

Τμήμα Επιστήμης Υπολογιστών  
Πανεπιστήμιο Κρήτης

# **Οικονομία και Συνέπεια Θησαυρών Όρων**

Ντοάς Δημήτρης

Μεταπτυχιακή Εργασία

Οκτώβριος 1999



# Οικονομία και Συνέπεια σε Θησαυρούς Όρων

Εργασία που υποβλήθηκε από τον

Δημήτριο Ν. Ντοά

ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΞΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

---

Δημήτριος Ν. Ντοάς

Τμήμα Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

Εισηγητική Επιτροπή:

---

Πάνος Κωνσταντόπουλος, Καθηγητής, Επόπτης

---

Ευάγγελος Μαρκάτος, Καθηγητής, Μέλος

---

Δημήτριος Πλεξουσάκης, Επίκουρος Καθηγητής, Μέλος

Δεκτή:

---

Πάνος Κωνσταντόπουλος  
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Οκτώβριος 1999



# Οικονομία και Συνέπεια Θησαυρών Όρων

Δημήτρης Ν. Ντοάς

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

## Περίληψη

Ένας θησαυρός όρων είναι μια εννοιολογική δομή που περιγράφει έννοιες ενός πεδίου γνώσης με τη χρήση ενός λεξιλογίου όρων και τριών τύπων συσχετίσεων μεταξύ αυτών: συσχετίσεις ισοδυναμίας, ιεραρχικές συσχετίσεις και συσχετίσεις συνάφειας. Υπάρχουν πολλοί θησαυροί όρων που ασχολούνται με το ίδιο πεδίο γνώσης, με αποτέλεσμα να είναι επιθυμητή η συγχώνευσή τους με χρήση σημασιολογικών συσχετίσεων που θα δηλώνουν το κατά πόσο δυο έννοιες των θησαυρών αναπαριστούν την ίδια έννοια. Τόσο η συνένωση θησαυρών, όσο και η κατασκευή από το μηδέν ενός θησαυρού είναι μια πολύ επίπονη διαδικασία, η οποία μπορεί να προκαλέσει δηλώσεις συσχετίσεων που να οδηγήσουν σε ασυνέπεια το θησαυρό.

Επομένως είναι αναγκαία η έρευνα για τον σχεδιασμό και την υλοποίηση ενός συστήματος το οποίο θα ανιχνεύει τις δηλώσεις συσχετίσεων που οδηγούν το θησαυρό σε ασυνεπή κατάσταση. Στην εργασία αυτή, εντοπίσαμε τις σημαντικότερες περιπτώσεις ασυνέπειας σε έναν θησαυρό και προσπαθήσαμε να προτείνουμε τις αντίστοιχες λύσεις. Στην προσπάθειά μας αυτή συνέβαλε και η Περιγραφική Λογική. Η Περιγραφική Λογική παράγει εννοιολογικές γλώσσες οι οποίες είναι αρκετά εκφραστικές και αρκετά κοντά στη φυσική γλώσσα. Επιπλέον η περιγραφική λογική παρέχει ένα τυπικό πλαίσιο μέσα στο οποίο είναι δυνατή η αντιμετώπιση των προβλημάτων που έχουμε περιγράψει. Αν και στην υλοποίηση η απόδοση των συστημάτων περιγραφικής λογικής δεν ήταν η αναμενόμενη, μας βοήθησαν να αναπτύξουμε το θεωρητικό υπόβαθρο, το οποίο θα μπορεί να εφαρμοστεί στην περίπτωση των θησαυρών όρων.

Ιδιαίτερη βαρύτητα δώσαμε στην αντιμετώπιση ασυνεπειών κατά τη παραγωγή σύνθετων όρων. Έπρεπε να θέσουμε περιορισμούς ως προς το ποιοί όροι μπορούν να συνδυαστούν με ποιούς και πώς θα γίνει αυτός ο συνδυασμός. Για το λόγο αυτό ορίσαμε μια σειρά από συσχετίσεις μεταξύ των

κατηγοριών όρων οι οποίες ορίζουν ποιοι όροι μπορούν να συνδυαστούν και με ποιό τρόπο. Για την υλοποίηση ενός τέτοιου συστήματος χρησιμοποιήσαμε το μοντέλο δεδομένων της γλώσσας Telos του συστήματος Σημασιολογικού Ευρετηριασμού (Semantic Index System, SIS) και ένα απλό εννοιολογικό σχήμα. Χρησιμοποιώντας το SIS ως πλατφόρμα ανάπτυξης, υλοποιήσαμε σε C++ τις διαδικασίες σύνθεσης δυναμικών όρων καθώς και κάποιων λειτουργιών σχετικών με αυτούς. Επίσης πραγματοποιήσαμε ένα πείραμα για την αντικατάσταση των μη συνεπών σύνθετων όρων του θησαυρού όρων AAT (Arts & Architecture Thesaurus) με τους αντίστοιχους που δημιουργεί το σύστημά μας.

Επόπτης: Πάνος Κωνσταντόπουλος  
Καθηγητής Τμήμα Επιστήμης Υπολογιστών  
Πανεπιστήμιο Κρήτης

# Economy and Consistency in Thesauri

Dimitris N. Ntoas

Master of Science Thesis

Department of Computer Science

University of Crete

## Abstract

A thesaurus is a conceptual structure representing concepts from a particular domain of discourse using a controlled vocabulary and three types of relationships between concepts: equivalence, hierarchical and associative relationships. There are cases that several thesauri concern the same domain of knowledge, which we would like to merge using relationships denoting the degree to which two terms represent the same concept. The merging of thesauri as well as the construction of a thesaurus is an exceptionally hard and time consuming task, which can result in an inconsistent thesaurus.

Thus, research is necessary concerning the design and implementation of a system which would be able to find any relationship or term declaration that may lead a thesaurus to an inconsistent state. During our research, we traced the most important cases of a thesaurus' inconsistency and tried to propose a solution. In this task we used Description Logic. Description Logic produces conceptual languages that are expressive and quite close to natural language. Further more, Description Logic provides a formal framework for addressing the above issues. Although the performance of systems based on Description Logic wasn't satisfactory, Description Logic helped us build a theoretical background which if adjusted to the framework of thesauri can confront the cases of inconsistency that we have defined.

Maintaining consistency during the formation of compound terms was one of the most important parts of our research. We had to define a set of constraints on whether two terms can be combined and on the way that such a combination can be accomplished. We used the Telos data model of the Semantic Index System for the implementation of such a system. Using SIS as a platform, we have implemented the formation of compound terms and a set of relevant operations in C++. We have conducted an experiment of replacing inconsistent compound terms of the AAT thesaurus with those that our system produced and the results were encouraging.

Supervisor: Panos Constantopoulos  
Professor of Computer Science Department  
University of Crete

## Ευχαριστίες

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τον επόπτη μου κ. Κωνσταντόπουλο που μου εμπιστεύτηκε την εργασία αυτή καθώς και για την καθοδήγησή του. Η συνεργασία μας τα δυο αυτά χρόνια υπήρξε καθοριστική για τον ερευνητικό και επαγγελματικό προσανατολισμό μου.

Επίσης ευχαριστώ τα μέλη της εισηγητικής επιτροπής κ. Ευάγγελο Μαρκάτο και κ. Δημήτρη Πλεξουσάκη για τις εποικοδομητικές τους παρατηρήσεις και διορθώσεις στην εργασία αυτή. Δε θα μπορούσα να παραβλέψω ένα μεγάλο ευχαριστώ στον Dr. Martin Doerr του οποίου η συμβολή στην παρούσα εργασία είναι ανεκτίμητη. Είχαμε μια άψογη συνεργασία και οι συμβουλές που μου παρείχε ήταν ουσιαστικές και συνέβαλαν στην ολοκλήρωση αυτής της εργασίας.

Επιπλέον θα ήθελα να ευχαριστήσω:

Τους φίλους μου, Σέντη Πέτρο, Χατζηνικόλα Κώστα, Νόνα Δημήτρη, Λιάπη Σπύρο, Κούβαρη Κώστα, Χαλκιαδάκη Γιώργο, Ναούμ Νίκο, Μπερτέ Νίκο, Τουρτούνη Σωτήρη, Τσίγγο Δημήτρη και Πετράκη Μανόλη για τις ατελείωτες οινόποσιες οι οποίες με ενέπνευσαν για τη συγγραφή αυτής της εργασίας. Ένα ευχαριστώ αξίζει και για τον Σωτήρη Στογιαννόπουλο, που η διαρκής αρνησή του να νυχτοπερπατήσει μου δημιουργούσε (σπανίως βέβαια) τύψεις και με ανάγκαζε να τον μιμηθώ.

Ακόμη θα ήθελα να ευχαριστήσω την αδελφή μου Σταυρούλα, για την ηθική συμπαράσταση κατά τις δύσκολες στιγμές της συγγραφής της εργασίας αυτής.

Το μεγαλύτερο ευχαριστώ ωστόσο ανήκει στους γονείς μου. Χάρis στις θυσίες τους και τις συμβουλές τους βρίσκομαι σ' αυτό το σημείο σήμερα. Σε δείγμα ελάχιστης ευγνωμοσύνης τους αφιερώνω την παρούσα εργασία και ελπίζω στο μέλλον να φανώ αντάξιος των προσδοκιών τους.

**Δημήτρης Ν.  
Ντοάς**



# ΠΕΡΙΕΧΟΜΕΝΑ

<b>ΕΙΣΑΓΩΓΗ.....</b>	<b>14</b>
1.1 Πεδίο εφαρμογής.....	14
1.2 Ορισμός του προβλήματος.....	16
<b>ΘΗΣΑΥΡΟΙ ΟΡΩΝ – ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ .....</b>	<b>18</b>
2.1 Όροι.....	18
<i>Σύνθετοι όροι</i> .....	18
2.2 Σημασιολογικές Συσχετίσεις Μονόγλωσσου Θησαυρού.....	19
2.2.1 Συσχέτιση ισοδυναμίας .....	20
2.2.2. Συσχέτιση Ιεραρχίας.....	21
2.2.3 Συσχέτιση συνάφειας .....	22
2.3 Σημασιολογικές συσχετίσεις Πολύγλωσσου Θησαυρού .....	23
<b>ΤΟ ΠΡΟΒΛΗΜΑ .....</b>	<b>25</b>
3.1 Οικονομία και Συνέπεια σε Θησαυρούς Όρων.....	25
3.1.1 Εισαγωγή Περιττών Συσχετίσεων .....	26
3.1.2 Κυκλικές Συσχετίσεις .....	27
3.1.3 Εισαγωγή Συσχετίσεων που οδηγούν σε Ασυνέπεια.....	27
3.1.4 Συσχετίσεις με Δυναμικούς Όρους.....	28
3.2 Συστήματα που αντιμετωπίζουν παρόμοια προβλήματα .....	29
<b>Περιγραφική Λογική .....</b>	<b>30</b>
4.1 Περιγραφή Μοντέλου .....	30
4.1.1 Αντικείμενα.....	30
4.1.2 Ρόλοι.....	31
4.2 Το Γενικό Πλαίσιο της Περιγραφικής Λογικής.....	31
<i>TBox-Εντασιακή (Intensional) γνώση</i> .....	31
<i>Αρχικές έννοιες (primitive concepts)</i> .....	31
<i>Παράγωγες Έννοιες</i> .....	32
<i>Abox – Εκτασιακή (Extensional) γνώση</i> .....	34
4.3 Ερμηνεία .....	35
<i>Βάση Γνώσης</i> .....	35
<i>Ερμηνεία</i> .....	36
4.4 Σχέση Υπαγωγής .....	36
4.5 Υποθέσεις της Περιγραφικής Λογικής.....	36
4.6 Σύνταξη και σημασιολογία των εννοιολογικών γλωσσών.....	38
4.7 Συλλογισμοί της Περιγραφικής Λογικής .....	40
4.8 Εφαρμογή των ελέγχων στους Θησαυρούς .....	42
<b>Ερμηνεία των στοιχείων ενός Θησαυρού Όρων σε Περιγραφική Λογική.....</b>	<b>43</b>
5.1 Θησαυροί Όρων και Περιγραφική Λογική .....	43
5.1.1 Ερμηνεία της συσχέτισης ιεραρχίας σε Περιγραφική Λογική.....	44
5.1.2 Ερμηνεία των σχέσεων ισοδυναμίας σε Περιγραφική Λογική.....	44
5.1.3 Μεταφορά του Θησαυρού σε Περιγραφική Λογική .....	45
<i>NeoClassic</i> .....	46
<i>FaCT</i> .....	47
<i>Άλλα συστήματα</i> .....	49
<i>Συμπεράσματα</i> .....	50
<i>Galen</i> .....	50
<b>Δυναμικοί Όροι.....</b>	<b>52</b>
6.1 Δημιουργία σύνθετων όρων με χρήση συσχετίσεων.....	52
6.2 Η χρήση λογικών τελεστών στη δημιουργία δυναμικών όρων.....	54

6.3 Άμεση υπερκλάση .....	57
6.4 Λειτουργίες του συστήματος .....	58
6.5 Αναπαράσταση των δυναμικών όρων στο σύστημα .....	59
6.6 Σχεδιασμός Διεργασιών .....	60
6.6.1 Σύνθεση Δυναμικού Όρου .....	60
6.6.2 Εύρεση άμεσης υπερκλάσης/υποκλάσης .....	62
6.6.2 Δημιουργία Ιεραρχίας Δυναμικών Όρων .....	63
6.7 Αποτελέσματα .....	64
<b>Επίλογος.....</b>	<b>65</b>
7.1 Συμπεράσματα.....	65
7.2 Μελλοντικές κατευθύνσεις .....	66

---

# ΚΕΦΑΛΑΙΟ 1ο

## ΕΙΣΑΓΩΓΗ

Κάθε πεδίο γνώσης έχει ένα *λεξιλόγιο* (vocabulary), ένα σύνολο λέξεων ή φράσεων της φυσικής γλώσσας, που χρησιμοποιείται για να περιγράψει *έννοιες* (concepts) – άτομα, υλικά, αφηρημένες οντότητες – που σχετίζονται με το συγκεκριμένο πεδίο γνώσης. Υπάρχουν διάφοροι ορισμοί σχετικά με το τι είναι έννοια. Στην περίπτωση μας υιοθετούμε την άποψη ότι μια έννοια είναι μια αναφορά με την οποία κάποιοι άνθρωποι συμφωνούν να κατηγοριοποιήσουν ένα σύνολο αντικειμένων του πραγματικού κόσμου που έχουν τις ίδιες ιδιότητες αλλά χωρίς απαραίτητα να ορίζουν ποιες είναι αυτές [DF98a], ενώ όρος είναι μια φράση με την οποία αναπαρίσταται μια έννοια. Ένα λεξιλόγιο λέγεται ότι είναι *ελεγχόμενο* (controlled) όταν είναι ένα κλειστό, εγκεκριμένο (authorized) προς χρήση, σύνολο όρων ενός πεδίου γνώσης. Στην περίπτωση ενός ελεγχόμενου λεξιλογίου, οι όροι συχνά διακρίνονται σε *δόκιμους* (preferred terms) και *αδόκιμους* (non-preferred terms). Οι δόκιμοι όροι είναι εγκεκριμένοι προς χρήση, ενώ οι αδόκιμοι είναι εναλλακτικοί όροι που περιγράφουν την ίδια έννοια με κάποιο δόκιμο όρο. Οι όροι και οι έννοιες συνδέονται μεταξύ τους με συσχετίσεις που ονομάζονται *σημασιολογικές συσχετίσεις*.

