



ΤΕΙ ΚΑΛΑΜΑΤΑΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΜΟΝΑΔΩΝ ΥΓΕΙΑΣ ΠΡΟΝΟΙΑΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΣΗΜΑΣΙΟΛΟΓΙΚΟ ΠΕΡΙΒΑΛΛΟΝ ΑΝΑΠΤΥΞΗΣ ΚΑΙ
ΜΕΘΟΔΟΙ ΚΑΙ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΟ
ΧΩΡΟ ΤΗΣ ΥΓΕΙΑΣ**

ΣΠΟΥΔΑΣΤΡΙΕΣ: ΚΟΥΚΟΥΒΑΓΙΑ ΚΑΤΕΡΙΝΑ, ΑΜ: 2008073
ΝΙΚΟΛΑΚΟΠΟΥΛΟΥ ΜΑΡΙΑ, ΑΜ: 2008026

ΕΠΙΒΛΕΠΟΥΣΑ ΚΑΘΗΓΗΤΡΙΑ: ΧΡΙΣΤΟΠΟΥΛΟΥ ΣΤΕΛΛΑ

ΚΑΛΑΜΑΤΑ 2013

ΕΥΧΑΡΙΣΤΙΕΣ

Στο σημείο αυτό κρίνουμε ότι θα θέλαμε να ευχαριστήσουμε την επιβλέπουσα καθηγήτριά μας κα. Χριστοπούλου Στέλλα, για την συνεχή επίβλεψη, καθοδήγηση και υποστήριξη της έρευνάς μας, προκειμένου να ολοκληρωθεί με επιτυχία η παρούσα πτυχιακή εργασία.

Επίσης, ευχαριστούμε θερμά τις οικογένειές μας για την υλική και ψυχολογική συμπαράσταση που μας παρείχαν σε όλη τη διάρκεια των σπουδών στο ΤΕΙ Καλαμάτας.

ΠΕΡΙΛΗΨΗ

Ο σκοπός της παρούσας πτυχιακής εργασίας είναι να εξετάσει τις δυνατότητες στην εξόρυξη και στη παρουσίαση των αποτελεσμάτων της εξόρυξης δεδομένων μέσω σημασιολογικού ιστού και εφαρμόζοντας ευφυή πληροφοριακά συστήματα και συστήματα διαχείρισης περιεχομένου.

Αρχικά παρουσιάζονται οι βάσεις γνώσης, η διαχείριση των συστημάτων γνώσης με βάση τις οντολογίες, καθώς και τα έμπειρα συστήματα. Στη συνέχεια, παρουσιάζεται η έννοια της εξόρυξης δεδομένων, οι βασικές κατηγορίες εξόρυξης, καθώς και πως γίνεται η εξόρυξη από ιατρικά δεδομένα. Κατόπιν, περιγράφεται η τεχνολογία του σημασιολογικού ιστού και η αρχιτεκτονική στη οποία στηρίζεται η τεχνολογία αυτή. Ακολούθως, παρουσιάζεται η έννοια της οντολογίας και πως αυτή χρησιμοποιείται στο χώρο της υγείας, με τη βοήθεια κάποιων ενδεικτικών παραδειγμάτων. Τέλος, παρουσιάζεται μια πρακτική εφαρμογή οντολογίας υγείας, που δημιουργήθηκε στα πλαίσια της παρούσας εργασίας.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ.....	2
ΠΕΡΙΛΗΨΗ.....	3
ΠΕΡΙΕΧΟΜΕΝΑ.....	4
ΠΕΡΙΕΧΟΜΕΝΑ ΕΙΚΟΝΩΝ.....	8
ΕΙΣΑΓΩΓΗ.....	10
1 ΚΕΦΑΛΑΙΟ: ΒΑΣΕΙΣ ΓΝΩΣΗΣ ΚΑΙ ΕΜΠΕΙΡΑ ΣΥΣΤΗΜΑΤΑ.....	11
1.1 ΒΑΣΕΙΣ ΓΝΩΣΗΣ.....	11
1.2 ΣΥΣΤΗΜΑΤΑ ΒΑΣΙΣΜΕΝΑ ΣΕ ΓΝΩΣΗ.....	11
1.3 ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΓΝΩΣΗΣ ΒΑΣΙΣΜΕΝΑ ΣΕ ΟΝΤΟΛΟΓΙΕΣ.....	13
1.3.1 Διαδικασία γνώσης.....	14
1.3.2 Μετά – Διαδικασία γνώσης.....	15
1.4 ΕΜΠΕΙΡΑ ΣΥΣΤΗΜΑΤΑ.....	16
2 ΚΕΦΑΛΑΙΟ: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ.....	17
2.1 ΕΙΣΑΓΩΓΗ.....	17
2.2 ΟΡΙΣΜΟΙ.....	17
2.3 ΚΥΚΛΟΣ ΖΩΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ.....	19

2.4	ΒΑΣΙΚΕΣ ΚΑΤΗΓΟΡΙΕΣ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	21
2.4.1	Ταξινόμηση (Classification)	22
2.4.2	Ομαδοποίηση (Clustering).....	22
2.4.3	Ένωση (Association).....	23
2.4.4	Παλινδρόμηση (Regression).....	23
2.4.5	Πρόβλεψη (Forecasting)	24
2.4.6	Ανάλυση ακολουθίας (Sequence Analysis).....	24
2.5	ΟΙ ΒΑΣΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	25
2.5.1	Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks).....	25
2.5.2	Τα δέντρα αποφάσεων (Decision Trees)	25
2.5.3	Οι γενετικοί αλγόριθμοι (Genetic Algorithms).....	26
2.5.4	Η επαγωγή κανόνων (Rule Induction).....	26
2.5.5	Η μέθοδος του κοντινότερου γείτονα (Nearest Neighbor Method).....	26
2.5.6	Η απεικόνιση δεδομένων (Data Visualization).....	26
2.6	ΠΡΟΤΥΠΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ.....	26
2.6.1	Εξόρυξη δεδομένων OLE DB.....	26
2.6.2	Εξόρυξη δεδομένων API ΤΗΣ JAVA.....	27
2.6.3	Πρότυπα PMML	27
2.7	ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ.....	28
3	ΚΕΦΑΛΑΙΟ: ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ	31
3.1	ΣΗΜΑΣΙΟΛΟΓΙΚΟΣ ΙΣΤΟΣ.....	31
3.1.1	Εισαγωγή	31

3.1.2	Η σχέση του σημασιολογικού ιστού με τον παγκόσμιο ιστό.....	32
3.1.3	Αναπαράσταση γνώσης στο σημασιολογικό ιστό	32
3.1.4	Τα συστατικά του σημασιολογικού ιστού	33
3.1.4.1	Η ΓΛΩΣΣΑ XML.....	34
3.1.4.2	ΤΕΧΝΟΛΟΓΙΑ XML SCHEMA.....	36
3.1.4.3	Η ΓΛΩΣΣΑ RDF (RESOURCE DESCRIPTION FRAMEWORK).....	38
3.1.4.4	RDF SCHEMA.....	39
3.1.4.5	Η Η ΓΛΩΣΣΑ OWL (WEB ONTOLOGY LANGUAGE).....	39
3.1.5	Αναζήτηση δεδομένων στον σημασιολογικό ιστό.....	41
3.1.6	Αυτοματοποίηση και ολοκλήρωση υπηρεσιών	41
3.1.7	Αρχιτεκτονική του σημασιολογικού ιστού.....	42
4	ΚΕΦΑΛΑΙΟ: ΟΝΤΟΛΟΓΙΕΣ ΣΤΟΝ ΤΟΜΕΑ της ΥΓΕΙΑΣ.....	45
4.1	Οντολογίες και Υγεία.....	46
4.2	Οντολογία – Τι είναι	47
4.2.1	Ανάγκη Ανάπτυξης και χρήσης οντολογιών	49
4.2.2	Κατηγοριοποίηση Οντολογιών.....	50
4.2.3	Βασική Δομή της Οντολογίας.....	54
4.2.4	Γλώσσες Οντολογιών.....	55
4.2.5	Περιβάλλοντα Οντολογιών.....	57
4.2.6	Οντολογίες στο χώρο της Υγείας – Υφιστάμενη Κατάσταση.....	62
4.3	Παραδείγματα οντολογιών υγείας.....	69

4.3.1	Οντολογία Παραδοσιακής Αφρικανικής Ιατρικής (ΑΤΜΟ)	69
4.3.2	Οντολογία Δυσπλασίας των Οστών.....	70
4.3.3	Οντολογία Συμπεριφοράς Κυττάρων (CBO).....	72
4.3.4	Οντολογία Ανατομίας (ΑΕΟ)	73
4.3.5	Οντολογία Νεύρο – Συμπεριφοράς (NBO)	74
4.4	ΙΑΤΡΟΛΕΞΗ: ΕΛΛΗΝΙΚΗ ΟΝΤΟΛΟΓΙΑ ΥΓΕΙΑΣ	76
5	ΚΕΦΑΛΑΙΟ: ΑΝΑΛΥΣΗ MDS ONTOLOGY	78
5.1	ΤΙ ΕΙΝΑΙ ΤΟ MDS ONTOLOGY	78
5.2	MDS ONTOLOGY	79
5.3	ΚΩΔΙΚΑΣ ΟΝΤΟΛΟΓΙΑΣ	91
6	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	96
7	ΒΙΒΛΙΟΓΡΑΦΙΑ	97
	ΠΑΡΑΡΤΗΜΑΤΑ	103
	ΠΑΡΑΡΤΗΜΑ Α: IDENTIFICATION INFORMATION	104
	ΠΑΡΑΡΤΗΜΑ Β: HEARING, SPEECH, AND VISION	105

ΠΕΡΙΕΧΟΜΕΝΑ ΕΙΚΟΝΩΝ

Εικόνα 1.1 Διαδικασία γνώσης - Μετα-διαδικασία γνώσης: Δύο Ορθογώνιες, Αναδραστικές Διαδικασίες	13
Εικόνα 1.2 Διαδικασία ανάπτυξης ενός συστήματος διαχείρισης γνώσης ή Διαδικασία γνώσης	14
Εικόνα 1.3 Η Διαδικασία Ανάπτυξης Οντολογίας ή Μετά-διαδικασία Γνώσης	15
Εικόνα 2.1 Διαδικασίες που εμπλέκονται στην εξόρυξη δεδομένων	21
Εικόνα 3.1 Αρχιτεκτονική επτά επιπέδων του Σημασιολογικού Ιστού	42
Εικόνα 4.1 Protégé-Frames	58
Εικόνα 4.2 Protégé OWL	59
Εικόνα 4.3 Περιβάλλον Chimaera	60
Εικόνα 4.4 Υποδομή οντολογίας ΚΑΟΝ	61
Εικόνα 4.5 Διεθνής Ταξινόμηση των Νόσων – ICD 11	66
Εικόνα 4.6 Οντολογία Παραδοσιακής Αφρικανικής Ιατρικής	70
Εικόνα 4.7 Οντολογία Δυσπλασίας των Οστών	71
Εικόνα 4.8 Οντολογία Συμπεριφοράς Κυττάρων	72
Εικόνα 4.9 Οντολογία Ανατομίας	74
Εικόνα 4.10 Οντολογία Νεύρο - Συμπεριφοράς	75
Εικόνα 4.11 ΙΑΤΡΟΛΕΞΗ – Η Ελληνική Οντολογία	77
Εικόνα 5.1 Όλες οι κλάσεις	85
Εικόνα 5.2 Κλάση A και οι υποκλάσεις της	87
Εικόνα 5.3 Υποκλάσεις της A	88

Εικόνα 5.4 Κλάση Β	89
Εικόνα 5.5 Υποκλάσεις της Β.....	90

ΕΙΣΑΓΩΓΗ

Η επιστήμη της εξόρυξης δεδομένων έχει ευρύ φάσμα εφαρμογών και αποτελεί ένα σημαντικό εργαλείο της στατιστικής, της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων.

Στόχος της πτυχιακής αυτής είναι να εξετάσει τις δυνατότητες στην εξόρυξη και στη παρουσίαση των αποτελεσμάτων της εξόρυξης δεδομένων μέσω σημασιολογικού ιστού, καθώς και να ερευνήσει και να ενημερώσει σχετικά με τη δημιουργία, τη δομή, τη μορφή και τη χρήση των οντολογιών στην παροχή υπηρεσιών υγείας.

Για την καλύτερη κατανόηση του περιεχομένου, η πτυχιακή αποτελείται από 5 κεφάλαια. Ειδικότερα, το **1^ο Κεφάλαιο** παρουσιάζει τις βάσεις γνώσης, τη διαχείριση των συστημάτων γνώσης με βάση τις οντολογίες, καθώς και τα έμπειρα συστήματα. Στο **2^ο Κεφάλαιο** παρουσιάζεται η έννοια της εξόρυξης δεδομένων καθώς και οι βασικές κατηγορίες που χρησιμοποιούνται για τον σκοπό αυτό. Επίσης, παρουσιάζονται οι βασικοί αλγόριθμοι που χρησιμεύουν στην εξόρυξη δεδομένων, τα πρότυπα που εφαρμόζονται και πώς γίνεται η εξόρυξη από ιατρικά δεδομένα. Στο **Κεφάλαιο 3** περιγράφονται οι τεχνολογίες του διαδικτύου και συγκεκριμένα, η τεχνολογία του σημασιολογικού ιστού, οι γλώσσες που χρησιμοποιούνται και η αρχιτεκτονική στη οποία στηρίζεται η τεχνολογία αυτή. Στο **4^ο Κεφάλαιο** παρουσιάζεται η έννοια της οντολογίας και πώς αυτή χρησιμοποιείται στο χώρο της υγείας. Επίσης, παρουσιάζονται και κάποια ενδεικτικά παραδείγματα χρήσης οντολογιών σε θέματα υγείας. Τέλος, στο **5^ο Κεφάλαιο** παρουσιάζονται τα MDS και μια πρακτική εφαρμογή οντολογίας υγείας, που δημιουργήθηκε στα πλαίσια της παρούσας εργασίας.

1 ΚΕΦΑΛΑΙΟ: ΒΑΣΕΙΣ ΓΝΩΣΗΣ ΚΑΙ ΕΜΠΕΙΡΑ ΣΥΣΤΗΜΑΤΑ

1.1 ΒΑΣΕΙΣ ΓΝΩΣΗΣ

Οι Βάσεις Γνώσης είναι βάσεις δεδομένων που αποθηκεύουν τη ρητή γνώση που παράγεται σε έναν οργανισμό. Κύριος στόχος των βάσεων αυτών είναι η συλλογή και καταγραφή λύσεων σε προβλήματα που παρουσιάστηκαν και αντιμετωπίστηκαν στο παρελθόν, με σκοπό να αποφευχθεί η επανάληψη των ίδιων λαθών και να γίνουν πιο γρήγορα οι απαραίτητες διαδικασίες. Εκτός αυτού του τρόπου χρήσης τους, οι Βάσεις Γνώσης μπορούν να χρησιμοποιηθούν και για την εκπαίδευση νέου προσωπικού.

Μία Βάση Γνώσης έχει τα εξής βασικά γνωρίσματα:

- η μορφή των άρθρων που περιλαμβάνει είναι ομοιόμορφη,
- το περιεχόμενο των άρθρων της μπορεί να επικαιροποιηθεί,
- διαθέτει μηχανή αναζήτησης
- και το περιεχόμενό της είναι δομημένο.

Αυτό το τελευταίο στοιχείο είναι πολύ βασικό για τις Βάσεις Γνώσης καθώς σε αυτή ακριβώς τη δομή στηρίζονται τα υπόλοιπα χαρακτηριστικά που αναφέραμε παραπάνω.¹

1.2 ΣΥΣΤΗΜΑΤΑ ΒΑΣΙΣΜΕΝΑ ΣΕ ΓΝΩΣΗ

Ο όρος «συστήματα βασισμένα σε γνώση» (knowledge based systems) τις περισσότερες φορές αναφέρεται σε υπολογιστικά συστήματα τα οποία χρησιμοποιώντας συγκεκριμένη γνώση ενός πεδίου εφαρμογής θέλουν να παίξουν το ρόλο του ειδικού πάνω στο συγκεκριμένο τομέα. Ο κύριος σκοπός των συστημάτων

¹ Δάρλας, Γ.

αυτών είναι η δημιουργία ενός μοντέλου της διαδικασίας λήψης αποφάσεων του ειδικού και η προσομοίωση της συμπεριφοράς του σε καταστάσεις που αφορούν το πεδίο της εφαρμογής.

Σημαντικό πλεονέκτημα των συστημάτων βασισμένων σε γνώση είναι ο συνδυασμός των περισσοτέρων του ενός ειδικών για την λήψη αποφάσεων. Η συνδυασμένη γνώση υποδηλώνει πως το σύστημα έχει στην διάθεση του μεγαλύτερο όγκο γνώσης γεγονός που μπορεί να οδηγήσει σε ποιοτικότερη και πιο αντικειμενική λήψη αποφάσεων. Εκτός αυτού είναι σημαντικό να αναφερθεί πως η καταγραφή της γνώσης μεγιστοποιεί τις δυνατότητες διαχείρισης και μεταφοράς της σε μεγαλύτερο πλήθος ανθρώπων. Παράλληλα, μιας και η διαδικασία λήψης αποφάσεων είναι καθορισμένη, τα συστήματα αυτά λαμβάνουν υπόψη τους όλες τις παραμέτρους που έχουν προσδιοριστεί ως κρίσιμες για τη λήψη μίας απόφασης.²

Σε ιδανικές συνθήκες, ένα σύστημα βασισμένο σε γνώση θα ήταν ικανό να αντικαταστήσει τον ειδικό του συγκεκριμένου πεδίου. Παρόλα αυτά, η πρακτική εφαρμογή αντίστοιχων συστημάτων έχει δείξει πως υπάρχουν αρκετοί περιορισμοί για την επίτευξη του παραπάνω στόχου.

Είναι αποδεδειγμένο ότι ο ανθρώπινος εγκέφαλος είναι ικανότερος από τον υπολογιστή στη συνδυαστική σκέψη. Το γεγονός αυτό οφείλεται στο ότι ο άνθρωπος μπορεί και προσαρμόζεται γρήγορα και με επιτυχία σε αλλαγές καταστάσεων. Εκτός αυτού, σε αντίθεση με τα συστήματα διαχείρισης γνώσης όπου η γνώση βρίσκεται περιορισμένη στο πεδίο της εφαρμογής τους, η γνώση ενός ανθρώπου δεν περιορίζεται μόνο σε ένα συγκεκριμένο πεδίο εφαρμογής, γεγονός που αρκετές φορές είναι σημαντικός παράγοντας για την σωστή λήψη μίας κρίσιμης απόφασης. Συμπερασματικά, θα μπορούσαμε να πούμε ότι η γενική ικανότητα της ανθρώπινης αντίληψης δύσκολα προσομοιώνεται από ένα υπολογιστικό σύστημα και ειδικότερα από ένα σύστημα που είναι προσανατολισμένο στην επίλυση συγκεκριμένων προβλημάτων.³

² Παπαταξιάρχης, Β., (2008).

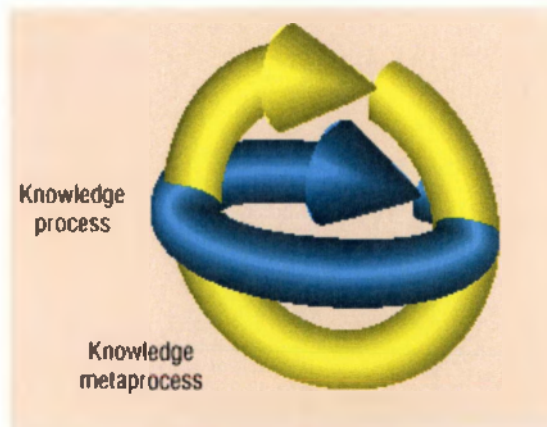
³ Παπαταξιάρχης, Β., (2008).

1.3 ΣΥΣΤΗΜΑΤΑ ΔΙΑΧΕΙΡΙΣΗΣ ΓΝΩΣΗΣ ΒΑΣΙΣΜΕΝΑ ΣΕ ΟΝΤΟΛΟΓΙΕΣ

Οι οντολογίες διαθέτουν έναν ικανοποιητικό μηχανισμό αναπαράστασης και διαμοίρασης της γνώσης σε περιπτώσεις όπου η γνώση χρειάζεται να μοντελοποιηθεί, να δομηθεί και να διασυνδεθεί.

Στη παράγραφο αυτή παρουσιάζεται ένα μοντέλο με βάση το οποίο θα μπορούσαν τα συστήματα διαχείρισης γνώσης να στηρίζονται σε οντολογίες. Το μοντέλο αυτό περιλαμβάνει δύο ξεχωριστές διαδικασίες καθεμία από τις οποίες αποτελείται από επιμέρους φάσεις. Σύμφωνα με το μοντέλο αυτό γίνεται διάκριση μεταξύ διαδικασίας γνώσης και μετά-διαδικασίας γνώσης. Οι δύο αυτές διαδικασίες είναι αναδραστικές, δηλαδή οι διάφορες φάσεις από τις οποίες αποτελούνται επαναλαμβάνονται κατά τη διάρκεια του χρόνου, και είναι ορθογώνιες μεταξύ τους.⁴

Η διαδικασία η οποία ακολουθείται με σκοπό να έχουμε ένα σύστημα διαχείρισης γνώσης βασισμένο στην οντολογία και εμφανίζεται στο μοντέλο ως οριζόντιος άξονας ονομάζεται διαδικασία γνώσης, ενώ η διαδικασία που πρέπει να ακολουθηθεί ώστε να δημιουργηθεί η κατάλληλη οντολογία και εμφανίζεται στο μοντέλο ως κάθετος άξονας ονομάζεται μετά-διαδικασία γνώσης.



Εικόνα 1.1 Διαδικασία γνώσης - Μετά-διαδικασία γνώσης: Δύο Ορθογώνιες, Αναδραστικές Διαδικασίες

Πηγή: Κοκκινίδης, Γ., Κοφφινά, Ι., Παπαγγελής, Μ.

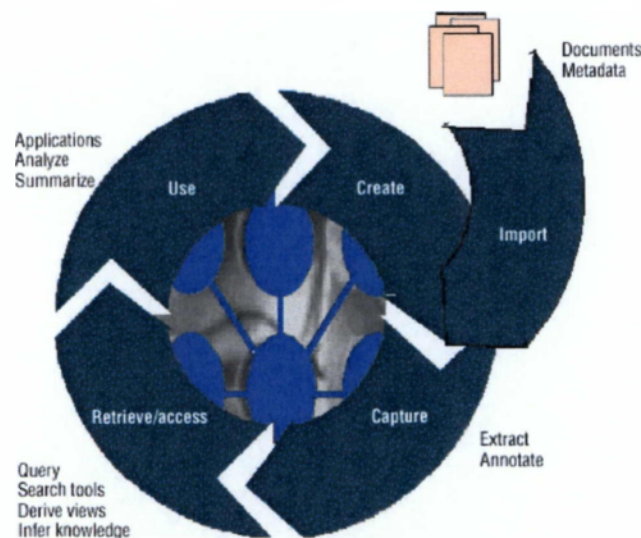
⁴ Κοκκινίδης, Γ., Κοφφινά, Ι., Παπαγγελής, Μ.

1.3.1 Διαδικασία γνώσης

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

- Δημιουργία Γνώσης
- Εισαγωγή Γνώσης
- Δέσμευση Γνώσης
- Ανάκτηση και Προσπέλαση Γνώσης
- Χρησιμοποίηση Γνώσης

Στην παρακάτω εικόνα παρουσιάζονται τα πέντε βήματα για την ανάπτυξη ενός συστήματος διαχείρισης γνώσης.



**Εικόνα 1.2 Διαδικασία ανάπτυξης ενός συστήματος διαχείρισης γνώσης ή
Διαδικασία γνώσης**

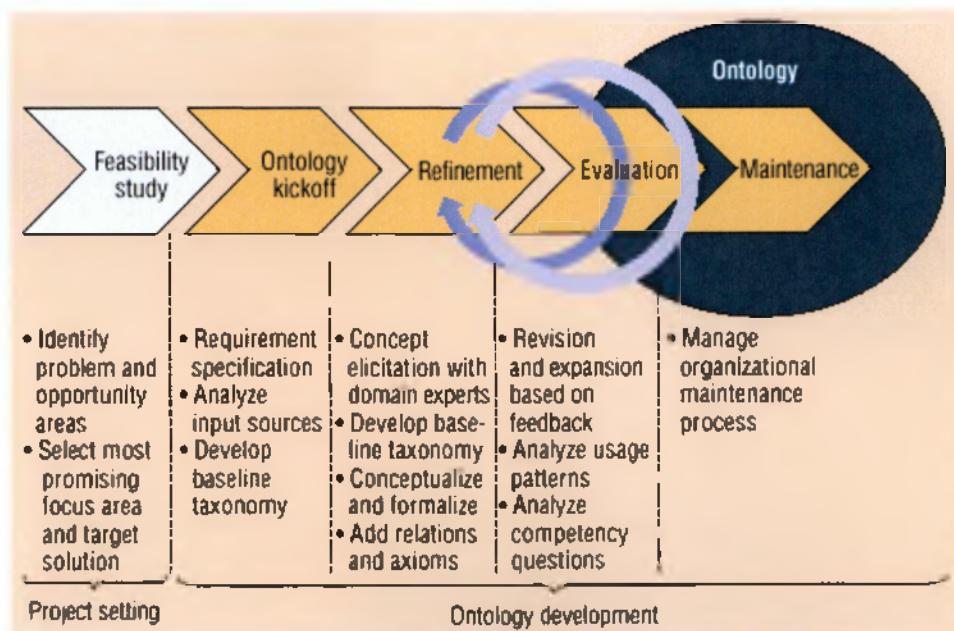
Πηγή: Κοκκινίδης, Γ., Κοφφινά, Ι., Παπαγγελής, Μ.

1.3.2 Μετά – Διαδικασία γνώσης

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

- Φάση Επιτευξιμότητας (Feasibility Study)
- Φάση Έναρξης (Kickoff Phase)
- Φάση τελειοποίησης (Refinement Phase)
- Φάση Αξιολόγησης (Evaluation Phase)
- Φάση Συντήρησης (Maintenance Phase)

Η παρακάτω εικόνα παρουσιάζει τις επιμέρους φάσεις για την μετα-διαδικασία γνώσης.



Εικόνα 1.3 Η Διαδικασία Ανάπτυξης Οντολογίας ή Μετά-διαδικασία Γνώσης

Πηγή: Κοκκινίδης, Γ., Κοφφινά, Ι., Παπαγγελής, Μ.

1.4 ΕΜΠΕΙΡΑ ΣΥΣΤΗΜΑΤΑ

Η αρχική προσπάθεια που έγινε από τους επιστήμονες στο πεδίο της Τεχνητής Νοημοσύνης είχε σαν στόχο της τη δημιουργία συστημάτων γενικής επίλυσης προβλημάτων. Μέσα από αυτή την προσπάθεια, έγινε εμφανές ότι η αποδοτική και αποτελεσματική επίλυση ρεαλιστικών προβλημάτων είναι άμεσα συνδεδεμένη με τη χρήση συγκεκριμένης γνώσης, γεγονός το οποίο οδήγησε στη δημιουργία των συστημάτων βάσης γνώσης (knowledge based systems).

Τα έμπειρα συστήματα είναι συστήματα βάσης γνώσης, τα οποία μπορούν να οργανωθούν ως συστήματα παραγωγής, πλαισίων κτλ. Σε γενικό επίπεδο ένας κοινός αποδεκτός ορισμός για το έμπειρο σύστημα είναι: *«έμπειρο σύστημα είναι ένα υπολογιστικό σύστημα, το οποίο μπορεί να επιλύσει αποδοτικά και αποτελεσματικά ρεαλιστικά προβλήματα, η επίλυση των οποίων εκ μέρους του ανθρώπου συνεπάγεται την ύπαρξη κάποιας μορφής εμπειρογνωμοσύνης»*. Γενικός σκοπός της τεχνολογίας των Έμπειρων Συστημάτων είναι η εξάπλωση, σε ευρεία κλίμακα, της ικανότητας επίλυσης προβλημάτων, που απαιτούν εμπειρογνωμοσύνη, για κοινωνικούς, οικονομικούς ή άλλους λόγους.⁵

⁵ Κεραύνου, Ε., 2000.

2 ΚΕΦΑΛΑΙΟ: Η ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

2.1 ΕΙΣΑΓΩΓΗ

Η εξόρυξη δεδομένων αποτελεί ένα σημαντικό εργαλείο εξεύρεσης πληροφοριών ή προτύπων από μεγάλες βάσεις δεδομένων συνδυάζοντας

- αρχές της στατιστικής,
- της τεχνητής νοημοσύνης,
- της μηχανικής μάθησης
- και των συστημάτων βάσεων δεδομένων.

Η επιστήμη της εξόρυξης δεδομένων έχει ευρύ φάσμα εφαρμογών. Στο παρόν κεφάλαιο αναλύεται η έννοια της εξόρυξης δεδομένων, παρατίθεται ο κύκλος ζωής της διαδικασίας και αναλύονται οι βασικοί αλγόριθμοι που χρησιμοποιούνται. Επίσης, παρουσιάζονται οι βασικές κατηγορίες μεθόδων εξόρυξης δεδομένων και τα πρότυπα εξόρυξης αυτών. Τέλος, γίνεται μία σύντομη αναφορά στην εξόρυξη δεδομένων που σχετίζονται με την ιατρική και τους τομείς της υγείας.

2.2 ΟΡΙΣΜΟΙ

Τα τελευταία χρόνια, οι δυνατότητές μας για τη δημιουργία και τη συλλογή δεδομένων έχουν αυξηθεί ταχύτατα. Αυτή η εκρηκτική ανάπτυξη των βάσεων δεδομένων έχει δημιουργήσει μια επείγουσα ανάγκη για τη δημιουργία νέων τεχνικών και εργαλείων, τα οποία μπορούν έξυπνα και αυτόματα να μετατρέψουν τα πρωτογενή δεδομένα σε χρήσιμες πληροφορίες και γνώσεις. Ως εκ τούτου, η εξόρυξη δεδομένων έχει καταστεί ένα ερευνητικό πεδίο με ολοένα και μεγαλύτερη σημασία.

Η εξόρυξη δεδομένων είναι η διαδικασία εξαγωγής έμμεσων, παλαιότερα γνωστών και δυνητικά χρήσιμων πληροφοριών από τα δεδομένα της βάσης δεδομένων.⁶ Κάποιοι άλλοι ερευνητές υποστηρίζουν ότι η εξόρυξη δεδομένων είναι η διαδικασία της εξερεύνησης και ανάλυσης πολλών δεδομένων, με αυτόματα ή ημιαυτόματα μέσα, προκειμένου να ανακαλυφθούν σημαντικά πρότυπα και κανόνες. Αυτές οι νέες πληροφορίες που προκύπτουν μπορούν να χρησιμοποιηθούν σε διάφορους τομείς, όπως για παράδειγμα στην υποστήριξη της λήψης αποφάσεων, στις προβλέψεις και στις εκτιμήσεις σημαντικών επιχειρηματικών αποφάσεων, κτλ.

Υπάρχουν, επίσης, πολλοί άλλοι όροι που εμφανίζονται σε ορισμένα επιστημονικά άρθρα οι οποίοι έχουν την ίδια ή ελαφρώς διαφορετική έννοια. Τέτοιοι όροι είναι οι ακόλουθοι: ανακάλυψη γνώσης σε βάσεις δεδομένων (knowledge discovery in databases), εξόρυξη γνώσης από βάσεις δεδομένων (knowledge mining from databases), εξαγωγή γνώσης (knowledge extraction), αρχαιολογία δεδομένων (data archaeology), εκβάθυνση δεδομένων (data dredging), ανάλυση δεδομένων (data analysis) και άλλα.

Ανέκαθεν, διάφοροι αναλυτές ασχολούνταν με την διαδικασία της εξαγωγής χρήσιμων πληροφοριών από καταγεγραμμένα δεδομένα. Στην σύγχρονη εποχή, ο αυξημένος όγκος των δεδομένων απαιτεί η εξαγωγή γνώσης να γίνεται με χρήση υπολογιστικών προσεγγίσεων. Πράγματι, ο όγκος των δεδομένων αυτών, τα οποία ονομάζονται σύνολα δεδομένων (datasets), έχει αυξηθεί τόσο σε μέγεθος, όσο και σε πολυπλοκότητα. Το αποτέλεσμα είναι οι παραδοσιακές μέθοδοι απ' ευθείας ανάλυσης να δίνουν την θέση τους σε αυτόματες μεθόδους ανάλυσης, οι οποίες χρησιμοποιούν πιο πολύπλοκα και εξεζητημένα εργαλεία.⁷

Η μέθοδος της εξόρυξης δεδομένων χρησιμοποιεί τα δεδομένα ως "πρώτη ύλη" και χρησιμοποιώντας έναν προκαθορισμένο αλγόριθμο ομαδοποιεί τις τεράστιες ποσότητες αυτών σύμφωνα με τα επιθυμητά κριτήρια που μπορεί να είναι χρήσιμα.

⁶ Frawley, W. J., Piatetsky-Shapiro G., and Matheus, C. J., 1992.

⁷ Γούλου Ζ., (2012).

2.3 ΚΥΚΛΟΣ ΖΩΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

Ο κύκλος ζωής εξόρυξης δεδομένων αποτελείται από τα ακόλουθα βήματα⁸:

ΒΗΜΑ 1: Κατανόηση του προβλήματος: Το πρώτο βήμα είναι να γίνει κατανοητό το πρόβλημα και να λάβει μια μορφή. Αυτό είναι αναγκαίο για να μπορούν να παραχθούν χρήσιμα συμπεράσματα και να επιλεγθούν οι κατάλληλες μέθοδοι.

ΒΗΜΑ 2: Συλλογή δεδομένων (data collection): Το δεύτερο βήμα της εξόρυξης δεδομένων είναι η συλλογή δεδομένων. Τα δεδομένα αποθηκεύονται σε ένα σύστημα βάσης δεδομένων. Τα δεδομένα κάποιες φορές δεν είναι πολλά και πιθανόν να πρέπει να συλλεχθούν περισσότερα καθώς επίσης μπορεί να μην περιέχουν όλες τις απαιτούμενες πληροφορίες και να πρέπει να εμπλουτιστούν.

ΒΗΜΑ 3: Καθαρισμός των δεδομένων (data cleaning): Είναι η διαδικασία ανίχνευσης και διόρθωσης (ή αφαίρεσης) ανακριβών, λανθασμένων ή ελλιπών στοιχείων από τη βάση δεδομένων. Είναι ένα βήμα το οποίο απαιτεί συνήθως αρκετό χρόνο και βαθιά γνώση των δεδομένων.

ΒΗΜΑ 4: Εξαγωγή των χαρακτηριστικών (feature construction/extraction): Αφορά στην απλούστευση των πόρων που απαιτούνται για να περιγραφεί ένα σύνολο δεδομένων με ακρίβεια. Το κύριο ζητούμενο κατά τη διαδικασία ευρέσεως των κατάλληλων χαρακτηριστικών είναι να ανακαλυφθεί το κατάλληλο υποσύνολο των χαρακτηριστικών παραμέτρων, τέτοιο ώστε να αντιπροσωπεύει με τον πληρέστερο τρόπο όλες τις περιπτώσεις του δείγματος. Αυτό έχει σαν αποτέλεσμα την περαιτέρω απλούστευση των προς επεξεργασία στοιχείων με τη μείωση της διάστασης τους. Οι

⁸ Bao, H. T.

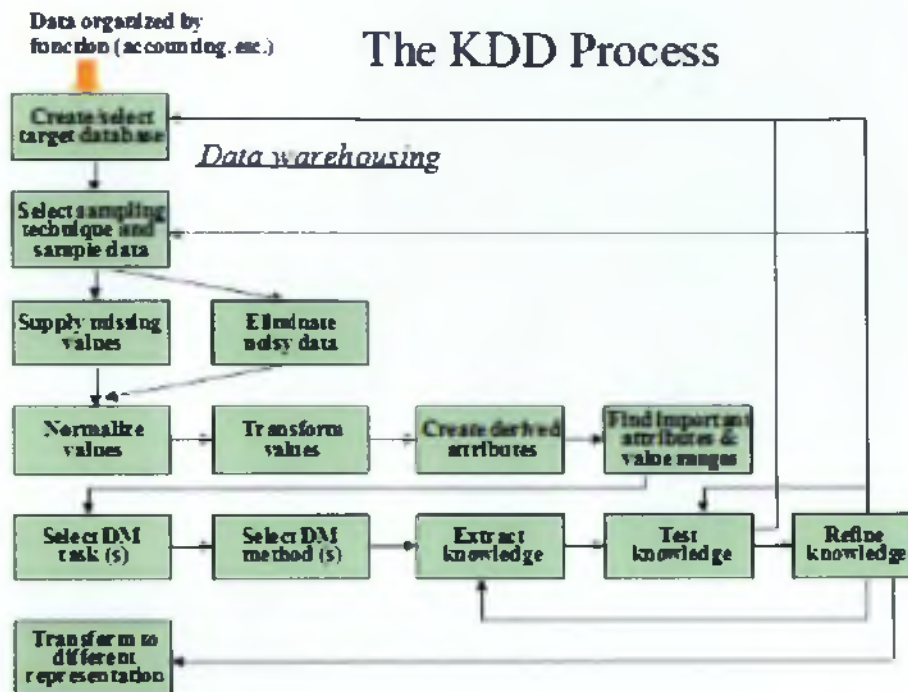
μέθοδοι ευρέσεως κατάλληλων χαρακτηριστικών είναι ανεξάρτητες από τις μεθόδους εξόρυξης δεδομένων, αν και κάποιες από αυτές έχουν ενσωματωθεί σε κάποιες μεθόδους εξόρυξης δεδομένων.

ΒΗΜΑ 5: Επιλογή αλγορίθμου και παραμέτρων (algorithm and parameter selection): Γίνεται η επιλογή του κατάλληλου αλγορίθμου εξόρυξης δεδομένων και ο καθορισμός των απαιτούμενων παραμέτρων. Σε αυτό το στάδιο, χρειάζεται συνεργασία με αυτούς που κατέχουν τη γνώση του συγκεκριμένου αντικείμενου. Η ανάπτυξη του μοντέλου είναι ο πυρήνας της διαδικασίας εξόρυξης δεδομένων, αν και δεν είναι τόσο απαιτητικό στάδιο σε πόρους και χρόνο όσο ο καθαρισμός των δεδομένων.

ΒΗΜΑ 6: Υποβολή έκθεσης αποτελεσμάτων: Σε αυτό το στάδιο παραδίδονται οι εκθέσεις των συμπερασμάτων ώστε να παρθούν αποφάσεις όπου χρειάζεται.

