή και ως MILO, Οντολογία υψηλού επιπέδου γραμμένη σε γλώσσα KIFF

Ζ Παράρτημα – Εγκατάσταση Reasoner ή Metrics

Στα παρακάτω βήματα εξηγείται η διαδικασία που πρέπει να ακολουθηθεί για να γίνει εγκατάσταση- επέκταση ενός νέου Reasoner ή μιας νέας Metrics μέσα στο σύστημα.

- ❖ Εγκατάσταση νέου Reasoner:
 1. Στο αρχείο *GUIdm.tabOAssociationRule.java* ενημερώνεται το πεδίο `protected String[] se_reasoner = {"Jena OWLReasoner", "Pellet via Jena"};` για την ονομασία του Reasoner που θα βλέπει ο χρήστης
 2. Στο αρχείο *GUIdm.tabOAssociationRule.java* ενημερώνεται η μέθοδος `setReasoner(String rsnr)` με ποιον τρόπο δημιουργείται το αντικείμενο του Reasoner
 3. Στο αρχείο *GUIdm.tabOAssociationRule.java* ενημερώνεται η μέθοδος `setOntologyReasoner(String rsnr)` με ποια κλάση υλοποιεί την προγονική της *OntologyReasoner*

4. Δημιουργείται ένα .java αρχείου (όπως το JenaReasoner) το οποίο επεκτείνει το *OntologyReasoner* μέσα στο πακέτο *ontology*.
5. Εισάγονται νέες βιβλιοθήκες, αν απαιτούνται.
- ❖ Εγκατάσταση νέας Μετρικής (Metrics)
 1. Στο αρχείο *GUIdm.tabOAssociationRule.java* ενημερώνεται το πεδίο


```
protected String[] se_metrics = {"Jena findShortestPath", "New Metrics"}
```

 για την ονομασία της Μετρικής που θα βλέπει ο χρήστης
 2. Στο αρχείο *GUIdm.tabOAssociationRule.java* ενημερώνεται η μέθοδος


```
setMetrics(String rsnr)
```

 με ποια κλάση υλοποιείται η νέα Μετρική.
 3. Δημιουργείται ένα νέο .java αρχείου (όπως το JenaMetrics) το οποίο επεκτείνει την *Metrics* μέσα στο πακέτο *ontology*.
 4. Εισάγονται νέες βιβλιοθήκες, αν απαιτούνται.

Η Παράρτημα – Εγκατάσταση PatternMiner

Ο παρών οδηγός εγκατάστασης είναι επέκταση του οδηγού εγκατάστασης του PatternMiner v.2 που βρίσκεται στον διαδικτυακό τόπο <http://isl.cs.unipi.gr/PatternMiner/>. Οι κύριες αλλαγές που έχουν πραγματοποιηθεί είναι της τιμής Ελάχιστης και Προτεινόμενης RAM, η δημιουργία ενός νέου αρχείου του *ontology_file* στο αρχείο εγκατάστασης *C:\patternMiner* και τέλος η εισαγωγή νέων βιβλιοθηκών. Η προτεινόμενη απαίτηση της RAM εξακολουθεί να είναι 256MB όταν δε γίνεται χρήση Οντολογιών. Σε περίπτωση που ενσωματώνονται Οντολογίες πρέπει η τιμή της RAM να είναι ανάλογη αυτής του όγκου της Οντολογίας.

Το σύστημα PatternMiner v.2 είναι σχεδιασμένο ούτως ώστε να εκτελείται σε λειτουργικό σύστημα Windows. Οι απαιτήσεις που πρέπει να διαθέτει ο υπολογιστής που θα εγκατασταθεί το σύστημα αναγράφονται στον παρακάτω πίνακα:

Πίνακας Η-5. Απαιτήσεις για την εγκατάσταση του συστήματος PatternMiner v.2

Απαίτηση	Ελάχιστη	Προτεινόμενη
CPU	1000MHz	1400+ MHz
RAM	128MB	256 MB χωρίς χρήση Οντολογίας 1500MB με χρήση Οντολογίας όπως openCYC
Λειτουργικό Σύστημα	Windows NT4, SP 6 Windows 2000 Windows XP, Windows Vista	Windows NT4, SP 6 Windows 2000 Windows Vista
Hard Disk space (software)	10MB	20MB
Hard Disk space (Berkeley DB XML)	300MB	500MB
Hard Disk space (PatternMiner v.2 & data)	200MB	1 GB

Η εγκατάσταση του συστήματος περιλαμβάνει τα ακόλουθα πέντε βήματα:

1. Εγκατάσταση του Java Runtime Environment (JRE)

Ο υπολογιστής στον οποίο θα εγκατασταθεί το σύστημα θα πρέπει να διαθέτει JRE version 1.4.2_03 ή νεότερη. Το JRE είναι διαθέσιμο στην ακόλουθη ηλεκτρονική διεύθυνση: http://java.sun.com/products/archive/j2se/1.4.2_03/index.html

2. Εγκατάσταση της βάσης δεδομένων Berkeley DB XML

Το σύστημα είναι συμβατό με τη βάση δεδομένων Berkeley DB XML release 2.2.13, η οποία διατίθεται δωρεάν στην ακόλουθη ηλεκτρονική διεύθυνση:

<http://www.oracle.com/technology/software/products/berkeley-db/xml/index.html>.