Ένας *θησαυρός όρων* (thesaurus) είναι ένα ελεγχόμενο λεξιλόγιο ενός πεδίου γνώσης, επαυξημένο με ρητές σημασιολογικές συσχετίσεις μεταξύ των εννοιών που αυτό περιγράφει [ISO86]. Αν το λεξιλόγιο ενός θησαυρού αποτελείται από όρους που χρησιμοποιούνται αποκλειστικά στο πλαίσιο μιας συγκεκριμένης φυσικής γλώσσας, εξαιρουμένων φυσικά αυτών που είναι δανεισμένοι από άλλες γλώσσες, τότε ονομάζεται *μονόγλωσσος* (monolingual thesaurus) ενώ στην αντίθετη περίπτωση ονομάζεται *πολύγλωσσος* (multilingual thesaurus).

### 1.1 Πεδίο εφαρμογής

Οι θησαυροί χρησιμοποιούνται κυρίως για την καλύτερη πρόσβαση σε δεδομένα (κείμενα, εικόνες, [Kris93], [Spin94], [Spin96], [BMT96], [Smea96]) που μπορεί να υπάρχουν σε πολλές ετερογενείς

και πολυγλωσσικές πηγές πληροφορίας. Σ' αυτά τα πλαίσια χρήσης, τέσσερις είναι οι βασικές απαιτήσεις από ένα θησαυρό.

- Η παροχή ενός πρότυπου λεξιλογίου για ένα συγκεκριμένο πεδίο γνώσης
- Η δημιουργία αμφιμονοσήμαντων αντιστοιχίσεων μεταξύ των όρων του λεξιλογίου και των εννοιών του πεδίου γνώσης, κάτι που δε συμβαίνει σ' ένα λεξιλόγιο φυσικής γλώσσας.
- Ο ρητός ορισμός των συσχετίσεων μεταξύ των εννοιών που αναπαριστούν οι όροι του θησαυρού προκειμένου να βελτιωθεί η απόδοση του ευρετηριασμού και της ανάκλησης πληροφορία μέσω του θησαυρού.

Από την πλευρά της ανάκλησης πληροφορίας, η επίτευξη των παραπάνω στόχων είναι εξαιρετικά σημαντική. Το βασικό πρόβλημα των περισσότερων τεχνικών ανάκλησης πληροφορίας είναι η μη ικανοποιητική απόδοση εκφραζόμενη σε όρους *βαθμού ανάκλησης* (recall) και *βαθμού ακρίβειας* (precision) [Pai91],[Sal89],[KS93]. Ο βαθμός ανάκλησης εκφράζει το ποσοστό σχετικής με μια ερώτηση πληροφορίας το οποίο ανακλήθηκε από την ερώτηση, ενώ ο βαθμός ακρίβειας εκφράζει το ποσοστό της πληροφορίας που ανακλήθηκε από μια ερώτηση το οποίο είναι σχετικό με αυτήν. Η χαμηλή απόδοση των συστημάτων ανάκλησης πληροφορίας οφείλεται κυρίως στην ύπαρξη ομώνυμων και συνώνυμων όρων σε μια φυσική γλώσσα και στην έλλειψη συσχετίσεων μεταξύ τους. Έτσι μια ερώτηση για τα “Αγγεία” μπορεί να επιστρέψει πληροφορία σχετικά τόσο με τα αντικείμενα “Αγγεία” όσο και με τα “Αγγεία” του ανθρώπινου σώματος, ενώ μια ερώτηση για “Αεροσκάφη” μπορεί να μην ανακαλέσει πληροφορία σχετική με “Αεροπλάνα”, μειώνοντας έτσι το βαθμό ανάκλησης. Αν ωστόσο είχαμε δηλώσει με μια συσχέτιση την ισοδυναμία των όρων “Αεροσκάφη” και “Αεροπλάνα”, τότε δε θα υπήρχε μειωμένος βαθμός ανάκλησης. Επιπλέον, με τη χρήση ιεραρχικών συσχετίσεων βελτιώνεται η απόδοση των συστημάτων ανάκλησης πληροφορίας. Για παράδειγμα μια ερώτηση για “Οχήματα”, θ' ανακαλέσει μεταξύ άλλων και πληροφορία για “Αυτοκίνητα” κάτι που δε γίνεται στα συστήματα ανάκλησης πληροφορίας που δεν υποστηρίζουν θησαυρούς. Πειραματικά αποτελέσματα δείχνουν πως η χρήση θησαυρών σε συστήματα ανάκλησης πληροφορίας βελτιώνουν σημαντικά την απόδοσή τους [Kri94],[LKL94]. Η χρήση θησαυρών υψηλής ποιότητας είναι ιδιαίτερα χρήσιμη για την κατηγοριοποίηση μουσειακών αντικειμένων, εικόνων και μουσικής [Smea96]. Η χρησιμότητα των θησαυρών επιβεβαιώθηκε και σε πρόσφατες μελέτες [Spink96],[BMT96]. Εκτός από την επιβεβαίωση μέσω ερευνών, συστήματα και προϊόντα ανάκτησης πληροφορίας χρησιμοποιούν θησαυρούς όρων (MISTRAL-x).

## 1.2 Ορισμός του προβλήματος

Στην εργασία αυτή ασχολούμαστε με το πρόβλημα της οικονομίας και συνέπειας θησαυρών, ένα θέμα το οποίο δεν έχει απασχολήσει σε ευρεία κλίμακα τους ερευνητές/επιστήμονες πληροφορικής, γεγονός που ίσως να οφείλεται στο ότι σχετικά πρόσφατα έχει αναγνωριστεί η αξία των θησαυρών στην ανάκληση πληροφορίας ([CLBD93], [Pai91], [LKL94], [Kri94]) και έχει αρχίσει η ευρεία εφαρμογή τους ([MSH],[AAT],[MERIMEE],RCHME,[UMLS],[DDC] κ.α.). Επιπλέον έχουν σημειωθεί προσπάθειες να ενοποιηθούν θησαυροί που έχουν κατασκευαστεί ανεξάρτητα, με σκοπό να κατασκευαστεί ένας νέος θησαυρός, ο οποίος διαισθητικά θ' αποτελεί την "ένωση" των αρχικών θησαυρών. Ωστόσο, τόσο η κατασκευή ενός θησαυρού από το μηδέν όσο και η ενοποίηση θησαυρών είναι μια πολύ δύσκολη εργασία ακόμη και για θησαυρούς μέσου μεγέθους [ISO86], [ISO85].

Τα πρότυπα [ISO86] και [ISO85] παρέχουν ένα σύνολο από τύπους κατασκευών (όρους και τύπους συσχετίσεων) και κανόνες για το συνδυασμό τους με σκοπό την ανάπτυξη θησαυρών. Ωστόσο, τόσο κατά τη ένωση θησαυρών όρων όσο και κατά τη δημιουργία τους είναι πολύ πιθανόν να δηλωθούν συσχετίσεις που να οδηγήσουν είτε σε έλλειψη συνέπειας είτε σε έλλειψη οικονομίας. Έλλειψη συνέπειας έχουμε όταν μετά από την εισαγωγή μιας σειράς συσχετίσεων, δεν μπορούμε να βγάλουμε συμπεράσματα για τη σχετική θέση δυο ή περισσότερων όρων σε μια ιεραρχία. Έλλειψη οικονομίας έχουμε όταν εισάγονται συσχετίσεις, οι οποίες προκύπτουν από τις ήδη υπάρχουσες.

Στα πλαίσια αυτής της εργασίας, εντοπίζουμε αρχικά τις δηλώσεις συσχετίσεων που οδηγούν σε ασυνέπεια και έλλειψη οικονομίας. Στη συνέχεια, μελετάμε πώς άλλα συστήματα αντιμετωπίζουν παρόμοια προβλήματα και προτείνουμε τρόπους αντιμετώπισης των προβλημάτων που έχουμε θέσει.

Στις περιπτώσεις που υλοποιήσαμε τις προτάσεις μας χρησιμοποιήσαμε το Σύστημα Σημασιολογικού Ευρετηριασμού (Semantic Index System, SIS) και το μοντέλο δεδομένων της γλώσσας Telos[MBJK90] το οποίο υποστηρίζεται από το SIS [DKT95]. Οι λόγοι για τους οποίους επιλέξαμε το SIS είναι οι εξής:

- Η παράσταση θησαυρών με μορφή δικτύου είναι πιο φυσική
- Διαθέτει ένα εύχρηστο Application Programming Interface για την επικοινωνία μέσα από εφαρμογές με τη βάση γνώσης.
- Το SIS ξεπερνά την ταχύτητα απόκρισης γνωστών εμπορικών DBMS σε αναδρομικές ερωτήσεις κατά μια έως δυο τάξεις μεγέθους
- Το SIS ικανοποιεί όλες τις προδιαγραφές [Mil91] ώστε να χρησιμοποιηθεί για τη διαχείριση θησαυρών
- Σε αρκετές περιπτώσεις οι εμπειρίες από τη χρήση άλλων συστημάτων σε σχετικές εργασίες, όπως του συστήματος Ingres στην εργασία [CLBD93] δεν ήταν ενθαρρυντικές.

Τα κυριότερα χαρακτηριστικά του SIS-Telos είναι: η ιεραρχία ταξινόμησης, η οποία δεν έχει περιορισμούς ως προς το μεγεθός της (εκτεταμένη ιεραρχία σχημάτων και μετασχημάτων), η πολλαπλή ταξινόμηση (ένα αντικείμενο μπορεί να ανήκει σε περισσότερες από μια κλάσεις), πολλαπλή κληρονομικότητα, καθώς και προαιρετικά ή πλειότιμα γνωρίσματα τα οποία μπορούν επίσης να έχουν τα δικά τους γνωρίσματα.

---

## ΚΕΦΑΛΑΙΟ 2ο

### ΘΗΣΑΥΡΟΙ ΟΡΩΝ – ΓΕΝΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ

Τα βασικά στοιχεία ενός θησαυρού όρων είναι οι όροι και οι συσχετίσεις μεταξύ των εννοιών που αντιπροσωπεύουν οι όροι. Στην ενότητα αυτή θα παρουσιάσουμε τα βασικά χαρακτηριστικά των θησαυρών όρων, δίνοντας ιδιαίτερη έμφαση σε όσα σχετίζονται άμεσα με την παρούσα εργασία.

#### 2.1 Όροι

Η βασική έννοια ενός θησαυρού σύμφωνα με το [ISO86] είναι ο *Όρος*. Όρος είναι η αναπαράσταση μιας έννοιας. Συχνά οι όροι ενός θησαυρού οργανώνονται σε πολύ γενικές, εννοιολογικά συγγενείς κλάσεις εννοιών τις οποίες ονομάζουμε *κατηγορίες όρων (facets)*. Κάθε όρος του θησαυρού πρέπει να ανήκει τουλάχιστον σε μια κατηγορία. Ορίζοντας τις κατηγορίες στις οποίες ένας όρος ανήκει, δίνουμε πληροφορία για την υπόσταση της έννοιας την οποία ο όρος αναπαριστά. Οι όροι που ανήκουν σε μια κατηγορία, χωρίζονται σε διαφορετικές *ιεραρχίες (hierarchies)*. Δυο όροι για να ανήκουν στην ίδια ιεραρχία θα πρέπει είτε να έχουν κάποιο όρο από τον οποίο να έχουν και οι δυο ειδικότερη έννοια είτε ο ένας να είναι ειδικότερος του άλλου. Όπως είναι φυσικό μια κατηγορία όρων μπορεί να αποτελείται από πολλές ιεραρχίες.

Όπως έχουμε προαναφέρει, οι όροι διακρίνονται σε δόκιμους και αδόκιμους. Από τους δόκιμους όρους επιλέγεται ένας ο οποίος ονομάζεται *περιγραφέας (Descriptor)*, ο οποίος θεωρούμε ότι περιγράφει μοναδικά μια έννοια στο θησαυρό.

#### *Σύνθετοι όροι*

Όροι οι οποίοι αποτελούνται από περισσότερες από μια λέξεις, λέγονται *σύνθετοι όροι (compound terms)* σε αντίθεση με τους *απλούς όρους* οι οποίοι αποτελούνται από μια μονάχα λέξη. Ένα σημαντικό θέμα που εισάγουν οι σύνθετοι όροι είναι η σύνταξή τους, η διάταξη δηλαδή των λέξεων

που αποτελούν τον όρο [Sve89]. Υπάρχουν δυο προσεγγίσεις στο θέμα της σύνταξης των σύνθετων όρων.

- Σύνταξη φυσικής γλώσσας, για παράδειγμα “Information Systems”.
- Αντίστροφη σύνταξη, για παράδειγμα “Systems, Information”.

Αν και η πρώτη προσέγγιση συνιστάται [ISO86], μπορεί να χρησιμοποιείται και η δεύτερη με το επιχείρημα ότι σε μια αλφαβητικά ταξινομημένη παράθεση όρων του θησαυρού σχετικοί όροι βρίσκονται ταξινομημένοι. Για παράδειγμα:

...  
“Systems, Distributed”  
“Systems, Information”  
...

Εκτός από τη σειρά παράθεσης των λέξεων ενός σύνθετου όρου, η σύνταξή τους επηρεάζεται και από τη συντακτική κατηγορία τους. Σύμφωνα με το [ISO86] ένας όρος μπορεί να είναι ουσιαστικό, ρήμα, επίρρημα, επίθετο, επιθετική φράση ή προθετική φράση. Προφανώς οι σύνθετοι όροι μπορούν να ανήκουν μόνο στις δυο τελευταίες κατηγορίες. Ωστόσο η επιλογή της συντακτικής κατηγορίας με την οποία θα αποδοθεί ο όρος επηρεάζει και τη σύνταξή του. Για παράδειγμα η διαχείριση πληροφοριακών συστημάτων μπορεί να αποδοθεί στην Αγγλική είτε από τον όρο “Information Systems Management” είτε από τον όρο “Management of Information Systems”.