ΒΗΜΑ 7: Πρόβλεψη: Τα αποκαλυπτόμενα πρότυπα χρησιμοποιούνται ώστε να προβλεφθούν μελλοντικές συμπεριφορές, τάσεις και μοτίβα. Για να δώσουμε τις προβλέψεις, πρέπει να έχουμε εκτός από το εκπαιδευμένο πρότυπο και ένα σύνολο νέων υποθέσεων.

ΒΗΜΑ 8: Αποτίμηση: Πρόκειται για το τελευταίο στάδιο της δημιουργία του πλήρους προβλεπτικού συστήματος. Σε αυτό χρησιμοποιούνται κατάλληλες μετρικές μέθοδοι για την αποτίμηση της απόδοσης των αλγορίθμων εξόρυξης δεδομένων. Έχει αναπτυχθεί ένα πλήθος από αυτές τις μεθόδους με αντιπροσωπευτικότερη και πλέον αξιόπιστη τη γραφική μέθοδο αποτίμησης των καμπυλών Receiving Operating Characteristics (ROC). Η αποτίμηση με τις καμπύλες ROC βασίζεται στον πίνακα συσχέτισης (confusion matrix).



Εικόνα 2.1 Διαδικασίες που εμπλέκονται στην εξόρυξη δεδομένων

Πηγή: Bao, H. T.

2.4 ΒΑΣΙΚΕΣ ΚΑΤΗΓΟΡΙΕΣ ΜΕΘΟΔΩΝ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

Οι δύο κύριοι στόχοι της εξόρυξης δεδομένων είναι η περιγραφή και η πρόβλεψη. Η περιγραφή αφορά στην αναπαράσταση των δεδομένων μιας πολύπλοκης βάσης δεδομένων με ένα κατανοητό και αξιοποιήσιμο τρόπο και η πρόβλεψη στην ανεύρεση κρυμμένων προτύπων, αποκάλυψη μη αναμενόμενων σχέσεων και πρόβλεψη μελλοντικών συνθηκών, συμπεριφορών και τάσεων.⁹

⁹ Bao, H. T.

2.4.1 Ταξινόμηση (Classification)

Η ταξινόμηση είναι μία από τις δημοφιλέστερες κατηγορίες εξόρυξης δεδομένων. Επιχειρησιακά προβλήματα όπως η διαχείριση κινδύνου επιλύονται συνήθως με την ταξινόμηση. Η ταξινόμηση αφορά στην κατάταξη των περιπτώσεων σε κατηγορίες με βάση μια προβλέψιμη ιδιότητα. Κάθε περίπτωση έχει ένα σύνολο ιδιοτήτων, μια από τις οποίες χαρακτηρίζεται ως ιδιότητα κατηγορίας (προβλέψιμη ιδιότητα). Σκοπός είναι να παραχθεί ένα μοντέλο το οποίο θα περιγράφει την ιδιότητα κατηγορίας σαν συνάρτηση των ιδιοτήτων εισόδου. Χαρακτηριστικοί αλγόριθμοι ταξινόμησης είναι τα δέντρα απόφασης, τα τεχνητά νευρωνικά δίκτυα, τα Naïve Bayes κ.α..¹⁰

Οι βασικοί αντικειμενικοί σκοποί των ταξινομητών δέντρων αποφάσεων είναι¹¹:

- 1 Να ταξινομήσουν σωστά όσο το δυνατόν περισσότερο ποσοστό από το σύνολο εκπαίδευσης (training set)
- 2 Να γενικεύσουν πέρα από το δείγμα εκπαίδευσης, έτσι ώστε ένα νέο και άγνωστο δείγμα να μπορεί να ταξινομηθεί με όσο το δυνατό μεγαλύτερη ακρίβεια.
- 3 Να μπορούν να κάνουν ενημέρωση (update), όταν γίνονται διαθέσιμα περισσότερα δεδομένα.
- 4 Να έχουν όσο πιο απλή δομή γίνεται.

2.4.2 Ομαδοποίηση (Clustering)

Η ομαδοποίηση ή αλλιώς κατάτμηση χρησιμοποιείται για να προσδιορίσει τους φυσικούς σχηματισμούς ομάδων από τα δεδομένα με βάση ένα σύνολο κοινών ιδιοτήτων. Οι περιπτώσεις μέσα στην ίδια ομάδα έχουν περισσότερες ή λιγότερες παρόμοιες τιμές ιδιοτήτων. Συγκεκριμένα, έστω ένα σύνολο δεδομένων πελατών που περιέχει δύο ιδιότητες: ηλικία και εισόδημα. Ένας αλγόριθμος ομαδοποίησης συγκεντρώνει το σύνολο δεδομένων με βάση αυτές τις ιδιότητες. Η ομάδα 1 περιέχει το νεότερο πληθυσμό με χαμηλό εισόδημα. Η ομάδα 2 περιέχει τους μέσης ηλικίας πελάτες με υψηλό εισόδημα. Η ομάδα 3 τους μεγαλύτερης ηλικίας με χαμηλό

¹⁰ MacLennan, J., Zhao, H. T., (2005).

¹¹ Safavian, S.R., and Landgrebe, D., (1991).

εισόδημα και ούτω καθεξής. Οι αλγόριθμοι ομαδοποίησης χαρακτηρίζονται ως μη επιβλέψιμοι (non supervised). Καμία ιδιότητα δεν χρησιμοποιείται για να καθοδηγήσει τη διαδικασία κατάρτισης, όλες οι ιδιότητες εισόδου αντιμετωπίζονται εξίσου.¹²

2.4.3 Ένωση (Association)

Η ένωση είναι επίσης μια από τις δημοφιλείς μεθόδους εξόρυξης δεδομένων. Μια χαρακτηριστική εφαρμογή της ένωσης είναι η ανάλυση του πίνακα πωλήσεων μιας επιχείρησης ώστε προσδιορισθούν εκείνα τα προϊόντα τα οποία πωλούνται συχνά μαζί στο ίδιο καλάθι αγορών. Η συνήθης χρήση της ένωσης είναι να προσδιορίσει τα κοινά σύνολα στοιχείων (συχνά σετ στοιχείων) και τους κανόνες δημιουργίας αυτών με σκοπό την στοχευόμενη πώληση.

Από την άποψη της ένωσης, κάθε προϊόν, ή γενικότερα κάθε ιδιότητα/αξία θεωρείται στοιχείο. Οι περισσότεροι αλγόριθμοι τύπου ένωσης βρίσκουν τα συχνά σετ στοιχείων με την ανίχνευση του συνόλου δεδομένων πολλές φορές. Το κατώτατο όριο συχνότητας (υποστήριξης) καθορίζεται από το χρήστη πριν την επεξεργασία. Παραδείγματα τέτοιων αλγορίθμων είναι οι αλγόριθμοι που βασίζονται στην υποστήριξη κανόνα (support rule) και στην εμπιστοσύνη κανόνα (confidence rule).¹³

2.4.4 Παλινδρόμηση (Regression)

Η μέθοδος της παλινδρόμησης είναι παρόμοια με την ταξινόμηση.

Οι τεχνικές παλινδρόμησης χρησιμοποιούνται ευρέως στον τομέα της στατιστικής. Η γραμμική συµμεταβολή και η λογιστική παλινδρόμηση αποτελούν δημοφιλέστερες μεθόδους παλινδρόμησης. Άλλες μέθοδοι παλινδρόμησης περιλαμβάνουν τα δέντρα παλινδρόμησης και τα τεχνητά νευρωνικά δίκτυα. Η παλινδρόμηση μπορεί να εφαρμοστεί σε διάφορους

¹² MacLennan, J., Zhao, H. T., (2005).

¹³ MacLennan, J., Zhao, H. T., (2005).

τομείς, όπως στη μετεωρολογία για να προβλεφθούν οι ταχύτητες ανέμου με βάση τη θερμοκρασία, την πίεση αέρα, και την υγρασία.

2.4.5 Πρόβλεψη (Forecasting)

Η πρόβλεψη είναι μια ακόμα σημαντική μέθοδος εξόρυξης δεδομένων. Μπορεί να βοηθήσει στην απάντηση ερωτημάτων όπως ποια θα είναι η αξία ενός αποθεματικού αύριο; Ποιό θα είναι το ποσοστό πωλήσεων αναψυκτικών για τον επόμενο μήνα; Τα δεδομένα εισόδου είναι τύπου χρονικής σειράς. Οι μέθοδοι πρόβλεψης εξετάζουν γενικές τάσεις και περιοδικότητα. Η δημοφιλέστερη μέθοδος πρόβλεψης θεωρείται η ARIMA, η οποία υλοποιεί τη μεθοδολογία Auto Regressive Integrated Moving Average Model.

2.4.6 Ανάλυση ακολουθίας (Sequence Analysis)

Η ανάλυση ακολουθίας χρησιμοποιείται για τη δημιουργία μοντέλων σε μια διακριτή σειρά. Μια ακολουθία αποτελείται από μια σειρά διακριτών τιμών (ή καταστάσεων). Μια ακολουθία DNA είναι μια μακρά ακολουθία από τέσσερα διαφορετικά μέρη: Αδενίνη (Adenine), Θυμίνη (Thymine), Κυτοσίνη (Cytosine) και Γουανίνη (Guanine). Μια επιλογή στον Παγκόσμιο Ιστό είναι μια ακολουθία από ιστοσελίδες. Οι αγορές πελατών μπορούν επίσης να διαμορφωθούν ως στοιχεία ακολουθίας. Χαρακτηριστικό παράδειγμα αποτελεί ένας πελάτης που αγοράζει αρχικά έναν υπολογιστή, έπειτα ένα μικρόφωνο και τελικά μια Web κάμερα. Το κοινό των μεθόδων ανάλυσης ακολουθίας και ένωσης (association) είναι ότι κάθε μεμονωμένη περίπτωση περιέχει ένα σύνολο στοιχείων ή καταστάσεων. Η διαφορά τους είναι ότι οι μέθοδοι ανάλυσης ακολουθίας αναλύουν τις μεταβάσεις καταστάσεων ενώ η μέθοδος ένωσης θεωρεί κάθε στοιχείο ίσο και ανεξάρτητο.

Σύμφωνα με τη μέθοδο ανάλυσης ακολουθίας, το να αγοράσει κάποιος έναν υπολογιστή προτού αγοράσει μικρόφωνο είναι μια διαφορετική ακολουθία από το να αγοράσει μικρόφωνο πριν από έναν υπολογιστή. Για έναν αλγόριθμο ένωσης, αυτά θεωρούνται όμοια. Η ανάλυση ακολουθίας είναι ένας σχετικά νέος τρόπος εξόρυξης δεδομένων. Είναι αρκετά σημαντική σε δύο κυρίως τύπους εφαρμογών: Ανάλυση

Παγκόσμιου Ιστού και ανάλυση DNA. Υπάρχουν διάφορες τεχνικές ανάλυσης ακολουθίας διαθέσιμες όπως οι αλυσίδες Markov και άλλες.

2.5 ΟΙ ΒΑΣΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΤΗΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

Οι κύριοι τύποι αλγορίθμων που χρησιμοποιούνται στην εξόρυξη δεδομένων είναι οι εξής¹⁴:

- Τα τεχνητά νευρωνικά δίκτυα (artificial neural networks)
- Τα δέντρα αποφάσεων (decision trees)
- Οι γενετικοί αλγόριθμοι (genetic algorithms)
- Η επαγωγή κανόνων (rule induction)
- Η μέθοδος του κοντινότερου γείτονα (nearest neighbor method)
- Η απεικόνιση δεδομένων (data visualization)

2.5.1 Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks)

Είναι μη γραμμικά προγνωστικά μοντέλα, τα οποία μαθαίνουν μέσω εκπαίδευσης και δομικά μοιάζουν με βιολογικά νευρωνικά δίκτυα.

2.5.2 Τα δέντρα αποφάσεων (Decision Trees)

Είναι δεντρικές δομές που περιέχουν σύνολα αποφάσεων. Οι αποφάσεις αυτές παράγουν κανόνες, με σκοπό την ταξινόμηση κάποιου συνόλου δεδομένων.

¹⁴ Ahmed, S. R., (2004).

2.5.3 Οι γενετικοί αλγόριθμοι (Genetic Algorithms)

Είναι αλγόριθμοι βελτιστοποίησης, οι οποίοι χρησιμοποιούν τεχνικές όπως οι γενετικοί συνδυασμοί, η μετάλλαξη και η φυσική επιλογή.

2.5.4 Η επαγωγή κανόνων (Rule Induction)

Είναι η εξαγωγή κανόνων από δεδομένα χρήσιμων if - then κανόνων, με βάση τη στατιστική τους σπουδαιότητα.

2.5.5 Η μέθοδος του κοντινότερου γείτονα (Nearest Neighbor Method)

Η τεχνική αυτή ταξινομεί κάθε αντικείμενο κάποιου συνόλου δεδομένων, με βάση κάποιο συνδυασμό των κλάσεων των πιο όμοιων αντικειμένων. Συχνά ονομάζεται και knearest neighbor τεχνική.

2.5.6 Η απεικόνιση δεδομένων (Data Visualization)

Γίνεται οπτική ερμηνεία πολύπλοκων συναφειών σε πολυδιάστατα δεδομένα. Για τις απεικονίσεις των συναφειών χρησιμοποιούνται εργαλεία γραφικών.

2.6 ΠΡΟΤΥΠΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

2.6.1 Εξόρυξη δεδομένων OLE DB

Το πρότυπο Object Linking Embedding (OLE) DB για Data Mining παρουσιάστηκε από την εταιρεία Microsoft το 1999. Ο πυρήνας του προτύπου είναι η DMX (Data Mining eXtensions) μια γλώσσα 4ης γενιάς για τη δημιουργία των μοντέλων

εξόρυξης δεδομένων. Το API αυτό μπορεί να χρησιμοποιηθεί για τη δημιουργία αυτόνομων εφαρμογών εξόρυξης δεδομένων χρησιμοποιώντας οποιαδήποτε πλατφόρμα ανάπτυξης. Υποστηρίχθηκε από διάφορους προμηθευτές συμπεριλαμβανομένων των Angoss, KXEN, και Megaruter. Ουσιαστικά ενοποιεί δύο τεχνολογίες, της OLE DB και της SQL.

2.6.2 Εξόρυξη δεδομένων API ΤΗΣ JAVA

Το Application Interface (API) της Java για την εξόρυξη δεδομένων ανέπτυξε η εταιρία Oracle με την εμφάνιση της έκδοσης Oracle 10g. Ο σκοπός είναι να βοηθήσει στην ανάπτυξη αυτόνομων εφαρμογών με χρήση της γλώσσας Java. Το πρότυπο χωρίζει τα στοιχεία πηγής (φυσικά στοιχεία) και τα λογικά στοιχεία έννοιες. Τα στοιχεία πηγής μπορούν να είναι οποιαδήποτε αρχεία δεδομένων σχεσιακών βάσεων ή κειμένων.

2.6.3 Πρότυπα PMML

Η Predictive Model Markup Language (PMML) είναι μια γλώσσα σήμανσης βασισμένη στην XML. Αναπτύχθηκε από το Data Mining Group (DMG) για να επιτρέψει το σχεδιασμό μοντέλων για χρήση τους σε προβλεπτική ανάλυση και εξόρυξη δεδομένων που να μπορούν να χρησιμοποιούνται και σε άλλες εφαρμογές λογισμικού. Η PMML παρέχει μια μέθοδο ορισμού προτύπων ανεξάρτητη από τον κατασκευαστή της εφαρμογής, έτσι δεν υπάρχουν περιορισμοί ασυμβατότητας λογισμικού για την ανταλλαγή των μοντέλων μεταξύ των εφαρμογών. Επιτρέπει στους χρήστες να κάνουν χρήση ενός λογισμικού, να παράγουν ένα μοντέλο και να το χρησιμοποιήσουν σε ένα άλλο λογισμικό για να κάνουν την απεικόνιση, την ανάλυση και εκτίμηση του μοντέλου. Η PMML υποστηρίζει τον καθορισμό του περιεχομένου για τους ακόλουθους αλγορίθμους εξόρυξης δεδομένων, Polynomial regression, Logistic regression, General regression, Center-based clusters, Density-based clusters, Trees, Associations, Neural nets, Naïve Bayes, Sequences, Text model, Vector machine. Όλοι οι σημαντικοί προμηθευτές μοντέλων εξόρυξης δεδομένων υποστηρίζουν τα πρότυπα PMML.

2.7 ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΙΑΤΡΙΚΑ ΔΕΔΟΜΕΝΑ

Ο ρόλος της πληροφορικής έχει εδραιωθεί πλέον στα περισσότερα συστήματα υγείας ανά το κόσμο. Η χρησιμοποίηση ηλεκτρονικών υπολογιστών στα περισσότερα νοσοκομεία, αλλά και σε υπόλοιπους οργανισμούς που έχουν σχέση με την ιατρική περίθαλψη των ανθρώπων έδωσε τη δυνατότητα να αποθηκευτεί μεγάλος όγκος ιατρικών δεδομένων και να υπάρχει εύκολη πρόσβαση σε αυτά (medical databases). Τα δεδομένα αυτά, που αποθηκεύονται πλέον σε ψηφιακή μορφή, αφορούν εγγραφές ασθενών στο αρχείο, τις ασθένειες του κάθε ανθρώπου, τα φάρμακα που το χορηγούνται, τη θεραπεία που έχει ακολουθηθεί, δημογραφικά στοιχεία κτλ.

Ωστόσο, όλος αυτός ο όγκος των ιατρικών δεδομένων παρότι είναι πολύ χρήσιμος, παρουσιάζει δυσκολίες η μελέτη τους έτσι ώστε να εξαχθεί κάποια χρήσιμη πληροφορία για να πάρουμε μια απόφαση. Με τις υπάρχουσες μεθόδους ανάλυσης είναι εξαιρετικά δύσκολο να υπάρξει κάποια γνώση, γι' αυτό και είναι αναγκαία η υλοποίηση μεθόδων μέσα από υπολογιστικά συστήματα για να πραγματοποιηθεί μία σωστή ανάλυση των δεδομένων.¹⁵

Αυτό, έχει σαν αποτέλεσμα να είναι προτιμότερη η χρήση των τεχνικών εξόρυξης γνώσης για τα ιατρικά δεδομένα. Η εξόρυξη από ιατρικά δεδομένα είναι ένα από τα πιο ενδιαφέροντα και δύσκολα πεδία της εξόρυξης γνώσης. Τα δεδομένα που χρειάζονται για την εφαρμογή της εξόρυξης είναι τόσα πολλά, περίπλοκα και ετερογενή που καθιστούν την εξόρυξη μία πρόκληση για κάθε αναλυτή.¹⁶

Τα ιατρικά δεδομένα περιλαμβάνουν αποτελέσματα απεικονιστικά, γραφήματα, κείμενο από την κλινική εξέταση, εργαστηριακές μετρήσεις σε αριθμούς καθώς και δεδομένα σε άλλες μορφές. Έτσι, η εξόρυξη της κρυμμένης γνώσης από αυτά πρέπει να γίνει από συνδυασμό εικόνων, σχημάτων, κειμένων, αριθμών, το οποίο είναι δυσκολότερο από τις κλασσικές περιπτώσεις επεξεργασίας δεδομένων σε αριθμούς και κατηγορίες. Οι σύγχρονες τεχνολογίες εξόρυξης δεδομένων επιτρέπουν πλέον τη διαχείριση της ετερογενούς φύσης των ιατρικών δεδομένων. Πρώτιστα η δυνατότητα

¹⁵ Lavrac, N., (1999).

¹⁶ Delen, D., (2009).

επεξεργασίας της φυσικής γλώσσας και οι τεχνικές εξόρυξης δεδομένων από κείμενα επιτρέπουν την εξαγωγή πληροφορίας και γνώσης από τις ιατρικές σημειώσεις και τις κλινικές εξετάσεις.¹⁷ Ιατρική οντολογία και ορολογία μπορούν να ανιχνευθούν με τη χρήση μεθόδων εξόρυξης δεδομένων του παγκόσμιου ιστού και με τεχνικές εκμάθησης οντολογίας.

Η έμφαση που δίνεται σήμερα σε ιατρικές πράξεις στηριζόμενες σε τεκμήρια (Evidence-based Medicine - EBM) είναι ένας από τους κύριους λόγους που ενισχύουν την εφαρμογή μεθόδων εξόρυξης δεδομένων στην ιατρική πρακτική. Αυτές οι ιατρικές πράξεις συνήθως αποτελούν οδηγίες κλινικής πρακτικής ή κανόνες κλινικών αποφάσεων. Η διασύνδεση της εξόρυξης δεδομένων με την ιατρική πρακτική υλοποιείται με την υπάρχουσα προσπάθεια ανάπτυξης αυτόματων συστημάτων υποστήριξης απόφασης.¹⁸

Η ύπαρξη συγκεκριμένης οντολογίας επιτρέπει να ξεκαθαριστεί ποιοι κανόνες αποφάσεων επιβεβαιώνουν την ιατρική γνώση. Τα ιατρικά συστήματα υποστήριξης απόφασης υποστηρίζουν τους ιατρούς στη λήψη ιατρικών αποφάσεων.

Οι πιο σημαντικές δυσκολίες που συναντούν οι ερευνητές στο τομέα της εξόρυξης γνώσης στο τομέα της ιατρικής είναι:¹⁹

- 1 Η ετερογένεια των ιατρικών δεδομένων. Τα ιατρικά δεδομένα χωρίς να έχουν υποστεί κάποια προ-επεξεργασία είναι ογκώδη και ετερογενή. Τα δεδομένα αυτά αποθηκεύονται ύστερα από εξετάσεις του ασθενούς, από εικόνες (π.χ. ακτινογραφίες) και εργαστηριακά δεδομένα. Ο συνδυασμός αυτών των δεδομένων μπορεί να είναι αναγκαίος για την πρόγνωση, τη διάγνωση και τη περίθαλψη ενός ασθενούς, παρότι είναι τελείως διαφορετικά μεταξύ τους. Γι' αυτό το λόγο δεν μπορούν να αγνοηθούν. Τα κυριότερα προβλήματα που σχετίζονται με την ετερογένεια των δεδομένων είναι:
 - a. Το μέγεθος και η περιπλοκότητα των δεδομένων.
 - b. Η ερμηνεία του κάθε ιατρού.

¹⁷ Savova, G. K., Ogren, P. V., Duffy, P. H., Buntrock, J. D., Chute, C. G., (2008).

¹⁸ Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., Richardson, W. S., (2004).

¹⁹ Delen, D., (2009).

- c. Οι μετρικές όπως το sensitivity και το specificity που μετρούν την πιθανότητα επιτυχίας ή λάθους.
 - d. Η δυσκολία να χαρακτηριστούν μαθηματικά αυτά τα δεδομένα.
 - e. Η κανονικοποίηση των δεδομένων. Είναι μια μέθοδος μετασχηματισμού των δεδομένων, όπου οι συνεχείς μεταβλητές διαμορφώνονται έτσι ώστε να βρίσκονται στο διάστημα $[0,1]$. Μια τεχνική κανονικοποίησης είναι η min-max normalization.
- 2 Τα ηθικά, νομικά και κοινωνικά ζητήματα. Τα ιατρικά δεδομένα που συλλέγονται στις βάσεις δεδομένων αφορούν ανθρώπινα θέματα υγείας γι' αυτό το λόγο υπάρχει ένα μεγάλο ηθικό και νομικό πλαίσιο έτσι ώστε να καλύπτει τη προσβολή του ασθενούς ή την άσκοπη χρήση των δεδομένων του. Τα κύρια σημεία αυτών των ζητημάτων μπορούν να χωρισθούν σε 5 κατηγορίες :
- a. Η ιδιοκτησία των δεδομένων.
 - b. Οι νόμοι που διέπουν αυτά τα δεδομένα.
 - c. Η ιδιωτικότητα και η ασφάλεια των προσωπικών δεδομένων.
 - d. Τα αναμενόμενα προνόμια.
 - e. Τα ζητήματα διαχείρισης των δεδομένων.

3 ΚΕΦΑΛΑΙΟ: ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΔΙΚΤΥΟΥ

3.1 ΣΗΜΑΣΙΟΛΟΓΙΚΟΣ ΙΣΤΟΣ

3.1.1 Εισαγωγή

Η αξία και η σημασία του Παγκόσμιου Ιστού όσον αφορά στην σχέση ανθρώπου με υπολογιστή είναι πάρα πολύ μεγάλη. Παρόλα αυτά, ο Παγκόσμιος Ιστός έχει ένα βασικό ελάττωμα: είναι εξολοκλήρου βασισμένος στην συντακτική περιγραφή του περιεχομένου (content), κύριος στόχος της οποίας είναι η κατανόησή του από τον άνθρωπο. Η ανάπτυξη του Παγκόσμιου Ιστού ή World Wide Web έγινε έχοντας σαν κύριο στόχο της την δημιουργία ενός πλέγματος πληροφοριών οι οποίες απευθύνονται αποκλειστικά και μόνο στους ανθρώπους. Στην πραγματικότητα οι σελίδες του Παγκόσμιου Ιστού απευθύνονται αποκλειστικά στους ανθρώπους μιας και η πληροφορία που περιέχουν είναι αναγνωρίσιμη και κατανοητή μόνο από αυτούς.

Αυτό το χαρακτηριστικό του WWW δεν επιτρέπει την αυτόματη επεξεργασία του περιεχομένου η οποία προϋποθέτει μια καταλληλότερη περιγραφή αυτού καθώς και αλγόριθμους που να προσδίδουν την επιθυμητή «ευφυΐα» στο υπολογιστικό περιβάλλον. Τον περιορισμό αυτό καθώς και τη λύση για την άρση του, εμπνεύστηκε πρώτος ο Tim Berners-Lee, ο οποίος όρισε ένα εξελικτικό στάδιο του WWW, το Σημασιολογικό Ιστό (Semantic Web).

Ο Σημασιολογικός Ιστός έχει σαν στόχο του να προσδώσει δομή στο νόημα του περιεχομένου των ιστοσελίδων, δημιουργώντας ένα περιβάλλον όπου οι πράκτορες λογισμικού (software agents) περιπλανώμενοι από σελίδα σε σελίδα θα μπορούν να εκτελούν προηγμένες εργασίες για τους χρήστες.

Με άλλα λόγια, ο βασικός στόχος του Σημασιολογικού Ιστού είναι η επίλυση του προβλήματος της αναπαράστασης και κατανόησης της γνώσης από τους υπολογιστές. Κύριο συστατικό του είναι ο μηχανισμός επεξεργασίας της γνώσης. Ο μηχανισμός αυτός διαχειρίζεται λογικά τις πληροφορίες και στοχεύει στην εξαγωγή

συμπερασμάτων, στη δημιουργία νέας γνώσης, στην υποστήριξη στη λήψη αποφάσεων, και τέλος στην αυτόματη εκτέλεση ενεργειών.

3.1.2 Η σχέση του σημασιολογικού ιστού με τον παγκόσμιο ιστό

Ο Παγκόσμιος Ιστός βασίζεται κυρίως σε έγγραφα γραμμένα σε HTML (Hypertext Markup Language), μια γλώσσα η οποία περιγράφει το σώμα ενός δομημένου κειμένου δίνοντας έμφαση στην οπτική παρουσίαση, διανθίζοντάς το με αντικείμενα πολυμέσων όπως εικόνες και φόρμες διαλόγου. Για παράδειγμα με τη χρήση της HTML και ενός προγράμματος πλοήγησης μπορούμε να δημιουργήσουμε και να παρουσιάσουμε μια ιστοσελίδα που απαριθμεί στοιχεία κάποιων προς πώληση βιβλίων.

Όμως με την HTML δεν μπορεί να γίνει αντιληπτό ότι το στοιχείο “The Da Vinci Code” χαρακτηρίζει ένα βιβλίο αφού αναφέρεται στον τίτλο του ή ότι το στοιχείο “€20” αναφέρεται στην τιμή του. Δεν υπάρχει επίσης κανένας τρόπος να εκφραστεί το γεγονός ότι αυτά τα κομμάτια πληροφορίας είναι αλληλένδετα στην περιγραφή ενός συγκεκριμένου στοιχείου (δηλαδή ενός βιβλίου και μόνο), ευδιάκριτου από άλλα που ίσως απαριθμούνται στη σελίδα.

Ο Σημασιολογικός Ιστός αντιμετωπίζει την αδυναμία αυτή χρησιμοποιώντας γλώσσες που περιγράφουν δεδομένα και τη σχέση που έχουν αυτά μεταξύ τους. Δύο από αυτές τις γλώσσες είναι η RDF (Resource Description Framework) και OWL (Web Ontology Language) οι οποίες περιγράφονται στην συνέχεια και είναι κατανοητές από τους υπολογιστές.

3.1.3 Αναπαράσταση γνώσης στο σημασιολογικό ιστό

Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί μια επέκταση του σημερινού Ιστού. Η επέκταση αυτή έχει ως σκοπό την αυτοματοποίηση των λειτουργιών και των εφαρμογών του διαδικτύου, όπως είναι οι μηχανές αναζήτησης και οι πράκτορες (agents). Η αυτοματοποίηση αυτή μπορεί να επιτευχθεί μόνον εφόσον η γνώση και η

πληροφορία που υπάρχει αποθηκευμένη και δημοσιευμένη αυτή τη στιγμή στο σημερινό Παγκόσμιο Ιστό αποκτήσι τυπικό νόημα (formal meaning) και σημασιολογία (semantics) και δομηθεί με ένα τέτοιο τρόπο ώστε να γίνεται κατανοητή από τις μηχανές που την επεξεργάζονται (machine understandable).

Από τη στιγμή που η πληροφορία θα είναι δομημένη με έναν σημασιολογικά πλούσιο τρόπο θα ενισχύεται ο διαμοιρασμός (sharing) και η επαναχρησιμοποίησή της (reusability) επιτυγχάνοντας τη διαλειτουργικότητα (interoperability) και συνδεσιμότητα (interconnectivity) ετερογενών (heterogeneous) συστημάτων και των εφαρμογών.

Προκειμένου η γνώση και η πληροφορία να περιγραφεί με έναν τυπικό (formal) τρόπο ο οποίος θα δηλώνει τη σημασία της πρέπει να χρησιμοποιήσουμε γλώσσες αναπαράστασης γνώσης. Προκειμένου όμως να χρησιμοποιήσουμε τέτοιες τεχνολογίες στο διαδίκτυο θα πρέπει να αναθεωρήσουμε και να τροποποιήσουμε κάποια από τα συστατικά τους.

Πιο συγκεκριμένα, όπως είναι γνωστό, στο σημερινό Ιστό κατά ένα πολύ μεγάλο ποσοστό η πληροφορία δομείται με τη χρήση της γλώσσας XML. Έτσι λοιπόν αφενός στο Σημασιολογικό Ιστό πρέπει να περιγράψουμε γνώση με τη χρήση κάποιας γλώσσας αναπαράστασης γνώσης αλλά αφετέρου η σύνταξη της γλώσσας που θα χρησιμοποιήσουμε θα πρέπει να βασίζεται στη γλώσσα XML. Για το λόγο αυτό η W3C, η οποία είναι ο οργανισμός που ασχολείται με την ανάπτυξη και προτυποποίηση τεχνολογιών για τον Παγκόσμιο Ιστό, έχει αναπτύξει δύο γλώσσες αναπαράστασης γνώσης. Οι γλώσσες αυτές είναι η RDF(S) και η OWL.

3.1.4 Τα συστατικά του σημασιολογικού ιστού

Ο Σημασιολογικός Ιστός στηρίζεται από τις ακόλουθες γλώσσες και πρότυπα:

- 1 XML (Extensible Markup Language):** Είναι μια γλώσσα περιγραφής δεδομένων τα οποία είναι εύκολο να διαβαστούν και να επεξεργαστούν από ανθρώπους και προγράμματα. Δεν επιβάλλει κανέναν σημασιολογικό περιορισμό στα δεδομένα που περιγράφει.

- 2 **XML Schema:** Είναι μια γλώσσα η οποία περιορίζει τη δομή των XML εγγράφων.
- 3 **RDF:** Είναι ένα μοντέλο περιγραφής και επεξεργασίας μεταδεδομένων.
- 4 **RDF Schema:** Είναι ένας μηχανισμός περιγραφής πόρων και των σχέσεων ανάμεσα τους και αποτελεί σημασιολογική επέκταση του RDF.
- 5 **OWL:** Παρέχει έναν τρόπο περιγραφής όρων και σχέσεων γύρω από ένα πεδίο ενδιαφέροντος, προσφέροντας πιο ισχυρό συνακτικό από τις RDF και RDF Schema καθώς και πιο ισχυρή σημασιολογία που βασίζεται στη λογική (logic-based semantics).

3.1.4.1 Η ΓΛΩΣΣΑ XML

Το 1990, δημιουργήθηκε η HTML από τους Tim Berners-Lee και Anders Berglund, οι οποίοι καθόρισαν έναν τύπο SGML εγγράφου, για έγγραφα υπερκειμένου, το οποίο ήταν συναφές και αποτελεσματικό. Είναι πολύ εύκολο να υλοποιηθεί λογισμικό που να υποστηρίζει HTML, ενώ η δημιουργία ενός HTML εγγράφου είναι ακόμα πιο εύκολη. Για το λόγο αυτό η HTML χρησιμοποιείται ευρέως.

Για να επιτευχθεί η απαραίτητη απλότητα από την HTML, μερικές από τις βασικές αρχές γενικού προγραμματισμού θυσιάστηκαν. Για παράδειγμα ένας τύπος εγγράφου χρησιμοποιείται για όλους τους σκοπούς με αποτέλεσμα να είναι απαραίτητη η υπερβολική χρήση ετικετών αντί για τον καθορισμό νέων με συγκεκριμένο σκοπό. Επιπλέον οι περισσότερες από τις ετικέτες σκοπεύουν μόνο στη διαμόρφωση της εμφάνισης του εγγράφου. Για την επίλυση αυτών των προβλημάτων έγινε προσπάθεια υιοθέτησης της SGML για χρήση στο διαδίκτυο, αλλά κάτι τέτοιο αποδείχτηκε πολύπλοκο και πολύ δύσκολο. Επομένως ήταν αναγκαίος ο καθορισμός μιας νέας γλώσσας, η οποία να είναι υποσύνολο της SGML αλλά να διατηρεί τα βασικά χαρακτηριστικά της. Η γλώσσα αυτή είναι η XML. Οι σχεδιαστικοί στόχοι της XML είναι οι εξής:

- 1 να είναι εύχρηστη στο Internet
- 2 να υποστηρίζει πλήθος εφαρμογών
- 3 να είναι συμβατή με την SGML
- 4 ευκολία ανάπτυξης προγραμμάτων που επεξεργάζονται XML έγγραφα

- 5 ο αριθμός των προαιρετικών χαρακτηριστικών στην XML να είναι όσο το δυνατόν πιο μικρός, ιδανικό επίπεδο το μηδέν
- 6 τα XML έγγραφα θα πρέπει να είναι ευανάγνωστα
- 7 ο σχεδιασμός XML θα πρέπει να προετοιμάζεται γρήγορα
- 8 ο σχεδιασμός XML θα πρέπει να είναι τυπικός και περιεκτικός
- 9 τα XML έγγραφα θα πρέπει να δημιουργούνται εύκολα
- 10 η περιεκτικότητα στον XML συμβολισμό είναι μικρής σημασίας

Τα κύρια χαρακτηριστικά της XML είναι:

- 1 παρέχει ακριβή καθορισμό του περιεχομένου, ώστε τα αποτελέσματα μιας αναζήτησης να είναι ικανοποιητικά.
- 2 παρέχει δυνατότητα καθορισμού των στοιχείων του εγγράφου (π.χ. Τίτλος, Συγγραφέας, Εκδότης ...)
- 3 επιτρέπει τη μεταφορά μέσω διαδικτύου, δομημένης πληροφορίας ανεξάρτητα από πλατφόρμες και συστήματα

Η XML συμπληρώνει και δεν αντικαθιστά την HTML. Ενώ η HTML χρησιμοποιείται για διατύπωση και εμφάνιση των δεδομένων η XML αναπαριστά τη συναφή έννοια των δεδομένων. Στην HTML οι ετικέτες είναι προκαθορισμένες ενώ η XML παρέχει τη δυνατότητα να καθορίζουν οι χρήστες τις ετικέτες και τις δομημένες μεταξύ τους σχέσεις.

Επομένως, ο χρήστης με τη βοήθεια της XML μπορεί να δημιουργήσει μια νέα γλώσσα σήμανσης καθορίζοντας ένα νέο σύνολο από ετικέτες ή να χρησιμοποιήσει ένα σύνολο από ετικέτες το οποίο είναι καθορισμένο από κάποιον άλλο. Αυτό έχει ως αποτέλεσμα να υπάρχει απεριόριστος αριθμός γλωσσών σήμανσης που προέρχονται από την XML, και για το λόγο αυτό η XML είναι μια μετά-γλώσσα. Τα πλεονεκτήματα της γλώσσας σήμανσης XML είναι τα εξής:

- 1 Πιο ικανοποιητικές αναζητήσεις
- 2 Ανάπτυξη ευέλικτων εφαρμογών
- 3 Ενσωμάτωση δεδομένων από ανόμοιες πηγές πληροφορίας
- 4 Δεδομένα από πολλαπλές εφαρμογές
- 5 Υπολογισμός και διαχείριση δεδομένων τοπικά
- 6 Πολλαπλοί τρόποι εμφάνισης των δεδομένων

- 7 Ενημέρωση πεδίων
- 8 Διαθέσιμα Πρότυπα
- 9 Μορφοποίηση εμφάνισης για το διαδίκτυο

Η γλώσσα προγραμματισμού XML περιγράφει μια κατηγορία πληροφοριών που καλούνται XML έγγραφα (documents) καθώς επίσης περιγράφει τμηματικά τη συμπεριφορά των προγραμμάτων που τα επεξεργάζονται. Τα XML έγγραφα είναι αυστηρά δομημένα και μπορούν να αναπαρασταθούν με δενδρική δομή. Το πρώτο στοιχείο του εγγράφου, το οποίο περικλείει και όλα τα άλλα, ονομάζεται «ρίζα» (root). Κάθε στοιχείο που περιέχει στοιχεία ή δεδομένα ονομάζεται «πατέρας» (parent), ενώ κάθε στοιχείο ή δεδομένο το οποίο περιέχεται σε κάποιο άλλο στοιχείο ονομάζεται «παιδί» (child). Κάθε στοιχείο ή δεδομένο πρέπει να ανήκει σε κάποιο άλλο ή να είναι παιδί της ρίζας.²⁰

3.1.4.2 ΤΕΧΝΟΛΟΓΙΑ XML SCHEMA

Τα DTD παρέχουν μια βασική γραμματική για τον καθορισμό ενός εγγράφου XML από την άποψη των μεταδεδομένων που περιλαμβάνουν τη μορφή του εγγράφου. Ένα XML Schema το παρέχει αυτό, και επιπλέον παρέχει έναν λεπτομερή τρόπο για να καθοριστεί τι μπορούν ή δεν μπορούν να περιέχουν τα δεδομένα. Παρέχει πολύ περισσότερο έλεγχο για τον σχεδιαστή του τύπου εγγράφου, και παρέχει μια αντικειμενοστρεφή προσέγγιση, με όλα τα οφέλη που αυτό συνεπάγεται.