Αφού εγκατασταθεί η βάση δεδομένων στον υπολογιστή γίνεται έλεγχος για το εάν προστέθηκαν στη μεταβλητή περιβάλλοντος των Windows CLASSPATH οι βιβλιοθήκες της DBXML. Για παράδειγμα, εάν η βάση εγκαταστάθηκε στον κατάλογο C:\DBXML, τότε το CLASSPATH θα πρέπει να περιέχει τις τέσσερις ακόλουθες βιβλιοθήκες:

- c:\DBXML\jar\dbxml.jar;
 - c:\DBXML\jar\db.jar;
 - c:\DBXML\jar\dbxmlexamples.jar;
 - c:\DBXML\jar\dbexamples.jar
3. Εγκατάσταση του Standard Widget Toolkit (SWT)

Το SWT είναι ένα σύστημα βιβλιοθηκών για την ανάπτυξη Διεπαφών Χρηστών και Γραφικών Παρουσιάσεων και χρησιμοποιείται για την υλοποίηση του γραφικού περιβάλλοντος που διαθέτει το σύστημα. Το εν λόγω σύστημα διατίθεται δωρεάν από την ηλεκτρονική διεύθυνση: <http://archive.eclipse.org/eclipse/downloads/drops/R-3.1.2-200601181600/download.php?dropFile=swt-3.1.2-win32-win32-x86.zip>.

Για την εγκατάσταση του πρέπει να δημιουργηθεί ο κατάλογος «C:\swt» στον οποίο θα αποσυμπίεστεί το αρχείο. Έπειτα, θα πρέπει να προστεθεί ο συγκεκριμένος κατάλογος στη μεταβλητή περιβάλλοντος των Windows PATH.

4. Εγκατάσταση του εργαλείου data mining WEKA

Το σύστημα είναι συμβατό με το εργαλείο εξόρυξης γνώσης WEKA version 3.4.7, το οποίο διατίθεται δωρεάν στην ηλεκτρονική διεύθυνση:

<http://kent.dl.sourceforge.net/sourceforge/weka/weka-3-4-7.exe>.

Αφού γίνει η εγκατάσταση του εργαλείου, προστίθεται στη μεταβλητή περιβάλλοντος των Windows CLASSPATH ο κατάλογος εγκατάστασης του weka και του weka.jar. Για παράδειγμα, αν το εργαλείο έχει εγκατασταθεί στον κατάλογο «C:\Weka», τότε θα πρέπει να προστεθούν στο CLASSPATH οι ακόλουθοι κατάλογοι:

- C:\Weka;
- C:\Weka\Weka.jar;

5. Δημιουργία του περιβάλλοντος του συστήματος PattenMiner v.2 και εγκατάσταση του λογισμικού του

Αρχικά επιλέγεται το αρχείο «run_me_first.bat», το οποίο περιέχει εντολές οι οποίες εκτελούνται από τη γραμμή εντολών των Windows. Με την εκτέλεση των συγκεκριμένων εντολών δημιουργούνται στο σύστημα αρχείων των Windows οι απαιτούμενοι κατάλογοι και τα απαραίτητα αρχεία για τη λειτουργία του συστήματος PatternMiner v.2. Παρακάτω παρατίθεται το περιεχόμενο του αρχείου «run_me_first.bat»:

Πίνακας Η-6. Οι εντολές του αρχείου «run_me_first.bat»

```
echo off
rem From command prompt run this batch file

echo creating patternMiner environment
mkdir C:\patternMiner
copy patternMiner.jar C:\patternMiner /V
copy pmm1-3-2.xsd C:\patternMiner /V
copy Run_patternMiner.bat C:\patternMiner /V
mkdir C:\patternMiner\lib
copy lib\db.jar C:\patternMiner\lib /V
copy lib\dbxml.jar C:\patternMiner\lib /V
copy lib\jdom.jar C:\patternMiner\lib /V
copy lib\swt.jar C:\patternMiner\lib /V
copy lib\org-netbeans-api-visual.jar C:\patternMiner\lib /V
copy lib\PANDA.jar C:\patternMiner\lib /V

copy lib\weka.jar C:\patternMiner\lib /V

copy lib\arq-2.8.7.jar C:\patternMiner\lib /V
```

```
copy lib\icu4j-3.4.4.jar C:\patternMiner\lib /V
copy lib\iri-0.8.jar C:\patternMiner\lib /V
copy lib\jena-2.6.4.jar C:\patternMiner\lib /V
copy lib\jena-2.6.4-tests.jar C:\patternMiner\lib /V
copy lib\junit-4.5.jar C:\patternMiner\lib /V
copy lib\log4j-1.2.13.jar C:\patternMiner\lib /V
copy lib\lucene-core-2.3.1.jar C:\patternMiner\lib /V
copy lib\slf4j-api-1.5.8.jar C:\patternMiner\lib /V
copy lib\slf4j-log4j12-1.5.8.jar C:\patternMiner\lib /V
copy lib\stax-api-1.0.1.jar C:\patternMiner\lib /V
copy lib\wstx-asl-3.2.9.jar C:\patternMiner\lib /V
copy lib\xercesImpl-2.7.1.jar C:\patternMiner\lib /V

copy lib\aterm-java-1.6.jar C:\patternMiner\lib /V
copy lib\pellet-cli.jar C:\patternMiner\lib /V
copy lib\pellet-core.jar C:\patternMiner\lib /V
copy lib\pellet-datatypes.jar C:\patternMiner\lib /V
copy lib\pellet-dig.jar C:\patternMiner\lib /V
copy lib\pellet-el.jar C:\patternMiner\lib /V
copy lib\pellet-explanation.jar C:\patternMiner\lib /V
copy lib\pellet-jena.jar C:\patternMiner\lib /V
copy lib\pellet-modularity.jar C:\patternMiner\lib /V
copy lib\pellet-owlapi.jar C:\patternMiner\lib /V
copy lib\pellet-owlapiv3.jar C:\patternMiner\lib /V
copy lib\pellet-pellint.jar C:\patternMiner\lib /V
copy lib\pellet-query.jar C:\patternMiner\lib /V
copy lib\pellet-rules.jar C:\patternMiner\lib /V
copy lib\pellet-test.jar C:\patternMiner\lib /V
copy lib\servlet.jar C:\patternMiner\lib /V

copy lib\jgrapht-jdk1.6.jar C:\patternMiner\lib /V

C:
cd\
cd patternMiner
mkdir data_files
mkdir dbxml
mkdir pmml_files
mkdir ontology_files
mkdir monitoring_results
mkdir result_arff
cd result_arff
mkdir arff_with_clusters
mkdir variances
cd..

cd dbxml
rem echo on
echo at the prompt "dbxml>" write exit
rem echo off
dbxml -c

echo patternMiner environment created
echo now create Containers "AssociationRules.dbxml" and "Clustering.dbxml"
and "Trees.dbxml"
```