## 2.2 Σημασιολογικές Συσχετίσεις Μονόγλωσσου Θησαυρού

Όπως αναφέρθηκε, ένας θησαυρός πρέπει να ορίζει ρητά τις σημασιολογικές συσχετίσεις μεταξύ των όρων του. Σύμφωνα τόσο με το πρότυπο για την οργάνωση θησαυρών του ISO [ISO86], όσο και με άλλες εργασίες [CLBD93], [MR88], [Mil91], [Sve89], [Pai91] υπάρχουν τρεις τύποι συσχετίσεων: η *συσχέτιση ισοδυναμίας (equivalence relationship)*, η *ιεραρχική συσχέτιση (hierarchical relationship)* και η *συσχέτιση συνάφειας (associative relationship)*.

### 2.2.1 Συσχέτιση ισοδυναμίας

Η συσχέτιση ισοδυναμίας συνδέει δυο όρους όταν αναπαριστούν την ίδια έννοια. Από ένα σύνολο ισοδύναμων όρων, ένας επιλέγεται ως αντιπρόσωπος του συνόλου και χαρακτηρίζεται ως δόκιμος όρος (περιγραφέας), ενώ οι υπόλοιποι ως αδόκιμοι. Η συσχέτιση ισοδυναμίας συνδέει πάντα έναν αδόκιμο όρο και έναν ή περισσότερους δόκιμους όρους, ανάλογα με τον τύπο της. Υπάρχουν τρεις τύποι συσχετίσεων ισοδυναμίας.

- Συσχετίσεις που συνδέουν έναν αδόκιμο και τον ισοδύναμο δόκιμο όρο. Για παράδειγμα:  
atmospheric pollution *USE* air pollution.
- Συσχετίσεις ισοδυναμίας που συνδέουν έναν αδόκιμο όρο και ένα σύνολο ισοδύναμων δόκιμων όρων. Για παράδειγμα:  
lifts *USE* ELEVATORS or LIFTING EQUIPMENT.
- Συσχετίσεις ισοδυναμίας που συνδέουν ένα σύνθετο αδόκιμο όρο με τα συνθετικά του που πρέπει να είναι δόκιμοι όροι. Για παράδειγμα:  
Aircraft engines *USE* AIRCRAFTS and ENGINES.

Αυτή η μορφή συσχέτισης ισοδυναμίας, ονομάζεται συντακτική παραγοντοποίηση (syntactic factoring).

Οι συσχετίσεις ισοδυναμίας είναι ο μοναδικός τύπος συσχέτισης στον οποίο μπορεί να λάβει μέρος ένας αδόκιμος όρος. Από την πλευρά του αδόκιμου όρου η συσχέτιση συμβολίζεται με το σύμβολο *USE*, ενώ από την πλευρά του δόκιμου όρου με το σύμβολο *UF* (Used For).

#### ΠΑΡΑΔΕΙΓΜΑ 2.1

ROCKS

UF basalt  
granite  
slate

basalt

USE ROCKS

granite

USE ROCKS

slate

Θα πρέπει ωστόσο να παρατηρήσουμε ότι αν και πολλές φορές θεωρούμε ότι δυο όροι είναι ισοδύναμοι (π.χ. συνώνυμοι), στην πραγματικότητα δεν αναπαριστούν την ίδια ακριβώς έννοια [Sve89]. Γιαυτό το λόγο το [ISO86] θεωρεί ότι μια συσχέτιση ισοδυναμίας μπορεί μεταξύ άλλων να συνδέει:

- Τον όρο που περιγράφει μια κλάση αντικειμένων και τα μέλη αυτής (παράδειγμα 2.1).
- Δυο όρους που εκφράζουν έννοιες οι οποίες είναι διαφορετικές τιμές σε μια κλίμακα. Για παράδειγμα οι όροι “Wetness” και “Dryness” συνδέονται με συσχέτιση ισοδυναμίας. Σε αυτή την περίπτωση η συσχέτιση ισοδυναμίας έχει την ερμηνεία ότι οι δυο όροι αναφέρονται στην ίδια κλίμακα και όχι ότι αναπαριστούν την ίδια έννοια.
- Διαφορετικές μορφές γραφής ενός όρου. Για παράδειγμα “Roumania” και “Romania”.
- Επίσημα και κοινά ονόματα. Για παράδειγμα “Polyethylene” και “Polythene”.

### 2.2.2. Συσχέτιση Ιεραρχίας

Οι συσχετίσεις ιεραρχίας είναι το χαρακτηριστικό που ξεχωρίζει ένα θησαυρό από ένα λεξικό ή μια λίστα όρων [Sve89] και ο βασικός μηχανισμός δόμησης θησαυρών [ISO86]. Δυο όροι βρίσκονται σε ιεραρχική συσχέτιση όταν ο ένας από αυτούς, ο *γενικότερος όρος* (*broader term*) αντιπροσωπεύει μια κλάση αντικειμένων, ενώ ο δεύτερος, ο *ειδικότερος όρος* (*narrower term*) αντιπροσωπεύει ένα μέλος ή υποσύνολο ή τμήμα του γενικότερου όρου. Γενικότερα, δυο όροι συνδέονται με μια ιεραρχική συσχέτιση όταν ο ένας υπάγεται στον άλλο κατά μια έννοια. Η συσχέτιση ιεραρχίας υποδηλώνεται από την πλευρά του ειδικότερου όρου με το σύμβολο *BT*, ενώ από την πλευρά του γενικότερου όρου με το σύμβολο *NT*. Οι ιεραρχικές συσχετίσεις διακρίνονται σε τρεις τύπους.

- *Ιεραρχική συσχέτιση γενίκευσης* (*generic*) η οποία συνδέει ένα όρο που περιγράφει ένα σύνολο οντοτήτων και ένα όρο που περιγράφει ένα υπερσύνολο αυτού του συνόλου και υποδεικνύεται από τα σύμβολα *BTG/NTG* (*Broader/Narrower Term Generic*). Ουσιαστικά δηλαδή η συσχέτιση αυτή έχει την έννοια συνόλου-υποσυνόλου. Παράδειγμα:

Cars **BTG** Vehicles ή Vehicles **NTG** Cars

- *Ιεραρχική συσχέτιση μέρους-όλου (part-whole)*, η οποία συνδέει τον όρο που περιγράφει ένα τμήμα μιας οντότητας και τον όρο που περιγράφει την οντότητα αυτή, η οποία υποδεικνύεται από τα σύμβολα *BTP/NTP* (Broader/Narrower Term Partitive).  
Παράδειγμα:

Central nervous system **BTP** Nervous system ή  
Nervous system **NTP** Nervous system

- *Ιεραρχική συσχέτιση περίπτωσης (instance of)* η οποία συνδέει τον όρο που περιγράφει ένα σύνολο οντοτήτων και τον όρο που περιγράφει μια οντότητα αυτού του συνόλου. Για παράδειγμα

Alps **BT** Mountain regions.

Στους περισσότερους θησαυρούς οι ιεραρχικές συσχετίσεις έχουν τη σημασιολογία της ιεραρχικής συσχέτισης γενίκευσης, αν και μπορεί να υποδηλώνονται με τα σύμβολα BT/NT. Για το λόγο αυτό στα επόμενα θα αναφερόμαστε σε μια ιεραρχική συσχέτιση με τον όρο συσχέτιση γενίκευσης/ειδίκευσης ή απλά συσχέτιση γενίκευσης. Η σχέση γενίκευσης κατέχει τη μεταβατική και την αντισυμμετρική ιδιότητα. Κατά συνέπεια, σχηματίζει μια ιεραρχία η οποία μπορεί να παρασταθεί μ' έναν ακυκλικό κατευθυνόμενο γράφο (Directed Acyclic Graph, DAG).

### 2.2.3 Συσχέτιση συνάφειας

Δυο όροι συνδέονται με μια συσχέτιση συνάφειας όταν στην καθημερινή τους χρήση συνδέονται με μια συσχέτιση η οποία ωστόσο δεν είναι ούτε συσχέτιση ισοδυναμίας ούτε συσχέτιση ιεραρχίας. Η σχέση συνάφειας σε ένα θησαυρό είναι κατά συνέπεια η ένωση ενός συνόλου σχέσεων. Για το λόγο αυτό η σχέση συνάφειας κατέχει μονάχα την ιδιότητα της συμμετρίας. Για παράδειγμα οι όροι “TEMPERATURE CONTROL” και “THERMOSTATS” είναι συναφείς κατά την έννοια ότι ο έλεγχος της θερμοκρασίας εκτελείται από θερμοστάτες και οι θερμοστάτες εκτελούν τον έλεγχο θερμοκρασίας. Η συσχέτιση συνάφειας εδώ εκφράζει την ύπαρξη της συσχέτισης “εκτελεί και εκτελείται από” μεταξύ των δυο όρων. Η συσχέτιση συνάφειας μεταξύ των δυο όρων υποδηλώνεται με το σύμβολο RT (Related Term).

### ΠΑΡΑΔΕΙΓΜΑ 2.2

TEMPERATURE CONTROL

RT THERMOSTATS

THERMOSTATS

RT TEMPERATURE CONTROL

## 2.3 Σημασιολογικές συσχετίσεις Πολύγλωσσου Θησαυρού

Όπως επιβεβαιώνουν τα [ISO85],[ISO86] για την κατασκευή θησαυρών, οι όροι που επιλέγονται από διαφορετικές γλώσσες ποικίλουν ως προς το κατά πόσο αναπαριστούν την ίδια έννοια. Συνεπώς είναι αναγκαίο να ορισθούν συσχετίσεις που θα περιγράφουν το κατά πόσο δυο όροι σε διαφορετικές γλώσσες περιγράφουν την ίδια έννοια. Αυτή την ανάγκη προσπάθησε να καλύψει το [ISO86], ορίζοντας τις ακόλουθες συσχετίσεις μεταξύ όρων που προέρχονται από διαφορετικές φυσικές γλώσσες:

**Ακριβής Ισοδυναμία:** Η συσχέτιση αυτή συνδέει δυο όρους οι οποίοι έχουν ακριβώς το ίδιο νόημα, δηλαδή αναπαριστούν την ίδια ακριβώς έννοια. Αυτό σημαίνει ότι τα αντικείμενα που είναι ταξινομημένα στον ένα όρο είναι ταυτόχρονα ταξινομημένα και στον άλλο και αντίστροφα. Για παράδειγμα, σύμφωνα με το θησαυρό MERIMEE, ο γαλλικός όρος “menhir” συνδέεται μέσω της σχέσης exact equivalence με τον αγγλικό όρο “standing stone”.

**Μερική Ισοδυναμία:** Η συσχέτιση αυτή συνδέει δυο όρους που δεν είναι δυνατόν να συνδεθούν με συσχέτιση exact equivalence, διότι ο ένας έχει είτε ευρύτερη είτε στενότερη έννοια από τον άλλο. Ωστόσο η συσχέτιση αυτή έχει το μειονέκτημα ότι δεν ξεχωρίζει την περίπτωση που ο όρος έχει ευρύτερη σημασία από τον άλλο και την περίπτωση που έχει ευρύτερη σημασία. Αυτό σημαίνει ότι δεν μπορούμε να χρησιμοποιήσουμε αυτή τη συσχέτιση στην ανάκτηση πληροφορίας, αφού δεν γνωρίζουμε αν η απάντηση που θα πάρουμε θα αφορά ένα ευρύτερο ή στενότερο σύνολο από την αρχική. Γιαυτό το λόγο στο [DF98b], προτείνεται να διασπαστεί η συσχέτιση αυτή σε δυο άλλες: την narrower equivalence και την broader equivalence. Αν ο Όρος<sup>1</sup> συνδέεται με τον Όρος<sup>2</sup> μέσω της narrower equivalence, σημαίνει ότι ο δεύτερος έχει στενότερη ερμηνεία από τον πρώτο. Το αντίθετο συμβαίνει όταν συνδέονται με τη συσχέτιση broader equivalence. Για παράδειγμα ο γαλλικός όρος

“cromlech” συνδέεται με τη συσχέτιση narrower equivalence με τον αγγλικό όρο “cromlechs” ενώ με τον όρο “funerary buildings” συνδέεται με τη συσχέτιση broader equivalence.

**Ισοδυναμία ενός προς πολλά:** Η συσχέτιση αυτή χρησιμοποιείται όταν ένας όρος σε μια γλώσσα αναπαριστά μια έννοια η οποία στην άλλη γλώσσα είναι δυνατόν να αναπαρασταθεί από το συνδυασμό δυο ή περισσότερων δόκιμων όρων. Ο συνδυασμός αυτών των όρων θεωρείται ως μια *δυναμική έννοια*, την οποία συχνότερα ονομάζουμε σύνθετη έννοια. Ωστόσο και αυτή η συσχέτιση έχει δυο διαφορετικές ερμηνείες, καθώς στο θησαυρό MERIMEE ο συνδυασμός των όρων σε άλλες περιπτώσεις είναι η ένωσή τους και σε άλλες η τομή τους. Πιο συγκεκριμένα, ο γαλλικός όρος “roche ornee” είναι ισοδύναμος με την ένωση των αγγλικών όρων του RCHME “tune stone” και “rock carving”, ενώ ο όρος “academie” είναι ισοδύναμος με την τομή των όρων του AAT “academies” και “buildings”.

**Μη ακριβής Ισοδυναμία :** Η συσχέτιση αυτή συνδέει δυο όρους των οποίων οι έννοιες δεν είναι ίδιες ούτε όμως περιέχεται η μια στην άλλη. Ουσιαστικά αυτό σημαίνει ότι η τομή των δυο εννοιών δεν είναι το κενό.

**Μη Ισοδυναμία:** Η συσχέτιση αυτή συνδέει δυο όρους που δε σχετίζονται με κανέναν από τους παραπάνω τρόπους. Ουσιαστικά αυτή η συσχέτιση δε χρησιμοποιείται αφού εννοείται, όταν δυο όροι δεν έχουν καμία άλλη συσχέτιση ισοδυναμίας.

Έχοντας παρουσιάσει τα κυριότερα χαρακτηριστικά ενός θησαυρού όρων, μπορούμε τώρα να παρουσιάσουμε το θέμα με το οποίο ασχολείται η παρούσα εργασία μας.