Το XML Schema αποτελεί από το Μάιο του 2001 μια επίσημη σύσταση της W3C και οι λειτουργίες που παρέχει μπορούν να περιγραφτούν ως εξής:

- 1 καθορίζει τα στοιχεία που μπορούν να εμφανιστούν σε ένα έγγραφο
- 2 καθορίζει τις ιδιότητες που μπορούν να εμφανιστούν σε ένα έγγραφο
- 3 καθορίζει ποια στοιχεία είναι στοιχεία-παιδιά
- 4 καθορίζει τη διάταξη των στοιχείων-παιδιών
- 5 καθορίζει τον αριθμό των στοιχείων-παιδιών
- 6 καθορίζει εάν ένα στοιχείο είναι κενό ή αν μπορεί να περιέχει κείμενο ή να είναι

²⁰ Σαμψών Δ., (2003).

- 7 καθορίζει τους τύπους δεδομένων (data types) για τα στοιχεία και τις ιδιότητες
- 8 καθορίζει τις προεπιλογές (default) και τις σταθερές τιμές (fixed values) για τα στοιχεία και τις ιδιότητες

Τα Schemas γράφονται ως αρχεία XML, αποφεύγοντας την ανάγκη για το λογισμικό να μπορεί να επεξεργαστεί τη δηλωτική σύνταξη της XML μέσω του DTD. Για την ακρίβεια, αυτό θα καθιστούσε ευκολότερη τη σύνταξη και το χειρισμό του σχήματος με τα συνηθισμένα εργαλεία χειρισμού εγγράφων. Αφετέρου, θεωρήθηκε ότι η παραδοσιακή σύνταξη του DTD δεν επέτρεπε στους σχεδιαστές μοντέλων εγγράφων τη δυνατότητα να επιβάλουν αρκετούς περιορισμούς στα δεδομένα (παραδείγματος χάριν, η δυνατότητα να δηλωθεί ότι ένας ορισμένος τύπος στοιχείου πρέπει πάντα να έχει μια θετική τιμή ακέραιων αριθμών, ότι δεν μπορεί να είναι κενός, ή ότι οι τιμές που δέχεται να προέρχονται από ένα σύνολο τιμών). Αυτό διευκολύνει την ανάπτυξη του λογισμικού χρησιμοποιώντας αυτά τα δεδομένα επειδή ο προγραμματιστής πρέπει να γράψει λιγότερο κώδικα για τον έλεγχο λαθών. Οι σημαντικότεροι λόγοι για την αποφυγή της χρήσης DTD είναι οι εξής:

- 1 Το DTD δε χρησιμοποιεί τη σύνταξη της
- 2 Αναμειγνύονται με την προδιαγραφή της XML 1.0 (θα υπήρχε πολύ λιγότερη σύγχυση εάν οριζόταν ξεχωριστά. Ακόμη και οι επεξεργαστές μη-επικύρωσης πρέπει να εξετάσουν το DTD)
- 3 Δεν υπάρχει κανένας περιορισμός στα δεδομένα χαρακτήρων (character data)
- 4 Έχουν πολύ απλά μοντέλα τιμών για τις ιδιότητες (οι απαριθμημένες τιμές είναι σαφώς ανεπαρκείς)
- 5 Δεν μπορούν να αναμείξουν τα δεδομένα χαρακτήρων με άλλο μοντέλο περιεχομένου (και τα μοντέλα περιεχομένων είναι γενικά δύσκολο να χρησιμοποιηθούν για τις σύνθετες απαιτήσεις)
- 6 Δεν παρέχουν καμία υποστήριξη για ονοματοδοσίες (Namespaces) (φυσικά, η XML 1.0 καθορίστηκε πριν από τα Namespaces)
- 7 Πολύ περιορισμένη υποστήριξη για συνδυασμό υπομονάδων κώδικα και για την επαναχρησιμοποίησή τους (ο μηχανισμός των οντοτήτων (entity) είναι πάρα πολύ χαμηλού επιπέδου)
- 8 Καμία υποστήριξη για την εξέλιξη του σχήματος, την επέκταση, ή την κληρονομικότητα των δηλώσεων (δύσκολο να γραφτούν, να διατηρηθούν, και

να διαβάσουν μεγάλα DTDs, και να καθοριστούν οικογένειες σχετικών σχημάτων)

- 9 Πάρα πολύ απλός μηχανισμός της ιδιότητας ID (δε σημειώνει τις απαιτήσεις, τη μοναδικότητα, την εμβέλεια κ.λπ.)
- 10 Υπάρχουν προεπιλογές μόνο για τις ιδιότητες, όχι για τα στοιχεία (αλλά αυτό θα ήταν συχνά βολικό)
- 11 Οι προεπιλογές δεν μπορούν να προσδιοριστούν ξεχωριστά από τις δηλώσεις (θα ήταν βολικό να υπάρχουν οι προεπιλογές σε ξεχωριστές υπομονάδες)

3.1.4.3 Η ΓΛΩΣΣΑ RDF (RESOURCE DESCRIPTION FRAMEWORK)

Η RDF (Resource Description Framework) είναι ένα W3C πρότυπο με το οποίο περιγράφονται μεταδεδομένα στο δίκτυο. Η RDF είναι στην πραγματικότητα ένα μοντέλο δεδομένων παρά μία οντολογική γλώσσα και χρησιμοποιείται για την περιγραφή πληροφοριών έτσι ώστε να είναι δυνατή η ανάγνωση και η κατανόησή τους από τους υπολογιστές.

Το RDF μοντέλο δεδομένων αποτελείται από τρία συστατικά:

- 1 Resources (Πόροι): Μπορούμε να θεωρήσουμε τον πόρο ως ένα αντικείμενο, ένα πράγμα για το οποίο θέλουμε να μιλήσουμε (π.χ. βιβλίο, συγγραφέας, σπίτι, κλπ.). Η αναφορά σε ένα πόρο γίνεται με τη χρήση ενός URI (Universal Resource Identifier) το οποίο μπορεί να είναι ένα URL (Unified Resource Locator) ή οτιδήποτε άλλο μπορεί να προσδιορίσει μοναδικά έναν πόρο.
- 2 Properties (Ιδιότητες): Ορίζουν ιδιότητες και σχέσεις με τις οποίες περιγράφονται οι πόροι (π.χ. ηλικία, τίτλος, κατάγεται από, κλπ.). Και οι ιδιότητες αναγνωρίζονται με τη χρήση URIs.
- 3 Statements (Δηλώσεις): Εκχωρούν μια τιμή σε μία ιδιότητα για ένα συγκεκριμένο πόρο. Υπάρχουν τρεις τρόποι αναπαράστασης των δηλώσεων RDF:
 - a. με τη χρήση τριπλέτων
 - b. με τη χρήση κατευθυνόμενων γράφων με ετικέτες στις ακμές (directed labeled graphs) και

- c. με τη χρήση ενός συντακτικού παρόμοιου με την XML (XML-like syntax).

3.1.4.4 RDF SCHEMA

Το RDF παρέχει έναν τρόπο για την έκφραση απλών δηλώσεων για πόρους, χρησιμοποιώντας ιδιότητες και τιμές. Είναι απαραίτητο όμως οι κοινότητες των χρηστών του RDF να μπορούν να ορίσουν λεξικά για να χρησιμοποιήσουν σε αυτές τις δηλώσεις, για να δηλώσουν ότι περιγράφουν συγκεκριμένους τύπους κλάσεων ή πόρων, και τις ιδιότητες που αυτοί έχουν .

Δεν παρέχει το ίδιο το RDF ένα τρόπο για τον ορισμό κλάσεων και ιδιοτήτων, όμως αυτές μπορούν να περιγραφούν ως ένα RDF λεξικό, το οποίο αναφέρεται ως **RDF σχήμα (RDF Schema)**. Το RDF σχήμα δεν παρέχει ένα λεξικό που να περιγράφει συγκεκριμένες κλάσεις και ιδιότητες. Αντί αυτού παρέχει τα μέσα για την περιγραφή τέτοιων κλάσεων και ιδιοτήτων και το πως σχετίζονται αυτές μεταξύ τους. Το RDF σχήμα μας παρέχει ένα σύστημα τύπων για το RDF, με αρκετές ομοιότητες με τα συστήματα τύπων αντικειμενοστρεφών γλώσσων. Για παράδειγμα, το RDF σχήμα επιτρέπει στους πόρους να μπορούν να οριστούν ως στιγμιότυπα κλάσεων. Επιπλέον επιτρέπει στις κλάσεις να οργανώνονται ιεραρχικά. Το λεξικό του RDF σχήματος έχει URIs με το πρόθεμα: <http://www.w3.org/2000/01/rdf-schema#>.

3.1.4.5 Η Η ΓΛΩΣΣΑ OWL (WEB ONTOLOGY LANGUAGE)

Η Web Ontology Language αποτελεί μια αναθεώρηση της DAML+OIL που ενσωματώνει χρήσιμες πληροφορίες από το σχεδιασμό και την εφαρμογή της DAML+OIL. Καθώς αποτελεί μέρος του σημασιολογικού ιστού δε θα μπορούσε παρά να στοχεύει στη χρήση και επεξεργασία των πληροφοριών από εφαρμογές

παρά στις περιπτώσεις όπου το περιεχόμενο χρειάζεται απλώς να παρουσιαστεί σε ανθρώπους-χρήστες. Μπορεί να χρησιμοποιηθεί για να αναπαραστήσει τη σημασία των όρων που συναντώνται σε λεξικά και των σχέσεων μεταξύ τους. Αυτή η

αναπαράσταση των όρων και των σχέσεων που τους διέπουν και τους συνδέουν ονομάζεται οντολογία.

Καθώς ο σημασιολογικός ιστός χτίζεται στην ικανότητα της XML να ορίζει εξειδικευμένες ετικέτες και στην ευέλικτη προσέγγιση του RDF για την αναπαράσταση δεδομένων, το επόμενο επίπεδο πιο πάνω από το RDF απαιτούσε μια γλώσσα οντολογιών που θα μπορεί με τυπικό τρόπο να περιγράφει τη σημασία της ορολογίας που χρησιμοποιείται σε έγγραφα του διαδικτύου. Εάν οι μηχανές απαιτείται να εκτελούν χρήσιμες ορθολογικές εργασίες πάνω σε αυτά τα κείμενα, τότε η γλώσσα θα πρέπει να επεκτείνεται πέρα από τις απλές σημασιολογικές δυνατότητες του RDF Schema. Τη λειτουργία αυτή στοχεύει να εκπληρώσει η OWL καθώς :

- 1 Η XML παρέχει ένα συντακτικό για δημιουργία δομημένων εγγράφων αλλά δεν εισάγει σημασιολογικούς περιορισμούς στη σημασία αυτών των εγγράφων.
- 2 Το XML Schema είναι μια γλώσσα για τον προσδιορισμό/περιορισμό της δομής των XML εγγράφων και επιπλέον επεκτείνει την XML με τη χρήση των τύπων δεδομένων.
- 3 Το RDF είναι ένα μοντέλο δεδομένων για πόρους και σχέσεις μεταξύ τους παρέχοντας μια απλή σημασιολογία και μπορεί να αναπαρασταθεί με χρήση XML σύνταξης, της RDF/XML.
- 4 Το RDF Schema είναι ένα λεξικό για την περιγραφή ιδιοτήτων και κλάσεων RDF πόρων, που εισάγει κάποια σημασιολογία σχετικά με την ιεραρχία τέτοιων ιδιοτήτων και κλάσεων.
- 5 Η OWL προσθέτει περισσότερο λεξικό για την περιγραφή ιδιοτήτων και κλάσεων, όπως συμμετρικότητα και ισότητα ως χαρακτηριστικά των ιδιοτήτων, απαριθμημένες κλάσεις καθώς και τα στοιχεία που απαριθμήθηκαν στην αρχή της υποενότητας.

3.1.5 Αναζήτηση δεδομένων στον σημασιολογικό ιστό

Η ανάκτηση πληροφοριών τόσο από τον Παγκόσμιο Ιστό, όσο και από μία Βάση Δεδομένων είναι μία από τις προφανείς ιδιότητες του Σημασιολογικού Ιστού. Όταν η πληροφορία είναι οργανωμένη σημασιολογικά, το αποτέλεσμα είναι η βελτίωση των μηχανισμών αναζήτησης και η επιστροφή καλύτερων ποιοτικά αποτελεσμάτων κατά τις διαδικασίες ανάκτησης δεδομένων και πληροφοριών.

Παράλληλα, η χρήση μεταδεδωμένων στην περιγραφή Πόρων Ιστού δίνει την δυνατότητα αυτόματης ταξινόμησης και αποθήκευσής τους, με βάση κανόνες εξαγωγής συμπερασμάτων που περιλαμβάνονται σε σχετικές οντολογίες. Μία τέτοια προοπτική προσφέρει επιπλέον δυνατότητες στην οργάνωση του ψηφιακού περιεχομένου με σημασιολογικά κριτήρια, στην αναζήτηση πληροφοριών και στην εξέλιξη των Ψηφιακών Βιβλιοθηκών και των εφαρμογών Διαχείρισης Γνώσης (Knowledge Management).²¹

3.1.6 Αυτοματοποίηση και ολοκλήρωση υπηρεσιών

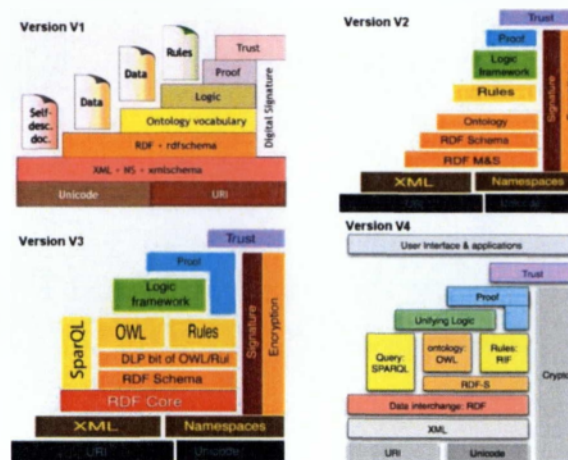
Η ύπαρξη οντολογιών που καλύπτουν ποικίλους τομείς της ανθρώπινης ζωής αποτελεί μία σημαντική υποδομή για την ανάπτυξη προηγμένων υπηρεσιών· η ελεύθερη πρόσβαση μέσω Ιστού σε οντολογίες στην ουσία αποτελεί ελεύθερη πρόσβαση σε γνώση που, επιπλέον, είναι οργανωμένη σύμφωνα με λογικά και σημασιολογικά κριτήρια. Ειδικά προγράμματα, βασιζόμενα σε αυτή τη γνώση, μπορούν να αναζητούν, να εκτελούν ή και να συνθέτουν ακόμη νέες υπηρεσίες, προς όφελος του χρήστη. Τα προγράμματα αυτά ονομάζονται -ανάλογα με την τεχνολογία στην οποία στηρίζονται- Πράκτορες (Agents) ή Υπηρεσίες Ιστού (Web Services). Και τα δύο είδη προγραμμάτων, αν και αποτελούν διαφορετικές προσεγγίσεις του Σημασιολογικού Ιστού, στηρίζονται στη γνώση που προσφέρουν οι οντολογίες.

²¹ Ζαλουνίδου, Σ., (2004).

Οδηγούνται μέσα στον Ιστό ή περιγράφονται από αυτές, προκειμένου να εκτελέσουν κάποια ενέργεια για λογαριασμό του χρήστη.²²

3.1.7 Αρχιτεκτονική του σημασιολογικού ιστού

Για να επιτευχθεί αυτή η εξέλιξη, το Semantic Web είναι οργανωμένο σε μία σειρά επιπέδων. Τα επίπεδα αυτά δεν ανταποκρίνονται αυστηρά στην έννοια της αρχιτεκτονικής λογισμικού, αλλά αποτελούν περισσότερο τεχνολογικά επίπεδα λειτουργικότητας. Οι τέσσερις διαφορετικές εκδοχές της διαστρωμάτωσης των τεχνολογιών στις οποίες βασίζεται η εξέλιξη του Semantic Web, όπως προκύπτουν με την εξέλιξη του όρου του Semantic Web από την πρώτη του εμφάνιση, παρουσιάζονται στην παρακάτω εικόνα.²³



Εικόνα 3.1 Αρχιτεκτονική επτά επιπέδων του Σημασιολογικού Ιστού

Πηγή: Gerber, A., Van der Merwe, A., & Barnard, A. (2008).

Η ανάπτυξη του Semantic Web ακολουθεί ένα ιεραρχικό σύστημα επιπέδων όπου το καθένα βρίσκεται πάνω από ένα άλλο, επεκτείνοντάς το. Κάθε επίπεδο αφορά έναν περιορισμένο τομέα του συνόλου των δραστηριοτήτων του Semantic Web, καθιστώντας έτσι παραγωγική τη συνεργασία ανάμεσα στις εκάστοτε επιστημονικές

²² Αριστογιάννης, Γ., (2010).

²³ Gerber, A., Van der Merwe, A., & Barnard, A. (2008).

ομάδες. Αν το Semantic Web δεν ακολουθούσε αυτή την αρχιτεκτονική και αντιμετωπιζόταν ως ένα ενιαίο και ευρύ ερευνητικό πεδίο, θα ήταν πολύ πιθανό να υπάρχουν πολλές αποκλίσεις ανάμεσα στις απόψεις των επιστημόνων που συμμετέχουν στην προσπάθεια δημιουργίας των απαραίτητων προτύπων.

Στην αρχιτεκτονική επιπέδων του Semantic Web ακολουθούνται δύο βασικές αρχές:

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

Ο Σημασιολογικός Ιστός έχει τις βάσεις του πάνω στην ήδη υπάρχουσα υποδομή του Παγκόσμιου Ιστού. Χρησιμοποιεί το πρωτόκολλο HTTP για τη μεταφορά των δεδομένων, τα URIs για την ονοματολογία, και την κωδικοποίηση Unicode για καθολική προσπέλαση. Ο Σημασιολογικός Ιστός στηρίζεται πάνω στην XML (eXtensible Markup Language – Επεκτάσιμη Γλώσσα Σήμανσης) που επιτρέπει την αναπαράσταση δομημένων εγγράφων χρησιμοποιώντας λεξιλόγια που ορίζονται από τους χρήστες. Η XML αποτελεί κατάλληλη λύση για διακίνηση εγγράφων στο χώρο του διαδικτύου. Πάνω στις τεχνολογίες αυτές ο Σημαντικός Ιστός προσθέτει συνολικά πέντε επίπεδα, ορισμένα από τα οποία έχουν ήδη υλοποιηθεί :

Μία κοινή αναπαράσταση για ημιδομημένα δεδομένα και μεταδεδομένα, το RDF (Resource Description Framework). Το RDF είναι το πρώτο επίπεδο που ασχολείται με τη σημασιολογία των δεδομένων. Είναι μία απλή γλώσσα αναπαράστασης μεταδεδομένων που χρησιμοποιεί URIs για την αναγνώριση διαδικτυακών πόρων καθώς και ένα μοντέλο γράφου για την περιγραφή των σχέσεων μεταξύ αυτών των πόρων.

Μία κοινή αναπαράσταση για τις οντολογίες, που επιτρέπουν στους όρους που χρησιμοποιούνται στο επίπεδο δεδομένων να ορίζονται και να συσχετίζονται μεταξύ τους (RDFS, DAML+OIL, OWL).

Το επίπεδο της λογικής (logic), παρέχει το υπόβαθρο για τη δυνατότητα αυτοματοποιημένου συλλογισμού και συμπερασμάτων βάσει των πληροφοριών που υπάρχουν σε μία οντολογία. Το επίπεδο αυτό επιπλέον καθιστά δυνατή τη χρήση τυπικών κανόνων, βάσει των οποίων γίνεται εφικτή η διαδικασία λήψης αποφάσεων από τις υπολογιστικές μηχανές.

Το επίπεδο της απόδειξης (proof), λειτουργεί με τέτοιο τρόπο ώστε τα αποτελέσματα που συμπεραίνονται από δεδομένα στο Σημασιολογικό Ιστό να μπορούν να οδηγήσουν πίσω στις υποθέσεις που τα προκάλεσαν.

Το επίπεδο της εμπιστοσύνης (trust), σε συνδυασμό με την τεχνολογία των ψηφιακών υπογραφών (digital signatures), θα εξασφαλίζει το βαθμό στον οποίο οι πληροφορίες που διακινούνται, επεξεργάζονται και συμπεραίνονται στο Σημαντικό Ιστό είναι αξιόπιστες, με αυτοματοποιημένο τρόπο (για παράδειγμα, στην επικοινωνία μεταξύ πρακτόρων).

4 ΚΕΦΑΛΑΙΟ: ΟΝΤΟΛΟΓΙΕΣ ΣΤΟΝ ΤΟΜΕΑ ΤΗΣ ΥΓΕΙΑΣ

Στην εποχή που διανύουμε οι διαρκείς μεταβολές και οι ραγδαίες τεχνολογικές, πολιτικές και επιστημονικές εξελίξεις έχουν άμεσο αντίκτυπο στον τρόπο λειτουργίας των διάφορων οργανισμών. Οι ολοένα αυξανόμενες απαιτήσεις και ο διεθνής ανταγωνισμός που κυριαρχεί έχουν ως αποτέλεσμα τη συνεχή αναζήτηση για την εξεύρεση λύσεων στο χώρο της οργάνωσης και της λειτουργίας των οργανισμών. Αποτέλεσμα αυτών των ενεργειών είναι η εκτίμηση του πλέον πολύτιμου πόρου κάθε οργανισμού που δεν είναι άλλος από την ήδη κεκτημένη γνώση και πληροφορία. Σε συνδυασμό λοιπόν με τη χρήση ηλεκτρονικών εργαλείων και τη σύσταση πληροφοριακών βάσεων της προϋπάρχουσας γνώσης, οι οργανισμοί επιδιώκουν τη σωστή διαχείρισή της, την επεξεργασία της, τη διάχυσή της και πρωτίστως τη δημιουργία νέας γνώσης. Ο αιώτερος στόχος αυτών των ενεργειών ενός οργανισμού, είναι η βελτιστοποίηση της ταχύτητας και της αποτελεσματικότητας στις εσωτερικές διαδικασίες του, όπως και η ενίσχυση της καινοτομίας και η ανάδειξή του μέσα από τις διεθνείς ανταγωνιστικές συνθήκες που επικρατούν.

Σε αυτή την προσπάθεια παίζει καθοριστικό ρόλο ο Παγκόσμιος Ιστός (World Wide Web) ο οποίος αποτελεί το κύριο μέσο για την αναζήτηση, την εύρεση, τη διάδοση και την αποθήκευση της πληροφορίας και των δεδομένων. Πλέον η επικοινωνία και η αναζήτηση πληροφοριών αποτελεί μία τυπική διαδικασία στην καθημερινότητα των χρηστών. Εντούτοις, εντοπίζονται αρκετές δυσκολίες στη συλλογή πληροφορίας από διαφορετικές πηγές, λόγω του ότι τα αποτελέσματα που επιστρέφονται είναι σε μορφή ιστοσελίδων και η διαφορετική ορολογία που μπορεί να υπάρχει δεν καλύπτεται από τις λέξεις-κλειδιά της αναζήτησης. Έτσι ο χρήστης αναγκάζεται να εκτελέσει πολλαπλές αναζητήσεις με διαφορετική ορολογία και λεξιλόγιο με σκοπό να συλλέξει όλη την πληροφορία που αφορά συγκεκριμένο θέμα. Το πρόβλημα έγκειται στο ότι η διαθέσιμη πληροφορία μέσω του Παγκόσμιου Ιστού είναι κατανοητή μόνο στους ανθρώπινους χρήστες και όχι στους υπολογιστές - μηχανές. Σε αυτό το ζήτημα που εντοπίζεται στη συσχέτιση δεδομένων κατά τον εντοπισμό τους

από διαφορετικές πηγές με στόχο τη συγκρότηση έγκυρης πληροφορίας, έρχεται να δώσει λύση ο Σημασιολογικός Ιστός (Semantic Web) ο οποίος αποτελεί προέκταση του Παγκόσμιου Ιστού. Μέσω του Σημασιολογικού Ιστού μπορεί πλέον να επιτευχθεί η δημιουργική αναζήτηση μέσα από ένα τεράστιο όγκο δεδομένων και να εξαχθεί δομημένη και επεξεργάσιμη πληροφορία η οποία θα οδηγήσει στη σύνθεση και τη δημιουργία νέας πληροφορίας. Αυτό πραγματοποιείται με τη χρήση μεταδεδομένων (metadata) και οντολογιών (ontology) τα οποία είναι υπεύθυνα για την περιγραφή πληροφοριών και την οργάνωση δεδομένων αντίστοιχα που διαχειρίζεται ο Παγκόσμιος Ιστός. Με τη χρήση αυτών των εργαλείων, τα δεδομένα που είναι διαθέσιμα μέσω του Παγκόσμιου Ιστού καθίστανται δομημένα και σημασιολογικά ορισμένα, γεγονός που συμβάλλει στην επεξεργασία, στη διαχείριση, στη συνένωση και στην κατανόηση τους από τους υπολογιστές - μηχανές. Έτσι επιτυγχάνεται η αποτελεσματική και αποδοτική αναζήτηση και προσπέλαση πληροφοριών. Στο συγκεκριμένο κεφάλαιο θα γίνει μία αναφορά στον καθοριστικό ρόλο που διαδραματίζουν οι οντολογίες για την υλοποίηση και ανάπτυξη του Σημασιολογικού Ιστού καθώς και στον τρόπο που χρησιμοποιούνται αυτές στον τομέα της υγείας.

4.1 ΟΝΤΟΛΟΓΙΕΣ ΚΑΙ ΥΓΕΙΑ

Τα τελευταία χρόνια γίνεται μία προσπάθεια στον τομέα της υγείας για ενοποίηση του ολοένα και αυξανόμενου όγκου μεμονωμένων ορολογιών που έχουν αναπτυχθεί από διαφορετικούς φορείς της υγείας, όπως ερευνητικά κέντρα υγείας, τομέας βιοπληροφορικής, τομέας βιοϊατρικής, τομέας βιολογίας, νοσοκομεία, πανεπιστήμια, Παγκόσμιος Οργανισμός Υγείας κ.α., για διάφορες ενέργειες όπως, έρευνες, ευρετηρίαση πληροφοριών, ηλεκτρονικά αρχεία ασθενών, στατιστικές εκθέσεις κ.α.. Έτσι επιδιώκεται η υλοποίηση κατάλληλων κοινών οντολογιών, εργαλείων και σημασιολογικών δικτύων με σκοπό την αποτελεσματική και κοινή διαχείριση της ιατρικής πληροφορίας. Στόχος είναι η ενσωμάτωση και η σημασιολογική ενοποίηση διαμοιραζόμενων ιατρικών δεδομένων από διαφορετικές πηγές που θα έχει ως αποτέλεσμα τη σύσταση ενός ιατρικού δικτύου που απαρτίζεται από πλήρως

τεκμηριωμένα και διασυνδεδεμένα δεδομένα με τυπική σημασιολογία και η μοντελοποίηση της ήδη αποκτηθείσας γνώσης σε τυπικά ορισμένες οντολογίες.

4.2 ΟΝΤΟΛΟΓΙΑ – ΤΙ ΕΙΝΑΙ

Όπως αναφέρθηκε προηγουμένως οι οντολογίες αποτελούν βασικό συστατικό στη διαδικασία ανάπτυξης του Σημασιολογικού Ιστού. Μέσω των οντολογιών επιτυγχάνεται η σωστή οργάνωση, αναπαράσταση και διαχείριση της γνώσης και αποτελούν θεμελιώδη λίθο στην ανάπτυξη, υλοποίηση και επικοινωνία διαφόρων συστημάτων εφαρμογών.

Είναι σκόπιμο σε αυτό το σημείο να αναφερθούμε σε κάποιους ορισμούς που υπάρχουν στη διεθνή βιβλιογραφία για την οντολογία.

Οι πρώτες αναφορές συναντώνται στον τομέα της αρχαίας ελληνικής φιλοσοφίας και συγκεκριμένα στη μεταφυσική. Σύμφωνα με την ετυμολογία του όρου (ον-λόγος) η λέξη οντολογία αναφέρεται στη θεωρία της ύπαρξης και εκφράζει τη μελέτη της φύσης και των κατηγοριών της ύπαρξης και των οντοτήτων, την κατηγοριοποίησή τους ανάλογα με τα κοινά τους στοιχεία και τις διαφορές τους αλλά και τον τρόπο που ιεραρχούνται.

Ο χώρος της επιστήμης δανείζεται τον όρο οντολογία από τη φιλοσοφία, με τη διαφορά ότι δεν αναφέρεται σε μία θεωρία, αλλά πλέον χρησιμοποιείται για την απόδοση ενός μοντέλου δεδομένων και εννοιών που αφορούν συγκεκριμένη θεματολογία και ορίζουν μία εννοιολογική σύλληψη, όπου αναπαριστώνται και διακρίνονται οι μεταξύ τους σχέσεις με σκοπό την αναπαράσταση της γνώσης και την αντιπροσώπευση της πληροφορίας για την εν λόγω θεματολογία.

Σύμφωνα με τον Thomas Gruber (1993)²⁴, «η οντολογία αποτελεί μία τυποποιημένη, κατηγορηματική προδιαγραφή μιας διαμοιρασμένης εννοιολογικής αναπαράστασης». Ο συγκεκριμένος ορισμός παρουσιάζει την οντολογία σαν μία περιγραφή (τυπική προδιαγραφή ενός προγράμματος) των εννοιών και των σχέσεων που μπορεί να αναφέρονται σε ένα πράκτορα (agent) ή σε μία κοινότητα πρακτόρων (community of agents).

Ο Fensel (2000)²⁵ ορίζει την οντολογία σαν μία κοινή και διαμοιραζόμενη αντίληψη ενός γνωστικού τομέα, η οποία είναι κατανοήσιμη και διαχειρίσιμη από ανθρώπους αλλά και από ετερογενή και κατανεμημένα συστήματα εφαρμογών.

Κατά τους Huhns και Singh (1997)²⁶ η οντολογία ορίζεται σαν ένα σύμφωνο μίας εννοιολογικής αναπαράστασης, το οποίο περιλαμβάνει ορισμένα πρότυπα και συγκεκριμένες προδιαγραφές για τη μοντελοποίηση και την παρουσίαση ενός γνωστικού πεδίου.

Στην ουσία ως οντολογία νοείται ένα σύνολο λεξιλογίου και ορολογίας τα οποία ορίζουν, περιγράφουν ιδιότητες και δηλώνουν τον τρόπο σύνδεσης και τις μεταξύ τους σχέσεις και προδιαγράφουν τις λογικές σχέσεις οντοτήτων, σε ένα θεματικό - γνωστικό πεδίο. Η αναπαράσταση γίνεται με κατάλληλη κωδικοποίηση έτσι ώστε να είναι επεξεργάσιμη από υπολογιστή – μηχανή. Οι όροι που χρησιμοποιούνται οριοθετούνται και ιεραρχούνται μέσω μηχανισμών σε κατηγορίες και υποκατηγορίες και εκφράζουν έννοιες και αντικείμενα. Για την εξασφάλιση της διατήρησης της σημασίας και της έννοιας των όρων μίας οντολογίας προαπαιτείται σχετική συμφωνία των εμπλεκομένων.

²⁴ Gruber, T. R., (1993)

²⁵ Fensel, D., (2000)

²⁶ Huhns, N., Singh, M. (1997)

4.2.1 Ανάγκη Ανάπτυξης και χρήσης οντολογιών

Παρόλο που η εποχή μας χαρακτηρίζεται από την ευρεία διάθεση της πληροφορίας, η εύρεση και η πρόσβαση σε αυτή παρουσιάζουν κάποιες δυσκολίες. Ο λόγος έγκειται στο ότι δεν υπάρχει σύνδεση μεταξύ της πληροφορίας και της σημασιολογικής της διάστασης. Εδώ έρχεται να δώσει λύση η ανάπτυξη των οντολογιών. Συγκεκριμένα με τη χρήση των οντολογιών κυρίως επιτυγχάνεται:

- η διάχυση μιας κοινής εννοιολογικής οριοθέτησης και αντίληψης της πληροφορίας ανάμεσα σε ανθρώπους ή σε υπολογιστικούς πράκτορες (Musen, 1992, Gruber, 1993)²⁷. Έτσι, μέσω κοινών οντολογιών πραγματοποιείται η ανάκτηση και η σύνθεση πληροφοριών από διάφορες πηγές, και βελτιώνεται η επικοινωνία μεταξύ πρακτόρων (ανθρώπων και υπολογιστικών συστημάτων)
- Επιπλέον είναι εφικτή η επαναχρησιμοποίηση των πληροφοριακών μοντέλων και γνωστικών σχημάτων ενός γνωστικού πεδίου μέσω της χρήσης μίας οντολογίας ή από το συνδυασμό ήδη υπάρχουσών οντολογιών και την εφαρμογή τους σε διαφορετικούς τομείς.
- Η διασαφήνιση παραδοχών και υποθέσεων σε μία γνωστική περιοχή έτσι ώστε να καταστούν σαφείς και ρητές. Μέσω αυτού επιτρέπεται η αλλαγή των πληροφοριακών μοντέλων που χρησιμοποιούνται σε περίπτωση μεταβολής της γνώσης για τη συγκεκριμένη θεματική περιοχή και διευκολύνεται η κατανόηση των εννοιών και όρων από ένα άπειρο χρήστη.
- Διάκριση της γνώσης μίας θεματικής περιοχής από την πρακτική – επιχειρησιακή της εφαρμογή.
- Ανάλυση της αποκτηθείσας γνώσης σε ένα γνωστικό πεδίο. Αυτό επιτυγχάνεται όταν υπάρχουν ακριβής προδιαγραφές των όρων που χρησιμοποιούνται και έτσι διευκολύνεται η διάχυση και επαναχρησιμοποίηση των οντολογιών.

²⁷ Musen, M.A.(1992), Gruber, T.R., (1993)

4.2.2 Κατηγοριοποίηση Οντολογιών

Λόγω της αναγκαιότητας ανάπτυξης και χρήσης των οντολογιών σε όλους τους επιστημονικούς τομείς, ο προσδιορισμός του τι πρεσβεύει η οντολογία δίδεται αναλόγως με τις ανάγκες του εκάστοτε επιστημονικού πεδίου. Έτσι υπάρχουν αρκετά είδη οντολογιών που έχουν καταγραφεί στην επιστημονική βιβλιογραφία. Παρακάτω γίνεται μία αναφορά σε μερικά από αυτά.

Ένα είδος κατηγοριοποίησης (Uschold and Gruninger 1996)²⁸ γίνεται σύμφωνα με την τυπικότητα και τη συνέπεια του λεξιλογίου που χρησιμοποιείται (φυσική ή τυπική-επίσημη γλώσσα) για την έκφραση και την αναπαράσταση του ορισμού των οντολογιών. Σύμφωνα με αυτή υπάρχουν τέσσερα είδη οντολογιών που παρατίθενται με αύξουσα σειρά ως προς το βαθμό τυπικότητας στην έκφραση.

1. **Highly Informal:** Οι οντολογίες εκφράζονται άτυπα και ορίζονται σε φυσική γλώσσα
2. **Semi-Informal:** Οι οντολογίες εκφράζονται σε φυσική γλώσσα η οποία είναι καταλλήλως δομημένη και οριοθετημένη με συγκεκριμένη ορολογία και λεξιλόγιο, επιτυγχάνοντας σε μεγάλο βαθμό τη σαφήνεια.
3. **Semi Formal:** Η γλώσσα που χρησιμοποιείται για τη διατύπωση των οντολογιών αποτελείται από τεχνητό και τυπικά ορισμένο λεξιλόγιο
4. **Rigorously Formal:** Αποτελεί την κατηγορία με το μεγαλύτερο βαθμό τυπικότητας στην έκφραση οντολογιών. Οι οντολογίες εκφράζονται με όρους που είναι σχολαστικά σημασιολογικά ορισμένοι μέσω τεχνητής και τυπικά καθορισμένης γλώσσας, θεωρημάτων, αξιωμάτων και αποδείξεων.

Η Ontoware είχε προτείνει τρεις κατηγορίες οντολογίας:

- **Classification according to Purpose**
 - **Οντολογία Εφαρμογής (Application Ontology):** Χρησιμοποιείται κατά τη διάρκεια εκτέλεσης ενός συγκεκριμένου προγράμματος εφαρμόζοντας μία οντολογία που θέτει περιορισμούς στα αξιώματα της ορολογίας και του λεξιλογίου. Επίσης αυτό το είδος οντολογίας μπορεί να χρησιμοποιηθεί για την περιγραφή συγκεκριμένων

²⁸ Uschold, M., Gruninger, M., (1996)

«οντοτήτων» (σημασιολογικές περιγραφές, βάση γνώσεων, μεταδεδομένα, σημασιολογικά δεδομένα κ.α.)

- Οντολογία Αναφοράς (Reference Ontology): Χρησιμοποιείται κατά τη διάρκεια ανάπτυξης των εφαρμογών έτσι ώστε να αναπτυχθεί κοινή αντίληψη μεταξύ (ανθρωπίνων ή υπολογιστικών) πρακτόρων που ανήκουν σε διαφορετικές κοινότητες, για τη σημασιολογική παραδοχή ενός νέου όρου ή απλά για την επεξήγηση μίας έννοιας σε ένα άπειρο χρήστη
- Classification according to Expressiveness
 - Οντολογία Βαριάς Μορφής (Heavyweight Ontology): Οι οντολογίες αυτής της κατηγορίας χαρακτηρίζονται από πλήθος αξιωμάτων και ως εκ τούτου αντιπροσωπεύουν ρητά την οντολογική δέσμευση. Η χρήση των αξιωμάτων αποκλείει τις ορολογικές και εννοιολογικές ασάφειες που μπορεί να οφείλονται σε μη ήθελημένες ερμηνείες. Κάθε οντολογία βαριάς μορφής μπορεί να έχει εκδοχή και στην υποκατηγορία ελαφριάς μορφής.
 - Οντολογία Ελαφριάς Μορφής (Lightweight Ontology): Οι οντολογίες αυτής της κατηγορίας είναι δομημένες ταξινομίες απλών ή σύνθετων όρων μαζί με τους σχετικούς ορισμούς.
- Classification according to Specificity:
 - Γενικές Οντολογίες (Generic Ontology): Οι οντολογίες που ανήκουν σε αυτή την κατηγορία θεωρούνται ότι είναι γενικές και ταιριάζουν σε αρκετά πεδία. Συνήθως, οι γενικές οντολογίες (είναι γνωστές και ως «upper level» ή «top-level» οντολογίες) ορίζουν έννοιες όπως κράτος, γεγονός, διαδικασία, δράση, συστατικό κ.λπ.
 - Βασικές Οντολογίες (Core Ontology): Οι οντολογίες που ανήκουν σε αυτή την κατηγορία ορίζουν γενικές έννοιες που μπορούν να χρησιμοποιηθούν από πολλά γνωστικά πεδία. Δεν είναι ξεκάθαρος ο διαχωρισμός των Γενικών και των Βασικών οντολογιών καθώς δεν υπάρχει εξαντλητική απαρίθμηση των τομέων και των εννοιών τους. Ωστόσο οι οντολογίες αυτής της κατηγορίας είναι χρήσιμες για τη δημιουργία βιβλιοθηκών.