Μετά την εκτέλεση του αρχείου «run_me_first.bat», θα έχει δημιουργηθεί το περιβάλλον του συστήματος PatternMiner v.2 καθώς και ο κατάλογος «C:\patternMiner\». Μέσα σε αυτόν τον κατάλογο δημιουργούνται οι εξής υποκατάλογοι:

- data_files: ο υποκατάλογος όπου θα αποθηκεύονται τα αρχεία τύπου arff τα οποία θα περιέχουν τα αρχικά δεδομένα.
 - dbxml: ο υποκατάλογος στον οποίο δημιουργήθηκε το περιβάλλον της βάσης προτύπων.
 - result_arff: ο υποκατάλογος όπου θα αποθηκεύονται τα παραγόμενα πρότυπα του WEKA.
 - rml_files: ο υποκατάλογος όπου θα αποθηκεύονται τα παραγόμενα πρότυπα σε δομή PMML.
 - Ontology_files: ο υποκατάλογος όπου θα αποθηκεύονται τα αρχεία τύπου owl και map τα οποία θα περιέχουν τα αρχικά δεδομένα σχετικά με την οντολογία και τις αντιστοιχίσεις των δεδομένων εκπαίδευσης με τις έννοιες της οντολογίας.
 - Monitoring_results: ο υποκατάλογος όπου θα αποθηκεύονται τα XML αρχεία με τα αποτελέσματα του αλγορίθμου επίβλεψης προτύπων.
 - lib: ο υποκατάλογος στον οποίο θα αποθηκεύονται όλες οι απαραίτητες βιβλιοθήκες και τα jar αρχεία για να λειτουργήσει το σύστημα.
- Μέσα στον υποκατάλογο result_arff δημιουργούνται δύο επιπλέον υποκατάλογοι:
- arff_with_clusters: ο υποκατάλογος στον οποίο αποθηκεύονται τα αρχεία που προκύπτουν από τη χρησιμοποίηση του φίλτρου «AddCluster» του εργαλείου WEKA.
 - variables: ο υποκατάλογος στον οποίο αποθηκεύονται οι τιμές διακύμανσης ανά χαρακτηριστικό και ανά συστάδα που υπολογίζονται βάσει των αποτελεσμάτων του φίλτρου «AddCluster» του εργαλείου WEKA.

Έπειτα, από τη γραμμή εντολών των Windows και τον κατάλογο «C:\patternMiner\dbxml>», πρέπει να εκτελεστεί η εντολή «C:\patternMiner\dbxml>dbxml, ώστε να γίνει εισαγωγή στο περιβάλλον εργασίας της Berkeley DB XML και στη συνέχεια να δημιουργηθούν τα απαραίτητα containers της βάσης προτύπων του συστήματος. Οι εντολές που πρέπει να εκτελεστούν είναι οι ακόλουθες:

- για τα πρότυπα τύπου κανόνων συσχέτισης, δημιουργείται το container «AssociationRules.dbxml» εκτελώντας την εντολή: «dbxml>createContainer AssociationRules.dbxml».
- για τα πρότυπα τύπου συνόλων συστάδων δημιουργείται το container «Clustering.dbxml» εκτελώντας την εντολή: «dbxml>createContainer Clustering.dbxml».
- για τα πρότυπα τύπου δέντρων απόφασης δημιουργείται το container «Trees.dbxml» εκτελώντας την εντολή: «dbxml> createContainer Trees.dbxml».

Τέλος, για να ξεκινήσει η λειτουργία του συστήματος PatternMiner v.2, επιλέγεται το αρχείο «Run_PatternMiner.bat», το οποίο εκτελεί στην γραμμή εντολών των Windows την κύρια κλάση του συστήματος PatternMiner δίνοντας του τις απαραίτητες παραμέτρους. Το περιεχόμενο του bat αρχείου απεικονίζεται στην ακόλουθη εικόνα:

Πίνακας Η-7. Οι εντολές του αρχείου «Run_PatternMiner.bat»

```
java -cp .\lib\swt.jar;.lib\jdom.jar;.lib\db.jar;.lib\dbxml.jar;.lib\org-netbeans-api-visual.jar;.lib\PANDA.jar;.lib\weka.jar;.lib\arq-2.8.7.jar;.lib\icu4j-3.4.4.jar;.lib\iri-0.8.jar;.lib\jena-2.6.4.jar;.lib\jena-2.6.4-tests.jar;.lib\junit-4.5.jar;.lib\log4j-1.2.13.jar;.lib\lucene-core-2.3.1.jar;.lib\slf4j-api-1.5.8.jar;.lib\slf4j-log4j12-1.5.8.jar;.lib\stax-api-1.0.1.jar;.lib\wstx-asl-3.2.9.jar;.lib\xercesImpl-2.7.1.jar;.lib\aterm-java-1.6.jar;.lib\pellet-cli.jar;.lib\pellet-core.jar;.lib\pellet-datatypes.jar;.lib\pellet-dig.jar;.lib\pellet-el.jar;.lib\pellet-explanation.jar;.lib\pellet-jena.jar;.lib\pellet-modularity.jar;.lib\pellet-owlapi.jar;.lib\pellet-owlapiv3.jar;.lib\pellet-pellint.jar;.lib\pellet-query.jar;.lib\pellet-rules.jar;.lib\pellet-test.jar;.lib\servlet.jar;.lib\jgrapht-jdk1.6.jar;patternMiner.jar; GUIDm.GUIDm
```

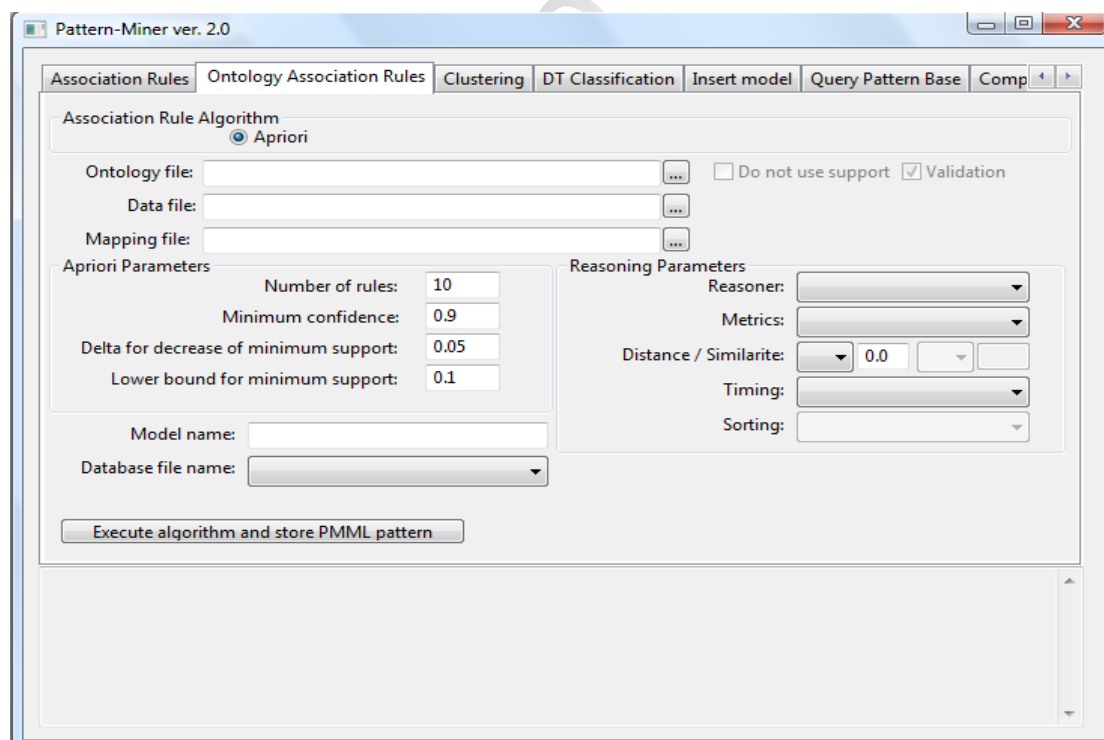
Θ Παράρτημα – Οδηγός Χρήσης

Αρχικά πρέπει να επιλεγεί η διεπαφή *Ontology Association Rules*. Έπειτα ο χρήστης θα πρέπει να αποφασίσει αν θα ενσωματώσει στην διαδικασία κάποια Οντολογία ή όχι. Στην περίπτωση όπου δεν θα γίνει ενσωμάτωση Οντολογίας θα χρειαστεί να συμπληρωθούν τα

πεδία του *Data file* με το αρχείο που περιέχει τα δεδομένα εκπαίδευσης, οι παράμετροι του αλγορίθμου Apriori, η ονομασία του μοντέλου, η βάση στην οποία θα αποθηκευτούν τα αποτελέσματα και τέλος να επιλεγεί το κουμπί για να εκτελεσθεί η διαδικασία. Όταν η διαδικασία ολοκληρωθεί τα μηνύματα της αριστερής στήλης του επόμενου πίνακα θα παρουσιαστούν στην οθόνη αιτιολόγησης.

Πίνακας Θ-8. Μηνύματα ορθής εκτέλεσης διαδικασίας

Run Apriori Without Ontology Intigration	Run Apriori With Ontology Intigration	Run Apriori With Ontology Intigration. It doesn't use Support Metric
Execute Apriori...	Ontology reading...	Ontology reading...
Executed	Ontology readied	Ontology readied
RunWeka_temp.bat	Ontology Validation...	Ontology Validation...
Creating PMML file...	Ontology is consistent	Ontology is consistent
Created PMML file...	Execute Apriori...	Execute Apriori...
PMML pattern was stored	Executed	Executed
Finished	RunWeka_temp.bat	RunWeka_temp.bat
=====	Creating PMML file...	Creating PMML file...
	Created PMML file...	Created PMML file...
	PMML pattern was stored	Created PMML file...
	Finished	PMML pattern was stored
	=====	Finished
		=====