---

## Κεφάλαιο 3

### ΤΟ ΠΡΟΒΛΗΜΑ

#### 3.1 Οικονομία και Συνέπεια σε Θησαυρούς Όρων

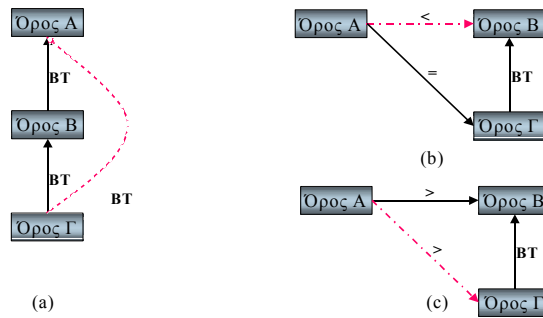
Όπως έχουμε ήδη αναφέρει, με τη φράση οικονομία σε θησαυρούς όρων εννοούμε τον εντοπισμό των περιττών συσχετίσεων, δηλαδή των συσχετίσεων αυτών που προκύπτουν από άλλες. Η συνέπεια θησαυρών όρων είναι να μην επιτρέψουμε δηλώσεις συσχετίσεων και όρων που μπορούν να οδηγήσουν το θησαυρό σε μη συνεπή κατάσταση, δηλαδή σε μια κατάσταση όπου δεν μπορούμε πλέον να ερμηνεύσουμε πλήρως όλες τις συσχετίσεις του θησαυρού και να μην μπορούμε να συμπεράνουμε τη σχετική θέση στην ιεραρχία δυο ή περισσότερων όρων. Αν για παράδειγμα, από τις συσχετίσεις που έχουν δηλωθεί προκύπτει ότι ο όρος A είναι ευρύτερος του όρου B και ταυτόχρονα ότι ο όρος B είναι ευρύτερος του όρου A, έχουμε έλλειψη συνέπειας. Στη συνέχεια προσπαθούμε να κατηγοριοποιήσουμε αυτές τις περιπτώσεις και να δώσουμε παραδείγματα από κάθε μια από τις κατηγορίες, ώστε να προσεγγίσουμε τον τρόπο με τον οποίο αυτές προκύπτουν.

Οι γενικές κατηγορίες στις οποίες μπορούμε να κατατάξουμε τις περιπτώσεις που θα μας απασχολήσουν είναι οι εξής:

- Εισαγωγή περιττών συσχετίσεων
- Δημιουργία κύκλων κατά την εισαγωγή συσχετίσεων μεταξύ όρων.
- Εισαγωγή συσχετίσεων που οδηγούν σε ασυνέπεια το θησαυρό.
- Συσχέτιση ενός όρου μιας γλώσσας με έναν δυναμικό σε μια άλλη.

### 3.1.1 Εισαγωγή Περιττών Συσχετίσεων

Περιττές θεωρούμε τις συσχετίσεις εκείνες που μας οδηγούν σε συμπεράσματα για το θησαυρό, τα οποία θα μπορούσαμε να τα έχουμε εξαγάγει και χωρίς αυτές.

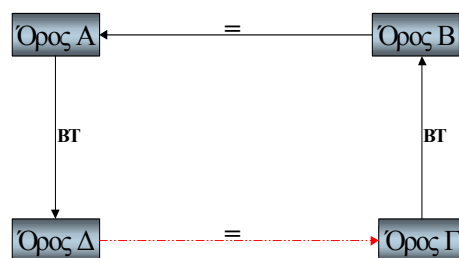


Εικόνα 3.1: Παραδείγματα περιττών συσχετίσεων

Στην εικόνα 3.1 παρουσιάζονται οι περιπτώσεις όπου έχουμε περιττές συσχετίσεις. Οι συσχετίσεις που παριστάνονται με διακεκομμένη γραμμή είναι αυτές που εισάγουμε στο σχήμα και οι οποίες προκύπτουν από τις ήδη υπάρχουσες. Με τη σχέση Όρος Β **BT** Όρος Α, εννοούμε ότι ο όρος Β έχει ευρύτερο τον όρο Α. Με τη συσχέτιση Όρος Α **broader equivalence** Όρος Β εννοούμε ότι ο όρος Α έχει ευρύτερο τον όρο Β ενός άλλου θησαυρού ενώ με τη σχέση **narrower equivalence** εννοούμε το αντίστροφο. Στην 3.1.a έχουμε αρχικά τις συσχετίσεις Όρος Β **BT** Όρος Α και Όρος Γ **BT** Όρος Β. Επειδή όπως έχουμε ήδη πει η συσχέτιση **BT** έχει την έννοια του υποσυνόλου, μπορούμε να ερμηνεύσουμε τις παραπάνω συσχετίσεις ως εξής: Όρος Γ  $\subseteq$  Όρος Β και Όρος Β  $\subseteq$  Όρος Α. Από αυτές τις συσχετίσεις προκύπτει ότι Όρος Γ  $\subseteq$  Όρος Α, δηλαδή Όρος Γ **BT** Όρος Α, επομένως είναι περιττό να εισάγουμε τη συσχέτιση Όρος Γ **BT** Όρος Α. Στην 3.1.b έχουμε τις συσχετίσεις Όρος Α **exact equivalence** Όρος Γ και Όρος Γ **BT** Όρος Β. Με τον ίδιο τρόπο έχουμε ότι Όρος Α = Όρος Γ και Όρος Γ  $\subseteq$  Όρος Β, επομένως Όρος Α  $\subseteq$  Όρος Β. Δηλαδή είναι περιττή η εισαγωγή της συσχέτισης Όρος Α **narrower equivalence** Όρος Β. Στην 3.1.c έχουμε ότι Όρος Γ **BT** Όρος Β και Όρος Β **broader equivalence** Όρος Α. Οι συσχετίσεις αυτές ερμηνεύονται ως Όρος Γ  $\subseteq$  Όρος Β και Όρος Β  $\subseteq$  Όρος Α. Επομένως είναι εμφανές ότι ισχύει Όρος Γ  $\subseteq$  Όρος Α και επειδή οι όροι Α και Γ ανήκουν σε διαφορετικούς θησαυρούς, Όρος Γ **broader equivalence** Όρος Α, δηλαδή είναι περιττή η εισαγωγή αυτής της συσχέτισης.

### 3.1.2 Κυκλικές Συσχετίσεις

Οι κύκλοι δημιουργούνται όταν με την εισαγωγή μιας νέας συσχέτισης δημιουργείται ένας κυκλικός γράφος, αν θεωρήσουμε ως κόμβους τους όρους και ως ακμές συσχετίσεις ίδιου τύπου μαζί με συσχετίσεις ισοδυναμίας.

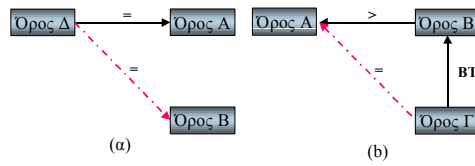


Εικόνα 3.2: Κυκλικές Συσχετίσεις

Στην εικόνα 3.2 έχουμε αρχικά τις δηλώσεις: Όρος Γ **BT** Όρος B, Όρος B **exact equivalence** Όρος A και Όρος A **BT** Όρος Δ. Δηλαδή έχουμε ότι Όρος Γ  $\subseteq$  Όρος B, Όρος A = Όρος B και Όρος A  $\subseteq$  Όρος Δ. Εισάγοντας τη συσχέτιση Όρος Δ **exact equivalence** Όρος Γ δημιουργείται ένας κύκλος μεταξύ των όρων A,B,Γ,Δ ο οποίος ισχύει προφανώς μόνο όταν οι όροι A,B,Γ,Δ είναι ταυτόσημοι. Σε αυτή όμως την περίπτωση θα έπρεπε οι όροι Γ,B και A,Δ να είναι περιγραφείς και να περιγράφουν στην ίδια γλώσσα την ίδια έννοια, κάτι που είναι άτοπο, αφού όπως έχουμε πει από ένα σύνολο δόκιμων όρων που αναπαριστούν την ίδια έννοια, μόνο ένας επιλέγεται ως περιγραφέας.

### 3.1.3 Εισαγωγή Συσχετίσεων που οδηγούν σε Ασυνέπεια

Στην εικόνα 3.3α έχουμε τις συσχετίσεις: Όρος Δ **exact equivalence** Όρος A και Όρος Δ **exact equivalence** Όρος B.



Εικόνα 3.3: Παραδείγματα περιττών συσχετίσεων

Όπως γνωρίζουμε η συσχέτιση *exact equivalence* συνδέει δυο όρους διαφορετικής γλώσσας που αναπαριστούν την ίδια ακριβώς έννοια. Με βάση τις συσχετίσεις που έχουμε θα έπρεπε οι όροι Α και Β του θησαυρού να είναι ισοδύναμοι. Ωστόσο όπως έχουμε πει, σε κάθε γλώσσα επιλέγουμε έναν μόνο όρο για να αναπαριστά μια έννοια και όλοι οι άλλοι σχετίζονται με αυτόν. Επομένως, αν ο όρος Α είναι ο επιλεγμένος όρος, δε θα έπρεπε ο όρος Β να συμμετέχει σε συσχέτιση με άλλο όρο πλην του Α. Στην 3.3b έχουμε αρχικά τις συσχετίσεις Όρος Γ **BT** Όρος Β, Όρος Β **broader equivalence** Όρος Α. Έχουμε δηλαδή Όρος Γ  $\subseteq$  Όρος Β και Όρος Β  $\subseteq$  Όρος Α. Από τις συσχετίσεις αυτές προκύπτει ότι ισχύει Όρος Γ  $\subseteq$  Όρος Α. Οπότε εισάγοντας τη συσχέτιση Όρος Γ = Όρος Α οδηγούμαστε σε ασυνεπή κατάσταση, γιατί θα έπρεπε να ισχύει ότι οι Όροι Α και Γ αναπαριστούν την ίδια έννοια, που όπως έχουμε ήδη πει, δεν γίνεται.

### 3.1.4 Συσχετίσεις με Δυναμικούς Όρους

Κατά τη σύνδεση δυο θησαυρών, αυτός που έχει αναλάβει τη σύνδεση, πρέπει να συνδέσει με τις κατάλληλες συσχετίσεις τους όρους των δυο θησαυρών. Υπάρχει περίπτωση να υπάρχει ένας όρος στον ένα θησαυρό που να μην έχει έναν ισοδύναμο στον άλλο. Ωστόσο είναι δυνατόν να συνδυάσουμε δυο ή περισσότερους όρους του θησαυρού και να δημιουργήσουμε έναν δυναμικό όρο ο οποίος θα έχει το ίδιο ακριβώς νόημα με τον άλλο όρο. Το πρόβλημα όμως είναι το πώς θα συνδυαστούν οι όροι για τη δημιουργία του δυναμικού. Για παράδειγμα, στην περίπτωση του [DF98a], οι όροι συνδέονται με τον τελεστή **&**. Ο τελεστής **&** έχει δυο δυνατές ερμηνείες: την τομή και την παράθεση. Για παράδειγμα μεταξύ του MEPIMEE και του AAT υπάρχει η συσχέτιση: TermFR'Usine de chapeaux **exact equivalence** EnTerm'hats & factories. Αν ο τελεστής **&** έχει την έννοια της τομής, ο όρος hats & factories είναι ισοδύναμος με τον hats  $\cap$  factories. Δηλαδή θεωρούμε την τομή των συνόλων των αντικειμένων που αντιπροσωπεύουν οι όροι hats, factories. Είναι εμφανές ότι η τομή τους είναι το κενό, οπότε δεν έχει νόημα μια τέτοια σύνδεση. Αν ο

τελεστής & έχει την έννοια της απλής παράθεσης όρων τότε είναι αποδεκτή μια τέτοια σύνδεση, αλλά έχουμε απώλεια πληροφορίας γιατί δεν γνωρίζουμε πώς ακριβώς συνδυάζονται οι όροι μεταξύ τους. Δηλαδή ενώ για τον όρο Usine de chapeaux γνωρίζουμε την ακριβή ερμηνεία του, για τον όρο hats & factories δεν μπορούμε να βγάλουμε ακριβή συμπεράσματα. Επομένως, θα πρέπει να βρούμε κάποιο τρόπο ώστε να μπορούμε να συνθέσουμε δυναμικούς όρους χωρίς απώλεια πληροφορίας.

### **3.2 Συστήματα που αντιμετωπίζουν παρόμοια προβλήματα.**

Μετά από έρευνα που κάναμε διαπιστώσαμε ότι παρόμοια προβλήματα αντιμετωπίζουν και τα συστήματα περιγραφικής λογικής, όπου έχουν αναπτυχθεί αλγόριθμοι ώστε να τα επιλύουν. Γιαυτό θεωρήσαμε σκόπιμο να δούμε το πώς αντιμετωπίζουν τα συστήματα αυτά τα παραπάνω θέματα ώστε να διαπιστώσουμε αν μπορούμε και εμείς να τα αντιμετωπίσουμε με παρόμοιο τρόπο. Στο επόμενο κεφάλαιο παρουσιάζουμε τα βασικά χαρακτηριστικά των συστημάτων περιγραφικής λογικής.

---

## Κεφάλαιο 4

### Περιγραφική Λογική

#### 4.1 Περιγραφή Μοντέλου

Η θεωρία της περιγραφικής λογικής είναι άμεσα συσχετισμένη με το μοντέλο δεδομένων που χρησιμοποιείται και το οποίο είναι αρκετά γενικό. Η Περιγραφική Λογική αποτελεί ένα φορμαλισμό για παράσταση γνώσης και συλλογισμών. Προήλθε από το έργο KL\_ONE[BS85] και παρέχει γλώσσες οι οποίες διαχειρίζονται έννοιες, ρόλους και ανεξάρτητα αντικείμενα.

##### 4.1.1 Αντικείμενα

Ένα αντικείμενο ορίζεται ως η υπολογιστική παράσταση μιας οντότητας του πραγματικού κόσμου στο σύστημα. Κάθε αντικείμενο είναι μοναδικό και χαρακτηρίζεται από κάποιο μοναδικό κωδικό που το διαφοροποιεί από όλα τα υπόλοιπα. Το σύνολο αντικειμένων θα συμβολίζεται με  $\mathcal{I}$ .

Σύνολα αντικειμένων του πραγματικού κόσμου με κάποια κοινή σημασία ή ιδιότητα, ομαδοποιούνται σε έννοιες (ή αλλιώς κλάσεις). Κάθε έννοια του μοντέλου παριστάνει μια έννοια του πραγματικού κόσμου. Οι όροι “έννοια” και “κλάση” από δω και στο εξής θα χρησιμοποιούνται ισοδύναμα. Για παράδειγμα, η έννοια “factories” περιγράφει το σύνολο των εργοστασίων. Στα αντικείμενα προσδίδονται ιδιότητες, με την κατάταξή τους κάτω από έννοιες (concept membership), (π.χ. το “Αρχαιολογικό Μουσείο Ηρακλείου” είναι Μουσείο και συνδέεται με άλλα αντικείμενα μέσω ρόλων (π.χ. “Αρχαιολογικό Μουσείο Ηρακλείου” “βρίσκεται στο” “Ηράκλειο”). Ένα αντικείμενο το οποίο ανήκει σε μια έννοια λέγεται ότι αποτελεί περίπτωση (instance) αυτής.

### 4.1.2 Ρόλοι

Ρόλος είναι μια αντιστοίχιση από τα αντικείμενα στα αντικείμενα. Κάθε ρόλος έχει ένα χαρακτηριστικό όνομα, μια αρχή και ένα τέλος. Πεδίο τιμών είναι το  $\Delta x \Delta$ .

Οι ρόλοι παριστάνουν συσχετίσεις μεταξύ αντικειμένων. Για παράδειγμα ο ρόλος “βρίσκεται” μπορεί να ενώνει ένα Μουσείο με έναν Τόπο και να δηλώνει το πού βρίσκεται το συγκεκριμένο μουσείο.