- **Οντολογίες Τομέα (Domain Ontology):** Οι οντολογίες αυτής της κατηγορίας εκφράζουν έννοιες που είναι ειδικές για συγκεκριμένους γλωσσολογικούς τομείς. Οι έννοιες στις οντολογίες τομέα ορίζονται συχνά σαν εξειδικεύσεις των Βασικών και Γενικών οντολογιών. Και σε αυτή την περίπτωση δεν είναι ξεκάθαρος ο διαχωρισμός των Βασικών οντολογιών και οντολογιών Τομέα διότι οι Βασικές οντολογίες παίζουν το ρόλο των Γενικών οντολογιών σε επίπεδο οντολογιών Τομέα. Μία έννοια όπως ένα συστατικό λογισμικού θα μπορούσε να τοποθετηθεί σε μία Βασική οντολογία διακομιστών εφαρμογών έτσι ώστε να επαναχρησιμοποιηθεί σε οποιαδήποτε οντολογία Τομέα.

Σύμφωνα με τον Guarino²⁹ οι οντολογίες κατηγοριοποιούνται με βάση το επίπεδο γενικότητας και το βαθμό συσχέτισης με μία συγκεκριμένη διαδικασία ή με μία άποψη:

- **Top-level ontologies:** Οι οντολογίες που ανήκουν σε αυτή την κατηγορία περιγράφουν πολύ γενικές έννοιες οι οποίες είναι ανεξάρτητες ενός συγκεκριμένου προβλήματος ή συγκεκριμένου πεδίου ενδιαφέροντος, όπως ο χώρος, ο χρόνος, η ύλη, το αντικείμενο, το συμβάν, η ενέργεια κ.α. Για αυτό το λόγο, θεωρητικά τουλάχιστον, οι συγκεκριμένες οντολογίες χρησιμοποιούνται από μεγάλες κοινότητες χρηστών.
- **Domain ontologies και task ontologies:** Σε αυτές τις κατηγορίες ανήκουν οι οντολογίες που χρησιμοποιούνται για την περιγραφή λεξιλογίου που σχετίζεται με ένα γενικό πεδίο (π.χ. ιατρική ή αυτοκινητοβιομηχανίες) ή με γενικές εργασίες (π.χ. διάγνωση ή πώληση) εξειδικεύοντας τους όρους που εισήχθησαν στις top-level ontologies.
- **Application ontologies:** Σε αυτή την κατηγορία ανήκουν οι οντολογίες οι οποίες περιγράφουν έννοιες που σχετίζονται ταυτόχρονα με ένα συγκεκριμένο πεδίο και με μία συγκεκριμένη εργασία και αποτελούν εξειδικεύσεις εννοιών των δύο συσχετιζόμενων οντολογιών (Domain ontologies και task ontologies). Συχνά αυτές οι έννοιες ανταποκρίνονται σε ρόλους που παίζουν οι οντότητες ενός πεδίου κατά τη διάρκεια εκτέλεσης μίας συγκεκριμένης δραστηριότητας, όπως αντικαταστάσιμες μονάδες ή ανταλλακτικά συστατικά.

²⁹ Guarino, N., (1998)

Ο Fensel³⁰ (2004) κατηγοριοποιεί τις οντολογίες ως εξής:

- **Domain ontologies:** Σε αυτή την κατηγορία ανήκουν οι οντολογίες που καταγράφουν και αποτυπώνουν έγκυρα τη γνώση για ένα συγκεκριμένο τομέα ενδιαφέροντος (π.χ. ηλεκτρονικό τομέα, ιατρικό τομέα, μηχανικό τομέα, ψηφιακό τομέα).
- **Metadata ontologies:** Σε αυτή την κατηγορία ανήκουν οι οντολογίες που παρέχουν λεξιλόγιο για την περιγραφή περιεχομένου για διαδικτυακές πηγές πληροφόρησης (π.χ. Dublin Core³¹(Weibel et al., 1995))
- **Generic ή Common Sense ontologies:** Σε αυτή την κατηγορία ανήκουν οι οντολογίες που στοχεύουν να αποδώσουν τη γενική γνώση για τον κόσμο παρέχοντας βασικές αντιλήψεις και έννοιες για όρους όπως ο χρόνος, ο χώρος, η κατάσταση, το συμβάν κ.α. ([Fridman-Noy & Hafner, 1997³²], [Pirlein & Studer, in press³³]), με αποτέλεσμα να είναι έγκυρες για πολλά πεδία ενδιαφέροντος. Για παράδειγμα, μία οντολογία που αφορά τη μετεωρολογία είναι εφαρμόσιμη σε πολλά τεχνικά πεδία (Borst & Akkermans, 1997³⁴).
- **Representational ontologies:** Οι οντολογίες που ανήκουν σε αυτή την κατηγορία δεν απευθύνονται σε ένα συγκεκριμένο πεδίο ενδιαφέροντος. Παρέχουν οντότητες χωρίς να συνιστούν τι θα μπορούσαν να αναπαραστήσουν και ορίζουν έννοιες για object-oriented ή framework-oriented προσεγγίσεις. Μία τέτοια οντότητα είναι η Frame Ontology (Gruber, 1993³⁵).
- **Method ontologies (Fensel & Groenboom, 1997³⁶, Studer et al., 1996³⁷):** Οι οντολογίες που ανήκουν σε αυτή την κατηγορία παρέχουν όρους για

³⁰ Fensel, D., (2004)

³¹ Weibel, S., Gridby, J., Miller, E., (1995)

³² Fridman-Noy, N., Hafner, C.D., (1997)

³³ Pirlein, Th., Studer, R., (1999)

³⁴ Borst, W.N., Akkermans, J.M., (1997)

³⁵ Gruber, T.R., (1993)

³⁶ Fensel, D., Groenboom, R., (1997)

³⁷ Studer, R., Eriksson, H., Gennari, J.H., Tu, S.W., Fensel, D., Musen, M., (1996)

συγκεκριμένες μεθόδους επίλυσης προβλημάτων (PSMs) (π.χ. ο όρος «correct state» ανήκει στη Propose-and-Revise method ontology).

- Task ontologies (Fensel & Groenboom, 1997, Studer et al., 1996): Οι οντολογίες που ανήκουν σε αυτή την κατηγορία παρέχουν όρους για συγκεκριμένες διεργασίες (π.χ. ο όρος «hypothesis» ανήκει στη diagnosis task ontology). Οι δύο τελευταίες κατηγορίες οντολογιών παρέχουν μία λογική σκοπιά σε ένα πεδίο ενδιαφέροντος.

4.2.3 Βασική Δομή της Οντολογίας

Οι οντολογίες ανεξάρτητα με την κατηγορία στην οποία ανήκουν, αποτελούνται από κοινά δομικά στοιχεία των οποίων οι ονομασίες μπορεί να διαφέρουν αναλόγως τη γλώσσα της οντολογίας, τις φιλοσοφικές πεποιθήσεις και το υπόβαθρο των δημιουργών τους. Σύμφωνα με τους M. Uschold, και M. Gruninger³⁸ (1996) τα δομικά στοιχεία των οντολογιών είναι τα παρακάτω:

Άτομα (Individuals): Τα συγκεκριμένα δομικά στοιχεία είναι γνωστά και ως Στιγμιότυπα (Instances) ή Λεπτομέρειες (Particulars) και αποτελούν τη βασική μονάδα μίας οντολογίας. Με τα Άτομα μπορούν να μοντελοποιηθούν συγκεκριμένα αντικείμενα (π.χ. μηχανήματα, πρωτεΐνες, άνθρωποι κ.α.), όπως και πιο αφηρημένα αντικείμενα (π.χ. ένα άρθρο, το επάγγελμα ενός ανθρώπου, μία λειτουργία κ.α.).

Έννοιες (Concepts): Τα συγκεκριμένα δομικά στοιχεία αναπαριστώνται από τις **Κλάσεις (Classes)**, και συναντώνται στη δομή των περισσότερων οντολογιών. Αποτελούνται από οτιδήποτε για το οποίο μπορεί να ειπωθεί κάτι, από περιγραφή μίας διενέργειας, μίας λειτουργίας, μίας ιδέας, μίας κρίσης κ.α. Μπορεί να καταταχθούν σε δύο κατηγορίες:

- Πρωταρχικές Έννοιες (Primitive Concepts): Εδώ ανάγονται οι έννοιες οι οποίες έχουν μόνο απαραίτητες συνθήκες ως προς τις ιδιότητές τους.

³⁸ Uschold, M., Gruninger, M., (1996)

- Έννοιες εξ ορισμού (Defined Concepts): Εδώ ανήκουν οι έννοιες των οποίων η περιγραφή είναι ικανή και αναγκαία συνθήκη για να ανήκει στην κλάση.

Σχέσεις (Relations): Οι Σχέσεις σε μία οντολογία περιγράφουν τον τρόπο με τον οποίο αλληλεπιδρούν τα Άτομα. Οι Σχέσεις εκφράζονται συνήθως απευθείας μεταξύ των Ατόμων (π.χ. αυτό το βιβλίο έχει συγγραφέα τον Πάολο Κοέλιο) ή μεταξύ των Εννοιών (ένα βιβλίο έχει συγγραφέα ένα άτομο). Στην τελευταία περίπτωση περιγράφεται η σχέση μεταξύ όλων των Ατόμων των Εννοιών. Παρόλο που εξαρτάται από τη γλώσσα της οντολογίας, συχνά είναι πιθανό να εκφράζονται διαφορετικές κατηγορίες Σχέσεων μεταξύ των Εννοιών, π.χ. «άτομο έχει πατέρα άτομο» .

Συναρτήσεις (Functions): Οι Συναρτήσεις εκφράζουν μία ειδική περίπτωση Σχέσης.

Αξιώματα (Axioms): Τα Αξιώματα εκφράζουν προτάσεις που είναι πάντοτε αληθείς. Χρησιμοποιούνται για την αναπαράσταση γνώσης όταν δεν είναι εφικτό αυτό με τα υπόλοιπα διαθέσιμα δομικά στοιχεία. Επίσης είναι χρήσιμα για την επιβεβαίωση της οντολογίας και της γνώσης που βρίσκονται σε μία βάση γνώσης (Gruber, 1993)³⁹.

Στιγμιότυπα (Instances): Χρησιμοποιούνται για την αναπαράσταση οντοτήτων ή Ατόμων (individuals) σε μία οντολογία.

4.2.4 Γλώσσες Οντολογιών

Για την ανάπτυξη και υλοποίηση οντολογιών χρησιμοποιούνται γλώσσες για τη μοντελοποίηση που να είναι διαχειρίσιμες από υπολογιστή. Η γλώσσα που χρησιμοποιείται είναι συνήθως logic-based έτσι ώστε να παρέχουν συνέπεια, διάκριση μεταξύ των τάξεων, των ιδιοτήτων και των σχέσεων. Διακρίνονται δύο κατηγορίες: οι κλασικές γλώσσες και οι γλώσσες σήμανσης. Παρακάτω γίνεται μία αναφορά σε κάποιες από αυτές.

³⁹ Gruber, T., (1993)

- **RDF (Resource Description Framework) –RDFS (RDF Schema)**⁴⁰: Το RDF αρχικά σχεδιάστηκε ως metadata data model και δημιουργήθηκε από το World Wide Web Consortium (W3C). Χρησιμοποιείται σαν γενική μέθοδος για την εννοιολογική περιγραφή ή μοντελοποίηση πληροφοριών που υλοποιούνται σε διαδικτυακούς πόρους, χρησιμοποιώντας μία ποικιλία συμβολισμών σύνταξης και σειριακές μορφές δεδομένων. Η RDFS αποτελεί επέκταση της RDF και αποτελεί ένα σύνολο κλάσεων με συγκεκριμένες ιδιότητες χρησιμοποιώντας την επέκταση της RDF, παρέχοντας βασικά συστατικά για την περιγραφή οντολογιών που καλούνται λεξιλόγιο RDF με σκοπό να δομηθούν οι RDF resources. Σε αυτές τις γλώσσες στηρίχθηκε η θεμελίωση του Σημασιολογικού Ιστού και απετέλεσαν τη βάση για την δημιουργία τριών ακόμα γλωσσών (OIL, DAML+OIL και OWL).
- **OWL (Web Ontology Language)**⁴¹: Η OWL είναι μία οικογένεια γλωσσών αναπαράστασης για συγγραφή οντολογιών. Οι γλώσσες χαρακτηρίζονται από τυπική σημασιολογία και είναι RDF/XML-based serializations για τον Σημασιολογικό Ιστό. Έχει εγκριθεί από το World Wide Web Consortium (W3C) και έχει προσελκύσει το ακαδημαϊκό, ιατρικό και εμπορικό ενδιαφέρον. Το 2009 ανακοινώθηκε από το W3C μία νέα έκδοση, η OWL2 η οποία χρησιμοποιείται από πολλούς σημασιολογικούς εκδότες όπως το Protégé.
- **XOL (XML-Based Ontology Exchange Language)**⁴²: Η XOL είναι μία γλώσσα ανταλλαγής οντολογιών που αναπτύχθηκε για να διευκολύνει τη δημιουργία διαμοιραζόμενων οντολογιών που είχαν χρήση αρχικά στον τομέα της βιοπληροφορικής. Προορίζεται να χρησιμοποιηθεί ως ενδιάμεση γλώσσα για τη μεταφορά οντολογιών μεταξύ διαφορετικών συστημάτων βάσεων δεδομένων, εργαλεία ανάπτυξης οντολογιών ή προγραμμάτων εφαρμογής.
- **LOOM**⁴³: Η LOOM είναι μία γλώσσα αναπαράστασης γνώσης που αναπτύχθηκε από ερευνητές στην Τεχνητή Νοημοσύνη στο Information Sciences Institute του University of Southern California. Στόχος του έργου

⁴⁰ <http://www.w3.org/RDF/>

⁴¹ <http://www.w3.org/TR/owl-features/>

⁴² <http://www.ai.sri.com/pkarp/xol/>

⁴³ <http://www.isi.edu/isd/LOOM/>

LOOM είναι η ανάπτυξη προηγμένων εργαλείων για την αναπαράσταση γνώσης στην Τεχνητή Νοημοσύνη. Επιτρέπει την αναπαράσταση των εννοιών, της ταξινόμιας των εννοιών, των σχέσεων n-ary, των συναρτήσεων, των αξιωμάτων και των κανόνων παραγωγής.

4.2.5 Περιβάλλοντα Οντολογιών

Τα περιβάλλοντα οντολογιών αποτελούν εργαλεία με σκοπό την υιοθέτηση οντολογιών από πραγματικά πληροφοριακά συστήματα, όπου μπορούν να εισαχθούν νέα δομικά στοιχεία μέσω component-based architectures. Παρακάτω γίνεται μία σύντομη αναφορά σε κάποια από αυτά.

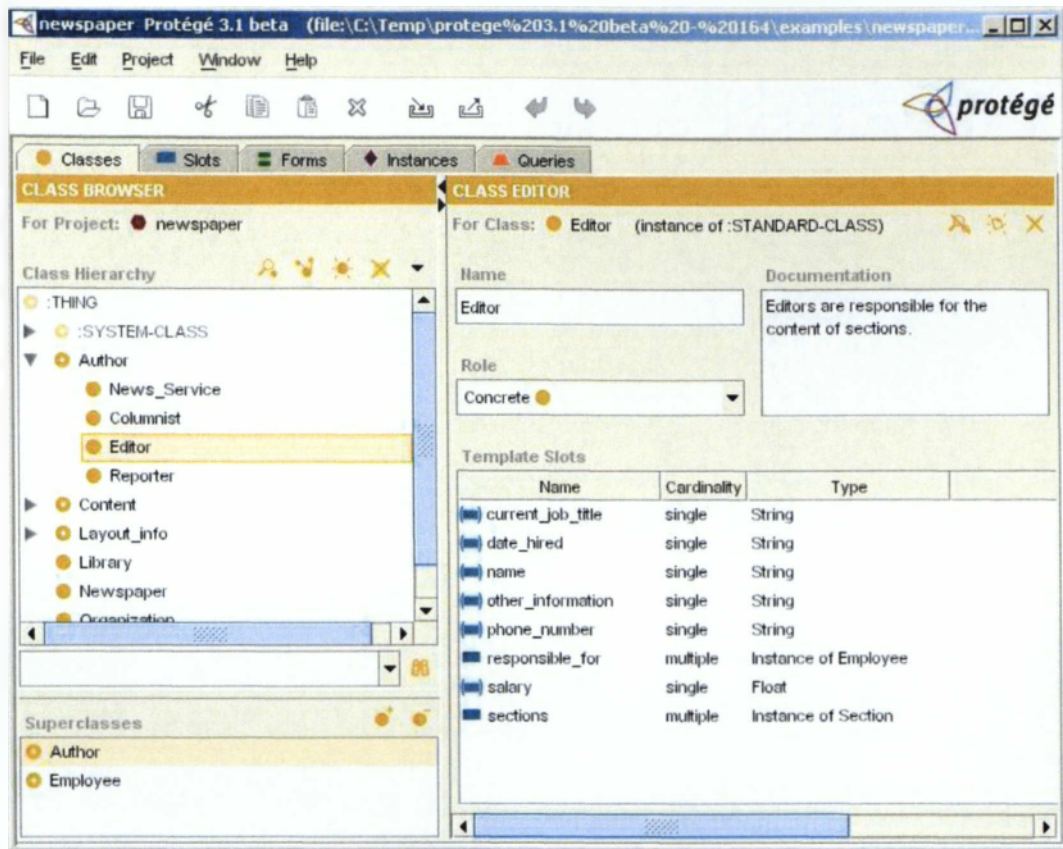
Protégé

Το Protégé⁴⁴ είναι ένας δωρεάν ανοικτού-κώδικα stand-alone συντάκτης οντολογιών με επεκτάσιμη αρχιτεκτονική. Αναπτύχθηκε από το Stanford Medical Informatics (SMI) του Stanford University σε συνεργασία με το University of Manchester σε Java, και παρέχει plug-and-play περιβάλλον που τον καθιστά ευέλικτη βάση για ταχεία προτυποποίηση και ανάπτυξη εφαρμογών.

Η πλατφόρμα Protégé υποστηρίζει δύο βασικούς τρόπους μοντελοποίησης οντολογιών μέσω των εφαρμογών επεξεργασίας Protégé-Frames και Protégé-OWL.

Η διαδραστική εφαρμογή επεξεργασίας Protégé-Frames έχει οργανωθεί σε μια σειρά από καρτέλες, που η κάθε μια εστιάζει σε μια ιδιαίτερη πτυχή της οικοδόμησης οντολογίας (εικόνα 4.1).

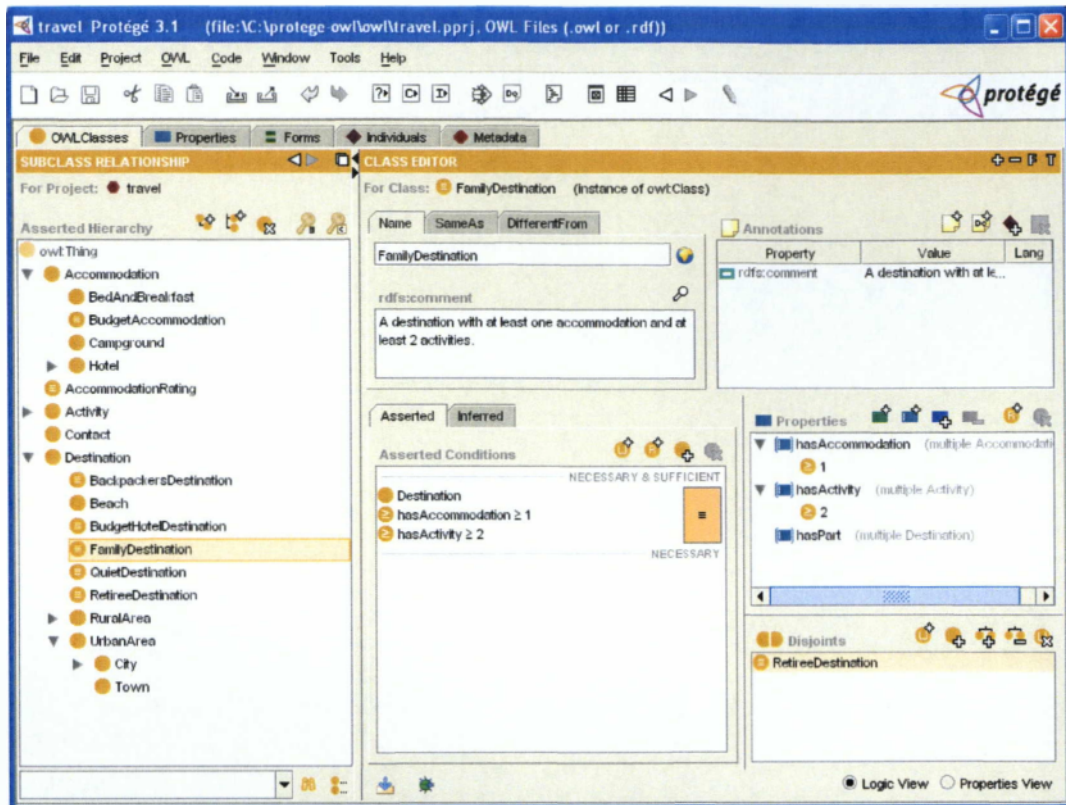
⁴⁴ <http://protege.stanford.edu/>



Εικόνα 4.1 Protégé-Frames

Πηγή: <http://protege.stanford.edu/>

Οι οντολογίες Protégé μπορούν να εξαχθούν σε διάφορες μορφές, συμπεριλαμβανομένων των RDF (S), OWL και XML Schema.



Εικόνα 4.2 Protégé OWL

Πηγή: <http://protege.stanford.edu/>

Η οπτική OWLClasses (εικόνα 4.2) μπορεί να χρησιμοποιηθεί για την επεξεργασία ιεραρχικής κατάταξης εννοιών. Οι λεπτομέρειες της επιλεγμένης κατηγορίας φαίνεται στο δεξί μέρος της οθόνης. Το επάνω μέρος της περιοχής αυτής επιτρέπει στους χρήστες να προσθέτουν σχόλια, ετικέτες και άλλα σχόλια. Το κάτω μέρος εμφανίζει τα λογικά χαρακτηριστικά της επιλεγμένης κατηγορίας.

Το Protégé υποστηρίζεται από μια ισχυρή κοινότητα των προγραμματιστών και των ακαδημαϊκών, την κυβέρνηση και τους εταιρικούς χρήστες, οι οποίοι χρησιμοποιούν αυτή τη πλατφόρμα για λύσεις γνώσης σε ποικίλους τομείς, όπως η βιοϊατρική, συλλογή πληροφοριών, και εταιρικών μοντέλων.

Chimaera

Η Chimaera⁴⁵ είναι ένα περιβάλλον συγχώνευσης και εύρεσης οντολογιών μέσω διαδικτυακής περιήγησης. Η σχεδίαση και υλοποίησή του βασίζεται σε γνωστικές διεπαφές και έχει ως στόχο την συνεργασία με πλήθος οντολογιών. Δέχεται πάνω από δεκαπέντε επιλογές σχεδιαστικών εισόδων (όπως ANSI KIF, Ontolingua, Protégé, CLASSIC, iXOL κ.α.) όπως και συμβατές μορφοποιήσεις OKBC. Παρέχει ένα απλό περιβάλλον επεξεργασίας και επιτρέπει στο χρήστη να χρησιμοποιήσει την πλήρη έκδοση του περιβάλλοντος περιήγησης/σύνταξης Ontolingua για πιο εκτεταμένη επεξεργασία. Διευκολύνει τη συγχώνευση, επιτρέποντας στους χρήστες να ανεβάσουν υπάρχουσες οντολογίες σε ένα νέο περιβάλλον εργασίας ή σε μία ήδη υπάρχουσα οντολογία. Μελλοντικά θα είναι σε θέση να είναι συμβατή και με άλλα πρότυπα συγχώνευσης όπως RDF και DAML.

Select KB content to upload:

Please select the file to upload and select "Do it" to merge the contents of the file into the current KB.

Upload a file:

Upload from a URL:

Load file on server:

Language:

Name of source KB:

Forms to upload:

- ANSI KIF
- CLASSIC
- CML
- COOL/CLIPS
- HPKB WITH ANSI KIF
- HPKB WITH KIF 3.0
- INDENTED INPUT
- KIF 3.0
- OCELOT
- OKBC
- OKBC WITH ANSI KIF
- ONTOLINGUA
- PROTEGE
- SNARK HPKB WITH ANSI KIF

Εικόνα 4.3 Περιβάλλον Chimaera

Πηγή: <http://www.ksl.stanford.edu/software/chimaera/>

⁴⁵ <http://www.ksl.stanford.edu/software/chimaera/>

OntoEdit

Το **OntoEdit**⁴⁶ είναι ένας συντάκτης οντολογιών που χρησιμοποιεί πρόσφατες μεθόδους ανάπτυξης οντολογιών και λογικά συμπεράσματα. Αποτελεί επεκτάσιμο και ευέλικτο περιβάλλον και ακολουθεί τρία βασικά βήματα (προσδιορισμός απαιτήσεων, βελτίωση και αξιολόγηση) για την ανάπτυξη των οντολογιών με σκοπό τη λειτουργικότητα. Αναπτύχθηκε από το Ινστιτούτο AIFB του Karlsruhe University. Είναι διαθέσιμο σε δύο εκδόσεις: το **OntoEditFree** και το **OntoEditProfessional**.

KAON και KAON2

Το **KAON**⁴⁷ (Karlsruhe Ontology) είναι μία υποδομή οντολογίας που αναπτύχθηκε από το University of Karlsruhe και το Research Center for Information Technologies της Karlsruhe. Αρχικά αναπτύχθηκε το 2002 και υποστήριζε μία βελτιωμένη έκδοση των RDF οντολογιών. Πολλά εργαλεία όπως ο γραφικός συντάκτης οντολογιών OIModeler ή ο KAON Server βασίστηκαν στο KAON.

```
<?xml version="1.0"?>
<!DOCTYPE Ontology [
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
]>
<owlx:Ontology owlx:name="http://www.example.org/wine"
  xmlns:owlx="http://www.w3.org/2003/05/owl-xml">

  <!--
    Contains optional header elements as well as
    any number of classes, properties, and individuals.
  -->

</owlx:Ontology>
```

Εικόνα 4.4 Υποδομή οντολογίας KAON

Πηγή: <http://kaon2.semanticweb.org/>

Το 2005 κυκλοφόρησε η πρώτη έκδοση του KAON2, προσφέροντας υποστήριξη στις OWL οντολογίες. Το KAON2 δεν είναι συμβατό με το KAON. Αναπτύχθηκε σαν μία κοινή προσπάθεια του Information Process Engineering (IPE) στο Research Center

⁴⁶ Sure, Y., Angele, J., Staab, S., (2002)

⁴⁷ <http://kaon2.semanticweb.org/>

for Information Technologies (FZI), του Institute of Applied Informatics and Formal Description Methods (AIFB) στο University of Karlsruhe και του Information Management Group (IMG) στο University of Manchester. Το KAON είναι ανοικτού κώδικα αλλά το KAON2 δεν είναι.

4.2.6 Οντολογίες στο χώρο της Υγείας – Υφιστάμενη Κατάσταση

Στο χώρο της υγείας έχουν γίνει αρκετές προσπάθειες ανάπτυξης οντολογιών που επιδιώκουν να ικανοποιήσουν το διαμοιρασμό και την επαναχρησιμοποίηση ιατρικών δεδομένων. Παρακάτω γίνεται αναφορά σε κάποιες από αυτές.

SNOMED CT

Η SNOMED CT⁴⁸ (Systematized Nomenclature Of Medicine Clinical Terms) (K. Spackman, K. Campbell) είναι μία συστηματικά οργανωμένη υπολογιστικά επεξεργάσιμη συλλογή ιατρικών - κλινικών όρων που παρέχει κώδικες, ορισμούς, εκθέσεις, συνώνυμα και επεξηγήσεις οι οποίοι καλύπτουν νόσους, ευρήματα, διαδικασίες, μικροοργανισμούς, ουσίες κλπ.. Οι έννοιες της συγκεκριμένης συλλογής είναι αντιπροσωπευτικές μονάδες που κατατάσσουν όλα τα αντικείμενα που χαρακτηρίζουν τις διαδικασίες υγειονομικής περίθαλψης και πρέπει να καταγράφονται σε αυτή. Επιτρέπει την έγκυρη κατηγοριοποίηση, αποθήκευση, ανάκτηση και ομαδοποίηση κλινικών δεδομένων όπως και διευκολύνει στην οργάνωση περιεχομένου ιατρικών αρχείων, μειώνοντας με αυτό τον τρόπο τις διαφορές που υπάρχουν στον τρόπο συλλογής, κωδικοποίησης και χρήσης δεδομένων που χρησιμοποιούνται για την ιατρική περίθαλψη ασθενών και την έρευνα. Η οντολογία SNOMED-CT προέκυψε από τη συγχώνευση της ορολογίας αναφοράς SNOMED (SNOMED RT) που αναπτύχθηκε από το Κολέγιο Αμερικανών Παθολόγων (CAP) και την Έκδοση Κλινικών Όρων 3(CTV3) που αναπτύχθηκε από την Εθνική Υπηρεσία Υγείας (NHS) του Ηνωμένου Βασιλείου.

⁴⁸ Spackman, K., Campbell, K., et al. (1997)

GALEN⁴⁹

Η GALEN⁵⁰ αποτελεί μία τεχνολογία που έχει σχεδιαστεί για την αναπαράσταση κλινικών πληροφοριών και παράγει ένα computer-based πολυγλωσσικό σύστημα κωδικοποίησης για την ιατρική, χρησιμοποιώντας μία ποιοτικά διαφορετική προσέγγιση από εκείνες που έχουν χρησιμοποιηθεί στο παρελθόν. Συγκεκριμένα, χρησιμοποιείται ένα ακριβές σημασιολογικό μοντέλο κλινικής ορολογίας που αναπαριστάται μέσω μίας τυπικά ορισμένης γλώσσας (GRAIL).

RxNorm

Η RxNorm (O. Bodenreider⁵¹) αναπτύχθηκε από την National Library of Medicine (NLM) και παρέχει κανονικοποιημένη ονοματολογία και μοναδικά αναγνωριστικά για τα φάρμακα και τις φαρμακευτικές ουσίες. Λόγω του ότι τα υπολογιστικά συστήματα που υπάρχουν στα νοσοκομεία και στους φαρμακευτικούς οργανισμούς, χρησιμοποιούν διαφορετική ορολογία για τα ονόματα των φαρμάκων, υπάρχει δυσκολία στην επικοινωνία μεταξύ αυτών των συστημάτων. Για το λόγο αυτό αναπτύχθηκε η RxNorm η οποία εξυπηρετεί δύο στόχους. Αρχικά επιδιώκει τη δημιουργία ενός συστήματος κοινής ορολογίας για τα γενόσημα και επώνυμα φάρμακα και επιπλέον την σύσταση ενός εργαλείου για την υποστήριξη της σημασιολογικής διαλειτουργικότητας μεταξύ της φαρμακευτικής ορολογίας και των φαρμακευτικών βάσεων δεδομένων. Απώτερος σκοπός της εν λόγω ορολογίας είναι να καταγράψει και να καλύψει όλες τις φαρμακευτικές συνταγογραφήσεις στις Ηνωμένες Πολιτείες.

MeSH

Το MeSH⁵² αναπτύχθηκε από την National Library of Medicine (NLM) και αποτελεί ένα ελεγχόμενο λεξιλόγιο προκαθορισμένων όρων (vocabulary thesaurus) που χρησιμοποιείται για την κατάταξη και εύρεση αναφορών της PubMed Central. Οι

⁴⁹ <http://www.opengalen.org/>

⁵⁰ Rector, A., Solomon, W., Nowlan, W., Rush, T., (1995)

⁵¹ Bodenreider, O., (2006)

⁵² <http://www.nlm.nih.gov/mesh/meshhome.html>

MeSH όροι έχουν σχεδιαστεί σε μια ιεραρχία όρων και φράσεων η οποία επιτρέπει με αυτόν τον τρόπο την ανάκτηση πληροφορίας που μπορεί να χρησιμοποιεί διαφορετικούς όρους για να περιγράψει την ίδια έννοια. Το λεξιλόγιο αυτό έχει επίσης σχεδιαστεί ως εγκυκλοπαιδικός θησαυρός, ώστε να διευκολύνεται η αναζήτηση. Οι MeSH περιγραφείς είναι αλφαβητικά και ιεραρχικά κατανεμημένοι. Στο γενικότερο επίπεδο της ιεραρχίας ανήκουν γενικές επικεφαλίδες όπως «Ανατομία» και «Διανοητικές Διαταραχές», ενώ σε χαμηλότερα επίπεδα της ιεραρχίας συναντώνται συγκεκριμένες επικεφαλίδες όπως «Αγκώνας» και «Διαταραχή Διεύθυνσης». Στο MeSH 2013 περιλαμβάνονται 26.853 περιγραφείς και 213.000 όρους εισόδου που εξυπηρετούν την εύρεση της πιο κατάλληλης MeSH επικεφαλίδας (π.χ. η «Vitamin C» αποτελεί όρο εισόδου για το «Ascorbic Acid»).

UMLS

Το UMLS⁵³ (Unified Medical Language System) σχεδιάστηκε και δημιουργήθηκε το 1986 από τη U.S. National Library of Medicine (NLM) και είναι μία συλλογή από πολλά ελεγχόμενα λεξιλόγια της βιοϊατρικής επιστήμης. Παρέχει μία χαρτογραφημένη δομή μεταξύ αυτών των λεξιλογίων και έτσι επιτρέπει τη μετάφραση μεταξύ διαφόρων συστημάτων ορολογίας. Θεωρείται και ως ένας ολοκληρωμένος θησαυρός και οντολογία των βιοϊατρικών εννοιών⁵⁴. Η UMLS⁵⁵ αποτελείται από διάφορες βάσεις δεδομένων και από ένα σύνολο εργαλείων λογισμικού και περιλαμβάνει τρεις πηγές γνώσης:

- Το Metathesaurus: περιέχει σημασιολογική πληροφορία για τις έννοιες της βιοϊατρικής, τις ορολογίες που χρησιμοποιούνται και τις μεταξύ τους σχέσεις.
- Το Σημασιολογικό Δίκτυο (Semantic Network): περιέχει γενικές κατηγορίες μέσω των οποίων προσδιορίζονται όλες οι έννοιες του Metathesaurus.
- Το Λεξικό Εξειδικευμένων όρων (Specialist lexicon): περιέχει συντακτική πληροφορία που αφορά βιοϊατρικούς όρους και επιχειρεί να καλύψει την πλειονότητα των όρων που συνθέτουν τις έννοιες που βρίσκονται στο Metathesaurus.

⁵³ <http://www.nlm.nih.gov/research/umls/>

⁵⁴ Pisanelli, D., Gangemi, A., Steve, G., (1998)

⁵⁵ *UMLS Knowledge Sources*, National Library of Medicine, (1999)

ICD - 10

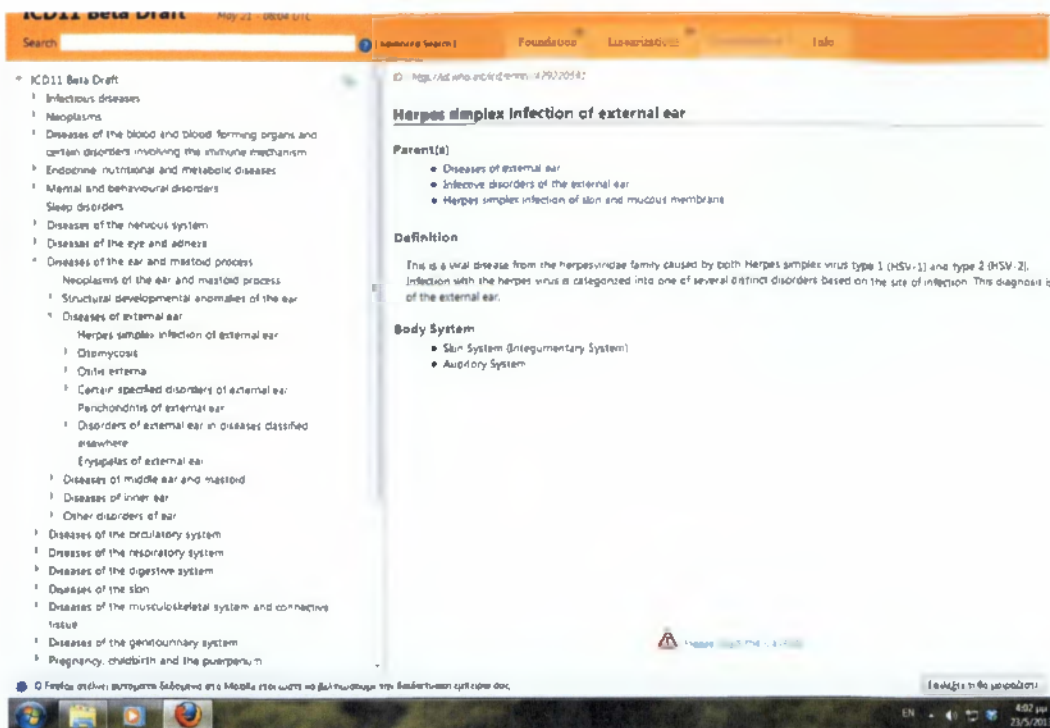
Η ICD⁵⁶ (International Classification of Diseases) είναι ένα πρότυπο διαγνωστικό εργαλείο για την επιδημιολογία, τη διαχείριση της υγείας και για κλινικούς σκοπούς. Αυτό περιλαμβάνει την ανάλυση της γενικής κατάστασης της υγείας πληθυσμιακών ομάδων.