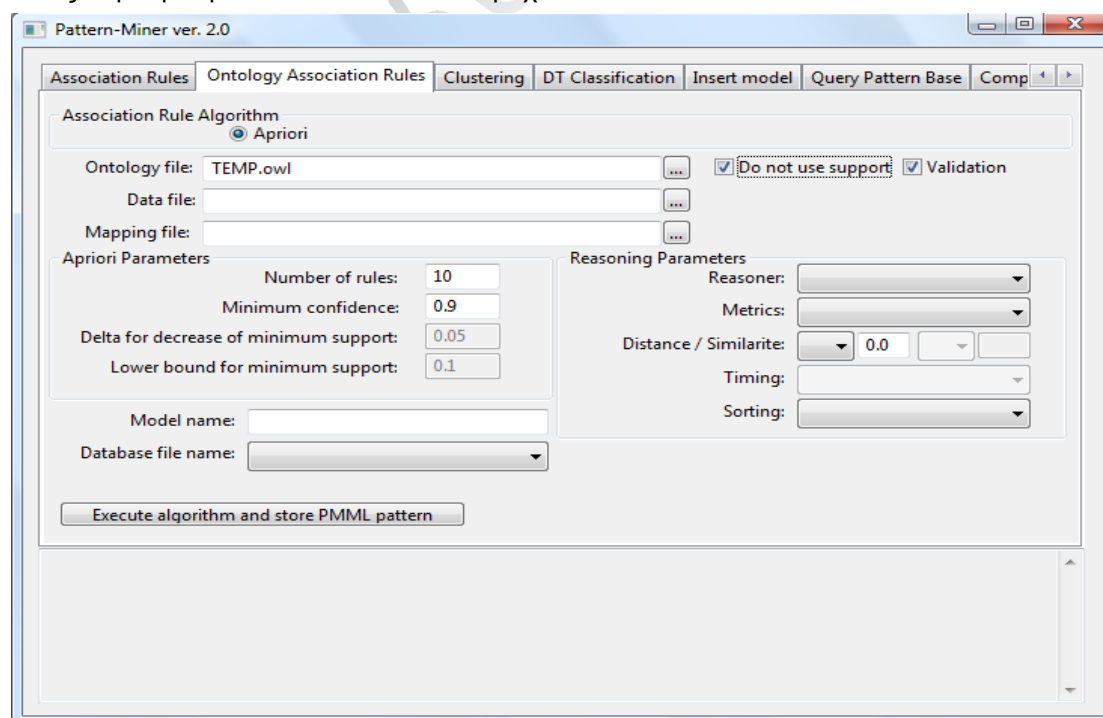
Εικόνα Θ-1. Διεπαφή Ontology Association Rules

Στην περίπτωση όπου ο χρήστης επιθυμεί την ενσωμάτωση Οντολογίας, έχει την δυνατότητα να τρέξει δύο διαφορετικά σενάρια. Το πρώτο είναι να αναζητήσει κανόνες με βάση το support, confidence και τις σημασιολογικές απόστασης της Οντολογίας και το δεύτερο είναι να χρησιμοποιήσει μόνο τα σημασιολογικά μέτρα αποστάσεων και το confidence.

Για την πρώτη περίπτωση θα πρέπει να συμπληρώσει τα στοιχεία όπως αναφέρθηκαν για την απλή εκτέλεση του Apriori και επιπλέον να επιλέξει στο πεδίο *Ontology file* το αρχείο της OWL Οντολογίας, στο πεδίο *Mapping file* το αρχείο που θα έχουν γίνει οι αντιστοιχίες μεταξύ των στοιχείων εκπαίδευσης και των εννοιών της Οντολογίας, και τέλος θα πρέπει να συμπληρώσει τις παραμέτρους για την διαδικασία του συμπερασμού. Στους *Reasoner Parameters* επιλέγει αρχικά την Μηχανή Συλλογιστικής Ανάλυσης, η οποία θα διαβάσει κατά κάποιο τρόπο την Οντολογία, έπειτα επιλέγει τη μετρική με την οποία θα υπολογίζονται οι αποστάσεις μεταξύ των εννοιών της Οντολογίας, στη συνέχεια θα επιλέξει σε πιο διάστημα θα ανήκει η ομοιότητα των εννοιών που συμμετέχουν σε ένα στοιχειοσύνολο ή κανόνα. Τέλος, επιλέγει στο πεδίο *Timing*, τότε ο αλγόριθμος Apriori θα ενσωματώνει την Οντολογία, όπου *During* είναι κατά την διάρκεια της εκτέλεσής του, *Ending* μετά το τέλος της εκτέλεσης του αλγορίθμου και *Over_all* για να ενσωματωθεί και κατά τη διάρκεια και στο τέλος της. Όταν η διαδικασία ολοκληρωθεί επιτυχώς τα μηνύματα της μεσαίας στήλης του πίνακα *Μηνύματα ορθής εκτέλεσης διαδικασίας* θα παρουσιαστούν στην οθόνη αιτιολόγησης.

Για τη δεύτερη περίπτωση ο χρήστης μετά την συμπλήρωση του πεδίου *Ontology file* θα πρέπει να τσεκάρει την επιλογή *Doesn't use support*. Έπειτα θα πρέπει να συμπληρώσει όλα τα ενεργά πεδία της εφαρμογής (Εικόνα Θ-2. Ενεργά πεδία για παραγωγή κανόνων χωρίς support). Στο πεδίο *Sorting* περιέχονται δύο επιλογές *max* και *min*, αυτά καθορίζουν την τελική διάταξη των κανόνων σε σχέση με την σημασιολογική τους απόσταση. Δηλαδή, αν χρησιμοποιηθεί το *max*, στην αρχή των αποτελεσμάτων θα παρουσιάζονται κανόνες με μεγάλη μέση σημασιολογική απόσταση. Όταν η διαδικασία ολοκληρωθεί επιτυχώς τα μηνύματα της δεξιάς στήλης του πίνακα *Μηνύματα ορθής εκτέλεσης διαδικασίας* θα παρουσιαστούν στην οθόνη αιτιολόγησης.