Οι γλώσσες οι οποίες διαχειρίζονται έννοιες, ρόλους και αντικείμενα ονομάζονται εννοιολογικές γλώσσες (concept languages), και εν γένει στηρίζονται στη λογική πρώτης τάξης.

Στις επόμενες ενότητες παρουσιάζονται οι βασικές αρχές σχετικές με τις εννοιολογικές γλώσσες, τις βάσεις γνώσεων που κατασκευάζονται από αυτές καθώς και μεθόδους για την εξαγωγή συμπερασμάτων, από τέτοιου είδους βάσεις, στηριζόμενοι στην Περιγραφική Λογική.

## 4.2 Το Γενικό Πλαίσιο της Περιγραφικής Λογικής

Μια βάση γνώσης, (συνήθως θα παριστάνεται με το σύμβολο  $\Sigma$ ), η οποία κατασκευάζεται με χρήση της Περιγραφικής Λογικής, αποτελείται από δυο τμήματα. Το πρώτο τμήμα περιέχει τις περιγραφές εννοιών δηλαδή την εντασιακή γνώση (Intensional Knowledge) και ονομάζεται TBox. Το δεύτερο περιέχει τα πραγματικά δεδομένα δηλαδή την εκτασιακή γνώση (Extensional Knowledge) και ονομάζεται Abox [Bor95].

### ***TBox-Εντασιακή (Intensional) γνώση***

Ακολουθώντας την προσέγγιση που παρουσιάστηκε στο [BDNS94], χωρίζουμε το TBox σε δυο μέρη. Το πρώτο θα αποκαλείται *μέρος σχήματος* (scheme part) και περιέχει τις λεγόμενες *αρχικές έννοιες* ενώ το δεύτερο, αποκαλείται *μέρος όψεων* (view part) και περιέχει τις λεγόμενες *παράγωγες έννοιες*.

### ***Αρχικές έννοιες (primitive concepts)***

Αρχικές έννοιες ονομάζονται οι βασικές έννοιες οι οποίες είναι ανεξάρτητες, αυτόνομες, δεν μπορούν να αναγνωριστούν από εξωτερικές ιδιότητές τους ή να παραχθούν από άλλες. Για να γίνει

αντιληπτή η ύπαρξή τους θα πρέπει να έχουν δηλωθεί ρητώς. Επίσης, τα αυτόνομα αντικείμενα (γεγονότα) θα πρέπει και αυτά με τη σειρά τους να εισαχθούν ρητώς ως μέλη εννοιών [Bor95].

Η Περιγραφική Λογική μέσω της εννοιολογικής γλώσσας, παρέχει τη δυνατότητα περιγραφής και ορισμού νέων εννοιών.

### *Παράγωγες Έννοιες*

Δοθείσας μιας εννοιολογικής γλώσσας  $L$ , οι έννοιες οι οποίες παράγονται μέσω των τελεστών της  $L$ , και των αρχικών εννοιών, ονομάζονται παράγωγες (derived).

Ο ορισμός νέων εννοιών γίνεται προσδιορίζοντας τις ικανές ή τις ικανές και αναγκαίες συνθήκες για τη συμμετοχή ενός αντικειμένου στις περιπτώσεις κάποιας έννοιας. Ανάλογος είναι και ο ορισμός των ρόλων.

Ο ορισμός του TBox γίνεται μέσω μιας σειράς αξιωματικών δηλώσεων οι οποίες μπορούν να πάρουν κάποια από τις παρακάτω μορφές:

#### 1. Προδιαγραφή αρχικής έννοιας (Primitive Concept Specification)

$$A \leq C$$

Που δηλώνει ρητώς ότι η έννοια  $A$  αποτελεί υπάλληλη έννοια της  $C$ , δηλαδή αντικείμενα που είναι μέλη της  $A$  είναι αναγκαστικά και μέλη της  $C$ . Η δήλωση αυτή έχει προφανώς την έννοια συνόλου-υποσυνόλου.

#### 2. Ορισμός παράγωγης έννοιας (Concept Definition)

$$A \equiv C$$

η οποία ορίζει την παράγωγη έννοια  $A$  και τον τρόπο παραγωγής της μέσω της έννοιας  $C$ .

#### 3. Ορισμός Ρόλου (Role Definition)

$$R \leq C \times D$$

Που δηλώνει ότι ο ρόλος  $R$  είναι μια αντιστοίχιση από αντικείμενα της έννοιας  $C$  σε αντικείμενα της  $D$ .

Οι ακόλουθες παρατηρήσεις είναι αναγκαίες:

- Μια αρχική έννοια ορίζει τις ικανές συνθήκες ώστε ένα αντικείμενο να αποτελεί μέλος της. Οι έννοιες οι οποίες ορίζονται στις βάσεις δεδομένων, για παράδειγμα, είναι όλες αρχικές έννοιες. Αυτό γιατί για να ανήκει ένα αντικείμενο στην έννοια αυτή θα πρέπει να έχει εισαχθεί από το χρήστη και να έχει δηλωθεί ρητώς ότι ανήκει στην έννοια αυτή. Όταν για παράδειγμα ορίζεται η έννοια Date, η βάση δεδομένων δεν είναι σε θέση από μόνη της να αποφασίσει αν ένα αντικείμενο αποτελεί ή όχι μέλος της έννοιας αυτής, παρά μόνο αν κάτι τέτοιο έχει δηλωθεί ρητώς από το χρήστη. Από τις τρεις παραπάνω μορφές οι (1) και (3) περιγράφουν ικανές συνθήκες. Η πληροφορία ότι ένα αντικείμενο αποτελεί μέλος της έννοιας που περιγράφεται στο αριστερό μέλος, είναι αρκετή για την εξαγωγή του συμπεράσματος ότι το αντικείμενο αυτό ανήκει στην έννοια που περιγράφεται στο δεξί μέλος. Δεν ισχύει όμως το αντίστροφο.

- Αντίθετα, μια παράγωγη έννοια ορίζει τις αναγκαίες και ικανές συνθήκες προκειμένου ένα αντικείμενο να ανήκει σε αυτή. Για παράδειγμα η έννοια DateOfCreation ορίζεται σαν η ημερομηνία η οποία περιγράφει το γεγονός μια δημιουργίας. Αν δηλαδή ένα αντικείμενο ημερομηνίας περιγράφει ένα γεγονός δημιουργίας, το σύστημα, έχοντας υπόψη του τον ορισμό αυτό, είναι σε θέση να συμπεράνει από μόνο του ότι το αντικείμενο αυτό αποτελεί μέλος της έννοιας DateOfCreation. Σε αντίθεση με τις (1) και (3), η (2) περιγράφει την ικανή και αναγκαία συνθήκη, δηλαδή η πληροφορία ότι ένα αντικείμενο ανήκει στην έννοια που περιγράφεται από κάποιο μέλος, συνεπάγεται ότι το αντικείμενο αυτό ανήκει και στην έννοια που περιγράφεται από το άλλο μέλος. Αυτή η παρατήρηση αποτελεί τη βάση για την εκτέλεση συλλογισμών πάνω στο σχήμα και στα δεδομένα της βάσης γνώσης, όπως θα εξηγηθεί παρακάτω.

- Η πρώτη φορά που μια έννοια A θα εμφανισθεί στο αριστερό μέλος μιας δήλωσης του TBox (TBox-statement), θα αποκαλείται *εισαγωγή* (Introduction) της A. Κάθε έννοια έχει το πολύ μια εισαγωγή.

- Το δεξιό μέλος κάποιας από τις (1),(2) και (3) μπορεί να είναι από μια αρχική έννοια έως μια σύνθετη εννοιολογική έκφραση της γλώσσας L. Μια έννοια που εμφανίζεται στο αριστερό μέλος μιας έκφρασης προδιαγραφών (δηλ. της μορφής (1)) είναι πάντοτε μια αρχική έννοια. Η έκφραση προδιαγραφών στην οποία η έννοια του αριστερού της μέλους εμφανίζεται πρώτη φορά, θα θεωρείται ο *ορισμός* της αρχικής έννοιας αυτής. Οι αρχικές έννοιες οι οποίες δεν εμφανίζονται στο αριστερό μέλος καμίας έκφρασης του TBox, ονομάζονται *ατομικές* και θεωρείται ότι εισάγονται από μια έκφραση της μορφής:

$$A \leq T,$$

όπου  $T$  είναι το σύνολο όλων των εννοιών που έχουν περιγραφεί στο σύστημα.

Μια έννοια που έχει χρησιμοποιηθεί στο αριστερό μέλος κάποιου ορισμού θα αποκαλείται *ορισμένη* (defined).

Το εννοιολογικό σχήμα (schema part) ορίζεται μέσω μιας σειράς προδιαγραφών δηλώσεων της μορφής (1) και (3) μόνο, ενώ το μέρος όψεων (view part) ορίζεται μέσω δηλώσεων της (2) μόνο.

*Κυκλικοί ορισμοί* στην παρούσα προσέγγιση δεν είναι επιτρεπτοί. Σύμφωνα με τον Nebel, όπως περιγράφει στο [Neb91], ένας κύκλος ορίζεται ως εξής: Μια έννοια  $A$  χρησιμοποιεί άμεσα μια άλλη έννοια  $B$  αν και μόνο αν η  $B$  εμφανίζεται στην εισαγωγή της  $A$ . Μια έννοια  $A_0$  χρησιμοποιεί τη έννοια  $A_n$  αν υπάρχει μια αλυσίδα  $A_0, A_1, \dots, A_n$  τέτοια ώστε η  $A_i$  να χρησιμοποιεί την  $A_{i+1}$ ,  $0 \leq i \leq n-1$ . Τέλος λέγεται ότι το TBox περιέχει ένα κύκλο, αν μια έννοια χρησιμοποιεί τον εαυτό της. Δηλαδή, το σύστημα ανιχνεύει τους κύκλους και τους απορρίπτει, όπως ακριβώς θα θέλαμε και στην περίπτωση μας.

•Ανάλογη σχέση με την (1) μπορεί να υπάρξει και μεταξύ ρόλων. Για παράδειγμα, στους θησαυρούς όρων είπαμε ότι η σχέση  $RT$  είναι ένα σύνολο σχέσεων. Οι σχέσεις στους θησαυρούς όρων είναι ισοδύναμες με τους ρόλους σε ένα σύστημα περιγραφικής λογικής. Αν λοιπόν συμβολίσουμε με  $RN$  το σύνολο των ρόλων ενός συστήματος περιγραφικής λογικής, τότε ισχύει ότι :  $RT \cong RN$ .

### ***ABox – Εκτασιακή (Extensional) γνώση***

Το ABox περιέχει τα γεγονότα της βάσης γνώσεων. Δυο είναι οι βασικοί τύποι δηλώσεων-αξιωμάτων για την κατασκευή του ABox. Οι τύποι αυτοί επιτρέπουν συσχέτιση αντικειμένων με έννοιες-κλάσεις του TBox όσο και συσχετίσεις αντικειμένων μεταξύ τους μέσω ρόλων. Οι δυο αυτοί τύποι δηλώσεων είναι οι εξής:

Δήλωση συμμετοχής αντικειμένου σε έννοια (Concept Membership Assertion): Συμβολίζεται  $C(a)$ , και δηλώνει ότι το αντικείμενο  $a$  αποτελεί μέλος της έννοιας  $C$ .

Δήλωση συμμετοχής αντικειμένου σε ρόλο (Role Membership Assertion): Συμβολίζεται  $R(a,b)$  και δηλώνει ότι τα αντικείμενα  $a$  και  $b$  συνδέονται μέσω του ρόλου  $R$ .

Ένα ABox είναι μια πεπερασμένη σειρά από δηλώσεις (statements) των δυο παραπάνω τύπων [Sch94].

### 4.3 Ερμηνεία

Γενικά, σε οποιαδήποτε τυπική λογική, ο κόσμος αποτελείται από ένα σύνολο από αντικείμενα που αποκαλείται *πεδίο* (domain) ή *σύμπαν αναφοράς* (universe of discourse), καθώς και μια σειρά από συσχετίσεις και λειτουργίες σε αυτό. Η σημασιολογία των εννοιών και των ρόλων στο σύμπαν αυτό, καθορίζεται με χρήση της έννοιας της *ερμηνείας* (interpretation)  $I$ .

**Ορισμός Σαν Ερμηνεία (interpretation)** [Bor96]  $I = \langle I(\Delta) \ I(\cdot) \rangle$ , ορίζεται ένα σύνολο από τιμές  $I(\Delta)$  και μια συνάρτηση  $I(\cdot)$  η οποία αντιστοιχίζει:

- Κάθε έννοια σε ένα υποσύνολο του  $I(\Delta)$
- Κάθε ρόλο σε μια δυαδική συνάρτηση που ανήκει στο  $I(\Delta) \times I(\Delta)$
- Κάθε ανεξάρτητο αντικείμενο σε ένα στοιχείο του  $I(\cdot)$  [Wil95].

Η ερμηνεία μιας εννοιολογικής έκφρασης ορίζεται με βάση την ερμηνεία των τελεστών και των εννοιών που αποτελούν την έκφραση.

Μια ερμηνεία  $I = \langle I(\Delta), I(\cdot) \rangle$  ικανοποιεί τις παρακάτω δηλώσεις σύμφωνα με το [Sch94]:

- $C_1 \sqsubseteq C_2$  αν και μόνο αν  $I(C_1) \subseteq I(C_2)$ .
- $A \sqsubseteq C$  αν και μόνο αν  $I(A) = I(C)$
- $R \sqsubseteq C_1 \times C_2$  αν και μόνο αν  $I(R) \subseteq I(C_1) \times I(C_2)$
- $C(a)$  αν και μόνο αν  $I(a) \subseteq I(C)$
- $R(a,b)$  αν και μόνο αν  $(I(a), I(b)) \subseteq I(R)$

#### *Βάση Γνώσης*

Δοθείσας μιας γλώσσας  $L$ , μια βάση γνώσης εννοιών (Concept-Based Knowledge Base ή CBKB για συντομία), θα αποκαλείται η τριάδα  $\Sigma = \langle T, A, L \rangle$ , όπου  $T$  και  $A$ , ένα TBox και ένα ABox αντίστοιχα.

## Ερμηνεία

Μια ερμηνεία  $I = \langle I(\Delta), I(\cdot) \rangle$  λέγεται ότι αποτελεί μοντέλο (model) του  $\Sigma$  αν κάθε αξίωμα του  $\Sigma$  ικανοποιείται από το  $I$ . Το  $\Sigma$  λέγεται ικανοποιημένο αν υπάρχει έστω και ένα μοντέλο αυτού.

Ειδικότερα, σύμφωνα με το [Wil95], ισχύει ότι:

Μια ερμηνεία  $I$  αποτελεί μοντέλο ενός TBox  $T$  αν και μόνο αν η  $I$  ικανοποιεί όλα τα εννοιολογικά αξιώματα στο  $T$ .