Χρησιμοποιείται για να παρακολουθεί τη συχνότητα της εμφάνισης και την διάρκεια των ασθενειών όπως και άλλων προβλημάτων υγείας. Επίσης έχει ως στόχο την ταξινόμηση νοσημάτων και προβλημάτων υγείας που είναι καταγεγραμμένα σε αρχεία ασθενών, σε ιατρικούς φακέλους ακόμα και σε πιστοποιητικά θανάτου.

Εκτός από τη δυνατότητα αποθήκευσης και ανάκτησης διαγνωστικών πληροφοριών για κλινικούς, επιδημιολογικούς σκοπούς και σκοπούς ποιότητας, τα αρχεία αυτά παρέχουν τη βάση για τη σύνταξη των εθνικών στατιστικών νοσηρότητας και θνησιμότητας των κρατών μελών του WHO (World Health Organization).

Η ICD-10 εγκρίθηκε από την Τεσσαρακοστή Τρίτη Παγκόσμια Συνέλευση Υγείας το Μάιο του 1990 και τέθηκε σε εφαρμογή από τα κράτη μέλη του WHO από το 1994.

⁵⁶ International Classification of Diseases, 10th Revision. <http://www.who.int/classifications/icd/en/>



Εικόνα 4.5 Διεθνής Ταξινόμηση των Νόσων – ICD 11

Πηγή: <http://www.who.int/classifications/icd/en/>

Η 11η αναθεώρηση της ταξινόμησης έχει ήδη ξεκινήσει και θα συνεχιστεί μέχρι το 2015. Η ICD-10 είναι διαθέσιμη σε 42 διαφορετικές γλώσσες όπου συμπεριλαμβάνεται και η ελληνική.

Gene Ontology (GO)

Η Gene Ontology⁵⁷, ή GO, είναι μία σημαντική πρωτοβουλία της βιοπληροφορικής για τη συνεπή αναπαράσταση των γονιδίων και των γονιδιακών χαρακτηριστικών όλων των γονιδιωμάτων. Πιο συγκεκριμένα, στόχοι του προγράμματος είναι:

- Η διατήρηση και η ανάπτυξη του ελεγχόμενου λεξιλογίου του γονιδίου και των χαρακτηριστικών του γονιδιακού προϊόντος.
- Η καταγραφή των γονιδίων και των γονιδιακών προϊόντων, η αφομοίωση και η διάδοση των καταγραφών

⁵⁷ The Gene Ontology: <http://www.geneontology.org>.

- Η παροχή εργαλείων για την εύκολη πρόσβαση σε όλες τις πτυχές των δεδομένων που παρέχονται από το πρόγραμμα

Η Gene Ontology αποτελείται από τρεις οντολογίες με τους εξής κορυφαίους κόμβους:

- Cellular Component
- Molecular Function
- Biological Process

Η οντολογία GO αποτελεί μέλος μίας μεγαλύτερης προσπάθειας ταξινόμησης, την OBO (Open Biomedical Ontologies).

OBO

Η OBO⁵⁸ (Open Biomedical Ontologies) αποτελεί μία προσπάθεια για τη δημιουργία ελεγχόμενων λεξιλογίων για κοινή χρήση σε διαφορετικούς βιολογικούς και ιατρικούς τομείς. Αυτή η πρωτοβουλία έχει ως σκοπό τη διαλειτουργικότητα αυτών των οντολογιών, βάσει μίας κοινής φιλοσοφίας σχεδιασμού και υλοποίησης, επιτρέποντας με αυτό τον τρόπο σε επιστήμονες και σε υπολογιστικά μέσα να επικοινωνούν μεταξύ τους με την ελάχιστη δυνατή ασάφεια. Μέσω αυτής της διαδικασίας τα δεδομένα που προκύπτουν κατά τη διάρκεια της βιοϊατρικής έρευνας, αποτελούν ένα ενιαίο, συνεπές, επεκτάσιμο και αλγοριθμικά διαχειρίσιμο σύνολο. Αυτός ο πυρήνας θα είναι γνωστός ως «OBO Foundry». Από το 2006 οι οντολογίες OBO αποτελούν μέρος των πόρων του U.S. National Center for Biomedical Ontology, και θα αποτελέσουν κύριο στοιχείο του περιεχομένου του BioPortal του NCBO (National Center for Biomedical Ontology).

FMA (Foundation Model Of Anatomy)

Το FMA⁵⁹ αναπτύχθηκε από το Structural Informatics Group του University of Washington και είναι μία εξελισσόμενη computer-based πηγή γνώσης για την

⁵⁸ <http://www.obofoundry.org/>

⁵⁹ <http://sig.biostr.washington.edu/projects/fm/index.html>

βιοϊατρική πληροφορική με αντικείμενο την παρουσίαση των κλάσεων, των ειδών και των σχέσεων (domain ontology) που είναι απαραίτητα για τη συμβολική αναπαράσταση της φαινοτυπικής δομής του ανθρώπινου σώματος σε μία μορφή που είναι κατανοητή από τον άνθρωπο και επίσης ερμηνεύσιμο και διαχειρίσιμο από υπολογιστικά συστήματα και μηχανές. Περιέχει πάνω από 80.000 έννοιες ανατομίας και αποτελεί μία από τις πιο συστηματικά δομημένες οντολογίες στη βιοϊατρική.

ON9

Η ON9⁶⁰ υλοποιήθηκε από το Τμήμα Ιατρικής Πληροφορικής του Ινστιτούτου Βιοϊατρικής Τεχνολογίας της Ρώμης σε μία προσπάθεια ενοποίησης της ορολογίας από διάφορες βάσεις δεδομένων. Οι γλώσσες υλοποίησης είναι η Loom και η Ontolingua και προς το παρόν περιλαμβάνει πέντε πηγές γνώσης: τη UMLS-SN, την ICD 10, την GMN, την SNOMED III και την CORE.

SBO (Systems Biology Ontology)

Η SBO⁶¹ αποτελεί ένα σύνολο ελεγχόμενων, συσχετιζόμενων λεξιλογίων όρων που χρησιμοποιούνται στα βιολογικά συστήματα και συγκεκριμένα στην υπολογιστική μοντελοποίηση. Αποτελείται από επτά orthogonal λεξιλόγια που καθορίζουν: ρόλους συμμετεχόντων, ποσοτικές παραμέτρους, ταξινόμηση μαθηματικών εκφράσεων που περιγράφουν το σύστημα, πλαίσιο μοντελοποίησης που χρησιμοποιούνται, η φύση του φορέα, τον τύπο αλληλεπίδρασης καθώς και μία διακλάδωση προς τον καθορισμό των διαφόρων τύπων των μεταδεδομένων που μπορεί να υπάρχουν μέσα σε ένα μοντέλο. Οι όροι της SBO μπορούν να χρησιμοποιηθούν για να εισάγουν ένα επίπεδο σημασιολογικής πληροφορίας στην τυπική περιγραφή ενός μοντέλου, ή για την καταγραφή και το σχολιασμό των αποτελεσμάτων βιοχημικών πειραμάτων, προκειμένου να διευκολυνθεί η αποτελεσματική επαναχρησιμοποίηση τους. Η SBO αποτελεί έργο της BioModels.com και αναπτύχθηκε μέσα από κοινοτική συνεργασία και αποτελεί υποψήφια οντολογία των Open Biomedical Ontologies (OBO).

⁶⁰ Pisanelli, D., Gangemi, A., Steve, G., (1997)

⁶¹<http://www.ebi.ac.uk/sbo/main/>

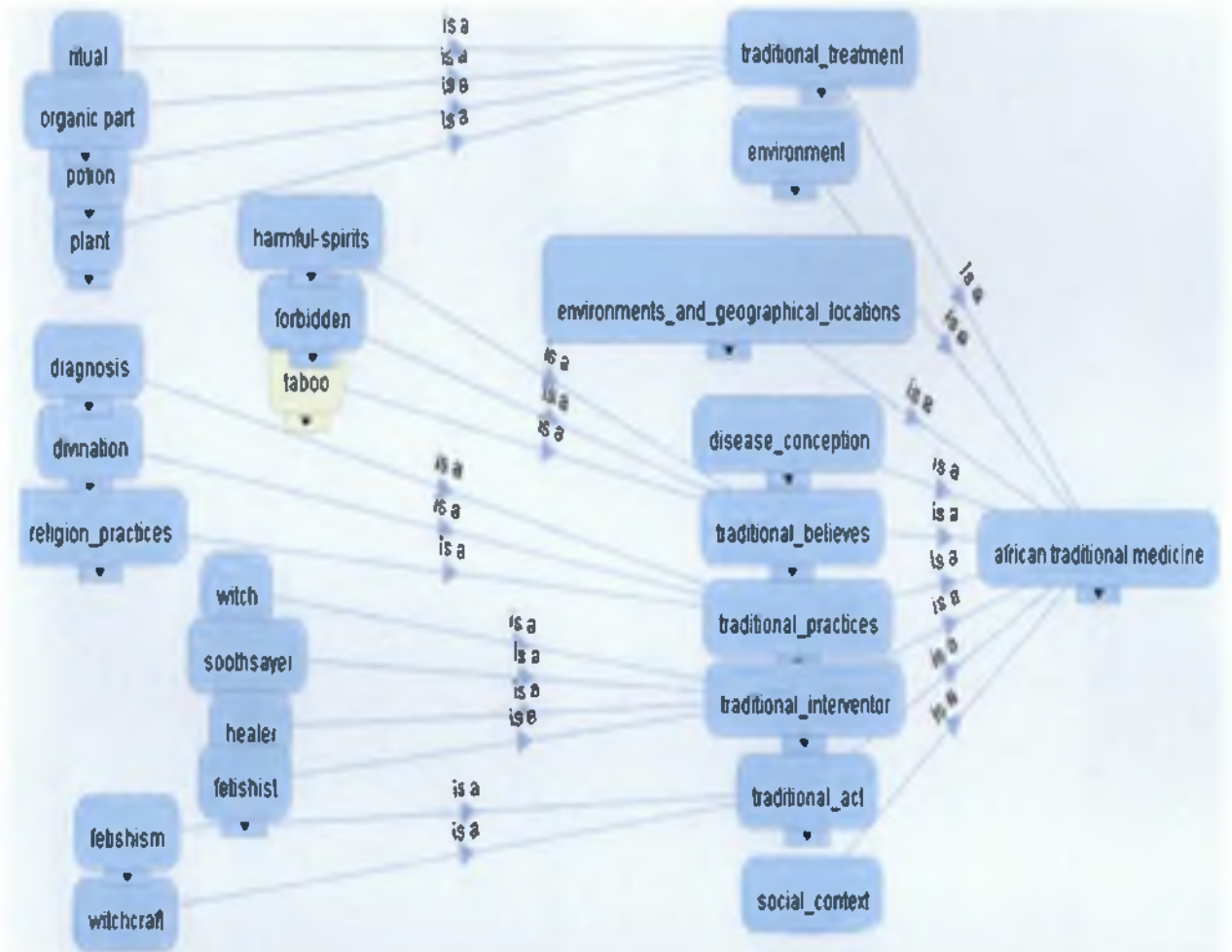
4.3 ΠΑΡΑΔΕΙΓΜΑΤΑ ΟΝΤΟΛΟΓΙΩΝ ΥΓΕΙΑΣ

Στην υποενότητα αυτή παρουσιάζονται κάποια ενδεικτικά παραδείγματα οντολογιών στο χώρο της υγείας.

4.3.1 Οντολογία Παραδοσιακής Αφρικανικής Ιατρικής (ΑΤΜΟ)

Η Οντολογία Παραδοσιακής Αφρικανικής Ιατρικής (African Traditional Medicine Ontology – ΑΤΜΟ) περιγράφει τη λειτουργία αυτών που την ασκούν (θεραπευτής, φετιχιστής ή μάντης). Επίσης περιγράφει τα διάφορα είδη των προτεινόμενων θεραπειών, τους ρόλους του συμπτωμάτων και την εξέταση της νόσου⁶².

⁶² <http://bioportal.bioontology.org/ontologies/1099>



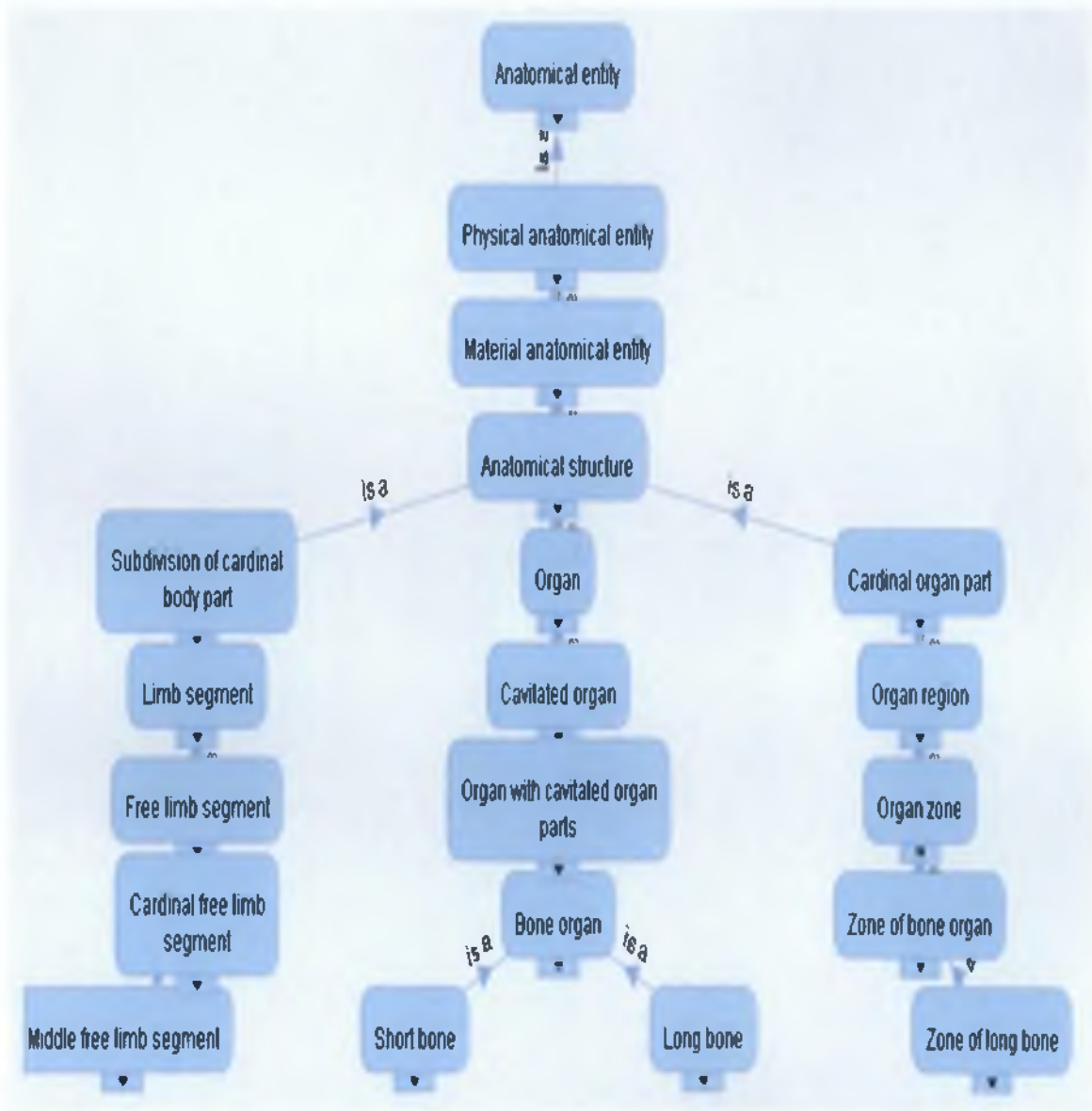
Εικόνα 4.6 Οντολογία Παραδοσιακής Αφρικανικής Ιατρικής

Πηγή: <http://bioportal.bioontology.org/ontologies/40223?p=terms>

4.3.2 Οντολογία Δυσπλασίας των Οστών

Η οντολογία για την δυσπλασία των οστών (Bone Dysplasia Ontology – BDO) παρέχει μια ολοκληρωμένη και επίσημη εκπροσώπηση των διαφόρων κύριων εννοιών που εμπλέκονται στην πλήρη τεκμηρίωση της πολυπλοκότητας του τομέα της σκελετικής δυσπλασίας. Συλλαμβάνει και συνδυάζει τα γενετικά χαρακτηριστικά που διακρίνουν τις δυσμορφίες των οστών με το πλήθος των φαινοτυπικών

χαρακτηριστικών που περιγράφεται από τους ασθενείς και πρέπει να ληφθούν υπόψη προκειμένου να υποστηριχθεί η διαδικασία της διάγνωσης⁶³.



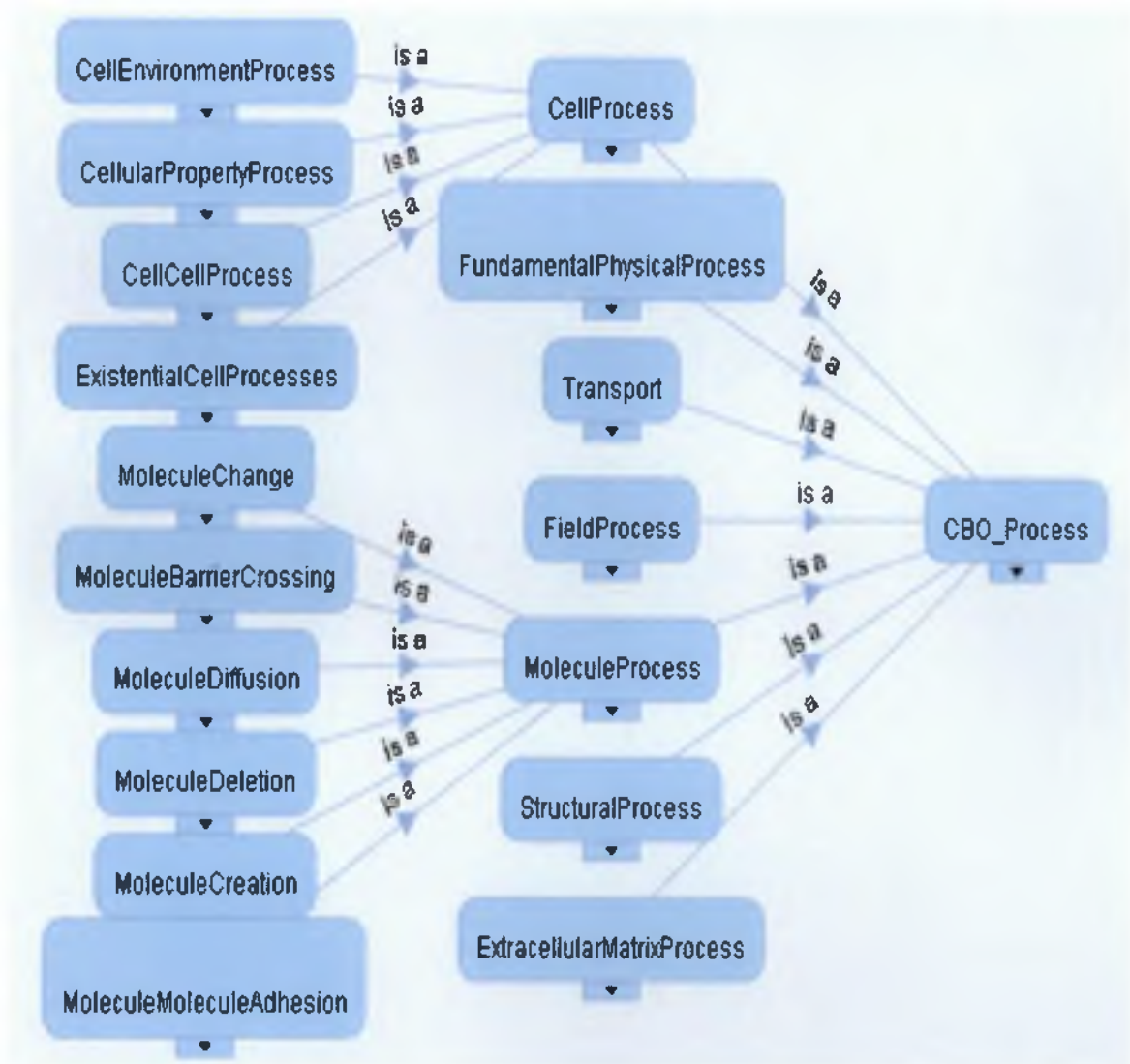
Εικόνα 4.7 Οντολογία Δυσπλασίας των Οστών

Πηγή: <http://bioportal.bioontology.org/ontologies/46875?p=terms>

⁶³ <http://bioportal.bioontology.org/ontologies/1613>

4.3.3 Οντολογία Συμπεριφοράς Κυττάρων (CBO)

Η Οντολογία για την συμπεριφορά των κυττάρων (Cell Behavior Ontology – CBO) περιγράφει πολυκυτταρικά υπολογιστικά μοντέλα. Ειδικότερα, σχεδιάστηκε για να περιγράψει τόσο τις υπαρκτικές συμπεριφορές των κυττάρων (χωρικότητα, την ανάπτυξη, την κυκλοφορία, την προσκόλληση, θάνατο κτλ) όσο και υπολογιστικά μοντέλα των εν λόγω συμπεριφορών⁶⁴.



Εικόνα 4.8 Οντολογία Συμπεριφοράς Κυττάρων

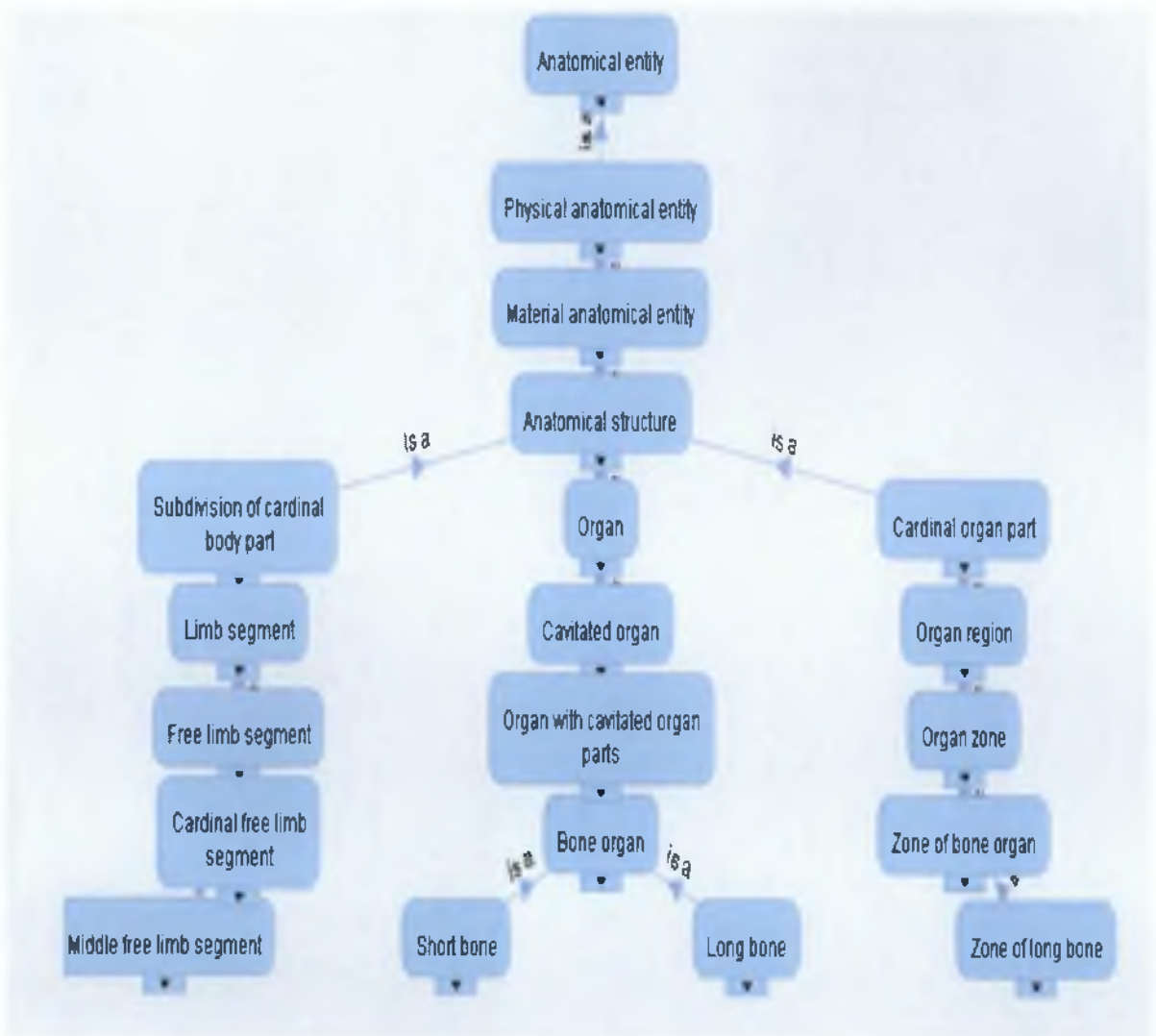
Πηγή: <http://bioportal.bioontology.org/ontologies/50108?p=terms>

⁶⁴ <http://bioportal.bioontology.org/ontologies/1158>

4.3.4 Οντολογία Ανατομίας (ΑΕΟ)

Η Οντολογία Ανατομίας (Anatomical Entity Ontology – ΑΕΟ) είναι μια οντολογία των ανατομικών δομών που επεκτείνει την Οντολογία Κοινής Αναφοράς Ανατομίας, σε περίπου 200 τάξεις χρησιμοποιώντας τη σχέση «is_a». Με αυτόν τον τρόπο παρέχεται μια λεπτομερής ταξινόμηση για τους ιστούς. Οι νέες τάξεις επιλέχθηκαν για τη χρήση τους στην κατηγοριοποίηση των κύριων οντολογιών ανατομίας σπονδυλωτών και ασπόνδυλων σε μια επαρκή ανάλυση για ιστούς ενός μόνο κυτταρικού τύπου. Η οντολογία είναι χρήσιμη για την αύξηση της ποσότητας των γνώσεων στις οντολογίες ανατομίας, τη διευκόλυνση του σχολιασμού και επιτρέπει διαλειτουργικότητα μεταξύ οντολογιών και ανατομίας⁶⁵.

⁶⁵ <http://bioportal.bioontology.org/ontologies/1568>



Εικόνα 4.9 Οντολογία Ανατομίας

Πηγή: <http://bioportal.bioontology.org/ontologies/47581?p=terms>

4.3.5 Οντολογία Νεύρο – Συμπεριφοράς (NBO)

Η Οντολογία Νεύρο – Συμπεριφοράς (Neurobehavior Ontology – NBO) αποτελείται από δύο κύριες συνιστώσες, μια οντολογία της συμπεριφοράς των διαδικασιών και μια οντολογία της συμπεριφοράς φαινοτύπων. Το σκέλος της διαδικασίας της συμπεριφοράς περιέχει μια ταξινόμηση των διαδικασιών συμπεριφοράς που συμπληρώνει και επεκτείνει προϋπάρχουσες οντολογίες πάνω σε αυτό το αντικείμενο. Ο κλάδος συμπεριφοράς φαινοτύπων αποτελείται από μια κατάταξη των

φυσιολογικών και μη φυσιολογικών χαρακτηριστικών της συμπεριφοράς των οργανισμών⁶⁶.



Εικόνα 4.10 Οντολογία Νεύρο - Συμπεριφοράς

Πηγή: <http://bioportal.bioontology.org/ontologies/49823?p=terms>

⁶⁶ <http://bioportal.bioontology.org/ontologies/1621>

4.4 ΙΑΤΡΟΛΕΞΗ: ΕΛΛΗΝΙΚΗ ΟΝΤΟΛΟΓΙΑ ΥΓΕΙΑΣ

Το έργο ΙΑΤΡΟΛΕΞΗ (<http://www.iatrolexi.gr>) είχε ως αντικείμενο τη δημιουργία της απαραίτητης γλωσσικής υποδομής για την Ελληνική γλώσσα, που επέτρεψε σε πρώτη φάση την καλύτερη διαχείριση και επεξεργασία της βιοϊατρικής πληροφορίας που υπήρχε σε ψηφιακή μορφή. Αυτό κατέστη δυνατό με τη δημιουργία και διάθεση στους χρήστες εξελιγμένων εργαλείων για την αναζήτηση, τη συσχέτιση και το χαρακτηρισμό των βιοϊατρικών κειμένων⁶⁷.

Σε επόμενη φάση (πέρα από τα πλαίσια του παρόντος), είναι δυνατή η επέκταση των αποτελεσμάτων του και η διαμόρφωσή του ώστε να εξυπηρετεί και άλλους σκοπούς (πχ. υλοποίηση δυνατοτήτων semantic web σε βιοϊατρικούς ιστοχώρους, κλπ.).

Το έργο χρηματοδοτήθηκε από τη ΓΓΕΤ και υλοποιήθηκε σε συνεργασία με την εταιρία Neurosoft ΑΕ. Ξεκίνησε το 04/2006 και ολοκληρώθηκε τον 12/2007.

⁶⁷ <http://www.iatrolexi.gr>

Εργαλεία

ΕΝΘΗΤΑ ΕΡΓΑΣΙΑΣ	ΤΙΤΛΟΣ	ΠΕΡΙΓΡΑΦΗ
ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Περιγητής Οντολογίας	Με το εργαλείο αυτό ο χρήστης μπορεί να πλοηγηθεί στην οντολογία των βιοϊατρικών όρων του ΙΑΤΡΟΛΕΞΗ και να αναζητήσει βιοϊατρικούς όρους μέσα σε αυτήν.
ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Μηχανή Αναζήτησης	Με το εργαλείο αυτό ο χρήστης μπορεί να κάνει αναζήτηση στο σώμα βιοϊατρικών κειμένων, τα οποία έχουν ευρετηριαστεί επί τη βάση σημασιολογικών κατηγοριών από την οντολογία, αλλά και σε κείμενα του Παγκόσμιου Ιστού.
ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Ορθογραφικός Διορθωτής Ιστού	Με το εργαλείο αυτό ο χρήστης μπορεί να ελέγξει την ορθογραφία μοναλεκτικού βιοϊατρικού όρου ή όρου της κοινής ελληνικής γλώσσας.
ΕΕ5. Υλοποίηση Εφαρμογών Ιστού	Μορφολογικός και Σημασιολογικός Σχολιαστής	Με το εργαλείο αυτό ο χρήστης μπορεί να εσάγει ένα κείμενο και να επισυνάψει μορφολογικά (από το Μορφολογικό Λεξικό) και σημασιολογικά (από την Οντολογία) σφάλματα σε λέξεις/στοιχεία σε αυτό.
ΕΕ2. Υλοποίηση Εργαλείων Διαχείρισης Σώματος Κειμένων και Αυτόματης Εξαγωγής Ορολογίας	Συμφραζόμενα όρων (concordances)	Το εργαλείο αυτό επιτρέπει την αναζήτηση όρων στη συλλογή κειμένων (corpus) και επιστρέφει τον όρο (είτε στη μορφή που δόθηκε, είτε σε κάθε κλιτικό τύπο εφόσον επιλεγεί η «λημματοποίηση») μαζί με τα συμφραζόμενά του (μια γραμμή) από το κείμενο που βρέθηκε. Ο χρήστης μπορεί στην συνέχεια να επιλέξει και να δει το κείμενο, όπου σημειώνεται που ακριβώς έχει βρεθεί ο όρος.

2482 Επισκέπτες από 13-08-2006

W3C HTML 4.01

Αρχική Σελίδα Επικοινωνία Δικαιώματα Πνευματικής Ιδιοκτησίας Χάρτης Πλοήγησης
 Σχεδίαση & Ανάπτυξη: Εργαστήριο ΗΡΩΗ - EAITY
 Για την ανάπτυξη αυτής της δικτυακής Πύλης χρησιμοποιήθηκε ανοιχτό λογισμικό υπό την άδεια GNU/GPL.

Εικόνα 4.11 ΙΑΤΡΟΛΕΞΗ – Η Ελληνική Οντολογία

Πηγή: <http://www.iatrolexi.gr>

Μια από τις εφαρμογές του ιστού που αναπτύχθηκε στα πλαίσια του έργου, είναι η «Μηχανή Αναζήτησης» μέσω της οποίας ο χρήστης μπορεί να αναζητήσει πληροφορίες στη συλλογή βιοϊατρικών κειμένων του ΙΑΤΡΟΛΕΞΗ.

5 ΚΕΦΑΛΑΙΟ: ΑΝΑΛΥΣΗ MDS ONTOLOGY

5.1 ΤΙ ΕΙΝΑΙ ΤΟ MDS ONTOLOGY

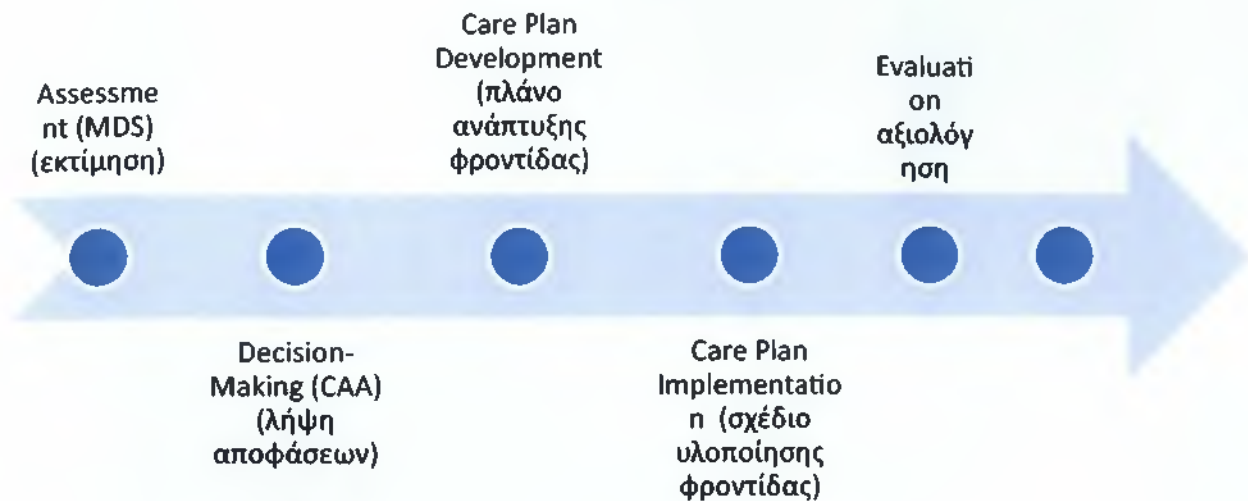
Το MDS (Minimum Data Set) ή αλλιώς Ελάχιστο Σύνολο Δεδομένων είναι ένα σημείο εκκίνησης. Το Ελάχιστο Σύνολο Δεδομένων (MDS) είναι ένα τυποποιημένο όργανο που χρησιμοποιείται για την αξιολόγηση των γηροκομείων. Πρόκειται για μια συλλογή βασικών φυσικών (π.χ., ιατρικές παθήσεις, τη διάθεση και το όραμα), λειτουργικών (π.χ., δραστηριότητες της καθημερινής ζωής, συμπεριφορά), και ψυχοκοινωνικών (π.χ., τις προτιμήσεις, τους στόχους και τα συμφέροντα) πληροφοριών σχετικά με τους κατοίκους. Όταν ολοκληρωθεί, το MDS παρέχει τη βάση για μια πιο διεξοδική αξιολόγηση και την ανάπτυξη της ένα εξατομικευμένο πρόγραμμα περίθαλψης. Τα εγχειρίδια MDS 3,0 εξηγούν λεπτομερώς πώς να ολοκληρωθεί το MDS⁶⁸.

Στην παρούσα εργασία ακολουθώντας την OWL μεθοδολογία ανάπτυξης οντολογιών δημιουργήσαμε την οντολογία MDS με τη χρήση του protégé.

Συγκεκριμένα η οντολογία αυτή αφορά στη καταγραφή του προφίλ του ασθενή σε σχέση με τη κατάσταση της υγείας του προκειμένου να του παρασχεθεί η κατάλληλη ιατρική νοσηλευτική και φαρμακευτική υποστήριξη. Αφορά κυρίως υπερήλικες που βρίσκονται σε γηροκομεία ή ασθενείς σε κέντρα φροντίδας ή άτομα τα οποία έχουν ανάγκη υποστήριξης φροντίδας από το σπίτι. Η οντολογία ουσιαστικά που αναπτύσσουμε κωδικοποιεί τα στοιχεία που ήδη χρησιμοποιούνται από μέλη της κοινότητας παροχής υπηρεσιών υγείας (κυρίως από την κυβέρνηση της USA (CMS)

⁶⁸<http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/MDS30RAIManual.html>

σε μορφή οντολογίας. Αυτό θα βοηθήσει στη καλύτερη διαχείριση, επεξεργασία και γενικά αξιοποίηση της σχετιζόμενης με την παροχή φροντίδας πληροφορίας⁶⁹.



5.2 MDS ONTOLOGY

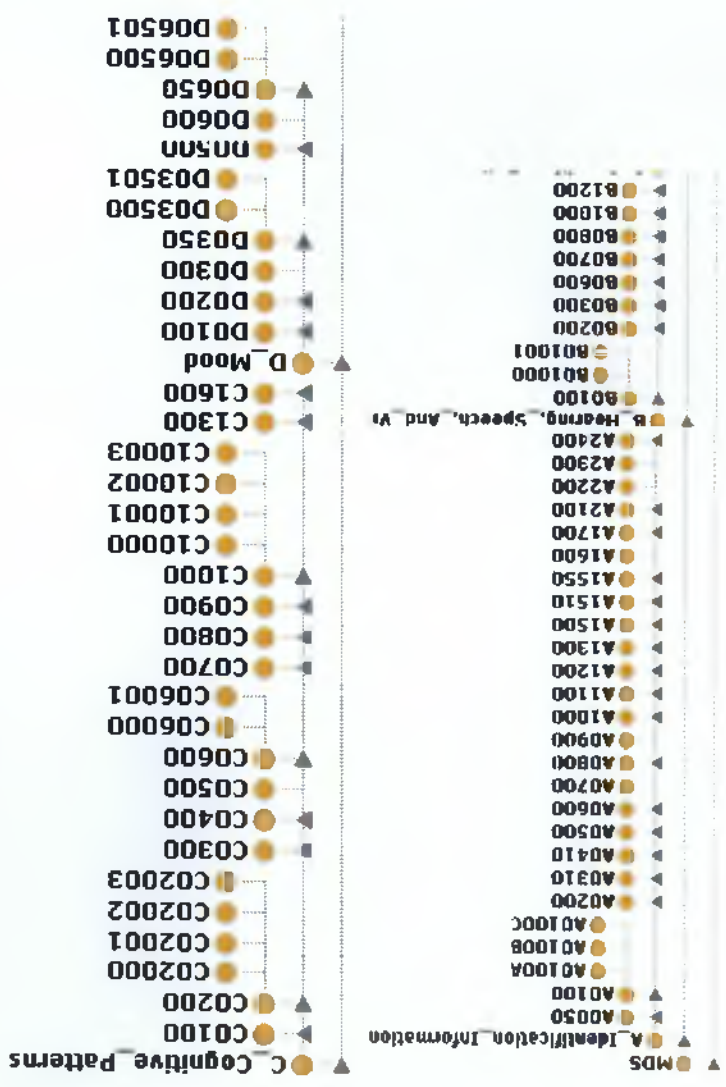
Η συγκεκριμένη οντολογία, που ονομάζεται MDS ontology έχει δημιουργηθεί με το πρόγραμμα Protégé και αποτελείται από συνολικά 20 κλάσεις. Κάποιες από αυτές έχουν δικές τους υποκλάσεις, ενώ κάποιες άλλες όχι.