Παρόλο που η αξιολόγηση της Οντολογίας για να διαπιστωθεί η συνέπειά της είναι ένα από τα αναγκαία και χρήσιμα βήματα, παρατηρήθηκε ότι σε μεγάλες Οντολογίες μπορεί να επιφέρει πρόβλημα καθώς η χρήση μεγάλου ποσοστού μνήμης προκαλεί εξαίρεση σχετικά με το μέγεθος της *cache memory*. Γι' αυτό και δίνεται η επιλογή *Validation* κατά την οποία ο χρήστης επιλέγει αν θα αξιολογηθεί ή όχι η Οντολογία. Καθώς η αξιολόγηση είναι ένα αναπόσπαστο κομμάτι της διαδικασίας, η αρχική τιμή του *Validation* είναι αληθής, σε περίπτωση που ο χρήστης επιθυμεί να το απενεργοποιήσει απλά κάνει κλικ πάνω στο κουτί επιλογής, βέβαια θα πρέπει να είναι σίγουρος για την συνέπεια της Οντολογίας, διαφορετικά λάθος συμπεράσματα είναι πιθανό να παραχθούν.



Εικόνα Θ-2. Ενεργά πεδία για παραγωγή κανόνων χωρίς support

I Παράρτημα – Scripts για την δημιουργία των δεδομένων εκπαίδευσης

Στο πρώτο σύνολο των Scripts γίνεται επεξεργασία των ακατέργαστων δεδομένων της GTD. Οι αλλαγές που πραγματοποιούνται είναι η εξάλειψη κενών εγγραφών, αντικατάσταση των ονομάτων των οργανώσεων με συγκεκριμένη αριθμητική κωδικοποίηση και την τροποποίηση των τύπων τιμών των *nkill* και *nwound* από αριθμητικούς σε κατηγορηματικούς.

```

/*Start imonth: Ενημερώνει τις null έγγραφες του "imonth" με την τιμή μηδέν (0) */
UPDATE RAWDATA
SET "imonth" = '0'
WHERE "imonth" IS NULL;
/*End imonth*/

/*Start natlty1: Ενημερώνει τις null έγγραφες του "natlty1" με την τιμή -9 */
UPDATE RAWDATA
SET "natlty1" = '-9'
WHERE "natlty1" IS NULL;
/*End natlty1*/

/* Start PERPETRATOR INFORMATION: Δημιουργεί έναν νέο πίνακα (T_GNAME) με τα ονόματα των ομάδων ("tgname") και έναν κωδικό ("tgid") που τους χαρακτηρίζει*/
DECLARE
temp NUMBER :=0;
counter varchar2(4000);
BEGIN
SELECT count(*) INTO temp FROM user_tables WHERE table_name='T_GNAME';

IF temp=1 THEN EXECUTE IMMEDIATE('DROP TABLE T_GNAME');
END IF;

EXECUTE IMMEDIATE('CREATE TABLE T_GNAME("tgname"
VARCHAR2(4000),"tgid" VARCHAR2(4000))');

counter:=1;
FOR r IN (SELECT "gname" FROM RAWDATA WHERE "gname" IS NOT NULL
UNION SELECT "gname2" FROM RAWDATA WHERE "gname2" IS NOT NULL
UNION SELECT "gname3" FROM RAWDATA WHERE "gname3" IS NOT NULL)
LOOP
INSERT INTO T_GNAME VALUES (r."gname", counter);
counter:=counter+1;
END LOOP;

INSERT INTO T_GNAME VALUES('-9','-9');
END;

/* Δημιουργεί έναν νέο πίνακα (T_GSUBNAME) με τα ονόματα των υποομάδων ("tgid") και έναν κωδικό ("tgid") που τους χαρακτηρίζει*/

```



```

DECLARE
  temp NUMBER :=0;
  counter varchar2(4000);
BEGIN
  SELECT count(*) INTO temp FROM user_tables WHERE table_name='T_GSUBNAME';

  IF temp=1 THEN EXECUTE IMMEDIATE('DROP TABLE T_GSUBNAME');
  END IF;

  EXECUTE IMMEDIATE('CREATE TABLE T_GSUBNAME("tgsubname"
  VARCHAR2(4000),"tgsubnameid" VARCHAR2(4000))');

  counter:=1;
  FOR r IN (SELECT "gsubname" FROM RAWDATA WHERE "gsubname" IS NOT NULL
  UNION SELECT "gsubname2" FROM RAWDATA WHERE "gsubname2" IS NOT
  NULL
  UNION SELECT "gsubname3" FROM RAWDATA WHERE "gsubname3" IS NOT
  NULL)
  LOOP
    INSERT INTO T_GSUBNAME VALUES (r."gsubname", counter);
    counter:=counter+1;
  END LOOP;

  INSERT INTO T_GSUBNAME VALUES('-9','-9');
  END;
/* End PERPETRATOR INFORMATION*/

/*Start gname: Αντικαθιστά το όνομα της ομάδας με έναν κωδικό*/
UPDATE RAWDATA
SET "gname"= (SELECT DISTINCT "tnameid"
  FROM T_GNAME
  WHERE "gname"="tname")
/*End gname*/

/*Start nkill: Ενημερώνει τις null έγγραφες του "nkill" με την τιμή -99 */
UPDATE RAWDATA
SET "nkill" = '-99'
WHERE "nkill" IS NULL;

/*Αντικαθιστά τις τιμές του nkill σύμφωνα με μια κλίμακα τριών διαβαθμίσεων
Κανένας, WP= Λίγοι, DP= Πολύ. Επίσης έχει την τιμή 'UNKNOWN' για τιμές -99*/
UPDATE RAWDATA
SET "nkill"= CASE WHEN "nkill"= -99 THEN 'UNKNOWN'
  WHEN "nkill"= 0 THEN 'I'
  WHEN "nkill"> 0 AND "nkill"<= 12 THEN 'WP'
  WHEN "nkill"> 12 AND "nkill"<= 1384 THEN 'DP'
  END;
/*End nkill*/