Μια ερμηνεία  $I$  αποτελεί μοντέλο ενός ABox  $A$ , ως προς το TBox  $T$ , αν και μόνο αν η  $I$  ικανοποιεί όλα τα δηλωτικά (assertion) αξιώματα στο  $A$ , και όλα τα εννοιολογικά (terminological) αξιώματα στο  $T$ .

## 4.4 Σχέση Υπαγωγής

Η σχέση  $\mathcal{E}$  αποκαλείται *υπαγωγή* (subsumption). Στο [Wil95] η υπαγωγή ορίζεται ως:

**Ορισμός** Έστω δυο έννοιες  $C_1$  και  $C_2$  και έστω  $T$  ένα TBox. Λέγεται ότι η  $C_1$  υπάγεται στην  $C_2$ , σύμφωνα με το  $T$  και συμβολίζεται  $C_1 \mathcal{E} C_2$ , αν και μόνο  $I(C_1) \subset I(C_2)$  για κάθε μοντέλο του  $T$ .

Η σχέση  $\mathcal{E}$  ορίζει μια σχέση μερικής διάταξης. Αν για δυο έννοιες ισχύει το ότι η μια υπάγει την άλλη, τότε οι δυο αυτές έννοιες θα θεωρούνται *ισοδύναμες* (equivalent). Δηλαδή:  $C_1 \equiv C_2 \Leftrightarrow C_1 \mathcal{E} C_2 \wedge C_2 \mathcal{E} C_1$  [BM95].

Η υπαγωγή για ρόλους μπορεί να οριστεί ανάλογα. Λέγεται επίσης ότι μια βάση γνώσης  $\Sigma$  λογικά συνεπάγεται μια δήλωση  $a$  του TBox ή του ABox, και γράφεται  $\Sigma \models a$  αν το  $a$  είναι αληθές για οποιοδήποτε μοντέλο του  $\Sigma$ .

## 4.5 Υποθέσεις της Περιγραφικής Λογικής

Οι βάσεις δεδομένων και τα συστήματα παράστασης γνώσεων, παριστάνουν και εξάγουν συμπεράσματα για γεγονότα του πραγματικού κόσμου. Τα Συστήματα Διαχείρισης Βάσεων Δεδομένων (ΣΔΒΔ) παρέχουν μια γλώσσα που χρησιμοποιείται για τον ορισμό του σχήματος και για

λειτουργίες πάνω στα δεδομένα. Η γλώσσα αυτή ονομάζεται *γλώσσα χειρισμού δεδομένων* (data manipulation language). Οι γλώσσες του είδους αυτού στηρίζουν τη λειτουργία τους, πάνω σε ορισμένες προϋποθέσεις σύμφωνα με το [Hus94]: Στη μοναδικότητα του ονόματος (Unique Name Assumption), στην υπόθεση κλειστού κόσμου (Closed World Assumption) και στην υπόθεση κλειστού πεδίου (Closed-Domain Assumption).

Απο την άλλη μεριά, οι υποθέσεις πάνω στις οποίες στηρίζονται οι γλώσσες επερωτήσεων και διαχείρισης γεγονότων των Συστημάτων Παράστασης Γνώσης και Συλλογισμών (ΣΠΓΣ), συμπίπτουν μόνο στην υπόθεση της μοναδικότητας του ονόματος. Κατά τα άλλα ισχύουν οι υποθέσεις:

**Υπόθεση ανοικτού κόσμου (Open World Assumption).** Σύμφωνα με αυτή την υπόθεση, μπορεί να υπάρχουν γεγονότα τα οποία να είναι αληθή και να μην περιέχονται στη βάση γνώσης

**Υπόθεση ανοικτού πεδίου (Open Domain Assumption).** Σύμφωνα με την υπόθεση αυτή μπορεί να υπάρξουν περισσότερα αντικείμενα στον κόσμο από όσα βρίσκονται αποθηκευμένα στη βάση.

Οι διαφορετικές αυτές υποθέσεις δηλώνουν ότι ακόμα και αν τα δεδομένα ενός ΣΔΒΔ και ενός ΣΠΓΣ συμπίπτουν, τα αποτελέσματα των ερωτήσεων σε αυτά θα είναι εντελώς διαφορετικά. Αναπόφευκτο γεγονός είναι ότι τόσο οι βάσεις δεδομένων όσο και τα συστήματα γνώσης, περιέχουν μόνο τμήμα της γνώσης του κόσμου και έτσι δεν είναι ξακάθαρο το τι είναι γνωστό και τι όχι.

Η δυνατότητα παράστασης μη πλήρους γνώσης, αποτελεί το βασικό χαρακτηριστικό που διαφοροποιεί τα συστήματα γνώσης από τις βάσεις δεδομένων. Η σημασιολογία των εννοιολογικών γλωσσών στηρίζεται στην υπόθεση του ανοικτού κόσμου. Έτσι η απάντηση σε με επερώτηση Q θα είναι “Yes” αν η Q είναι αληθής για όλα τα μοντέλα του Σ, “No” αν είναι ψευδής για όλα τα μοντέλα και “Unknown” στις υπόλοιπες περιπτώσεις. Στο [Bre95a] υποστηρίζεται ότι μια βάση γνώσης η οποία είναι πλήρης υπό την έννοια αυτή μπορεί να συμπεριφερθεί σαν μια βάση δεδομένων, και οι εννοιολογικές γλώσσες της Περιγραφικής Λογικής να χρησιμοποιηθούν σαν γλώσσες επερωτήσεων.

Για να επιτευχθεί μια τέτοια δυνατότητα εισάγεται ένας νέος *επιστημολογικός τελεστής* (epistemic operator), με τον τρόπο που ορίζεται στο [DLNN94], ο οποίος διαχωρίζει τη γνώση που αναφέρεται σε ολόκληρο τον κόσμο (world assumption), από τη γνώση που περιέχεται στη βάση (knowledge of the state of the base). Με προσεκτική χρήση του τελεστού αυτού, είναι δυνατόν να εκφράσει κανείς επερωτήσεις, η επεξεργασία των οποίων αναγκάζει το σύστημα να υποθέσει πλήρη γνώση για τη βάση γνώσης. Η παρατήρηση εδώ είναι ότι η προσέγγιση αυτή είναι διαφορετική από το να υποθέσει

κανείς *κλειστή γνώση* για τη βάση. Η κλειστότητα δεν αφορά τη σημασιολογία της βάσης, αλλά κατά την εκτέλεση των επερωτήσεων, τα είδη των συλλογισμών που λαμβάνουν χώρα, έχουν τα ίδια αποτελέσματα με αυτά που θα είχαμε αν ίσχυε η υπόθεση του κλειστού κόσμου. Αυτό πράγματι συμβαίνει στις καταναεμημένες, ετερογενείς και ομόσπονδες βάσεις. Κανείς δεν υποθέτει ότι κάποια βάση έχει πλήρη γνώση του κόσμου, αλλά κατά την εκτέλεση των επερωτήσεων, απαντήσεις της βάσης εξάγονται όπως εαν ίσχυε κάτι τέτοιο.

Ο επιστημολογικός αυτός τελεστής συμβολίζεται με το σύμβολο “**K**” (known) και εισάγεται μπροστά από έννοιες και ρόλους. Ο συμβολισμός **KC** δηλαδή, παριστάνει τα αντικείμενα τα οποία είναι γνωστό ότι ανήκουν στην C, δηλαδή όσα είναι αποθηκευμένα στη βάση μια συγκεκριμένη χρονική στιγμή, και ανήκουν στην *έκταση* (extension) της C. Η χρήση του τελεστή αυτού αποτελεί μια μορφή της υπόθεσης του κλειστού κόσμου.

Σύμφωνα με το [Sch94], όσον αφορά στις ερμηνείες των διαφόρων εννοιών, ισχύουν οι υποθέσεις:

- Κάθε ερμηνεία ορίζεται πάνω σε ένα σταθερό πεδίο, που ονομάζεται Δ.
- Για κάθε ερμηνεία, η αντιστοίχιση από τα ανεξάρτητα αντικείμενα στα αντικείμενα του Δ, είναι σταθερή και αποκαλείται γ.

Έτσι, το **KC** θα ερμηνεύεται σαν το σύνολο των αντικειμένων που είναι γνωστό ότι αποτελούν περιπτώσεις της C, σε οποιαδήποτε ερμηνεία. Αυτό έχει σαν αποτέλεσμα, για κάθε βάση Σ να θεωρείται η ύπαρξη ενός και μοναδικού μοντέλου. Από εδώ και στο εξής σαν ερμηνεία των διαφόρων εννοιών και ρόλων, θα θεωρείται πάντα αυτή που θα ήταν με τη χρήση του παραπάνω τελεστή, ο οποίος όμως, για λόγους ευκολίας, δε θα αναφέρεται.

## 4.6 Σύνταξη και σημασιολογία των εννοιολογικών γλωσσών

Οι εννοιολογικές γλώσσες χρησιμοποιούνται στη δημιουργία εννοιολογικών εκφράσεων, προκειμένου να οριστούν παράγωγες έννοιες, καθώς και οι συσχετίσεις τους. Μια εννοιολογική γλώσσα αποτελείται από τα παρακάτω βασικά συντακτικά σύμβολα όπως αναφέρονται στο [BMS<sup>+</sup>91]:

- **Εννοιες (concepts)**, το σύνολο των οποίων θα συμβολίζεται C.

- **Ρόλους (roles)**, που συμβολίζονται με RN.
- **Ανεξάρτητα αντικείμενα (individual objects)**, το σύνολο των οποίων είναι το  $\Delta$ .
- **Τελεστές παραγωγής εννοιών και ρόλων (concept & role forming operators)**.

Από εδώ και στο εξής, εκτός και αν αναφέρεται ρητώς κάτι διαφορετικό, τα κεφαλαία γράμματα A,B,C, θα συμβολίζουν έννοιες, τα κεφαλαία R,P ρόλους και τα μικρά  $I_j, k$  ανεξάρτητα αντικείμενα.

Name	Concrete Form	Mathematical Representation	Semantics
Concept Name	A	A	$I(A) \subseteq I(\Delta)$
Top	TOP	T	$\Delta$
Bottom	BOTTOM	$\perp$	$\emptyset$
Union	(OR C D)	$C \cup D$	$\{d_i \mid d_i \in I(C) \cup I(D)\}$
Intersect	(AND C D)	$C \cap D$	$\{d_i \mid d_i \in I(C) \cap I(D)\}$
Not	NOT C	$\neg C$	$\{d_i \mid d_i \notin I(C)\}$
Existential Quantification	(SOME R C)	$\exists R.C$	$\{d_i \mid \exists d_2 : (d_i, d_2) \in I(R) \wedge d_2 \in I(C)\}$
Universal Quantification	(ALL R C)	$\forall R.C$	$\{d_i \mid \forall d_2 : (d_i, d_2) \in I(R) \rightarrow d_2 \in I(C)\}$
At-Least Restriction	(AT LEAST n R)	$\leq n.R$	$\{d_i \mid \#\{d_2 \mid (d_i, d_2) \in I(R)\} \leq n\}$
At-Most Restriction	(AT MOST n R)	$\geq n.R$	$\{d_i \mid \#\{d_2 \mid (d_i, d_2) \in I(R)\} \geq n\}$
one of	(ONE OF $i_j$ )	$\{i_j, \dots\}$	$\{i_j, \dots\}$

**Πίνακας 5.1 Τελεστές σύνθεσης εννοιών στην Περιγραφική Λογική**

<i>Abstract Form</i>	<i>Concrete Form</i>	<i>Interpretation</i>
$R, S \rightarrow P$	$P$	$P^I$
$  T$	AnyRelation	$D \times D$
$  R \cap S$	(and R S)	$R^I \cap S^I$
$  R \cup S$	(or R S)	$R^I \cup S^I$
$  R^{-1}$	(inverse R)	$\{(d, d') \mid (d', d) \in R^I\}$
$  R _C$	(restr R C)	$\{(d, d') \in R^I \mid d' \in C^I\}$
$  {}_C R$	(domrestr R C)	$\{(d, d') \in R^I \mid d \in C^I\}$

**Πίνακας 5.2 Οι τελεστές σύνθεσης ρόλων**

Στους πίνακες 5.1, 5.2 περιγράφονται οι τελεστές εννοιών και ρόλων αντίστοιχα, καθώς και η ερμηνεία τους. Για τους πίνακες αυτούς θα πρέπει να σημειωθούν τα εξής:

1. Ο τελεστής  $T$  χρησιμοποιείται για να δηλώσει την έννοια, η ερμηνεία της οποίας περιλαμβάνει όλα τα αντικείμενα που είναι αποθηκευμένα στη βάση, σε οποιοδήποτε στιγμιότυπό της.
2. Η χρήση του τελεστή  $\perp$  γίνεται σε περιπτώσεις εννοιών των οποίων η ερμηνεία σε οποιοδήποτε στιγμιότυπο της βάσης είναι το κενό σύνολο. Κάτι τέτοιο είναι χρήσιμο όπως θα φανεί παρακάτω στις περιπτώσεις όπου μια έννοια δεν παρίσταται στη βάση, που σημαίνει ότι η βάση δεν έχει καθόλου γνώση για την ερμηνεία της.
3. Τα σύμβολα  $C$ ,  $D$ ,  $C_1$ ,  $C_2$  μπορεί να είναι τόσο αρχικές ή παράγωγες έννοιες, όσο και εννοιολογικές εκφράσεις.
4. Οι πράξεις ίσο, γνήσια μεγαλύτερο, ή γνήσια μικρότερο, δεν υποστηρίζονται άμεσα, μια και μπορούν να εξαχθούν από τους ήδη υπάρχοντες τελεστές  $\leq$  και  $\geq$  σε συνδυασμό με τους τελεστές τομής ή/και διαφοράς.

Παρακάτω περιγράφονται μερικά παραδείγματα για τον τρόπο χρήσης της Περιγραφικής Λογικής και ειδικότερα των εννοιολογικών εκφράσεων στον ορισμό νέων εννοιών.