Οι κλάσεις είναι:

- a. Identification information (στοιχεία ταυτότητας)
- b. Hearing, speech and vision (ακοή, ομιλία, όραση)
- c. Cognitive patterns (γνωστικά σχήματα)
- d. Mood (διάθεση)

⁶⁹<http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/MDS30RAIManual.html>

- e. Behavior (συμπεριφορά)
- f. Preferences for customary routine and activities (προτιμήσεις για τη συνήθη ρουτίνα (καθημερινός) & δραστηριότητες)
- g. Functional status (λειτουργική κατάσταση)
- h. Bladder and Bowel (ουροδόχος κύστη & έντερο)
- i. Active diagnoses (ενεργή διάγνωση)
- j. Health condition (κατάσταση υγείας)
- k. Swallowing/ Nutritional status (κατάποση/ διατροφική κατάσταση)
- l. Oral/ dental status (κατάσταση στόματος/ δοντιών)
- m. Skin conditions (κατάσταση δέρματος)
- n. Medications (φάρμακα)
- o. Special treatments, procedures and programs (ειδικές θεραπείες, διαδικασίες & προγράμματα)
- p. Restraints (περιορισμοί)
- q. Participation in assessment and goal setting (συμμετοχή στην αξιολόγηση και τον καθορισμό στόχου)
- v. Care area assessment summary (περίληψη της αξιολόγησης του τομέα φροντίδας)
- x. Correction request (αίτηση διόρθωσης)
- z. Assessment administration (αξιολόγηση διοίκησης)



- ▼ E_Behavior
 - ▶ E0100
 - ▼ E0200
 - ▶ E0200A
 - ▼ E0200B
 - E0200B0
 - E0200B1
 - E0200B2
 - E0200B3
 - ▶ E0200C
 - ▶ E0300
 - ▼ E0500
 - ▶ E0500A
 - ▶ E0500B
 - ▶ E0500C
 - ▶ E0600
 - ▼ E0800
 - E08000
 - E08001
 - E08002
 - E08003
 - ▶ E0900
 - ▼ E1000
 - ▶ E1000A
 - ▶ E1000B
 - ▶ E1100
- ▼ F_Preferences_for_Customary_Routine_and_Activities
 - ▼ F0300
 - F03000
 - F03001
 - ▶ F0400
 - ▶ F0500
 - ▶ F0600
 - ▼ F0700
 - F07000
 - F07001
 - ▶ F0800
- ▶ G_Functional_Status

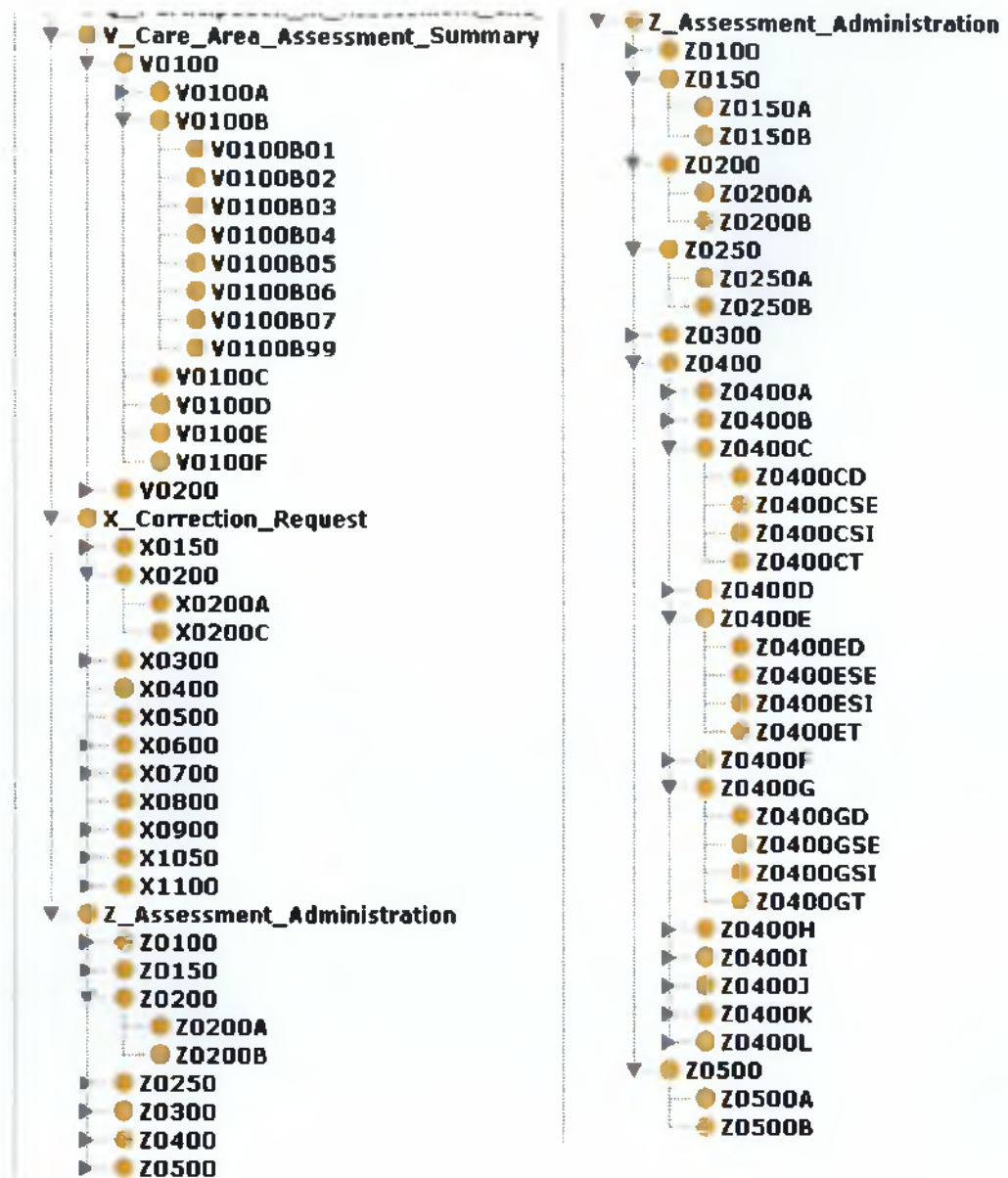
- ▼ G_Functional_Status
 - ▶ G0110
 - ▼ G0120
 - ▶ G0120A
 - G0120B
 - ▼ G0300
 - ▶ G0300A
 - ▶ G0300B
 - ▶ G0300C
 - ▶ G0300D
 - ▶ G0300E
 - ▶ G0400
 - ▼ G0600
 - G0600A
 - G0600B
 - G0600C
 - G0600D
 - G0600Z
 - ▶ G0900
- ▼ H_Bladder_and_Bowel
 - ▶ H0100
 - ▼ H0200
 - ▶ H0200A
 - H0200B0
 - H0200B1
 - H0200B2
 - H0200B9
 - ▶ H0200C
 - ▶ H0300
 - ▼ H0400
 - H04000
 - H04001
 - H04002
 - H04003
 - H04009
 - ▼ H0500
 - H05000
 - H05001
 - ▶ H0600

- I_Active_Diagnoses
- J_Health_Conditions
 - J0100
 - J0200
 - J02000
 - J02001
 - J0300
 - J0400
 - J0500
 - J0600
 - J0700
 - J07000
 - J07001
 - J0800
 - J0850
 - J1100
 - J1100A
 - J1100B
 - J1100C
 - J1100Z
 - J1300
 - J1400
 - J1550
 - J1700
 - J1800
 - J1900
- K_Swallowing/Nutritional_Status
 - K0100
 - K0200
 - K0300
 - K03000
 - K03001
 - K03002
 - K0310
 - K0510
 - K0510A
 - K0510B
 - K0510C
 - K0510D
 - K0510Z
 - K0700

- L_Oral/Dental_Status
 - L0200
 - L0200A
 - L0200B
 - L0200C
 - L0200D
 - L0200E
 - L0200F
 - L0200G
 - L0200Z
- M_Skin_Conditions
 - M0100
 - M0150
 - M0210
 - M02100
 - M02101
 - M0300
 - M0300A
 - M0300B
 - M0300C
 - M0300D
 - M0300E
 - M0300F
 - M0300G
 - M0610
 - M0610A
 - M0610B
 - M0610C
 - M0700
 - M0800
 - M0900
 - M0900A
 - M0900A0
 - M0900A1
 - M0900B
 - M0900C
 - M0900D
 - M1030
 - M1040
 - M1200

- ▼ N_Medications
 - N0300
 - N0350
 - N0410
- ▼ O_Special_Treatments_Procedures_And_Programs
 - ▼ O0100
 - O0250
 - ▶ O0250A
 - ▶ O0250B
 - ▼ O0250C
 - O0250C1
 - O0250C2
 - O0250C3
 - O0250C4
 - O0250C5
 - O0250C6
 - O0250C9
 - ▶ O0300
 - ▶ O0400
 - ▶ O0400A
 - ▼ O0400B
 - O0400B1
 - O0400B2
 - O0400B3
 - O0400B4
 - O0400B5
 - O0400B6
 - ▶ O0400C
 - ▼ O0400D
 - O0400D1
 - O0400D2
 - ▶ O0400E
 - ▼ O0400F
 - O0400F1
 - O0400F2
 - O0450
 - ▶ O0500
 - O0600
 - O0700

- ▼ P_Restraints
 - ▼ P0100
 - ▶ P0100A
 - ▼ P0100B
 - P0100B0
 - P0100B1
 - P0100B2
 - ▶ P0100C
 - ▼ P0100D
 - P0100D0
 - P0100D1
 - P0100D2
 - ▼ P0100E
 - P0100E0
 - P0100E1
 - P0100E2
 - ▶ P0100F
 - ▶ P0100G
 - P0100G0
 - P0100G1
 - P0100G2
 - ▶ P0100H
 - ▼ Q_Participation_In_Assessment_and_Goal_Setting
 - ▶ Q0100
 - ▼ Q0300
 - ▶ Q0300A
 - ▶ Q0300B
 - ▶ Q0400
 - ▶ Q0490
 - ▼ Q0500
 - ▼ Q0500B
 - Q0500B0
 - Q0500B1
 - Q0500B9
 - ▼ Q0550
 - ▶ Q0550A
 - ▶ Q0550B
 - ▼ Q0600
 - Q06000
 - Q06001
 - Q06002



Εικόνα 5.1 Όλες οι κλάσεις

Λαμβάνοντας υπόψη ότι ο αριθμός των υποκλάσεων είναι σημαντικά μεγάλος για το λόγο αυτό παρουσιάσουμε ενδεικτικά κάποιες υποκλάσεις από την Α κλάση και Β.

Κλάση Α: υπάρχουν οι υποκλάσεις

- Type of record (τύπος αρχείου)
- Facility provider numbers (αριθμός παρεχόμενων εγκαταστάσεων)

- Type of provider (τύπος παροχέα)
- Type of assessment (τύπος αξιολόγησης)

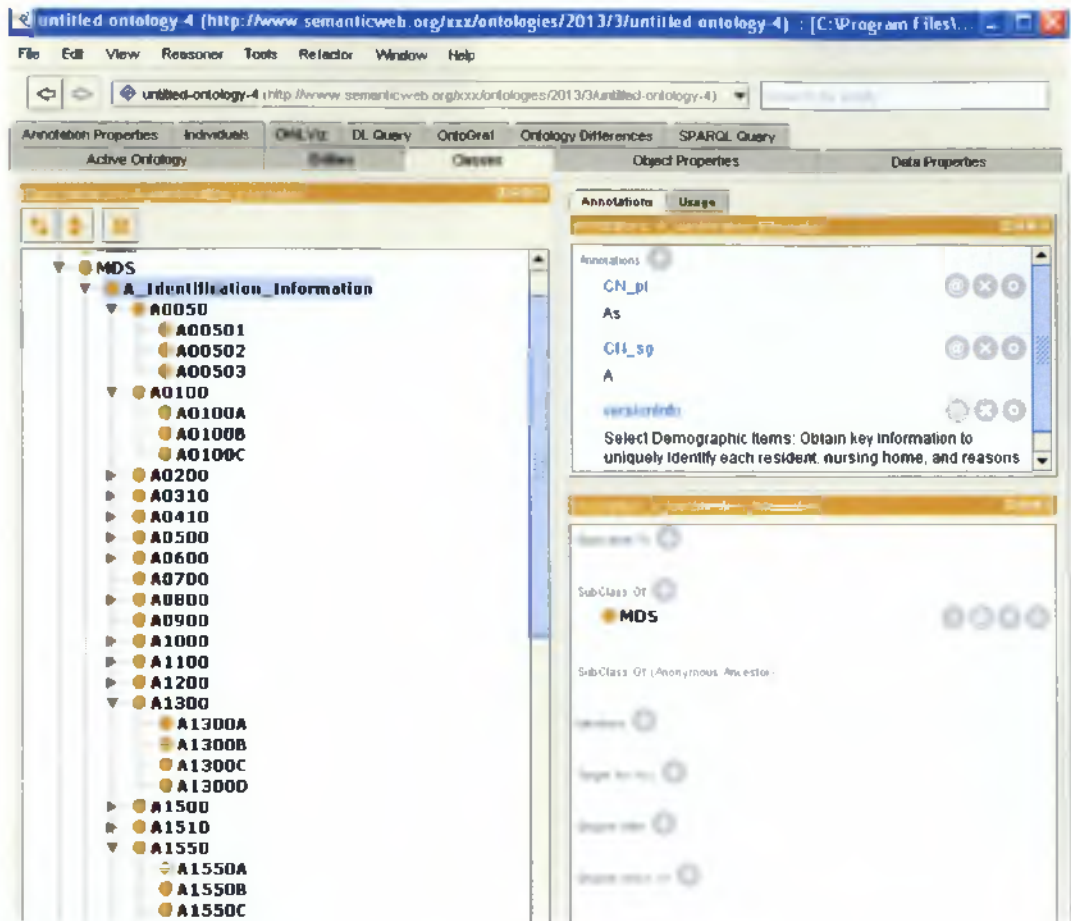
Κλάση B: υπάρχουν οι υποκλάσεις

- Comatose (κατάσταση κώματος)
- Hearing (ακοή)
- Hearing aid (ενίσχυση ακοής)
- Speech clarity (καθαρότητα ομιλίας)

Ενδεικτική ανάλυση κάποιων κλάσεων και υποκλάσεων

Κλάση A

Η κλάση αυτή κωδικοποιεί πληροφορίες του ερωτηματολογίου που προσδιορίζουν μοναδικά κάθε κάτοικο, το σπίτι που αυτός/η κατοικεί και ο λόγος για την αξιολόγηση.



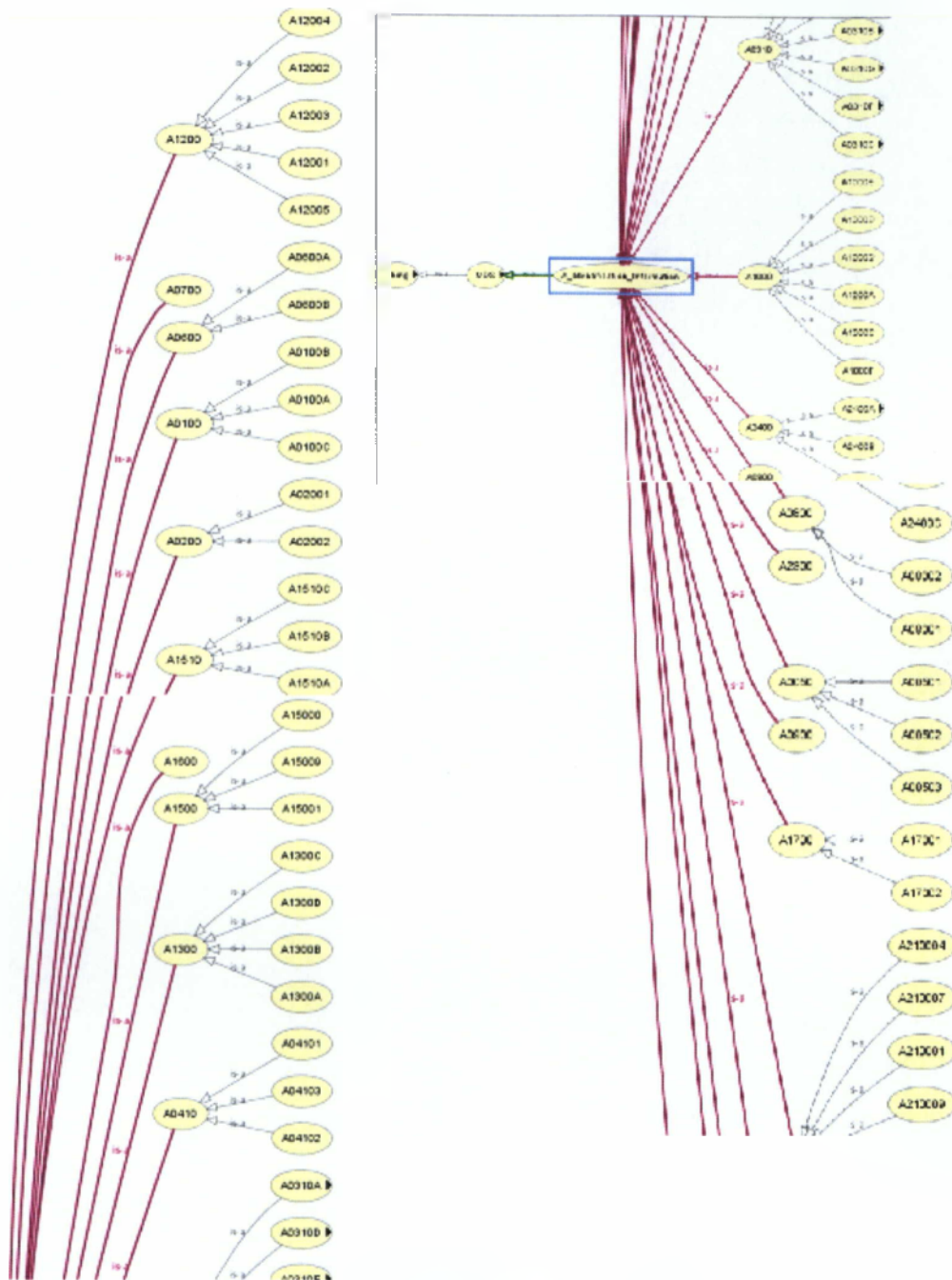
Εικόνα 5.2 Κλάση A και οι υποκλάσεις της

Η υποκλάση *type of record* της κλάσης A επιτρέπει αναγνώριση του τύπου φακέλου του κατοίκου και δίνει επιλογές όπως:

- Add new record (προσθήκη νέου φακέλου)
- Modify existing record (μετατροπή του ήδη υπάρχοντος φακέλου)
- Inactivate existing record (αδρανοποίηση του ήδη υπάρχοντος φακέλου)

Η υποκλάση *type of provider* επιτρέπει την αναγνώριση του είδους φορέα και δίνει επιλογές όπως

- Nursing home (γηροκομείο)
- Swing bed (νοσοκομειακό κρεβάτι)

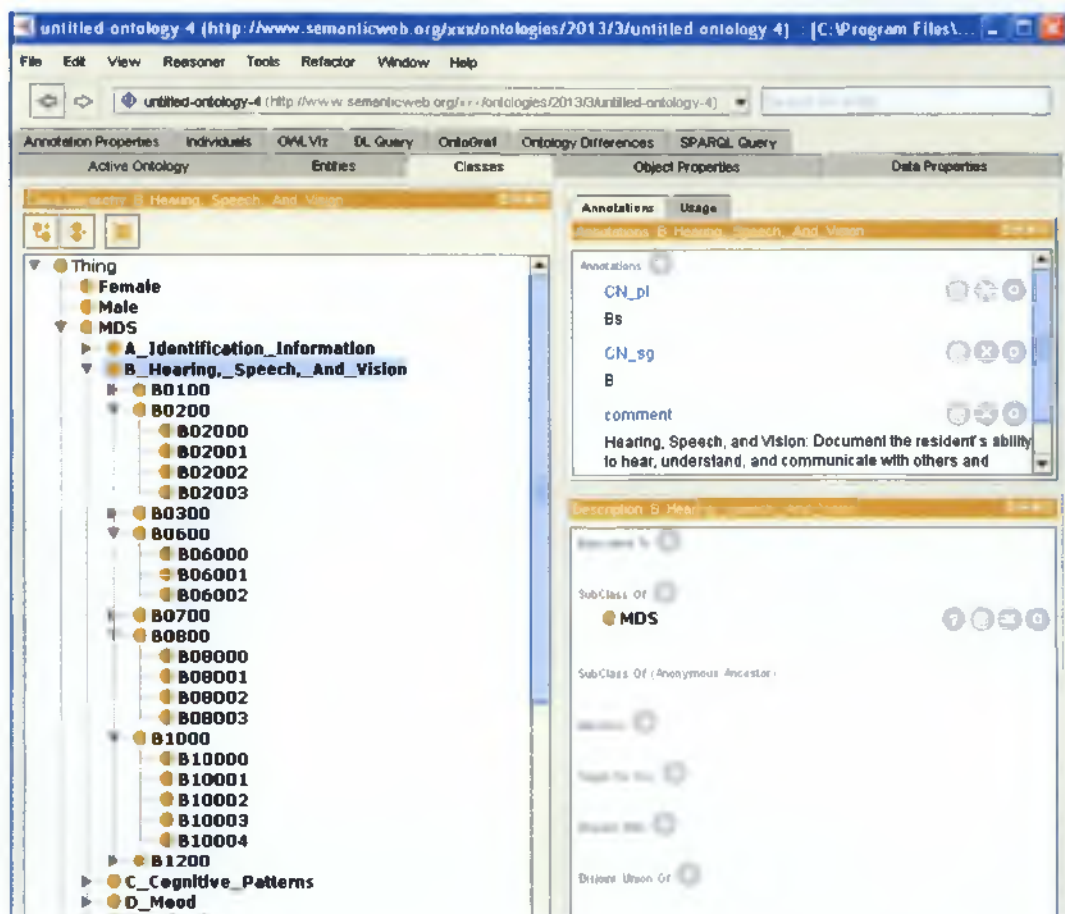


Εικόνα 5.3 Υποκλάσεις της Α

Κλάση Β

Η κλάση αυτή κωδικοποιεί πληροφορίες του ερωτηματολογίου που αφορούν την ικανότητα τους/της κατοίκου να ακούει (με βοηθητικές συσκευές ακοής, αν

χρησιμοποιούνται), να καταλαβαίνει και να επικοινωνεί με άλλους και αν ο κάτοικος βιώνει οπτικούς περιορισμούς ή δυσκολίες που σχετίζονται με ασθένειες, που είναι κοινές σε ηλικιωμένα άτομα.



Εικόνα 5.4 Κλάση B

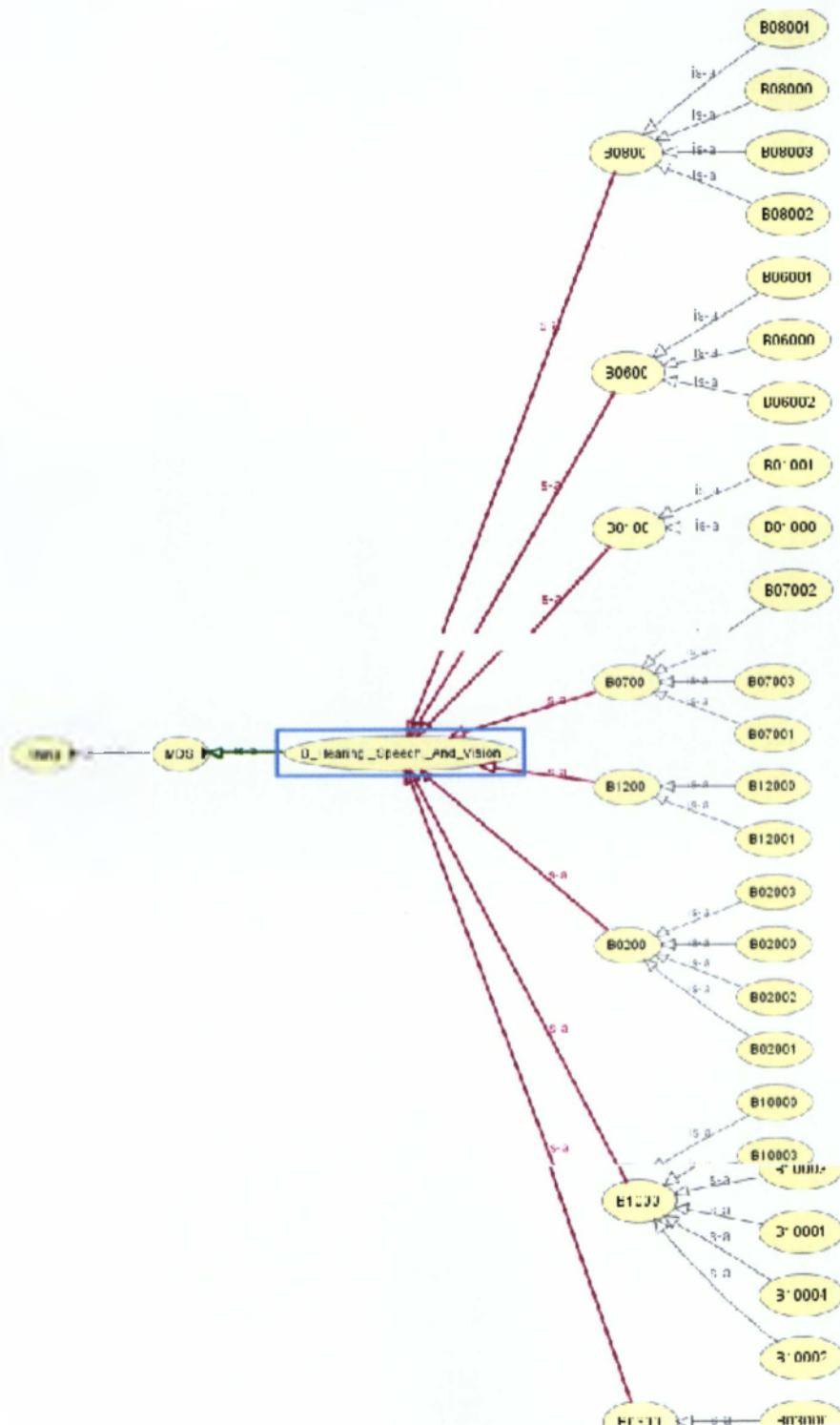
Η υποκλάση comatose αφορά το αν ο κάτοικος είναι σε κόμμα και δίνει επιλογές

- No (όχι)
- Yes (ναι)

Σχετικά με το αν ο κάτοικος έχει επιλογή κωματώδη κατάσταση/καμία ευδιάκριτη συνείδηση.

Η υποκλάση hearing αναφέρεται στην ικανότητα να ακούει (με ακουστικό ή συσκευές ενίσχυσης ακοής ανεπίσημα χρησιμοποιούμενες)

- Adequate (επαρκής)
- Minimal difficulty (ελάχιστη δυσκολία)
- Moderate difficulty (μέτρια δυσκολία)
- Highly impaired (εξαιρετικά μειωμένη)



Εικόνα 5.5 Υποκλάσεις της Β

5.3 ΚΩΔΙΚΑΣ ΟΝΤΟΛΟΓΙΑΣ

Η δημιουργία της MDS Ontology στηρίχθηκε στις οδηγίες χρήσης του λογισμικού, οι οποίες βρίσκονται σε αρχεία μορφής pdf, διαθέσιμα ελεύθερα στην ιστοσελίδα <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/MDS30RAIManual.html>

```
<?xml version="1.0"?>
<!DOCTYPE Ontology [
  <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
  <!ENTITY xml "http://www.w3.org/XML/1998/namespace" >
  <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.semanticweb.org/xxx/ontologies/2013/3/untitled-ontology-4"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  ontologyIRI="http://www.semanticweb.org/xxx/ontologies/2013/3/untitled-ontology-4">
```

```

    <Prefix name=""
IRI="http://www.semanticweb.org/xxx/ontologies/2013/3/untitled-ontology-4#"/>

    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>

    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>

    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>

    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>

    <Declaration>

    <Class IRI="#37"/>

    </Declaration>

    <Declaration>

        <Class IRI="#38"/>

    </Declaration>

    <Declaration>

    <!-- comment 1: declare of class A. Identification Information ..... -->

    <Class IRI="#A0050"/>

    </Declaration>

    <Declaration>

        <Class IRI="#A00501"/>

    </Declaration>

    <Declaration>

    <!-- comment 2: declare of class D. Hearing, Speech and Vision ..... -->

    <Class IRI="#D0100"/>

    </Declaration>

```

```
<Declaration>
    <Class IRI="#D01000"/>
</Declaration>
</Declaration>
<!-- comment 3: declare of class G. Functional Status ..... -->
    <Class IRI="#G0110"/>
</Declaration>
<Declaration>
    <Class IRI="#G0110A"/>
</Declaration>
<Declaration>
</Declaration>
<!-- comment 4: declare of class J. Active Diagnoses ..... -->
    <Class IRI="#J0100"/>
</Declaration>
<Declaration>
    <Class IRI="#J0100A"/>
</Declaration>
<Declaration>
</Declaration>
<!-- comment 5: declare of class M. Skin Conditions ..... -->
    <Class IRI="#M0300F2"/>
</Declaration>
<Declaration>
    <Class IRI="#M0300G"/>
```

```
</Declaration>

<Declaration>

<! -- comment 6: declare of class P. Restraints ..... -->

<Class IRI="#P0100"/>

</Declaration>

<Declaration>

    <Class IRI="#P0100A"/>

</Declaration>

<Declaration>

<! -- comment 7: declare of class V. Care area assessment summary ..... -->

<Class IRI="#V0100"/>

</Declaration>

<Declaration>

    <Class IRI="#V0100A"/>

</Declaration>

<Declaration>

<! -- comment 8: declare of class Z. Assessment administration ..... -->

<Class IRI="#Z0100"/>

</Declaration>

<Declaration>

    <Class IRI="#Z0100A"/>

</Declaration>

<Declaration>
```

(.....)

<!-- comment 9: -->

<AnnotationProperty IRI="http://attempto.ifi.uzh.ch/ace_lexicon#TV_vbg"/>

<IRI>#is_in_a_Comatose_state</IRI>

<Literal datatypeIRI="&rdf;PlainLiteral">is_in_a_Comatose_stated</Literal>

</AnnotationAssertion>

</Ontology>

<!-- Generated by the OWL API (version 3.4.2) <http://owlapi.sourceforge.net> -->⁷⁰

⁷⁰ Ο κώδικας είναι ιδιαίτερα μεγάλος και σε έντυπη έκδοση word εκτείνεται στις 681 σελίδες. Για το λόγο αυτό παρουσιάζεται κομμάτι αυτού.

6 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στη σημερινή εποχή, την «εποχή της πληροφορίας», οι πληροφορίες και κατ' επέκταση οι γνώσεις διαδίδονται με πολύ γρήγορους ρυθμούς. Αυτό κάνει επιβεβλημένη τη χρήση συστημάτων διαχείρισης γνώσης από κάθε οργανισμό, προκειμένου να μπορεί να γίνεται σωστή και αποδοτική διαχείριση του μεγάλου όγκου πληροφοριών, που καθημερινά εισέρχονται σε αυτόν. Οι οντολογίες μπορούν να βοηθήσουν στην επίλυση αυτού του προβλήματος.

Για το λόγο αυτό, η διεθνή αγορά δείχνει μεγάλο ενδιαφέρον για τις οντολογίες και πραγματοποιούνται πολλές έρευνες, που κατευθύνονται στην ανάπτυξη του Σημαιολογικού Ιστού, προκειμένου να χρησιμοποιηθεί στη διαχείριση γνώσης μέσω οντολογιών.

Οι οντολογίες έχοντας την δυνατότητα της προσεκτικής επιλογής κάποιων όρων, ιδίως όταν υπάρχουν μεταξύ τους μικρές διαφορές, δημιουργούν ένα σύνολο σχέσεων που καθορίζονται από διάφορες τεχνικές και οι οποίες αποτελούν τη σημασιολογική βάση για την ορολογία που επιλέχτηκε. Υπάρχουν πολλά εργαλεία, που μπορούν να χρησιμοποιηθούν για την δημιουργία οντολογιών. Ένα από αυτά είναι το πρόγραμμα ροιτέγέ, η χρήση του οποίου είναι αρκετά εύκολη και απλή και δεν απαιτεί ιδιαίτερες γνώσεις σε ανάπτυξη γλώσσας OWL.

Η οντολογία που δημιουργήθηκε στο πλαίσιο της παρούσας εργασίας αποτελεί μια προσπάθεια για την δημιουργία οντολογιών, οι οποίες θα είναι σχετικές με το χώρο της υγείας.

7 ΒΙΒΛΙΟΓΡΑΦΙΑ

Ahmed, S. R., (2004). Application of data mining in Retail Business. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04). IEEE Computer Society Washington, DC, USA 2044.

Bao, H. T. Introduction to knowledge discovery and data mining, course, Institute of Information Technology National Center for Natural Science and Technology. <http://www.scribd.com/doc/71900807/Ho-Tu-Bao-All-Chapters-2> [Web Accessed October 2012].

Bodenreider O., (2006) Visualization tools for the Unified Medical Language System (Sem-Nav), the Gene Ontology (GenNav), and RxNorm. In *Humans and the Semantic Web* HCIL Workshop.

Borst W. N. and Akkermans J. M. (1997): Engineering Ontologies, *International Journal of Human-Computer Studies*, 46 (2/3):365-406.

Delen, D., (2009). Analysis of cancer data: a data mining approach, Expert Systems, February, Vol. 26, No. 1 <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0394.2008.00480.x/full> [Web Accessed October 2012].

Duda-Gardiner, S., (2010), Resident Assessment Protocol, Care Area Assessment, Long – term Living, July 31, <http://www.ltlmagazine.com/article/resident-assessment-protocol-care-area-assessment>. [Web Accessed May 2013].

Fensel D. and Groenboom R. (1997): Specifying Knowledgebased Systems with Reusable Components. In *Proceedings 9th International Conference on Software Engineering and Knowledge Engineering (SEKE '97)*, Madrid.

Fensel D., (2000): Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, URL: <http://software.ucv.ro/~cbadica/didactic/sbc/documente/silverbullet.pdf>

Fensel D., (2004) “Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce”, Second Edition, Springer – Verlag, Berlin, Heidelberg.

Fridman-Noy N. and Hafner C.D. (1997): The State of the Art in Ontology Design, *AI Magazine*, 18(3):53-74.

Frawley, W. J., Piatetsky-Shapiro G., and Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview, *AI Magazine* Volume 13 Number 3. <http://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&cad=rja&ved=0CFYQFjAG&url=http%3A%2F%2Faaai.org%2Fpubs%2Fmagazines%2Farticle%2Fdownload%2F1011%2F929&ei=cp1vULiHemL4gTHvICAAG&usq=AFOjCNF5uK-7W0dvzZyWP3Msv0Xtwf9SdA> [Web Accessed October 2012].

Gerber, A., Van der Merwe, A., & Barnard, A. (2008). A Functional Semantic Web Architecture. European Semantic Web Conference 2008, ESWC'08.

Gruber, T. R., (1993) “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220.

Guarino N., (1998) “Formal Ontology in Information Systems”, In Proceedings of FOIS'98, Trento, Italy, IOS Press, Amsterdam.

Huhns N. & Singh M. (1997): Ontologies for Agents, *IEEE internet Computing*, Nov - Dec, <http://computer.org/internet>

Lavrac, N., (1999). Data mining in medicine: Selected techniques and Applications <http://sce.umkc.edu/~leeyu/MEDSS/References/dss-4.pdf> [Web Accessed October 2012].

MacLennan, J., Zhao, H. T., (2005). Data Mining with SQL Server 2005. s.l.: Wiley Publishing, Inc.

http://its.lnpu.edu.ua/edocs1/new_doc/en/Maclennan%20J..%20Tang%20Z.%20-%20Data%20Mining%20With%20SQL%20Server%202005%20%282005%29%28en%29.pdf [Web Accessed October 2012].

Musen, M. A. (1992) “Dimensions of knowledge sharing and reuse”. Computers and Biomedical Research, 25, 435-467.

Pirlein Th. and Studer R. (1999): Integrating the Reuse of Commonsense Ontologies and Problem-Solving Methods, *International Journal of Expert Systems: Research and Applications*.

Pisanelli D., Gangemi A., Steve G., (1997) “WWW-Available Conceptual Integration of Medical Terminologies: the ONIONS Experience”, In *Proceedings of AMIA 97 Conference*.

Pisanelli D., Gangemi A., Steve G., (1998) “An Ontological Analysis of the UMLS Metathesaurus”, In *Proceedings of AMIA 98 Conference*.

Rector A., Solomon W., Nowlan W., Rush T., (1995) “A Terminology Server for Medical Language and Medical Information Systems”. *Methods of Information in Medicine*, Vol. 34, pages 147-157.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R B., Richardson, W. S., (2004). Evidence based medicine: what it is and what it isn't. *BMJ*, 312 (7023), 71-2.

Safavian, S.R., and Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Trans. Systems, Man & Cybernetics*. 21 (3), pp.: 660-674.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2349778/pdf/bmj00524-0009.pdf>
[Web Accessed October 2012].

Savova, G. K., Ogren, P. V., Duffy, P. H., Buntrock, J. D., Chute, C. G., (2008). Mayo clinic NLP system for patient smoking status identification. *J Am Med Inform*

Assoc, 15(1), 25-8. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2274870/> [Web Accessed October 2012].

Spackman K., Campbell K., et al. (1997) SNOMED CT: a reference terminology for health care. In *Proceedings of the AMIA annual fall symposium*, page 640. American Informatics Association.

Studer R., Eriksson H., Gennari J. H., Tu S. W., Fensel D., and Musen M. (1996): Ontologies and the Configuration of Problem-Solving Methods. In B. R. Gaines and M. A. Musen, (eds.), *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada.