```

```

/*Start nbound: Ενημερώνει τις null έγγραφες του "nbound" με την τιμή -99 */
UPDATE RAWDATA
SET "nbound" = '-99'
WHERE "nbound" IS NULL;

```

```

/*Αντικαθιστά τις τιμές του nbound σύμφωνα με μια κλίμακα τριών διαβαθμίσεων
I= Κανένας, WP= Λίγοι, DP= Πολύ. Επίσης έχει την τιμή 'UNKNOWN' για τιμές -99*/
UPDATE RAWDATA
SET "nbound"= CASE WHEN "nbound"= -99 THEN 'UNKNOWN'
                  WHEN "nbound"= 0 THEN 'I'
                  WHEN "nbound"> 0 AND "nbound"<= 55 THEN 'WP'
                  WHEN "nbound"> 55 AND "nbound"<= 5500 THEN 'DP'
END;
/*End nbound*/

```

```

/*Start property: Ενημερώνει τις null έγγραφες του "property" με την τιμή -9 */
UPDATE RAWDATA
SET "property" = '-9'
WHERE "property" IS NULL;

```

```

/*Αντικαθιστά τις τιμές του property όταν αυτές είναι 2 με -9*/
UPDATE RAWDATA
SET "property" = '-9'
WHERE "property" =2;
/*End property*/

```

```

/*Start propextent: Ενημερώνει την μεταβλητή με μια νέα τιμή την '0',
όταν το property=0 θεωρούμε ότι δεν υπάρχει ζημιά*/
UPDATE RAWDATA
SET "propextent" = '0'
WHERE "property" =0 AND "propextent" IS NULL;

```

```

/*Αντικαθιστά τις τιμές του propextent με 4 (Unknown) όταν αυτές είναι NULL ή -9*/
UPDATE RAWDATA
SET "propextent" = '4'
WHERE "property" !=0 AND ("propextent" IS NULL OR "propextent"= -9);
/*End propextent*/

```

Στο επόμενο σειτ των Scripts δημιουργείται ένας πίνακας που περιέχει τα δεδομένα εκπαίδευσης που είναι χρήσιμα για την εκτέλεση της περίπτωσης χρήσης.

```

/*Δημιουργία πίνακα με μεταβλητές της GTD*/
CREATE TABLE TRAINDATA("imonth" VARCHAR2(4000),
                        "country" VARCHAR2(4000),
                        "region" VARCHAR2(4000),
                        "attacktype1" VARCHAR2(4000),
                        "targtype1" VARCHAR2(4000),

```

```
"natlty1" VARCHAR2(4000),  
"gname" VARCHAR2(4000),  
"weaptype1" VARCHAR2(4000),  
"nkill" VARCHAR2(4000),  
"nwound" VARCHAR2(4000),  
"propextent" VARCHAR2(4000));
```

*/*Εισαγωγή δεδομένων στον πίνακα σύμφωνα με τους κωδικούς των επιθυμητών οργανώσεων*/*

```
INSERT INTO TRAINDATA
```

```
SELECT cgv."imonth",cgv."country", cgv."region", cgv."attacktype1",  
cgv."targetype1",cgv."natlty1",  
cgv."gname",cgv."weaptype1", cgv."nkill", cgv."nwound", cgv."propextent"
```

```
FROM RAWDATA cgv
```

```
WHERE cgv."gname" IN (1147, 1096, 1149, 1148, 1182, 1188, 1266, 238, 1034, 617, 778,  
806, 63, 64, 65, 66, 67, 204, 563, 758, 14, 549, 567, 632 );
```

*/*Εισαγωγή νέων μεταβλητών στον πίνακα*/*

```
ALTER TABLE TRAINDATA
```

```
ADD ("terroristcountry" VARCHAR2(4000), "terroristregion" VARCHAR2(4000), "ideology"  
VARCHAR(4000));
```

*/*Ενημέρωση νέων μεταβλητών*/*

*/*Σύμπτυξη μερικών οργανώσεων*/*

```
UPDATE TRAINDATA
```

```
SET "gname"= CASE WHEN "gname"=1149 THEN '1148'  
WHEN "gname"=1188 THEN '1182'  
WHEN "gname"=563 THEN '204'  
ELSE "gname"  
END;
```