## 4.7 Συλλογισμοί της Περιγραφικής Λογικής

Η γνώση η οποία περιέχεται σε ένα TBox και σε ένα ABox είναι περισσότερη από αυτή που έχει ρητώς εισαχθεί και μπορεί να εξαχθεί με χρήση της σημασιολογίας των εννοιών και των τελεστών που ορίστηκε προηγουμένως. Για την ανάκτηση της γνώσης αυτής έχουν αναπτυχθεί διάφοροι

μηχανισμοί (reasoning services). Οι πιο βασικοί από αυτούς, όπως ορίζονται στα [DLNS96] και [DLNN95], είναι οι παρακάτω:

Δοθείσας μιας βάσης γνώσης  $\Sigma = \langle T, A, L \rangle$ , δυο εννοιών  $C$  και  $A$  και ενός αντικειμένου  $a$ , ονομάζουμε:

- **Ικανοποίηση Εννοιών (Concept Satisfiability)**, που συμβολίζεται  $\Sigma \models C \equiv \perp$ , το πρόβλημα ελέγχου της ικανοποίησης ή μη της έννοιας  $C$  στο  $\Sigma$ . Με άλλα λόγια αν  $I(C) \neq \emptyset$ .
- **Έλεγχος συνέπειας (Consistency)**, που περιγράφεται ως  $\Sigma \models$ , τον έλεγχο αν υπάρχει έστω και ένα μοντέλο στο  $\Sigma$  ή όχι, αν δηλαδή όλες οι έννοιες καταλήγουν τελικά στην  $\perp$ .
- **Έλεγχος υπαγωγής (Subsumption)**, το οποίο γράφεται ως  $\Sigma \models C \subseteq D$ , το πρόβλημα του ελέγχου αν η έννοια  $C$  εμπεριέχεται στην  $D$ , δηλαδή, αν  $I(C) \subseteq I(D)$  σε κάθε μοντέλο  $I$  του  $\Sigma$ .
- **Έλεγχος περιπτώσεως (Instance Checking)**, με συμβολισμό  $\Sigma \models C(a)$ , τον έλεγχο αν το  $C(a)$  ικανοποιείται από κάθε μοντέλο του  $\Sigma$ .
- **Μη συσχέτιση και ισοδυναμία (Disjointness and Equivalence)**, η οποία συμβολίζεται  $\Sigma \models C \neq D$  και  $\Sigma \models C \equiv D$  αντίστοιχα, τον έλεγχο αν η έννοια  $C$  είναι εντελώς ανεξάρτητη από την έννοια  $D$ , δηλαδή αν  $I(C) \cup I(D) \neq \emptyset$ , ή αν η  $C$  είναι ισοδύναμη με την  $D$ .

Ο έλεγχος για την ικανοποίηση μιας έννοιας είναι βασικός μια και μπορεί να χρησιμοποιηθεί για τον έλεγχο καταλληλότητας κάποιων σημείων πρόσβασης.

Ο έλεγχος συνέπειας αποτελεί βασική διαδικασία για τους μηχανισμούς γνώσης που θέλουν να ελέγχουν τη συνέπεια των μοντέλων τους [BDNS94].

Ο έλεγχος υπαγωγής θεωρείται η βασικότερη λειτουργία από αυτές που μπορεί να επιτελέσει η Περιγραφική Λογική. Δοσμένης μιας έννοιας  $C$  και ενός TBox  $T$ , μπορεί για κάθε έννοια  $D$  του  $T$  να ελεγχθεί αν  $D \subseteq C$  ή  $C \subseteq D$  ώστε η  $C$  να μπορεί να κατηγοριοποιηθεί σωστά στην ιεραρχία των εννοιών του  $T$  [DLNS96]. Ακριβώς αυτός είναι ο έλεγχος που θα θέλαμε να γίνεται για τον έλεγχο ακεραιότητας των θησαυρών, ώστε να αντιμετωπίζονται όλες οι περιπτώσεις ασυνέπειας που έχουμε περιγράψει.

Η υπαγωγή είναι βασικό στοιχείο για πιο περίπλοκες δυνατότητες κατηγοριοποίησης. Με δεδομένο ότι τα σημεία πρόσβασης ενός προφίλ αντιστοιχούν σε έννοιες του TBox, η ιεράρχηση των εννοιών

συνεπάγεται και ιεράρχηση των αντίστοιχων σημείων πρόσβασης. Σαν αποτέλεσμα, έχουμε την εννοιολογική δόμηση των σημείων πρόσβασης του προφίλ.

Με τον έλεγχο ισοδυναμίας μπορεί κανείς να βρει ταυτόσημα σημεία πρόσβασης σε διαφορετικά ή όχι προφίλ και να διαχειρίζεται ερωτήσεις σε αυτά ισοδύναμα. Η λειτουργία αυτή μπορεί να χρησιμοποιηθεί επίσης και για κάποια μορφή αξιολόγησης του προφίλ.

Τέλος, ο έλεγχος περιπτώσεως, ελέγχει αν σε μια βάση, υπάρχει ή όχι κάποιο αντικείμενο σαν περίπτωση κάποιας συγκεκριμένης έννοιας. Αποτελεί κατεξοχήν διαδικασία για την ανάκτηση πληροφοριών από βάσεις γνώσης.

#### **4.8 Εφαρμογή των ελέγχων στους Θησαυρούς**

Στην περιγραφική λογική μόλις ο χρήστης εισάγει στο σύστημα τις έννοιες που επιθυμεί, υπάρχει ένας αλγόριθμος ταξινόμησης, ο οποίος αναλαμβάνει να ταξινομήσει τις έννοιες σε μια ιεραρχία και να εντοπίσει τυχόν ασυνέπειες με τη βοήθεια του ελέγχου υπαγωγής. Παράλληλα αναλαμβάνει να διαγράψει τις περιττές συσχετίσεις. Θα ήταν λοιπόν χρήσιμο να ελέγξουμε το αν υπάρχουν συστήματα περιγραφικής λογικής που να αντιμετωπίζουν με επιτυχία τα θέματα που έχουμε θέσει.

Για να επιτευχθεί αυτό, θα πρέπει να εισαγάγουμε τους όρους ενός θησαυρού καθώς και τις αντίστοιχες συσχετίσεις τους σε ένα σύστημα περιγραφικής λογικής. Όπως είναι προφανές κάθε όρος του θησαυρού αντιστοιχεί σε μια έννοια της περιγραφικής λογικής. Αυτό που πρέπει να κάνουμε είναι να βρούμε τις αντιστοιχίες μεταξύ των διαφόρων συσχετίσεων στο θησαυρό και των τελεστών που μας παρέχουν τα συστήματα περιγραφικής λογικής. Στο επόμενο κεφάλαιο δίνουμε τις αντιστοιχίσεις αυτές και μεταφέρουμε έναν θησαυρό σε επιλεγμένα συστήματα περιγραφικής λογικής ώστε να δούμε πώς αντιμετωπίζουν τα θέματα οικονομίας και συνέπειας σε ένα θησαυρό.

---

## Κεφάλαιο 5

### Ερμηνεία των στοιχείων ενός Θησαυρού Όρων σε Περιγραφική Λογική

#### 5.1 Θησαυροί Όρων και Περιγραφική Λογική

Για να μπορέσουμε να μεταφέρουμε τους όρους που υπάρχουν σε ένα θησαυρό όρων σε ένα οποιοδήποτε σύστημα Περιγραφικής Λογικής θα πρέπει να ορίσουμε την αντιστοίχιση που υπάρχει μεταξύ των ιδιαίτερων χαρακτηριστικών ενός θησαυρού με αυτή ενός συστήματος Περιγραφικής Λογικής.

Πρώτα απ' όλα θεωρούμε ότι όλα τα αντικείμενα του κόσμου τον οποίο θέλουμε να περιγράψουμε αποτελούν υπερσύνολο των αντικειμένων που περιγράφονται στον υπολογιστή μας. Αν δηλαδή  $\Delta$  είναι το σύνολο των αντικειμένων του πραγματικού κόσμου και  $\Delta'$  είναι το σύνολο των αντικειμένων που περιγράφονται στον υπολογιστή μας, ισχύει  $\Delta' \subseteq \Delta$ . Το σύστημα που θα αναπτύξουμε θα αναπαριστά με σύμβολα εννοιών μόνο τους περιγραφείς ενός θησαυρού και όχι όλους τους όρους. Αυτό το επιλέξαμε γιατί οι περιπτώσεις που θέλουμε να μελετήσουμε αφορούν συσχετίσεις που συνδέουν μόνο περιγραφείς. Στη συνέχεια θα ερμηνεύσουμε τις διάφορες συσχετίσεις μεταξύ των περιγραφέων ενός θησαυρού, χρησιμοποιώντας την Περιγραφική Λογική.

### 5.1.1 Ερμηνεία της συσχέτισης ιεραρχίας σε Περιγραφική Λογική

Όπως έχουμε ήδη πει, σύμφωνα με το [ISO86], στους θησαυρούς όρων έχουμε τρία είδη ιεραρχικής συσχέτισης: ιεραρχική συσχέτιση γενίκευσης, ιεραρχική συσχέτιση μέρους-όλου, ιεραρχική συσχέτιση περίπτωσης/μέλους. Η ιεραρχική συσχέτιση γενίκευσης είναι μεταβατική και αντισυμμετρική. Συνεπώς η ιεραρχία που σχηματίζει μπορεί να παρασταθεί με έναν ακυκλικό κατευθυνόμενο γράφο ( DAG – Directed Acyclic Graph). Έχει την ίδια ακριβώς ερμηνεία με την έννοια της υπαλληλίας στην Περιγραφική Λογική. Συνεπώς κάθε συσχέτιση ιεραρχίας στο θησαυρό όρων μεταφράζεται σε υπαλληλία στην Περιγραφική Λογική. Η ιεραρχική συσχέτιση μέρους-όλου είναι επίσης μεταβατική και αντισυμμετρική. Σε πολλές εργασίες έχει υποστηριχθεί η άποψη ότι δεν αποτελεί μια συσχέτιση ιεραρχίας, αλλά περισσότερο μια συσχέτιση μεταξύ δυο αντικειμένων. Συνεπώς στην Περιγραφική Λογική μπορεί να αναπαρασταθεί με έναν ρόλο τον οποίο ονομάζουμε **BTP**. Η ιεραρχική συσχέτιση περίπτωσης/μέλους έχει την ίδια ερμηνεία με το να δηλώσουμε στην Περιγραφική Λογική ότι ένα αντικείμενο είναι μέλος μιας έννοιας. Για παράδειγμα με το σύμβολισμό  $C(i)$ , δηλώνουμε ότι το αντικείμενο  $i$  ανήκει στην έννοια  $C$ .

### 5.1.2 Ερμηνεία των σχέσεων ισοδυναμίας σε Περιγραφική Λογική

Εκτός από τις σχέσεις που συνδέουν έννοιες που ανήκουν σε μια γλώσσα, υπάρχουν και σχέσεις που συνδέουν έννοιες που ανήκουν σε διαφορετικούς θησαυρούς. Οι σχέσεις αυτές περιγράφουν το κατά πόσο είναι εννοιολογικά ισοδύναμες δυο οντότητες που περιγράφονται σε διαφορετικές γλώσσες. Όπως έχουμε δει, οι σχέσεις αυτές είναι: *exact equivalence*, *partial equivalence*, *inexact equivalence*, *single-to-multiple equivalence*. Τη συσχέτιση *exact equivalence* μπορούμε να την περιγράψουμε σε Περιγραφική Λογική με το ορολογικό αξίωμα  $C = D$ , το οποίο δηλώνει ότι η ερμηνεία της έννοιας  $C$  είναι ισοδύναμη με αυτή της έννοιας  $D$ . Η συσχέτιση *partial equivalence* διαχωρίζεται όπως έχουμε ήδη πει στις *broader equivalence* και *narrower equivalence*. Η πρώτη έχει την ίδια ερμηνεία με αυτή της ιεραρχικής συσχέτισης γενίκευσης, μόνο που αφορά όρους σε διαφορετικές γλώσσες. Συνεπώς μπορούμε να περιγράψουμε την σχέση αυτή με υπαλληλία ( $\leq$ ). Η δεύτερη συσχέτιση έχει την αντίστροφη ερμηνεία, και μπορεί να παρασταθεί με το σύμβολο ( $\geq$ ). Η συσχέτιση *inexact equivalence* ουσιαστικά περιγράφει την τομή-επικάλυψη δυο όρων, θεωρώντας ότι το σύνολο των αντικειμένων που ανήκουν στην τομή είναι αρκετά μεγάλο. Συνεπώς η σχέση αυτή μπορεί να περιγραφεί σε ένα σύστημα Περιγραφικής Λογικής ως  $C \cap D \neq \emptyset$ . Τέλος, η συσχέτιση *single-to-*

multiple equivalence μπορεί να περιγραφεί με το ορολογικό αξίωμα  $C = D$ , όπου το  $D$  είναι ένας λογικός συνδυασμός εννοιών.

Στον πίνακα 5.4 έχουμε συγκεντρώσει όλες τις συσχετίσεις και τις πιθανές ερμηνείες τους.

World Symbol	World Semantics	Object Ids	Thesaurus Denotation
$\Delta$	All World Objects	None	None
IN	Described World	$i, j$	Instances
CN	Categories of Objects	$A, C, D$	Descriptors
RN	Relations between	$R, P$	$RT, BTP, \dots$

**Πίνακας 5.3 Ερμηνεία των συμβόλων της Περιγραφικής Λογικής**

Thesaurus Denotation	Thesaurus Symbol	Terminological Representation	SIS-TMS Symbol
Broader Term Generic	BTG		BT
Broader Term Part-Whole	BTP	BTP	BTP
Broader Term Instance	BTI	$C(i)$	BTI
Related Term	RT	RN	RT
exact equivalence	=	=	Exact_equivalence
single-to-multiple equivalence	$\_+ \_$	=	single_to_multiple_equivalence
narrower equivalence	>		narrower_equivalence
broader equivalence	<		broader_equivalence
inexact equivalence	+/-	$C \cap D \neq \emptyset$	inexact_equivalence

**Πίνακας 5.4 Ερμηνεία συσχετίσεων στην Περιγραφική Λογική**

### 5.1.3 Μεταφορά του Θησαυρού σε Περιγραφική Λογική

Αφού εντοπίσαμε τις αντιστοιχίες των χαρακτηριστικών ενός θησαυρού με τα χαρακτηριστικά της Περιγραφικής Λογικής, πρέπει τώρα να επιλέξουμε ένα συγκεκριμένο σύστημα Περιγραφικής Λογικής, ώστε να μεταφέρουμε ένα συγκεκριμένο θησαυρό και να δούμε αν το σύστημα που επιλέξαμε αντιμετωπίζει με επιτυχία τις περιπτώσεις που έχουμε περιγράψει στο κεφάλαιο 2. Ο

θησαυρός που επιλέξαμε να μεταφέρουμε σε Περιγραφική Λογική είναι ένα τμήμα του πολύγλωσσου θησαυρού που δημιουργήθηκε μέσω του εργαλείου TMS και αποτελείται από τον αμερικανικό θησαυρό όρων AAT, το γαλλικό θησαυρό MERIMEE και τον αγγλικό θησαυρό RCHME. Το τμήμα που θα χρησιμοποιήσουμε εμείς έχει το μεγαλύτερο τμήμα περιγραφών των θησαυρών AAT και MERIMEE, καθώς και τις μεταξύ τους συσχετίσεις. Συνολικά, το τμήμα που επιλέξαμε αποτελείται από 19.000 περίπου όρους, συσχετίσεις ισοδυναμίας, συσχετίσεις μερικής ισοδυναμίας, ιεραρχικές συσχετίσεις.

Για να επιλέξουμε κάποιο συγκεκριμένο σύστημα Περιγραφικής Λογικής θα πρέπει πρώτα να ορίσουμε τις προδιαγραφές που θέλουμε αυτό να πληροί.