Sure Y., Angele J., Staab S., (2002) OntoEdit: Guiding Ontology Development by Methodology and Inferencing, On the Move to Meaningful Internet Systems, 2002 - DOA/CoopIS/ODBASE 2002 Confederated International Conferences DOA, CoopIS and ODBASE 2002, p.1205-1222, October 30-November 01.

UMLS Knowledge Sources, National Library of Medicine, (1999) edition, available from the NLM, Bethesda, Maryland.

Uschold M. and Gruninger M. (1996), "Ontologies: principles, methods, and applications," *Knowledge Engineering Review*, Vol. 11:2, 93-136. Also available as AIAI-TR-191 from AIAI, The University of Edinburgh.

Weibel S., Gridby J., and Miller E. (1995): OCLC/NCSA Metadata Workshop Report, Dublin, EUA.

Αριστογιάννης, Γ., (2010). SEMANTIC WEB, Σημειώσεις για το μάθημα «Διαχείριση Πληροφορίας & Τεχνικές Κατηγοριοποίησης στο Διαδίκτυο»

Γούλου Ζ., (2012). Εφαρμογή μεθόδων εξόρυξης δεδομένων στη διαχείριση πελατειακών σχέσεων.

Δάρλας, Γ. Από τις Βάσεις Ορολογίας στις Βάσεις Γνώσης: Το Παράδειγμα του Συστήματος Διαχείρισης Ορολογίας της Εγνατία Οδός Α.Ε. (Οδο-λέξεις)

Ζαπουνίδου, Σ., (2004). Θεωρητική προσέγγιση του Σημασιολογικού Ιστού στο χώρο της πολιτισμικής πληροφορίας: μία πρότυπη εφαρμογή στη βιβλιοθηκονομία.

Επιχειρησιακό Πρόγραμμα Κοινωνία της Πληροφορίας, (2007), ΙΑΤΡΟΛΕΞΗ, Ινστιτούτο Τεχνολογίας Υπολογιστών – Neurosoft, Αθήνα.

Κανελλόπουλος, Δ., (2011). Τα οφέλη του Σημασιολογικού Ιστού στο e-Επιχειρείν, Σύγχρονη Τεχνική Επιθεώρηση.

Κεραύνου, Ε., (2000). Τεχνητή Νοημοσύνη και Έμπειρα Συστήματα.

Κοκκινίδης, Γ., Κοφινά, Ι., Παπαγγελής, Μ. Knowledge Management & Semantic Web.

Παπαταξιάρχης, Β., (2008). Τεχνολογίες Αναπαράστασης Γνώσης και Συμπερασμού: Η Περίπτωση ενός Έξυπνου Βοηθού Προπονητή Ποδοσφαίρου.

Σαμψών Δ., (2003). Πανεπιστημιακές Σημειώσεις, Η Γλώσσα Σήμανσης XML, Πανεπιστήμιο Πειραιώς, Τμήμα Διδακτικής της Τεχνολογίας και Ψηφιακών Συστημάτων.

- African Traditional Medicine Ontology (ATMO), διαθέσιμο στο <http://bioportal.bioontology.org/ontologies/1099>
- Anatomical Entity Ontology (AEO), διαθέσιμο στο <http://bioportal.bioontology.org/ontologies/47581?p=terms>
- Bone Dysplasia Ontology (BDO), διαθέσιμο στο <http://bioportal.bioontology.org/ontologies/46875?p=terms>

- Cell Behavior Ontology (CBO), διαθέσιμο στο <http://bioportal.bioontology.org/ontologies/1158>
- International Classification of Diseases (ICD), διαθέσιμο στο <http://www.who.int/classifications/icd/en/>
- Neurobehavior Ontology (NBO), διαθέσιμο στο <http://bioportal.bioontology.org/ontologies/1621>
- ΙΑΤΡΟΛΕΞΗ, διαθέσιμο στο <http://www.iatrolexi.gr>
- Οδηγίες χρήσης δημιουργίας και κωδικοποίησης MDS Ontology, διαθέσιμο στο <http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/MDS30RA1Manual.html>

ΠΑΡΑΡΤΗΜΑΤΑ

ΠΑΡΑΡΤΗΜΑ Α: IDENTIFICATION INFORMATION

SECTION A: IDENTIFICATION INFORMATION

Intent: The intent of this section is to obtain key information to uniquely identify each resident, the home in which he or she resides, and the reasons for assessment.

A0050: Type of Record

A0050. Type of Record	
Enter Code <input type="checkbox"/>	<ol style="list-style-type: none"> 1. Add new record → Continue to A0100, Facility Provider Numbers 2. Modify existing record → Continue to A0100, Facility Provider Numbers 3. Inactivate existing record → Skip to X0150, Type of Provider

Coding Instructions for A0050, Type of Record

- **Code 1, Add new record:** if this is a **new record** that has not been previously submitted and accepted in the QIES ASAP system. If this item is **coded as 1**, continue to A0100 Facility Provider Numbers.

If there is an existing database record for the same resident, the same facility, the same reasons for assessment/tracking, and the same date (assessment reference date, entry date, or discharge date), then the current record is a duplicate and not a new record. In this case, the submitted record will be rejected and not accepted in the QIES ASAP system and a “fatal” error will be reported to the facility on the Final Validation Report.
- **Code 2, Modify existing record:** if this is a **request to modify** the MDS items for a record that already has been submitted and accepted in the QIES ASAP system.

If this item is **coded as 2**, continue to A0100, Facility Provider Numbers.

When a modification request is submitted, the QIES ASAP System will take the following steps:

 1. The system will attempt to locate the existing record in the QIES ASAP database for this facility with the resident, reasons for assessment/tracking, and date (assessment reference date, entry date, or discharge date) indicated in subsequent Section X items.
 2. If the existing record is not found, the submitted modification record will be rejected and not accepted in the QIES ASAP system. A “fatal” error will be reported to the facility on the Final Validation Report.
 3. If the existing record is found, then the items in all sections of the submitted modification record will be edited. If there are any fatal errors, the modification record will be rejected and not accepted in the QIES ASAP system. The “fatal” error(s) will be reported to the facility on the Final Validation Report.
 4. If the modification record passes all the edits, it will replace the prior record being modified in the QIES ASAP database. The prior record will be moved to a history file in the QIES ASAP database.

A0050: Type of Record (cont.)

- **Code 3, Inactivate existing record:** if this is a request to inactivate a record that already has been submitted and accepted in the QIES ASAP system.

If this item is coded as 3, skip to X0150, Type of Provider.

When an inactivation request is submitted, the QIES ASAP system will take the following steps:

1. The system will attempt to locate the existing record in the QIES ASAP system for this facility with the resident, reasons for assessment/tracking, and date (assessment reference date, entry date, or discharge date) indicated in subsequent Section X items.
2. If the existing record is not found in the QIES ASAP database, the submitted inactivation request will be rejected and a "fatal" error will be reported to the facility on the Final Validation Report.
3. All items in Section X of the submitted record will be edited. If there are any fatal errors, the current inactivation request will be rejected and no record will be inactivated in the QIES ASAP system.
4. If the existing record is found, it will be removed from the active records in the QIES ASAP database and moved to a history file.

Identification of Record to be Modified/Inactivated

The Section X items from X0200 through X0700 identify the existing QIES ASAP database assessment or tracking record that is in error. In this section, reproduce the information **EXACTLY** as it appeared on the existing erroneous record, even if the information is incorrect. This information is necessary to locate the existing record in the database.

Example: A MDS assessment for Joan L. Smith is submitted and accepted by the QIES ASAP system. A data entry error is then identified on the previously submitted and accepted record. When the encoder "data entered" the prior assessment for Joan L. Smith, he typed "John" by mistake. To correct this data entry error, the facility will modify the erroneous record and complete the items in Section X including items under Identification of Record to be Modified/Inactivated. When completing X0200A, the Resident First Name, "John" will be entered in this item. This will permit the MDS system to locate the previously submitted assessment that is being corrected. If the correct name "Joan" were entered, the QIES ASAP system would not locate the prior assessment.

The correction to the name from "John" to "Joan" will be made by recording "Joan" in the "normal" A0500A, Resident First Name in the modification record. The modification record must include all items appropriate for that assessment, not just the corrected name. This modification record will then be submitted and accepted into the QIES ASAP system which causes the desired correction to be made.

A0100: Facility Provider Numbers

A0100. Facility Provider Numbers	
A. National Provider Identifier (NPI):	<input type="text"/>
B. CMS Certification Number (CCN):	<input type="text"/>
C. State Provider Number:	<input type="text"/>

Item Rationale

- Allows the identification of the nursing home submitting assessment.

Coding Instructions

- Nursing homes must have a National Provider Number (NPI) and a CMS Certified Number (CCN).
- Enter the nursing home provider numbers:
 - National Provider Identifier (NPI)
 - CMS Certified Number (CCN)
 - State Provider Number (optional)

DEFINITIONS

NATIONAL PROVIDER IDENTIFIER (NPI) A unique Federal number that identifies providers of health care services. The NPI applies to the long-term care hospital and all of its patients.

CMS CERTIFICATION NUMBER (CCN) Replaces the term "Medicare/Medicaid Provider Number" in survey, certification, and assessment-related activities.

STATE PROVIDER NUMBER Medicaid Provider Number established by a state.

A0200: Type of Provider

A0200. Type of Provider	
Enter Code	Type of provider
<input type="checkbox"/>	1. Nursing home (SNF/NF)
	2. Swing Bed

Item Rationale

- Allows designation of type of provider.

Coding Instructions

- Code 1, nursing home (SNF/NF):** if a Medicare skilled nursing facility (SNF) or Medicaid nursing facility (NF).
- Code 2, swing bed:** if a hospital with swing bed approval.

DEFINITION

SWING BED

A rural hospital with less than 100 beds that participates in the Medicare program that has CMS approval to provide post-hospital SNF care. The hospital may use its beds, as needed, to provide either acute or SNF care.

A0310: Type of Assessment

A0310. Type of Assessment	
Enter Code <input type="text"/>	A. Federal OBRA Reason for Assessment 01. Admission assessment (required by day 14) 02. Quarterly review assessment 03. Annual assessment 04. Significant change in status assessment 05. Significant correction to prior comprehensive assessment 06. Significant correction to prior quarterly assessment 99. None of the above
Enter Code <input type="text"/>	B. PPS Assessment PPS Scheduled Assessments for a Medicare Part A Stay 01. 5-day scheduled assessment 02. 14-day scheduled assessment 03. 30-day scheduled assessment 04. 60-day scheduled assessment 05. 90-day scheduled assessment 06. Readmission/return assessment PPS Unscheduled Assessments for a Medicare Part A Stay 07. Unscheduled assessment used for PPS (OMRA, significant or clinical change, or significant correction assessment) Not PPS Assessment 99. None of the above
Enter Code <input type="text"/>	C. PPS Other Medicare Required Assessment - OMRA 0. No 1. Start of therapy assessment 2. End of therapy assessment 3. Both Start and End of therapy assessment 4. Change of therapy assessment
Enter Code <input type="text"/>	D. Is this a Swing Bed clinical change assessment? Complete only if A0200 = 2 0. No 1. Yes
Enter Code <input type="text"/>	E. Is this assessment the first assessment (OBRA, Scheduled PPS, or Discharge) since the most recent admission/entry or reentry? 0. No 1. Yes
Enter Code <input type="text"/>	F. Entry/discharge reporting 01. Entry tracking record 10. Discharge assessment-return not anticipated 11. Discharge assessment-return anticipated 12. Death in facility tracking record 99. None of the above
Enter Code <input type="text"/>	G. Type of discharge - Complete only if A0310F = 10 or 11 1. Planned 2. Unplanned

For Comprehensive, Quarterly, and PPS Assessments, Entry and Discharge Records.

Item Rationale

- Allows identification of needed assessment content.

Coding Instructions for A0310, Type of Assessment

Enter the code corresponding to the reason or reasons for completing this assessment.

If the assessment is being completed for both Omnibus Budget Reconciliation Act (OBRA)–required clinical reasons (A0310A) and Prospective Payment System (PPS) reasons (A0310B and A0310C) all requirements for both types of assessments must be met. See Chapter 2 on assessment schedules for details of these requirements.

A0310: Type of Assessment (cont.)

Coding Instructions for A0310A, Federal OBRA Reason for Assessment

- Document the reason for completing the assessment, using the categories of assessment types. For detailed information on the requirements for scheduling and timing of the assessments, see Chapter 2 on assessment schedules.
- Enter the number corresponding to the OBRA reason for assessment. This item contains 2 digits. For codes 01-06, enter "0" in the first box and place the correct number in the second box. If the assessment is not coded 01-06, enter code "99".
 - 01.** Admission assessment (required by day 14)
 - 02.** Quarterly review assessment
 - 03.** Annual assessment
 - 04.** Significant change in status assessment
 - 05.** Significant correction to prior comprehensive assessment
 - 06.** Significant correction to prior quarterly assessment
 - 99.** None of the above

Coding Tips and Special Populations

- If a nursing home resident elects the hospice benefit, the nursing home is required to complete an MDS significant change in status assessment. The nursing home is required to complete a SCSA when they come off the hospice benefit (revoke). See Chapter 2 for details on this requirement.
- It is a CMS requirement to have a significant change in status assessment completed EVERY time the hospice benefit has been elected, even if a recent MDS was done and the only change is the election of the hospice benefit.

Coding Instructions for A0310B, PPS Assessment

- Enter the number corresponding to the PPS reason for completing this assessment. This item contains 2 digits. For codes 01-07, enter "0" in the first box and place the correct number in the second box. If the assessment is not coded as 01-07, enter code "99".
- See Chapter 2 on assessment schedules for detailed information on the scheduling and timing of the assessments.

PPS Scheduled Assessments for a Medicare Part A Stay

- 01.** 5-day scheduled assessment
- 02.** 14-day scheduled assessment
- 03.** 30-day scheduled assessment
- 04.** 60-day scheduled assessment
- 05.** 90-day scheduled assessment
- 06.** Readmission/return assessment

DEFINITION

PROSPECTIVE PAYMENT SYSTEM (PPS)

Method of reimbursement in which Medicare payment is made based on the classification system of that service (e.g., resource utilization groups, RUGs, for skilled nursing facilities).

A0310: Type of Assessment (cont.)

PPS Unscheduled Assessments for Medicare Part A Stay

- 07.** Unscheduled assessment used for PPS (OMRA, significant change, or significant correction assessment)
- 99.** None of the above

Coding Instructions for A0310C, PPS Other Medicare Required Assessment—OMRA

- **Code 0, no:** if this assessment is not an OMRA.
- **Code 1, Start of therapy assessment (OPTIONAL):** with an assessment reference date (ARD) that is 5 to 7 days after the first day therapy services are provided (except when the assessment is used as a short stay assessment, see Chapter 6). No need to combine with the 5-day assessment except for short stay. Only complete if therapy RUG (index maximized), otherwise the assessment will be rejected.
- **Code 2, End of therapy assessment:** with an ARD that is 1 to 3 days after the last day therapy services were provided.
- **Code 3, both the Start and End of therapy assessment:** with an ARD that is both 5 to 7 days after the first day therapy services were provided and that is 1 to 3 days after the last day therapy services were provided (except when the assessment is used as a short stay assessment, see Chapter 6).
- **Code 4, Change of therapy assessment:** with an ARD that is Day 7 of the COT observation period.

Coding Instructions for A0310D, Is This a Swing Bed Clinical Change Assessment?

- **Code 0, no:** if this assessment is not a swing bed clinical change assessment.
- **Code 1, yes:** if this assessment is a swing bed clinical change assessment.

Coding Instructions for A0310E, Is This Assessment the First Assessment (OBRA, PPS, or Discharge) since the Most Recent Admission/Entry or Reentry?

- **Code 0, no:** if this assessment is not the first assessment since the most recent admission/entry or reentry.
- **Code 1, yes:** if this assessment is the first assessment since the most recent admission/entry or reentry.

Coding Tips and Special Populations

- A0310E = 0 for any tracking record (entry or death in facility) because tracking records are not considered assessments.

A0310: Type of Assessment (cont.)

Coding Instructions for A0310F, Federal OBRA & PPS Entry/Discharge Reporting

- Enter the number corresponding to the reason for completing this assessment or tracking record. This item contains 2 digits. For code 01, enter “0” in the first box and place “1” in the second box. If the assessment is not coded as “01” or “10” or “11” or “12,” enter “99”:
 - 01.** Entry tracking record
 - 10.** Discharge assessment-return not anticipated
 - 11.** Discharge assessment-return anticipated
 - 12.** Death in facility tracking record
 - 99.** None of the above

Coding Instructions for A0310G, Type of Discharge

- **Code 1:** if type of discharge is a planned discharge.
- **Code 2:** if type of discharge is an unplanned discharge.

A0410: Submission Requirement

A0410. Submission Requirement	
Enter Code <input type="checkbox"/>	1. Neither federal nor state required submission 2. State but not federal required submission (FOR NURSING HOMES ONLY) 3. Federal required submission

Item Rationale

- There must be a federal and/or state authority to submit MDS assessment data to the MDS National Repository.
- Nursing homes must be certain they are submitting MDS assessments under the appropriate authority. With this item, the nursing home indicates the submission authority.

Steps for Assessment

1. Ask the nursing home administrator or representative which units in the nursing home are Medicare certified, if any, and which units are Medicaid certified, if any.
2. Identify all units in the nursing home that are not certified, if any.
 - If some or all of the units in the nursing home are neither Medicare nor Medicaid certified, ask the nursing home administrator or representative whether the State has authority to collect MDS information for residents on units that are neither Medicare nor Medicaid certified.

A0410: Submission Requirement (cont.)

Coding Instructions

- **Code 1, neither federal nor state required submission:** if the MDS record is for a resident on a unit that is neither Medicare nor Medicaid certified, and the state does not have authority to collect MDS information for residents on this unit. If the record is submitted, it will be rejected and all information from that record will be purged.
- **Code 2, State but not federal required submission:** if the MDS record is for a resident on a unit that is neither Medicare nor Medicaid certified, but the state has authority, under state licensure or other requirements, to collect MDS information for these residents.
- **Code 3, Federal required submission:** if the MDS record is for a resident on a Medicare and/or Medicaid certified unit. There is CMS authority to collect MDS information for residents on this unit.

A0500: Legal Name of Resident

A0500. Legal Name of Resident	
A. First name:	B. Middle initial:
<input type="text"/>	<input type="text"/>
C. Last name:	D. Suffix:
<input type="text"/>	<input type="text"/>

Item Rationale

- Allows identification of resident
- Also used for matching each of the resident's records

Steps for Assessment

1. Ask resident, family, significant other, guardian, or legally authorized representative.
2. Check the resident's name on his or her Medicare card, or if not in the program, check a Medicaid card or other government-issued document.

DEFINITION

LEGAL NAME

Patient's name as it appears on the Medicare card. If the patient is not enrolled in the Medicare program, use the patient's name as it appears on a Medicaid card or other government-issued document.

Coding Instructions

Use printed letters. Enter in the following order:

- A. First Name
- B. Middle Initial (if the resident has no middle initial, leave Item A0500B blank; if the resident has two or more middle names, use the initial of the first middle name)
- C. Last Name
- D. Suffix (e.g., Jr./Sr.)

A0600: Social Security and Medicare Numbers

A0600. Social Security and Medicare Numbers							
	<p>A. Social Security Number:</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td style="width: 25%;">□□□□</td> <td style="width: 5%;">-</td> <td style="width: 10%;">□□</td> <td style="width: 5%;">-</td> <td style="width: 55%;">□□□□□□</td> </tr> </table> <p>B. Medicare number (or comparable railroad insurance number):</p> <table border="1" style="width: 100%; text-align: center;"> <tr> <td>□□□□□□□□□□□□□□</td> </tr> </table>	□□□□	-	□□	-	□□□□□□	□□□□□□□□□□□□□□
□□□□	-	□□	-	□□□□□□			
□□□□□□□□□□□□□□							

Item Rationale

- Allows identification of the resident.
- Allows records for resident to be matched in system.

Coding Instructions

- Enter the Social Security Number (SSN) in A0600A, one number per space starting with the leftmost space. If no social security number is available for the resident (e.g., if the resident is a recent immigrant or a child) the item may be left blank.
- Enter Medicare number in A0600B exactly as it appears on the resident's documents.
- If the resident does not have a Medicare number, a Railroad Retirement Board (RRB) number may be substituted. These RRB numbers contain both letters and numbers. To enter the RRB number, enter the first letter of the code in the leftmost space followed by one letter/digit per space. If no Medicare number or RRB number is known or available, the item may be left blank.
- For PPS assessments (A0310B = 01, 02, 03, 04, 05, 06, and 07), either the SSN (A0600A) or Medicare number/RRB number (A0600B) must be present and both may not be blank.
- A0600B can only be a Medicare (HIC) number or a Railroad Retirement Board number.

DEFINITIONS

SOCIAL SECURITY NUMBER

A tracking number assigned to an individual by the U.S. Federal government for taxation, benefits, and identification purposes.

MEDICARE NUMBER (OR COMPARABLE RAILROAD INSURANCE NUMBER)

An identifier assigned to an individual for participation in national health insurance program. The Medicare Health Insurance identifier may be different from the resident's social security number (SSN), and may contain both letters and numbers. For example, many residents may receive Medicare benefits based on a spouse's Medicare eligibility.

A0700: Medicaid Number

A0700. Medicaid Number - Enter "+" if pending, "N" if not a Medicaid recipient		
	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>□□□□□□□□□□□□□□</td> </tr> </table>	□□□□□□□□□□□□□□
□□□□□□□□□□□□□□		

Item Rationale

- Assists in correct resident identification.

A0700: Medicaid Number (cont.)

Coding Instructions

- Record this number if the resident is a Medicaid recipient.
- Enter one number per box beginning in the leftmost box.
- Recheck the number to make sure you have entered the digits correctly.
- Enter a “+” in the leftmost box if the number is pending. If you are notified later that the resident does have a Medicaid number, just include it on the next assessment.
- If not applicable because the resident is not a Medicaid recipient, enter “N” in the leftmost box.

Coding Tips and Special Populations

- To obtain the Medicaid number, check the resident’s Medicaid card, admission or transfer records, or medical record.
- Confirm that the resident’s name on the MDS matches the resident’s name on the Medicaid card.
- It is not necessary to process an MDS correction to add the Medicaid number on a prior assessment. However, a correction may be a State-specific requirement.

A0800: Gender

A0800. Gender	
Enter Code <input type="checkbox"/>	1. Male 2. Female

Item Rationale

- Assists in correct identification.
- Provides demographic gender specific health trend information.

Coding Instructions

- **Code 1:** if resident is male.
- **Code 2:** if resident is female.

Coding Tips and Special Populations

- Resident gender on the MDS should match what is in the Social Security system.

A0900: Birth Date

A0900. Birth Date						
	<input type="text"/>	-	<input type="text"/>	-	<input type="text"/>	<input type="text"/>
	Month		Day		Year	

A0900: Birth Date (cont.)

Item Rationale

- Assists in correct identification.
- Allows determination of age.

Coding Instructions

- Fill in the boxes with the appropriate birth date. If the complete birth date is known, do not leave any boxes blank. If the month or day contains only a single digit, fill the first box in with a "0." For example: January 2, 1918, should be entered as 01-02-1918.
- Sometimes, only the birth year or the birth year and birth month will be known. These situations are handled as follows:
 - If only the birth year is known (e.g., 1918), then enter the year in the "year" portion of A0900, and leave the "month" and "day" portions blank. If the birth year and birth month are known, but the day of the month is not known, then enter the year in the "year" portion of A0900, enter the month in the "month" portion of A0900, and leave the "day" portion blank.

A1000: Race/Ethnicity

A1000. Race/Ethnicity	
↓ Check all that apply	
<input type="checkbox"/>	A. American Indian or Alaska Native
<input type="checkbox"/>	B. Asian
<input type="checkbox"/>	C. Black or African American
<input type="checkbox"/>	D. Hispanic or Latino
<input type="checkbox"/>	E. Native Hawaiian or Other Pacific Islander
<input type="checkbox"/>	F. White

Item Rationale

- This item uses the common uniform language approved by the Office of Management and Budget (OMB) to report racial and ethnic categories. The categories in this classification are social-political constructs and should not be interpreted as being scientific or anthropological in nature.
- Provides demographic race/ethnicity specific health trend information.
- These categories are NOT used to determine eligibility for participation in any Federal program.

A1000: Race/Ethnicity (cont.)

Steps for Assessment: Interview Instructions

1. Ask the resident to select the category or categories that most closely correspond to his or her race/ethnicity from the list in A1000.
 - Individuals may be more comfortable if this and the preceding question are introduced by saying, "We want to make sure that all our residents get the best care possible, regardless of their race or ethnic background. We would like you to tell us your ethnic and racial background so that we can review the treatment that all residents receive and make sure that everyone gets the highest quality of care" (Baker et al., 2005).
2. If the resident is unable to respond, ask a family member or significant other.
3. Category definitions are provided to resident or family only if requested by them in order to answer the item.
4. Respondents should be offered the option of selecting one or more racial designations.
5. Only if the resident is unable to respond and no family member or significant other is available, observer identification or medical record documentation may be used.

Coding Instructions

Check all that apply.

- Enter the race or ethnic category or categories the resident, family or significant other uses to identify him or her.

DEFINITIONS

RACE/ETHNICITY AMERICAN INDIAN OR ALASKA NATIVE

A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

ASIAN

A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, Vietnam.

BLACK OR AFRICAN AMERICAN

A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black" or "African American."

HISPANIC OR LATINO

A person of Cuban, Mexican, Puerto Rican, South or Central American or other Spanish culture or origin regardless of race. The term "Spanish Origin" can be used in addition to "Hispanic" or "Latino."

NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER

A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

WHITE

A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

A1100: Language

A1100. Language																					
Enter Code <input type="checkbox"/>	<p>A. Does the resident need or want an interpreter to communicate with a doctor or health care staff?</p> <p>0. No 1. Yes → Specify in A1100B, Preferred language 9. Unable to determine</p> <p>B. Preferred language:</p> <table border="1" style="width: 100%; height: 20px;"> <tr> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td> </tr> </table>																				

Item Rationale

Health-related Quality of Life

- Inability to make needs known and to engage in social interaction because of a language barrier can be very frustrating and can result in isolation, depression, and unmet needs.
- Language barriers can interfere with accurate assessment.

Planning for Care

- When a resident needs or wants an interpreter, the nursing home should ensure that an interpreter is available.
- An alternate method of communication also should be made available to help to ensure that basic needs can be expressed at all times, such as a communication board with pictures on it for the resident to point to (if able).
- Identifies residents who need interpreter services in order to answer interview items or participate in consent process.

Steps for Assessment

1. Ask the resident if he or she needs or wants an interpreter to communicate with a doctor or health care staff.
2. If the resident is unable to respond, a family member or significant other should be asked.
3. If neither source is available, review record for evidence of a need for an interpreter.
4. If an interpreter is wanted or needed, ask for preferred language.
5. It is acceptable for a family member or significant other to be the interpreter if the resident is comfortable with it and if the family member or significant other will translate exactly what the resident says without providing his or her interpretation.

Coding Instructions for A1100A

- **Code 0, no:** if the resident (or family or medical record if resident unable to communicate) indicates that the resident does not want or need an interpreter to communicate with a doctor or health care staff.
- **Code 1, yes:** if the resident (or family or medical record if resident unable to communicate) indicates that he or she needs or wants an interpreter to communicate with a doctor or health care staff. Specify preferred language. Proceed to 1100B and enter the resident's preferred language.
- **Code 9, unable to determine:** if no source can identify whether the resident wants or needs an interpreter.

A1100: Language (cont.)

Coding Instructions for A1100B

- Enter the preferred language the resident primarily speaks or understands after interviewing the resident and family, observing the resident and listening, and reviewing the medical record.

Coding Tips and Special Populations

- An organized system of signing such as American Sign Language (ASL) can be reported as the preferred language if the resident needs or wants to communicate in this manner.

A1200: Marital Status

A1200. Marital Status	
Enter Code <input type="checkbox"/>	1. Never married 2. Married 3. Widowed 4. Separated 5. Divorced

Item Rationale

- Allows understanding of the formal relationship the resident has and can be important for care and discharge planning.
- Demographic information.

Steps for Assessment

1. Ask the resident about his or her marital status.
2. If the resident is unable to respond, ask a family member or other significant other.
3. If neither source can report, review the medical record for information.

Coding Instructions

- Choose the answer that best describes the current marital status of the resident and enter the corresponding number in the code box:
 1. Never Married
 2. Married
 3. Widowed
 4. Separated
 5. Divorced

A1300: Optional Resident Items

A1300. Optional Resident Items	
A. Medical record number:	<input type="text"/>
B. Room number:	<input type="text"/>
C. Name by which resident prefers to be addressed:	<input type="text"/>
D. Lifetime occupation(s) - put "/" between two occupations:	<input type="text"/>

Item Rationale

- Some facilities prefer to include the nursing home medical record number on the MDS to facilitate tracking.
- Some facilities conduct unit reviews of MDS items in addition to resident and nursing home level reviews. The unit may be indicated by the room number.
- Preferred name and lifetime occupation help nursing home staff members personalize their interactions with the resident.
- Many people are called by a nickname or middle name throughout their life. It is important to call residents by the name they prefer in order to establish comfort and respect between staff and resident. Also, some cognitively impaired or hearing impaired residents might have difficulty responding when called by their legal name, if it is not the name most familiar to them.
- Others may prefer a more formal and less familiar address. For example, a physician might appreciate being referred to as "Doctor."
- Knowing a person's lifetime occupation is also helpful for care planning and conversation purposes. For example, a carpenter might enjoy pursuing hobby shop activities.
- These are optional items because they are not needed for CMS program function.

Coding Instructions for A1300A, Medical Record Number

- Enter the resident's medical record number (from the nursing home medical record, admission office or Health Information Management Department) if the nursing home chooses to exercise this option.

Coding Instructions for A1300B, Room Number

- Enter the resident's room number if the nursing home chooses to exercise this option.

Coding Instructions for A1300C, Name by Which Resident Prefers to Be Addressed

- Enter the resident's preferred name. This field captures a preferred nickname, middle name, or title that the resident prefers staff use.
- Obtained from resident self-report or family or significant other if resident is unable to respond.

A1300: Optional Resident Items (cont.)

Coding Instructions for A1300D, Lifetime Occupation(s)

- Enter the job title or profession that describes the resident’s main occupation(s) before retiring or entering the nursing home. When two occupations are identified, place a slash (/) between each occupation.
- The lifetime occupation of a person whose primary work was in the home should be recorded as “homemaker.” For a resident who is a child or a mentally retarded/developmentally delayed adult resident who has never had an occupation, record as “none.”

A1500: Preadmission Screening and Resident Review (PASRR)

A1500. Preadmission Screening and Resident Review (PASRR)	
Complete only if A0310A = 01, 03, 04, or 05	
Enter Code	Is the resident currently considered by the state level II PASRR process to have serious mental illness and/or intellectual disability ("mental retardation" in federal regulation) or a related condition?
<input type="checkbox"/>	0 No → Skip to A1550, Conditions Related to ID/DD Status
	1 Yes → Continue to A1510, Level II Preadmission Screening and Resident Review (PASRR) Conditions
	9. Not a Medicaid-certified unit → Skip to A1550, Conditions Related to ID/DD Status

Item Rationale

Health-related Quality of Life

- All individuals who are admitted to a Medicaid certified nursing facility must have a Level I PASRR completed to screen for possible mental illness (MI), intellectual disability (ID), (“mental retardation” (MR) in federal regulation), or related conditions regardless of the resident’s method of payment (please contact your local State Medicaid Agency for details regarding PASRR requirements and exemptions).
- Individuals who have or are suspected to have MI/ID or related conditions may not be admitted to a Medicaid-certified nursing facility unless approved through Level II PASRR determination. Those residents covered by Level II PASRR process may require certain care and services provided by the nursing home, and/or specialized services provided by the State.
- A resident with MI or ID must have a Resident Review (RR) conducted when there is a significant change in the resident’s physical or mental condition. Therefore, when a significant change in status MDS assessment is completed for a resident with MI or ID, the nursing home is required to notify the State mental health authority, intellectual disability or developmental delay disability authority (depending on which operates in their State) in order to notify them of the resident’s change in status. Section 1919(e)(7)(B)(iii) of the Social Security Act requires the notification or referral for a significant change.¹

¹ The statute may also be referenced as 42 USC 1396r(e)(7)(B)(iii). Note that as of this revision date the statute supersedes Federal regulations at 42 CFR 483.114(c), which still reads as requiring annual resident review. The regulation has not yet been updated to reflect the statutory change to resident review upon significant change in condition.

A1500: Preadmission Screening and Resident Review (PASRR) (cont.)

- Each State Medicaid agency might have specific processes and guidelines for referral, and which types of significant changes should be referred. Therefore, facilities should become acquainted with their own State requirements.
- Please see https://www.cms.gov/PASRR/01_Overview.asp for CMS information on PASRR.

Planning for Care

- The Level II PASRR determination and the evaluation report specify services to be provided by the nursing home and/or specialized services defined by the State.
- The State is responsible for providing specialized services to individuals with MI/ID. In some States specialized services are provided to residents in Medicaid-certified facilities (in other States specialized services are only provided in other facility types such as a psychiatric hospital). The nursing home is required to provide all other care and services appropriate to the resident's condition.
- The services to be provided by the nursing home and/or specialized services provided by the State that are specified in the Level II PASRR determination and the evaluation report should be addressed in the plan of care.
- Identifies individuals who are subject to Resident Review upon change in condition.

Steps for Assessment

1. Complete if A0310A = 01, 03, 04 or 05 (admission assessment, annual assessment, significant change in status assessment, significant correction to prior comprehensive assessment).
2. Review the Level I PASRR form to determine whether a Level II PASRR was required.
3. Review the PASRR report provided by the State if Level II screening was required.

Coding Instructions

- **Code 0, no:** and skip to A1550, Conditions Related to ID/DD Status, if any of the following apply:
 - PASRR Level I screening did not result in a referral for Level II screening, or
 - Level II screening determined that the resident does not have a serious mental illness and/or intellectual disability or related condition, or
 - PASRR screening is not required because the resident was admitted from a hospital after requiring acute inpatient care, is receiving services for the condition for which he or she received care in the hospital, and the attending physician has certified before admission that the resident is likely to require less than 30 days of nursing home care.

A1500: Preadmission Screening and Resident Review (PASRR) (cont.)

- **Code 1, yes:** if PASRR Level II screening determined that the resident has a serious mental illness and/or intellectual disability or related condition, and continue to A1510, Level II Preadmission Screening and Resident Review (PASRR) Conditions.
- **Code 9, not a Medicaid-certified unit:** if bed is not in a Medicaid-certified nursing home. Skip to A1550, Conditions Related to ID/DD Status. The PASRR process does not apply to nursing home units that are not certified by Medicaid (unless a State requires otherwise) and therefore the question is not applicable.
 - Note that the requirement is based on the certification of the part of the nursing home the resident will occupy. In a nursing home in which some parts are Medicaid certified and some are not, this question applies when a resident is admitted, or transferred to, a Medicaid certified part of the building.

A1510: Level II Preadmission Screening and Resident Review (PASRR) Conditions

A1510. Level II Preadmission Screening and Resident Review (PASRR) Conditions	
Complete only if A0310A = 01, 03, 04, or 05	
↓ Check all that apply	
<input type="checkbox"/>	A. Serious mental illness
<input type="checkbox"/>	B. Intellectual Disability ("mental retardation" in federal regulation)
<input type="checkbox"/>	C. Other related conditions

Steps for Assessment

1. Complete if A0310A = 01, 03, 04 or 05 (admission assessment, annual assessment, significant change in status assessment, significant correction to prior comprehensive assessment).
2. Check all that apply.

Coding Instructions

- **Code A, Serious mental illness:** if resident has been diagnosed with a serious mental illness.
- **Code B, Intellectual Disability ("mental retardation" in federal regulation):** if resident has been diagnosed with intellectual disability (or "mental retardation").
- **Code C, Other related conditions:** if resident has been diagnosed with other related conditions.

A1550: Conditions Related to Intellectual Disability/Developmental Delay (ID/DD) Status

A1550. Conditions Related to ID/DD Status	
If the resident is 22 years of age or older, complete only if A0310A = 01	
If the resident is 21 years of age or younger, complete only if A0310A = 01, 03, 04, or 05	
↓ Check all conditions that are related to ID/DD status that were manifested before age 22, and are likely to continue indefinitely	
	ID/DD With Organic Condition
<input type="checkbox"/>	A. Down syndrome
<input type="checkbox"/>	B. Autism
<input type="checkbox"/>	C. Epilepsy
<input type="checkbox"/>	D. Other organic condition related to ID/DD
	ID/DD Without Organic Condition
<input type="checkbox"/>	E. ID/DD with no organic condition
	No ID/DD
<input type="checkbox"/>	Z. None of the above

Item Rationale

- To document conditions associated with intellectual or developmental delay disabilities.

Steps for Assessment

- If resident is 22 years of age or older on the assessment reference date, complete only if A0310A = 01 (admission assessment).
- If resident is 21 years of age or younger on the assessment reference date, complete if A0310A = 01, 03, 04, or 05 (admission assessment, annual assessment, significant change in status assessment, significant correction to prior comprehensive assessment).

Coding Instructions

- Check all conditions related to ID /DD status that were present before age 22.
- When age of onset is not specified, assume that the condition meets this criterion AND is likely to continue indefinitely.
- Code A:** if Down syndrome is present.
- Code B:** if autism is present.
- Code C:** if epilepsy is present.
- Code D:** if other organic condition related to ID/DD is present.

DEFINITIONS

DOWN SYNDROME

A common genetic disorder in which a child is born with 47 rather than 46 chromosomes, resulting in developmental delays, intellectual disability, low muscle tone, and other possible effects.

AUTISM

A developmental disorder that is characterized by impaired social interaction, problems with verbal and nonverbal communication, and unusual, repetitive, or severely limited activities and interests.

EPILEPSY

A common chronic neurological disorder that is characterized by recurrent unprovoked seizures.

A1550: Conditions Related to Intellectual Disability/Developmental Delay (ID/DD) Status (cont.)

- **Code E:** if an ID/DD condition is present but the resident does not have any of the specific conditions listed.
- **Code Z:** if ID/DD condition is not present.

DEFINITION

OTHER ORGANIC CONDITION RELATED TO ID/DD

Examples of diagnostic conditions include congenital syphilis, maternal intoxication, mechanical injury at birth, prenatal hypoxia, neuronal lipid storage diseases, phenylketonuria (PKU), neurofibromatosis, microcephalus, macrocephaly, meningomyelocele, congenital hydrocephalus, etc.

A1600: Entry Date (date of this admission/entry or reentry into the facility)

A1600. Entry Date (date of this admission/entry or reentry into the facility)			
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Month	Day	Year	

Item Rationale

- To document the date of admission/entry or reentry into the nursing home.