*/*Ενημέρωση χώρας καταγωγής της οργάνωσης σύμφωνα με το country */*

```
UPDATE TRAINDATA
```

```
SET "terroristcountry"= CASE WHEN "gname"=1147 THEN '75'  
WHEN "gname"= 1096 THEN '155'  
WHEN "gname"= 1148 THEN '98'  
WHEN "gname"= 1182 THEN '78'  
WHEN "gname"= 1266 THEN '159'  
WHEN "gname"= 238 THEN '185'  
WHEN "gname"= 1034 THEN '155'  
WHEN "gname"= 617 THEN '96'  
WHEN "gname"= 778 THEN '296'  
WHEN "gname"= 806 THEN '186'  
WHEN "gname"= 63 THEN '422'  
WHEN "gname"= 64 THEN '95'  
WHEN "gname"= 65 THEN '60'  
WHEN "gname"= 66 THEN '173'  
WHEN "gname"= 67 THEN '6'  
WHEN "gname"= 204 THEN '101'
```

```
WHEN "gname"= 758 THEN '217'  
WHEN "gname"= 14 THEN '160'  
WHEN "gname"= 549 THEN '155'  
WHEN "gname"= 567 THEN '110'  
WHEN "gname"= 632 THEN '155'  
END;
```

/*Ενημέρωση περιοχή καταγωγής της οργάνωσης σύμφωνα με το region */

UPDATE TRAINDATA

```
SET "terroristregion"= CASE WHEN "gname"=1147 THEN '8'  
WHEN "gname"= 1096 THEN '10'  
WHEN "gname"= 1148 THEN '8'  
WHEN "gname"= 1182 THEN '8'  
WHEN "gname"= 1266 THEN '3'  
WHEN "gname"= 238 THEN '8'  
WHEN "gname"= 1034 THEN '10'  
WHEN "gname"= 617 THEN '8'  
WHEN "gname"= 778 THEN '10'  
WHEN "gname"= 806 THEN '6'  
WHEN "gname"= 63 THEN '6'  
WHEN "gname"= 64 THEN '10'  
WHEN "gname"= 65 THEN '10'  
WHEN "gname"= 66 THEN '10'  
WHEN "gname"= 67 THEN '10'  
WHEN "gname"= 204 THEN '4'  
WHEN "gname"= 758 THEN '1'  
WHEN "gname"= 14 THEN '5'  
WHEN "gname"= 549 THEN '10'  
WHEN "gname"= 567 THEN '10'  
WHEN "gname"= 632 THEN '10'  
END;
```

/*Ενημέρωση ιδεολογία οργάνωσης.

1=Σοσιαλιστές/ Κομμουνιστές

2=Θρησκευτικό-Πολιτικές

3=Εθνικό Απελευθερωτικό*/

UPDATE TRAINDATA

```
SET "ideology"= CASE WHEN "gname"=1147 THEN '1'  
WHEN "gname"= 1096 THEN '1'  
WHEN "gname"= 1148 THEN '1'  
WHEN "gname"= 1182 THEN '1'  
WHEN "gname"= 1266 THEN '1'  
WHEN "gname"= 238 THEN '2'  
WHEN "gname"= 1034 THEN '2'  
WHEN "gname"= 617 THEN '2'  
WHEN "gname"= 778 THEN '2'  
WHEN "gname"= 806 THEN '2'  
WHEN "gname"= 63 THEN '3'
```

```

WHEN "gname"= 64 THEN '3'
WHEN "gname"= 65 THEN '3'
WHEN "gname"= 66 THEN '3'
WHEN "gname"= 67 THEN '3'
WHEN "gname"= 204 THEN '3'
WHEN "gname"= 758 THEN '3'
WHEN "gname"= 14 THEN '3'
WHEN "gname"= 549 THEN '3'
WHEN "gname"= 567 THEN '3'
WHEN "gname"= 632 THEN '3'
END;

```

Κ Παράρτημα – Αποτελέσματα σεναρίων

Στο παρόν παράρτημα γίνεται σύνδεση των σεναρίων που έχουν τρέξει στη δεύτερη και τρίτη ενότητα με τα αρχεία αποτελεσμάτων που είναι αποθηκευμένοι οι παραγόμενοι κανόνες:

Πίνακας Κ-9. Αποτελέσματα σεναρίων δεύτερης ενότητας

Πρώτο σενάριο	patternMiner\extra\result_arff\AR_ConfBigWithOut.arff patternMiner\extra\result_arff\AR_ConfSmallWithOut.arff
Δεύτερο σενάριο	patternMiner\extra\result_arff\AR_ConfBigWithDur.arff patternMiner\extra\result_arff\AR_ConfBigWithEnd.arff patternMiner\extra\result_arff\AR_ConfBigWithOver.arff
Τρίτο σενάριο	patternMiner\extra\result_arff\AR_ConfSmallWithDur.arff patternMiner\extra\result_arff\AR_ConfSmallWithEnd.arff patternMiner\extra\result_arff\AR_ConfSmallWithOver.arff
Τέταρτο σενάριο	patternMiner\extra\result_arff\AR_ConfSmallWithMinMaxOver.arff patternMiner\extra\result_arff\AR_ConfSmallWithMinMaxDur.arff patternMiner\extra\result_arff\AR_ConfSmallWithMinMaxEnd.arff

Πίνακας Κ-10. Αποτελέσματα περιπτώσεων ανάλυσης τρίτης ενότητας

Πρώτη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_10000_09_005_01.arff
Δεύτερη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_1000_09_005_01_P_T_max00_x_D.arff
Τρίτη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max0223_min023_D.arff patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max0223_min023_E.arff
Τέταρτη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max0163_min017_D.arff patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max0163_min017_E.arff
Πέμπτη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max028_min0319_D.arff patternMiner\extra\result_arff\AR_TRAINDATA_1000_04_005_01_P_T_max028_min0319_E.arff
Έκτη περίπτωση ανάλυσης	patternMiner\extra\result_arff\AR_TRAINDATA_TRUE_1000_04_P_T_max0223_023_max.arff