Coding Instructions

- Enter the most recent date of admission/entry or reentry to this nursing home. Use the format: Month-Day-Year: XX-XX-XXXX. For example, October 12, 2010, would be entered as 10-12-2010.

DEFINITION

ENTRY DATE

The initial date of admission to the nursing facility, or the date the resident most recently returned to your nursing facility after being discharged.

A1700: Type of Entry

A1700. Type of Entry	
Enter Code <input type="checkbox"/>	1 Admission 2 Reentry

Item Rationale

- Captures whether date in A1600 is an admission/entry or reentry date.

Coding Instructions

- **Code 1, admission/entry:** when one of the following occurs:

A1700: Type of Entry (cont.)

1. resident has never been admitted to this facility before; OR
 2. resident has been in this facility previously and was discharged prior to completion of the OBRA admission assessment; OR
 3. resident has been in this facility previously and was discharged return not anticipated; OR
 4. resident has been in this facility previously and was discharged return anticipated and did not return within 30 days of discharge.
- **Code 2, reentry:** when all 3 of the following occurred prior to the this entry, the resident was:
 1. admitted to this nursing home (i.e., OBRA admission assessment was completed), AND
 2. discharged return anticipated, AND
 3. returned to facility within 30 days of discharge.

Coding Tips and Special Populations

- Swing bed facilities will always code the resident’s entry as an admission, ‘1’, since an OBRA Admission assessment must have been completed to code as a reentry. OBRA Admission assessments are not completed for swing bed residents.
- In determining if a resident returns to the facility within 30 days, the day of discharge from the facility is not counted in the 30 days. For example, a resident is discharged return anticipated on December 1 would need to return to the facility by December 31 to meet the “within 30 day” requirement.

A1800: Entered From

A1800. Entered From	
Enter Code	01. Community (private home/apt., board/care, assisted living, group home)
<input type="text"/>	02. Another nursing home or swing bed
<input type="text"/>	03. Acute hospital
	04. Psychiatric hospital
	05. Inpatient rehabilitation facility
	06. ID/DD facility
	07. Hospice
	09. Long Term Care Hospital (LTCH)
	99. Other

Item Rationale

- Understanding the setting that the individual was in immediately prior to nursing home admission informs care planning and may also inform discharge planning and discussions.
- Demographic information.

Steps for Assessment

1. Review transfer and admission records.
2. Ask the resident and/or family or significant others.

A1800: Entered From (cont.)

Coding Instructions

Enter the 2-digit code that corresponds to the location or program the resident was admitted from for this admission.

- **Code 01, community (private home/apt, board/care, assisted living, group home):** if the resident was admitted from a private home, apartment, board and care, assisted living facility or group home.
- **Code 02, another nursing home or swing bed:** if the resident was admitted from an institution (or a distinct part of an institution) that is primarily engaged in providing skilled nursing care and related services for residents who require medical or nursing care or rehabilitation services for injured, disabled, or sick persons. Includes swing beds.
- **Code 03, acute hospital:** if the resident was admitted from an institution that is engaged in providing, by or under the supervision of physicians for inpatients, diagnostic services, therapeutic services for medical diagnosis, and the treatment and care of injured, disabled, or sick persons.
- **Code 04, psychiatric hospital:** if the resident was admitted from an institution that is engaged in providing, by or under the supervision of a physician, psychiatric services for the diagnosis and treatment of mentally ill residents.
- **Code 05, inpatient rehabilitation facility (IRF):** if the resident was admitted from an institution that is engaged in providing, under the supervision of physicians, services for the rehabilitation of injured, disabled or sick persons. Includes IRFs that are units within acute care hospitals.
- **Code 06, ID/DD facility:** if the resident was admitted from an institution that is engaged in providing, under the supervision of a physician, any health and rehabilitative services for individuals who have intellectual or developmental delay disabilities.
- **Code 07, hospice:** if the resident was admitted from a program for terminally ill persons where an array of services is necessary for the palliation and management of terminal illness and related conditions. The hospice must be licensed by the State as a hospice provider and/or certified under the Medicare program as a hospice provider. Includes community-based or inpatient hospice programs.
- **Code 09, long term care hospital (LTCH):** if the patient was admitted from a hospital that is certified under Medicare as a short-term, acute-care hospital which has been excluded from the Inpatient Acute Care Hospital Prospective Payment System (IPPS) under §1886(d)(1)(B)(iv) of the Social Security Act. For the purpose of Medicare

DEFINITION

DEFINITIONS PRIVATE HOME OR APARTMENT

Any house, condominium, or apartment in the community whether owned by the resident or another person. Also included in this category are retirement communities and independent housing for the elderly.

BOARD AND CARE/ ASSISTED LIVING/ GROUP HOME

A non-institutional community residential setting that includes services of the following types: home health services, homemaker/personal care services, or meal services.

A1800: Entered From (cont.)

payment, LTCHs are defined as having an average inpatient length of stay (as determined by the Secretary) of greater than 25 days.

- **Code 99, other:** if the resident was admitted from none of the above.

Coding Tips and Special Populations

- If an individual was enrolled in a home-based hospice program enter **07, Hospice**, instead of **01, Community**.

A2000: Discharge Date

A2000. Discharge Date																	
Complete only if A0310F = 10, 11, or 12																	
<table border="1"> <tr> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> <td style="width: 20px; height: 20px;"></td> </tr> <tr> <td colspan="2" style="text-align: center;">Month</td> <td colspan="2" style="text-align: center;">Day</td> <td colspan="4" style="text-align: center;">Year</td> </tr> </table>									Month		Day		Year				
Month		Day		Year													

Item Rationale

- Closes case in system.

Coding Instructions

- Enter the date the resident was discharged (whether or not return is anticipated). This is the date the resident leaves the facility.
- For discharge assessments, the discharge date (A2000) and ARD (A2300) must be the same date.
- Do not include leave of absence or hospital observational stays less than 24 hours unless admitted to the hospital.
- Obtain data from the medical, admissions or transfer records.

Coding Tips and Special Populations

- If a resident was receiving services under SNF Part A PPS, the discharge date may be later than the end of Medicare stay date (A2400C).

A2100: Discharge Status

A2100. Discharge Status					
Complete only if A0310F = 10, 11, or 12					
<table border="1"> <tr> <td style="width: 50px;">Enter Code</td> <td style="width: 50px;"></td> </tr> <tr> <td style="width: 50px; height: 20px;"></td> <td style="width: 50px; height: 20px;"></td> </tr> </table>	Enter Code				<ul style="list-style-type: none"> 01 Community (private home/apt., board/care, assisted living, group home) 02 Another nursing home or swing bed 03 Acute hospital 04 Psychiatric hospital 05 Inpatient rehabilitation facility 06 ID/DD facility 07 Hospice 08 Deceased 09 Long Term Care Hospital (LTCH) 99 Other
Enter Code					

A2100: Discharge Status (cont.)

Item Rationale

- Demographic and outcome information.

Steps for Assessment

1. Review the medical record including the discharge plan and discharge orders for documentation of discharge location.

Coding Instructions

Select the 2-digit code that corresponds to the resident's discharge status.

- **Code 01, community (private home/apt., board/care, assisted living, group home):** if discharge location is a private home, apartment, board and care, assisted living facility, or group home.
- **Code 02, another nursing home or swing bed:** if discharge location is an institution (or a distinct part of an institution) that is primarily engaged in providing skilled nursing care and related services for residents who require medical or nursing care or rehabilitation services for injured, disabled, or sick persons. Includes swing beds.
- **Code 03, acute hospital:** if discharge location is an institution that is engaged in providing, by or under the supervision of physicians for inpatients, diagnostic services, therapeutic services for medical diagnosis, and the treatment and care of injured, disabled, or sick persons.
- **Code 04, psychiatric hospital:** if discharge location is an institution that is engaged in providing, by or under the supervision of a physician, psychiatric services for the diagnosis and treatment of mentally ill residents.
- **Code 05, inpatient rehabilitation facility:** if discharge location is an institution that is engaged in providing, under the supervision of physicians, rehabilitation services for the rehabilitation of injured, disabled or sick persons. Includes IRFs that are units within acute care hospitals.
- **Code 06, ID/DD facility:** if discharge location is an institution that is engaged in providing, under the supervision of a physician, any health and rehabilitative services for individuals who have intellectual or developmental delay disabilities.
- **Code 07, hospice:** if discharge location is a program for terminally ill persons where an array of services is necessary for the palliation and management of terminal illness and related conditions. The hospice must be licensed by the State as a hospice provider and/or certified under the Medicare program as a hospice provider. Includes community-based (e.g., home) or inpatient hospice programs.
- **Code 08, deceased:** if resident is deceased.
- **Code 09, long term care hospital (LTCH):** if the patient was discharged from a hospital that is certified under Medicare as a short-term, acute-care hospital which has

A2100: Discharge Status (cont.)

been excluded from the Inpatient Acute Care Hospital Prospective Payment System (IPPS) under §1886(d)(1)(B)(iv) of the Social Security Act. For the purpose of Medicare payment, LTCHs are defined as having an average inpatient length of stay (as determined by the Secretary) of greater than 25 days.

- **Code 99, other:** if discharge location is none of the above.

A2200: Previous Assessment Reference Date for Significant Correction

A2200. Previous Assessment Reference Date for Significant Correction			
Complete only if A0310A = 05 or 06			
	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	Month	Day	Year

Item Rationale

- To identify the ARD of a previous comprehensive or quarterly assessment (A0310A = 05 or 06) in which a significant error is discovered.

Coding Instructions

- Complete only if A0310A = 05 (Significant correction to prior comprehensive assessment) or A0310A = 06 (Significant correction to prior quarterly assessment).
- Enter the ARD of the prior comprehensive or quarterly assessment in which a significant error has been identified and a correction is required.

A2300: Assessment Reference Date

A2300. Assessment Reference Date			
Observation end date:			
	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
	Month	Day	Year

Item Rationale

- Designates the end of the look-back period so that all assessment items refer to the resident's status during the same period of time.
As the last day of the look-back period, the ARD serves as the reference point for determining the care and services captured on the MDS assessment. Anything that happens after the ARD will not be captured on that MDS. For example, for a MDS item with a 7-day look-back period, assessment information is collected for a 7-day period ending on and including the ARD which is the 7th day of this look-back period. For an item with a 14-day look-back period, the information is collected for a 14-day period ending on and including the ARD. The look-back period includes observations and events through the end of the day (midnight) of the ARD.

A2300: Assessment Reference Date (cont.)

Steps for Assessment

1. Interdisciplinary team members should select the ARD based on the reason for the assessment and compliance with all timing and scheduling requirements outlined in Chapter 2.

Coding Instructions

- Enter the appropriate date on the lines provided. Do not leave any spaces blank. If the month or day contains only a single digit, enter a "0" in the first space. Use four digits for the year. For example, October 2, 2010, should be entered as: 10-02-2010.
- For detailed information on the timing of the assessments, see Chapter 2 on assessment schedules.
- For discharge assessments, the discharge date item (A2000) and the ARD item (A2300) must contain the same date.

Coding Tips and Special Populations

- When the resident dies or is discharged prior to the end of the look-back period for a required assessment, the ARD must be adjusted to equal the discharge date.
- The look-back period may not be extended simply because a resident was out of the nursing home during part of the look-back period (e.g., a home visit, therapeutic leave, or hospital observation stay less than 24 hours when resident is not admitted). For example, if the ARD is set at day 13 and there is a 2-day temporary leave during the look-back period, the 2 leave days are still considered part of the look-back period.
- When collecting assessment information, data from the time period of the leave of absence is captured as long as the particular MDS item permits. For example, if the family takes the resident to the physician during the leave, the visit would be counted in Item O0600, **Physician Examination** (if criteria are otherwise met).

This requirement applies to all assessments, regardless of whether they are being completed for clinical or payment purposes.

DEFINITIONS

ASSESSMENT REFERENCE DATE (ARD)

The specific end-point for the look-back periods in the LTCH CARE Data Set assessment process. Almost all LTCH CARE Data Set items refer to the patient's status over a designated time period referring back in time from the Assessment Reference Date (ARD). Most frequently, this look-back period, also called the observation or assessment period, is a XX-day period ending on the ARD. Look-back periods may cover the XX days ending on this date, xx days ending on this date, etc.

A2400: Medicare Stay

A2400: Medicare Stay	
Enter Code <input type="checkbox"/>	<p>A. Has the resident had a Medicare-covered stay since the most recent entry?</p> <p>0. No → Skip to B0100, Comatose</p> <p>1. Yes → Continue to A2400B. Start date of most recent Medicare stay</p> <p>B. Start date of most recent Medicare stay:</p> <p><input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p style="text-align: center;">Month Day Year</p> <p>C. End date of most recent Medicare stay - Enter dashes if stay is ongoing</p> <p><input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> - <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p style="text-align: center;">Month Day Year</p>

A2400: Medicare Stay (cont.)

Item Rationale

- Identifies when a resident is receiving services under the scheduled PPS.
- Identifies when a resident's Medicare Part A stay begins and ends.
- The end date is used to determine if the resident's stay qualifies for the short stay assessment.

Coding Instructions for A2400A, Has the Resident Had a Medicare-covered Stay since the Most Recent Entry?

- **Code 0, no:** if the resident has not had a covered Medicare Part A covered stay since the most recent admission/entry or reentry. Skip to B0100, Comatose.
- **Code 1, yes:** if the resident has had a Medicare Part A covered stay since the most recent admission/entry or reentry. Continue to A2400B.

Coding Instructions for A2400B, Start of Most Recent Medicare Stay

- **Code the date of day 1** of this Medicare stay if A2400A is **coded 1, yes**.

Coding Instructions for A2400C, End Date of Most Recent Medicare Stay

- **Code the date of last day** of this Medicare stay if A2400A is **coded 1, yes**.
- If the Medicare Part A stay is ongoing there will be no end date to report. Enter dashes to indicate that the stay is ongoing.
- The end of Medicare date is coded as follows, whichever occurs first:
 - Date SNF benefit exhausts (i.e., the 100th day of the benefit); or
 - Date of last day covered as recorded on the effective date from the Generic Notice or
 - The last paid day of Medicare A when payer source changes to another payer (regardless if the resident was moved to another bed or not); or
 - Date the resident was discharged from the facility (see Item A2000, Discharge Date).

DEFINITIONS

MOST RECENT MEDICARE STAY

This is a Medicare Part A covered stay that has started on or after the most recent admission/entry or reentry to the nursing facility.

MEDICARE-COVERED STAY

Skilled Nursing Facility stays billable to Medicare Part A. Does not include stays billable to Medicare Advantage HMO plans.

CURRENT MEDICARE STAY

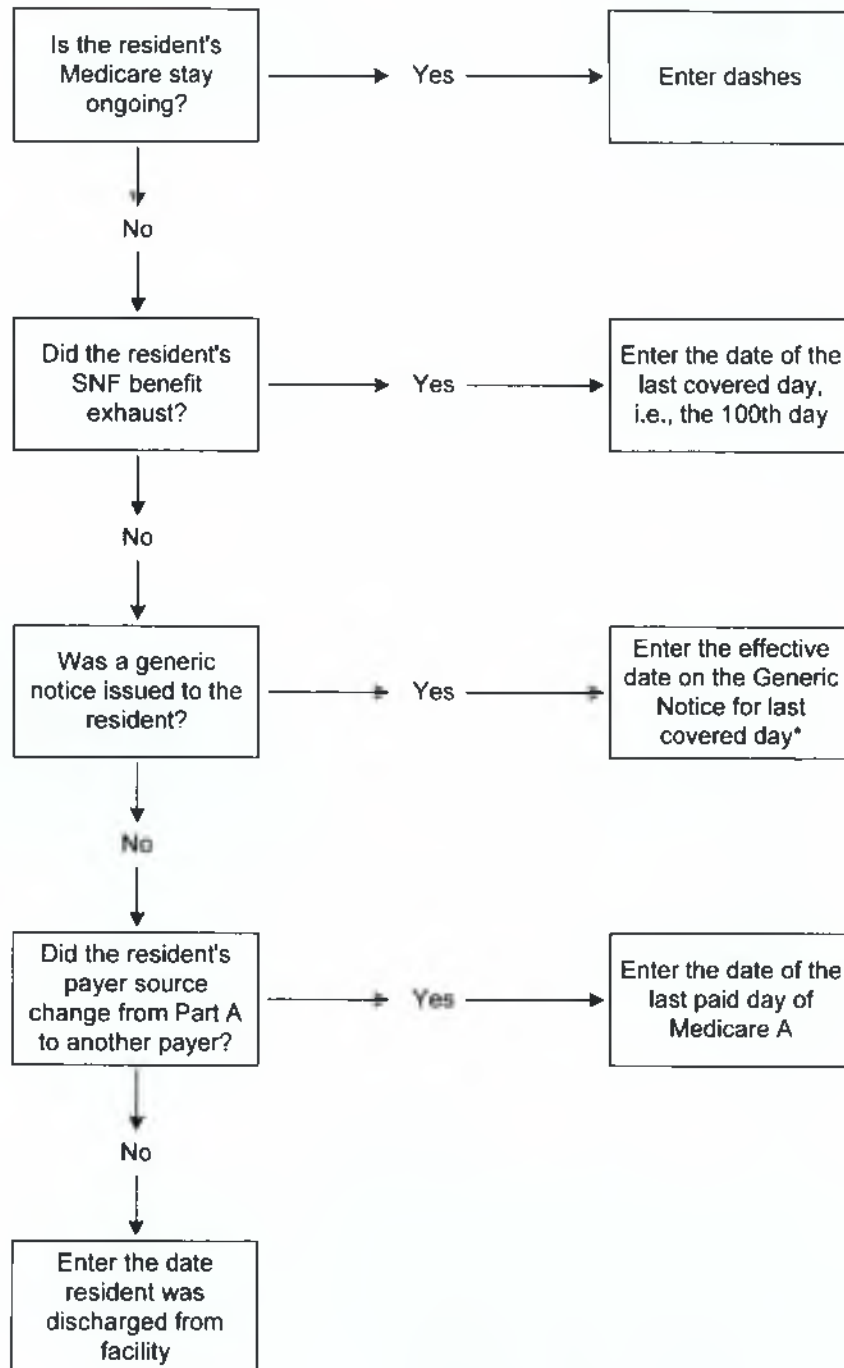
NEW ADMISSION:

Day 1 of Medicare Part A stay.

READMISSION:

Day 1 of Medicare Part A coverage after readmission following a discharge.

**Medicare Stay End Date Algorithm
A2400C**



*if resident leaves facility prior to last covered day as recorded on the generic notice, enter date resident left facility.

A2400: Medicare Stay (cont.)

Coding Tips and Special Populations

- When a resident on Medicare Part A returns following a therapeutic leave of absence or a hospital observation stay of less than 24 hours (without hospital admission), this is a continuation of the Medicare Part A stay, not a new Medicare Part A stay.
- The end date of the Medicare stay may be earlier than actual discharge date from the facility (Item A2000).

Examples

1. Mrs. G. began receiving services under Medicare Part A on October 14, 2010. Due to her stable condition and ability to manage her medications and dressing changes, the facility determined that she no longer qualified for Part A SNF coverage and issued an ABN with the last day of coverage as November 23, 2010. Mrs. G. was discharged from the facility on November 24, 2010. Code the following on her discharge assessment:
 - A2000 = 11-24-2010
 - A2400A = 1
 - A2400B = 10-14-2010
 - A2400C = 11-23-2010
2. Mr. N began receiving services under Medicare Part A on December 11, 2010. He was sent to the ER on December 19, 2010 at 8:30pm and was not admitted to the hospital. He returned to the facility on December 20, 2010, at 11:00 am. The facility completed his 14-day PPS assessment with an ARD of December 23, 2010. Code the following on his 14-day PPS assessment:
 - A2400A = 1
 - A2400B = 12-11-2010
 - A2400C = -----
3. Mr. R. began receiving services under Medicare Part A on October 15, 2010. He was discharged return anticipated on October 20, 2010, to the hospital. Code the following on his discharge assessment:
 - A2000 = 10-20-2010
 - A2400A = 1
 - A2400B = 10-15-2010
 - A2400C = 10-20-2010

ΠΑΡΑΡΤΗΜΑ Β': HEARING, SPEECH, AND VISION

SECTION B: HEARING, SPEECH, AND VISION

Intent: The intent of items in this section is to document the resident's ability to hear (with assistive hearing devices, if they are used), understand, and communicate with others and whether the resident experiences visual limitations or difficulties related to diseases common in aged persons.

B0100: Comatose

B0100. Comatose	
Enter Code	Persistent vegetative state/no discernible consciousness
<input type="checkbox"/>	0. No → Continue to B0200, Hearing
	1. Yes → Skip to G0110, Activities of Daily Living (ADL) Assistance

Item Rationale

Health-related Quality of Life

- Residents who are in a coma or persistent vegetative state are at risk for the complications of immobility, including skin breakdown and joint contractures.

Planning for Care

- Care planning should center on eliminating or minimizing complications and providing care consistent with the resident's health care goals.

DEFINITIONS

COMATOSE (coma)

A pathological state in which neither arousal (wakefulness, alertness) nor awareness exists. The person is unresponsive and cannot be aroused; he/she does not open his/her eyes, does not speak and does not move his/her extremities on command or in response to noxious stimuli (e.g., pain).

Steps for Assessment

- Review the medical record to determine if a neurological diagnosis of comatose or persistent vegetative state has been documented by a physician, or nurse practitioner, physician assistant, or clinical nurse specialist if allowable under state licensure laws.

Coding Instructions

- Code 0, no:** if a diagnosis of coma or persistent vegetative state is not present during the 7-day look-back period. Continue to B0200 **Hearing**.
- Code 1, yes:** if the record indicates that a physician, nurse practitioner or clinical nurse specialist has documented a diagnosis of coma or persistent vegetative state that is applicable during the 7-day look-back period. Skip to Section G0110, **Activities of Daily Living (ADL) Assistance**.

B0100: Comatose (cont.)

Coding Tips

- Only code if a diagnosis of coma or persistent vegetative state has been assigned. For example, some residents in advanced stages of progressive neurologic disorders such as Alzheimer's disease may have severe cognitive impairment, be non-communicative and sleep a great deal of time; however, they are usually not comatose or in a persistent vegetative state, as defined here.

DEFINITIONS

PERSISTENT VEGETATIVE STATE

Sometimes residents who were comatose after an anoxic-ischemic injury (i.e., not enough oxygen to the brain) from a cardiac arrest, head trauma, or massive stroke, regain wakefulness but do not evidence any purposeful behavior or cognition. Their eyes are open, and they may grunt, yawn, pick with their fingers, and have random body movements. Neurological exam shows extensive damage to both cerebral hemispheres.

B0200: Hearing

B0200. Hearing	
Enter Code <input type="checkbox"/>	Ability to hear (with hearing aid or hearing appliances if normally used) 0. Adequate - no difficulty in normal conversation, social interaction, listening to TV 1. Minimal difficulty - difficulty in some environments (e.g., when person speaks softly or setting is noisy) 2. Moderate difficulty - speaker has to increase volume and speak distinctly 3. Highly impaired - absence of useful hearing

Item Rationale

Health-related Quality of Life

- Problems with hearing can contribute to sensory deprivation, social isolation, and mood and behavior disorders.
- Unaddressed communication problems related to hearing impairment can be mistaken for confusion or cognitive impairment.

Planning for Care

- Address reversible causes of hearing difficulty (such as cerumen impaction).
- Evaluate potential benefit from hearing assistance devices.
- Offer assistance to residents with hearing difficulties to avoid social isolation.

B0200: Hearing (cont.)

- Consider other communication strategies for persons with hearing loss that is not reversible or is not completely corrected with hearing devices.
- Adjust environment by reducing background noise by lowering the sound volume on televisions or radios, because a noisy environment can inhibit opportunities for effective communication.

Steps for Assessment

1. Ensure that the resident is using his or her normal hearing appliance if they have one. Hearing devices may not be as conventional as a hearing aid. Some residents by choice may use hearing amplifiers or a microphone and headphones as an alternative to hearing aids. Ensure whatever hearing appliance is used, it is operational.
2. Interview the resident and ask about hearing function in different situations (e.g. hearing staff members, talking to visitors, using telephone, watching TV, attending activities).
3. Observe the resident during your verbal interactions and when he or she interacts with others throughout the day.
4. Think through how you can best communicate with the resident. For example, you may need to speak more clearly, use a louder tone, speak more slowly or use gestures. The resident may need to see your face to understand what you are saying, or you may need to take the resident to a quieter area for them to hear you. All of these are cues that there is a hearing problem.
5. Review the medical record.
6. Consult the resident's family, direct care staff, activities personnel, and speech or hearing specialists.

Coding Instructions

- **Code 0, adequate:** No difficulty in normal conversation, social interaction, or listening to TV. The resident hears all normal conversational speech and telephone conversation and announcements in group activities.
- **Code 1, minimal difficulty:** Difficulty in some environments (e.g., when a person speaks softly or the setting is noisy). The resident hears speech at conversational levels but has difficulty hearing when not in quiet listening conditions or when not in one-on-one situations. The resident's hearing is adequate after environmental adjustments are made, such as reducing background noise by moving to a quiet room or by lowering the volume on television or radio.
- **Code 2, moderate difficulty:** Speaker has to increase volume and speak distinctly. Although hearing-deficient, the resident compensates when the speaker adjusts tonal quality and speaks distinctly; or the resident can hear only when the speaker's face is clearly visible.

B0200: Hearing (cont.)

- **Code 3, highly impaired:** Absence of useful hearing. The resident hears only some sounds and frequently fails to respond even when the speaker adjusts tonal quality, speaks distinctly, or is positioned face-to-face. There is no comprehension of conversational speech, even when the speaker makes maximum adjustments.

Coding Tips for Special Populations

- Residents who are unable to respond to a standard hearing assessment due to cognitive impairment will require alternate assessment methods. The resident can be observed in their normal environment. Does he or she respond (e.g., turn his or her head) when a noise is made at a normal level? Does the resident seem to respond only to specific noise in a quiet environment? Assess whether the resident responds only to loud noise or do they not respond at all.

B0300: Hearing Aid

B0300. Hearing Aid	
Enter Code	Hearing aid or other hearing appliance used in completing B0200, Hearing
<input type="checkbox"/>	0 No
	1 Yes

Item Rationale

Health-related Quality of Life

- Problems with hearing can contribute to social isolation and mood and behavior disorders.
- Many residents without hearing aids or other hearing appliances could benefit from them.
- Many persons who benefit from and own hearing aids do not have them on arrival at the nursing home or the hearing aid is not functional.

Planning for Care

- Knowing if a hearing aid was used when determining hearing ability allows better identification of evaluation and management needs.
- For residents with hearing aids, use and maintenance should be included in care planning.
- Residents who do not have adequate hearing without a hearing aid should be asked about history of hearing aid use.
- Residents who do not have adequate hearing despite wearing a hearing aid might benefit from a re-evaluation of the device or assessment for new causes of hearing impairment.

Steps for Assessment

1. Prior to beginning the hearing assessment, ask the resident if he or she owns a hearing aid or other hearing appliance and, if so, whether it is at the nursing home.
2. If the resident cannot respond, write the question down and allow the resident to read it.

B0300: Hearing Aid (cont.)

3. If the resident is still unable, check with family and care staff about hearing aid or other hearing appliances.
4. Check the medical record for evidence that the resident had a hearing appliance in place when hearing ability was recorded.
5. Ask staff and significant others whether the resident was using a hearing appliance when they observed hearing ability (above).

Coding Instructions

- **Code 0, no:** if the resident did not use a hearing aid (or other hearing appliance) for the 7-day hearing assessment coded in **B0200, Hearing**.
- **Code 1, yes:** if the resident did use a hearing aid (or other hearing appliance) for the hearing assessment coded in **B0200, Hearing**.

B0600: Speech Clarity

B0600. Speech Clarity	
Enter Code:	Select best description of speech pattern
<input type="checkbox"/>	0 Clear speech - distinct intelligible words
	1 Unclear speech - slurred or mumbled words
	2 No speech - absence of spoken words

Item Rationale

Health-related Quality of Life

- Unclear speech or absent speech can hinder communication and be very frustrating to an individual.
- Unclear speech or absent speech can result in physical and psychosocial needs not being met and can contribute to depression and social isolation.

DEFINITIONS

SPEECH

The verbal expression of articulate words.

Planning for Care

- If speech is absent or is not clear enough for the resident to make needs known, other methods of communication should be explored.
- Lack of speech clarity or ability to speak should not be mistaken for cognitive impairment.

Steps for Assessment

1. Listen to the resident.
2. Ask primary assigned caregivers about the resident's speech pattern.
3. Review the medical record.

B0600: Speech Clarity (cont.)

4. Determine the quality of the resident's speech, not the content or appropriateness—just words spoken.

Coding Instructions

- **Code 0, clear speech:** if the resident usually utters distinct, intelligible words.
- **Code 1, unclear speech:** if the resident usually utters slurred or mumbled words.
- **Code 2, no speech:** if there is an absence of spoken words.

B0700: Makes Self Understood

B0700. Makes Self Understood	
Enter Code <input type="checkbox"/>	Ability to express ideas and wants, consider both verbal and non-verbal expression 0. Understood 1. Usually understood - difficulty communicating some words or finishing thoughts but is able if prompted or given time 2. Sometimes understood - ability is limited to making concrete requests 3. Rarely/never understood

Item Rationale

Health-related Quality of Life

- Problems making self understood can be very frustrating for the resident and can contribute to social isolation and mood and behavior disorders.
- Unaddressed communication problems can be inappropriately mistaken for confusion or cognitive impairment.

Planning for Care

- Ability to make self understood can be optimized by not rushing the resident, breaking longer questions into parts and waiting for reply, and maintaining eye contact (if appropriate).
- If a resident has difficulty making self understood:
 - Identify the underlying cause or causes.
 - Identify the best methods to facilitate communication for that resident.

DEFINITIONS

MAKES SELF UNDERSTOOD

Able to express or communicate requests, needs, opinions, and to conduct social conversation in his or her primary language, whether in speech, writing, sign language, gestures, or a combination of these. Deficits in the ability to make one's self understood (expressive communication deficits) can include reduced voice volume and difficulty in producing sounds, or difficulty in finding the right word, making sentences, writing, and/or gesturing.

B0700: Makes Self Understood (cont.)

Steps for Assessment

1. Assess using the resident's preferred language.
2. **Interact with the resident.** Be sure he or she can hear you or have access to his or her preferred method for communication. If the resident seems unable to communicate, offer alternatives such as writing, pointing or using cue cards.
3. Observe his or her interactions with others in different settings and circumstances.
4. Consult with the primary nurse assistant (over all shifts), if available, the resident's family, and speech-language pathologist.

Coding Instructions

- **Code 0, understood:** if the resident expresses requests and ideas clearly.
- **Code 1, usually understood:** if the resident has difficulty communicating some words or finishing thoughts **but** is able if prompted or given time. He or she may have delayed responses or may require some prompting to make self understood.
- **Code 2, sometimes understood:** if the resident has limited ability but is able to express concrete requests regarding at least basic needs (e.g., food, drink, sleep, toilet).
- **Code 3, rarely or never understood:** if, at best, the resident's understanding is limited to staff interpretation of highly individual, resident-specific sounds or body language (e.g., indicated presence of pain or need to toilet).

B0800: Ability to Understand Others

B0800. Ability To Understand Others	
Enter Code <input type="checkbox"/>	<p>Understanding verbal content, however able (with hearing aid or device if used)</p> <p>0. Understands - clear comprehension</p> <p>1. Usually understands - misses some part/intent of message but comprehends most conversation</p> <p>2. Sometimes understands - responds adequately to simple, direct communication only</p> <p>3. Rarely/never understands</p>

Item Rationale

Health-related Quality of Life

- Inability to understand direct person-to-person communication
 - Can severely limit association with others.
 - Can inhibit the individual's ability to follow instructions that can affect health and safety.

B0800: Ability to Understand Others (cont.)

Planning for Care

- Thorough assessment to determine underlying cause or causes is critical in order to develop a care plan to address the individual's specific deficits and needs.
- Every effort should be made by the facility to provide information to the resident in a consistent manner that he or she understands based on an individualized assessment.

Steps for Assessment

1. Assess in the resident's preferred language.
2. If the resident uses a hearing aid, hearing device or other communications enhancement device, the resident should use that device during the evaluation of the resident's understanding of person-to-person communication.
3. Interact with the resident and observe his or her understanding of other's communication.
4. Consult with direct care staff over all shifts, if possible, the resident's family, and speech-language pathologist (if involved in care).
5. Review the medical record for indications of how well the resident understands others.

DEFINITIONS

ABILITY TO UNDERSTAND OTHERS

Comprehension of direct person-to-person communication whether spoken, written, or in sign language or Braille. Includes the resident's ability to process and understand language. Deficits in one's ability to understand (receptive communication deficits) can involve declines in hearing, comprehension (spoken or written) or recognition of facial expressions.

Coding Instructions

- **Code 0, understands:** if the resident clearly comprehends the message(s) and demonstrates comprehension by words or actions/behaviors.
- **Code 1, usually understands:** if the resident misses some part or intent of the message **but** comprehends most of it. The resident may have periodic difficulties integrating information but generally demonstrates comprehension by responding in words or actions.
- **Code 2, sometimes understands:** if the resident demonstrates frequent difficulties integrating information, and responds adequately only to simple and direct questions or instructions. When staff rephrase or simplify the message(s) and/or use gestures, the resident's comprehension is enhanced.
- **Code 3, rarely/never understands:** if the resident demonstrates very limited ability to understand communication. Or, if staff have difficulty determining whether or not the resident comprehends messages, based on verbal and nonverbal responses. Or, the resident can hear sounds but does not understand messages.

B1000: Vision

B1000. Vision	
Enter Code <input type="checkbox"/>	Ability to see in adequate light (with glasses or other visual appliances) 0 Adequate - sees fine detail, including regular print in newspapers/books 1 Impaired - sees large print, but not regular print in newspapers/books 2 Moderately impaired - limited vision; not able to see newspaper headlines but can identify objects 3 Highly impaired - object identification in question, but eyes appear to follow objects 4 Severely impaired - no vision or sees only light, colors or shapes; eyes do not appear to follow objects

Item Rationale

Health-related Quality of Life

- A person's reading vision often diminishes over time.
- If uncorrected, vision impairment can limit the enjoyment of everyday activities such as reading newspapers, books or correspondence, and maintaining and enjoying hobbies and other activities. It also limits the ability to manage personal business, such as reading and signing consent forms.
- Moderate, high or severe impairment can contribute to sensory deprivation, social isolation, and depressed mood.

Planning for Care

- Reversible causes of vision impairment should be sought.
- Consider whether simple environmental changes such as better lighting or magnifiers would improve ability to see.
- Consider large print reading materials for persons with impaired vision.
- For residents with moderate, high, or severe impairment, consider alternative ways of providing access to content of desired reading materials or hobbies.

DEFINITIONS

ADEQUATE LIGHT

Lighting that is sufficient or comfortable for a person with normal vision to see fine detail.

Steps for Assessment

1. Ask direct care staff over all shifts if possible about the resident's usual vision patterns during the 7-day look-back period (e.g., is the resident able to see newsprint, menus, greeting cards?).
2. Then ask the resident about his or her visual abilities.
3. Test the accuracy of your findings:
 - Ensure that the resident's customary visual appliance for close vision is in place (e.g., eyeglasses, magnifying glass).
 - Ensure adequate lighting.

B1000: Vision (cont.)

- Ask the resident to look at regular-size print in a book or newspaper. Then ask the resident to read aloud, starting with larger headlines and ending with the finest, smallest print. If the resident is unable to read a newspaper, provide material with larger print, such as a flyer or large textbook.
- When the resident is unable to read out loud (e.g. due to aphasia, illiteracy), you should test this by another means such as, but not limited to:
 - Substituting numbers or pictures for words that are displayed in the appropriate print size (regular-size print in a book or newspaper)

Coding Instructions

- **Code 0, adequate:** if the resident sees fine detail, including regular print in newspapers/books.
- **Code 1, impaired:** if the resident sees large print, but not regular print in newspapers/books.
- **Code 2, moderately impaired:** if the resident has limited vision and is not able to see newspaper headlines but can identify objects in his or her environment.
- **Code 3, highly impaired:** if the resident's ability to identify objects in his or her environment is in question, but the resident's eye movements appear to be following objects (especially people walking by).
- **Code 4, severely impaired:** if the resident has no vision, sees only light, colors or shapes, or does not appear to follow objects with eyes.

Coding Tips and Special Populations

- Some residents have never learned to read or are unable to read English. In such cases, ask the resident to read numbers, such as dates or page numbers, or to name items in small pictures. Be sure to display this information in two sizes (equivalent to regular and large print).
- If the resident is unable to communicate or follow your directions for testing vision, observe the resident's eye movements to see if his or her eyes seem to follow movement of objects or people. These gross measures of visual acuity may assist you in assessing whether or not the resident has any visual ability. For residents who appear to do this, **code 3, highly impaired.**

B1200: Corrective Lenses

B1200. Corrective Lenses	
Enter Code	Corrective lenses (contacts, glasses, or magnifying glass) used in completing B1000, Vision
<input type="checkbox"/>	0 No
	1 Yes

B1200: Corrective Lenses (cont.)

Item Rationale

Health-related Quality of Life

- Decreased ability to see can limit the enjoyment of everyday activities and can contribute to social isolation and mood and behavior disorders.
- Many residents who do not have corrective lenses could benefit from them, and others have corrective lenses that are not sufficient.
- Many persons who benefit from and own visual aids do not have them on arrival at the nursing home.

Planning for Care

- Knowing if corrective lenses were used when determining ability to see allows better identification of evaluation and management needs.
- Residents with eyeglasses or other visual appliances should be assisted in accessing them. Use and maintenance should be included in care planning.
- Residents who do not have adequate vision without eyeglasses or other visual appliances should be asked about history of corrective lens use.
- Residents who do not have adequate vision, despite using a visual appliance, might benefit from a re-evaluation of the appliance or assessment for new causes of vision impairment.

Steps for Assessment

1. Prior to beginning the assessment, ask the resident whether he or she uses eyeglasses or other vision aids and whether the eyeglasses or vision aids are at the nursing home. Visual aids do not include surgical lens implants.
2. If the resident cannot respond, check with family and care staff about the resident's use of vision aids during the 7-day look-back period.
3. Observe whether the resident used eyeglasses or other vision aids during reading vision test (B1000).
4. Check the medical record for evidence that the resident used corrective lenses when ability to see was recorded.
5. Ask staff and significant others whether the resident was using corrective lenses when they observed the resident's ability to see.

B1200: Corrective Lenses (cont.)

Coding Instructions

- **Code 0, no:** if the resident did not use eyeglasses or other vision aid during the **B1000, Vision** assessment.
- **Code 1, yes:** if corrective lenses or other visual aids were used when visual ability was assessed in completing **B1000, Vision**.