Στην επιλογή του κατάλληλου συστήματος περιγραφικής λογικής μας βοήθησε το International Workshop in Description Logics, όπου παρουσιάστηκε μια σύγκριση διαφόρων συστημάτων και ανακοινώθηκαν συγκριτικά αποτελέσματα. Τα συστήματα που συμμετείχαν στη διαδικασία αυτή της σύγκρισης είναι τα NeoClassic (απόγονος του Classic), Kris, DLP, FaCT, HAM\_ALC, CRACK. Κατά τη διάρκεια του συνεδρίου διαπιστώθηκε ότι τα συστήματα περιγραφικής λογικής μπορούν να διακριθούν σε 3 κυρίως κατηγορίες.

- Στα συστήματα με περιορισμένη εκφραστικότητα αλλά ολοκληρωμένο αλγόριθμο υπαλληλίας (CLASSIC),
- στα συστήματα με μεγάλη εκφραστικότητα, αλλά ημιτελή αλγόριθμο υπαλληλίας (Loom), και
- στα συστήματα με μεγάλη εκφραστικότητα και με ολοκληρωμένο αλγόριθμο για τις πιο πιθανές περιπτώσεις (Kris).

Από τα αποτελέσματα που ανακοινώθηκαν, τις καλύτερες επιδόσεις είχαν τα συστήματα NeoClassic και FaCT. Επιλέξαμε λοιπόν να μεταφέρουμε το θησαυρό και στα δυο συστήματα, να συγκρίνουμε την απόδοσή τους και να βγάλουμε συμπεράσματα για το αν κάποιο από τα συστήματα αντιμετωπίζει με επιτυχία τα προβλήματα που θέσαμε.

### *NeoClassic*

Το NeoClassic είναι η μετεξέλιξη του συστήματος περιγραφικής λογικής Classic. Το πρώτο είναι υλοποιημένο σε C++, ενώ το δεύτερο σε Lisp. Το NeoClassic ανήκει στην οικογένεια των γλωσσών αναπαράστασης που είναι απόγονοι της KL-ONE. Τα συστήματα αυτά προσφέρουν τη δυνατότητα να ορίζονται οι όροι με τέτοιο τρόπο ώστε η σχέση υπαλληλίας να προκύπτει από τον ορισμό. Το

NeoClassic έχει ορισμένα επιπλέον χαρακτηριστικά: τη δυνατότητα απαρίθμησης με τη χρήση του ONE-OF, τη χρήση κανόνων, την υιοθέτηση της υπόθεσης ανοικτού κόσμου και τη δυνατότητα εκτέλεσης μιας πλειάδας ερωτήσεων προς τη βάση.

Η λογική που χρησιμοποιεί το NeoClassic, είναι η εξής: Αν C είναι μια έννοια και R ένας ρόλος, τότε μια έκφραση στο σύστημα μπορεί να είναι της μορφής:  $C \mid T \mid \perp \mid C \cap D \mid \forall R.C \mid \geq n R \mid \leq n R$ .

Για τη μεταφορά του θησαυρού στο σύστημα, χρησιμοποιήσαμε τις αντιστοιχίσεις του πίνακα 5.4. Αρχικά δηλώσαμε τις έννοιες του θησαυρού στο σύστημα μαζί με τις μεταξύ τους συσχετίσεις. Παρατηρήσαμε ότι το σύστημα δεν επιτρέπει τον κυκλικό ορισμό εννοιών, και αυτό επειδή δεν επιτρέπει να ορίσουμε μια έννοια ως προς κάποια που δεν έχει οριστεί πριν, ούτε επιτρέπει να συμπληρώσουμε τον ορισμό μιας έννοιας. Αυτοί οι περιορισμοί μας απαγόρευαν ουσιαστικά να εισάγουμε στο σύστημα συσχετίσεις που προκαλούσαν τα προβλήματα που θέλαμε να εξετάσουμε. Χωρίς τη δήλωση αυτών των συσχετίσεων, το σύστημα έκανε την ταξινόμηση των 19.000 εννοιών σε λογικό χρόνο (2 περίπου λεπτά). Ωστόσο δεν αντιμετώπισε όλα τα προβλήματα που είχαμε θέσει.

## ***FaCT***

Όπως έχουμε αναφέρει, τα συστήματα περιγραφικής λογικής υποστηρίζουν τη λογική περιγραφή των εννοιών και των ρόλων καθώς και τους συνδυασμούς τους χρησιμοποιώντας μια πλειάδα τελεστών για το σχηματισμό σύνθετων περιγραφών.

Η λογική που χρησιμοποιεί το FaCT είναι η  $ALCH_{R+}$  που είναι ένα υπερσύνολο της ALC. Η λογική ALC επιτρέπει τη δημιουργία περιγραφών με τη χρήση δεδομένων λογικών συνδέσμων καθώς και σχεσιακών τελεστών με καθολικό ή υπαρξιακό ποσοδείκτη. Αν C είναι μια έννοια και R ένας ρόλος τότε μια έκφραση ALC είναι της μορφής  $C \mid T \mid \perp \mid \neg C \mid C \cap D \mid C \cup D \mid \exists R.C \mid \forall R.C$ . Η λογική  $ALCH_{R+}$  εμπλουτίζει την ALC με κλειστούς μεταβατικούς βασικούς ρόλους και αξιώματα ρόλων, ο συνδυασμός των οποίων επιτρέπει τον ορισμό μιας ιεραρχίας μεταβατικών και μη μεταβατικών ρόλων. Για να είναι το FaCT εύχρηστο σε εφαρμογές Περιγραφικής Λογικής έχουν χρησιμοποιηθεί πολλές τεχνικές βελτιστοποίησης στην υλοποίηση του αλγόριθμου  $ALCH_{R+}$  για έλεγχο ικανοποίησης. Η σημασιολογία των εννοιολογικών εκφράσεων και η σημασιολογία των αξιωμάτων παρουσιάζονται στους παρακάτω πίνακες.

Σύνταξη	Ερμηνεία
CN	$CN^I \subseteq \Delta^I$
T	$\Delta^I$
$\perp$	$\emptyset$
$\neg C$	$\Delta^I - C^I$
$C \cap D$	$C^I \cap D^I$
$C \cup D$	$C^I \cup D^I$
$\exists R.C$	$\{d \in \Delta^I \mid R^I(d) \cap C^I \neq \emptyset\}$
$\forall R.C$	$\{d \in \Delta^I \mid R^I(d) \subseteq C^I\}$

**Πίνακας 5.5: Η σημασιολογία των εννοιολογικών εκφράσεων στη λογική  $ALCH_{R+}$**

Σύνταξη	Ερμηνεία
$C \subseteq D$	$C^I \subseteq D^I$
$C = D$	$C^I = D^I$
$R \subseteq S$	$R^I \subseteq S^I$
$R \in R_+$	$R^I = (R^I)^+$

**Πίνακας 5.6: Η σημασιολογία των αξιωμάτων στη λογική  $ALCH_{R+}$**

Το FaCT είναι ένας ορολογικός ταξινομητής (TBox) ο οποίος έχει αναπτυχθεί για την υλοποίηση του αλγόριθμου ελέγχου ικανοποίησης λογικής  $ALCH_{R+}$  χρησιμοποιώντας πολλές τεχνικές βελτιστοποίησης. Το FaCT έχει τη δυνατότητα συλλογισμού σχετικά με περιγραφές εννοιών, ρόλων και γνωρισμάτων και διατηρεί μια ιεραρχία εννοιών βασισμένη στη σχέση της υπαλληλίας. Ο αλγόριθμος που χρησιμοποιείται για τον έλεγχο υπαλληλίας είναι ο συνήθης: Η έννοια C είναι υπάλληλη της D αν και μόνο αν η  $D \cap \neg C$  δεν είναι ικανοποιήσιμη. Η αντιστοιχία που υπάρχει ανάμεσα στην τροπική και την προτασιακή λογική (Schild 1991) δίνει τη δυνατότητα στο FaCT να χρησιμοποιηθεί ως γεννήτρια απόδειξης θεωρημάτων για τις προτασιακές τροπικές λογικές  $K_{(m)}$ ,  $KT_{(m)}$ ,  $K4_{(m)}$ , και  $S4_{(m)}$ . Σύμφωνα με το [Hor98b] η απόδοση του FaCT είναι απόλυτα ικανοποιητική και πολλές φορές ξεχωρίζει από τα υπόλοιπα συστήματα. Το μεγαλύτερο πλεονέκτημά του είναι οι

τεχνικές βελτιστοποίησης που εφαρμόζει και οι οποίες μικραίνουν το χρόνο απόκρισης του συστήματος στην ταξινόμηση των εννοιών.

Για τη μεταφορά του θησαυρού στο FaCT χρησιμοποιήσαμε τις αντιστοιχίσεις του πίνακα 5.4. Στη συνέχεια ζητήσαμε από το σύστημα, αφού διαβάσει τις περιγραφές εννοιών και ρόλων, να τις ταξινομήσει. Κατά την ταξινόμηση, διαπιστώσαμε ότι το σύστημα έχει τη δυνατότητα εντοπισμού κύκλων καθώς και των ασυνεπών καταστάσεων που έχουμε περιγράψει. Ωστόσο, χρειάστηκε πάνω από 100 ώρες για να ολοκληρώσει την ταξινόμηση. Η καθυστέρηση αυτή οφείλεται στον μεγάλο όγκο εννοιών που καλείται το σύστημα να ταξινομήσει. Διαπιστώνουμε δηλαδή ότι παρότι το σύστημα χρησιμοποιεί τεχνικές βελτιστοποίησης, δεν καταφέρνει να ταξινομήσει τις έννοιες ενός συνήθη θησαυρού όρων σε λογικό χρονικό διάστημα.

### *Άλλα συστήματα*

Κατά τη διάρκεια έρευνας μας, μεταφέραμε το θησαυρό και σε άλλα συστήματα και πιο συγκεκριμένα στα CYCLOP, KRIS, Loom αλλά κανένα δεν κατάφερε να ολοκληρώσει τον αλγόριθμο ταξινόμησης των εννοιών, λόγω του πολύ μεγάλου Tbox (έννοιες και συσχετίσεις) που έπρεπε να εισάγουμε.

## *Συμπεράσματα*

Το κυριώτερο συμπέρασμα που προκύπτει είναι ότι όσα συστήματα αναγνωρίζουν μεγάλο αριθμό τελεστών, και κυρίως τα *or,not*, εκτελούν πολύπλοκους και χρονοβόρους αλγόριθμους, με αποτέλεσμα όταν εισάγουμε σε αυτά μεγάλο αριθμό εννοιών είτε να καθυστερούν υπερβολικά να ολοκληρώσουν την ταξινόμηση, είτε να μην την ολοκληρώνουν ποτέ. Από την άλλη μεριά, συστήματα που δέχονται περιορισμένο αριθμό τελεστών και συνεπώς έχουν απλούστερους αλγόριθμους, θέτουν και άλλους περιορισμούς που δεν είναι αποδεκτοί από τις προδιαγραφές που είχαμε θέσει. Γενικά ενώ πολλά από τα συστήματα περιγραφικής λογικής αντιμετωπίζουν με επιτυχία προβλήματα ασυνέπειας και κύκλων, δεν μπορούν να αντιμετωπίσουν το μεγάλο όγκο δεδομένων, με αποτέλεσμα να μην είναι εφικτή η αντιμετώπιση των προβλημάτων που είχαμε θέσει απλά με το να μεταφέρουμε το θησαυρό σε κάποιο σύστημα περιγραφικής λογικής.

Μια λύση στο πρόβλημά μας είναι να χρησιμοποιήσουμε τον αλγόριθμο του FaCT, αρκεί να εντοπίσουμε τα σημεία στα οποία ο μεγάλος όγκος δεδομένων κάνει τον αλγόριθμο μη αποδοτικό. Μια πολύ σημαντική βελτίωση είναι να μην κάνει ελέγχους για τους τελεστές *or* και *not*, τους οποίους δεν τους χρησιμοποιούμε στους θησαυρούς. Ωστόσο το θέμα δεν είναι απλό και χρειάζεται ακόμα πολλή έρευνα, ώστε να διαμορφώσουμε έναν αλγόριθμο που θα αντιμετωπίζει σε λογικό χρονικό διάστημα τα προβλήματα που έχουμε θέσει. Το χρονικό πλαίσιο της εργασίας μας δεν ήταν αρκετό ώστε να μπορέσουμε να σχεδιάσουμε και να υλοποιήσουμε έναν τέτοιο αλγόριθμο. Γιαυτόν ακριβώς το λόγο επικεντρώσαμε το ενδιαφέρον μας στην ανίχνευση ασυνεπειών κατά τη δημιουργία σύνθετων όρων.

Για το θέμα αυτό μελετήσαμε το Galen, ένα ερευνητικό πρόγραμμα στο πλαίσιο του οποίου έχει αναπτυχθεί μια θεωρία για τη σύνθεση δυναμικών όρων. Το Galen μας βοήθησε να αναπτύξουμε μια θεωρία σχετική με τη σύνθεση δυναμικών όρων, στην οποία βασιστήκαμε για να αναπτύξουμε ένα αντίστοιχο σύστημα.

## *Galen*

GALENείναι το όνομα που δόθηκε σε μια τεχνολογία που έχει σχεδιαστεί για να αναπαριστά ιατρική πληροφορία με έναν εντελώς νέο τρόπο. Το GALEN παράγει ένα πολυγλωσσικό κωδικοποιημένο

σύστημα για ιατρικούς σκοπούς([RBG<sup>+</sup>97],[RGG<sup>+</sup>94] Το ερευνητικό αυτό πρόγραμμα ανέπτυξε μια ορολογία (μοντέλο) κλινικών όρων στην οποία αναπαρίστανται ιατρικές έννοιες. Η γλώσσα αναπαράστασης που χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου αυτού ονομάζεται GRAIL [RSNR94],[ZHF<sup>+</sup>97]. Η γλώσσα αυτή είναι ουσιαστικά μια γλώσσα αναπαράστασης γνώσης, ένα σύστημα περιγραφικής λογικής, με κάποια ιδιαίτερα χαρακτηριστικά: από τους γνωστούς τελεστές σύνθεσης εννοιών (and, or, not) διαθέτει μόνο το and και μόνο για απλές περιπτώσεις, αλλά έχει τη δυνατότητα αυτόματης ταξινόμησης των εννοιών, έχει αναπτύξει μια πλειάδα συσχετίσεων μέρους-όλου και θέτει περιορισμούς για το πώς δυο βασικές έννοιες μπορούν να συνδυαστούν για το σχηματισμό σύνθετων εννοιών [RRP96].

Οι σύνθετες έννοιες παράγονται με τη βοήθεια του τελεστή which. Για παράδειγμα:

Hand which hasLaterality leftLaterality

Κάθε σύνθετη έννοια αποτελείται από μια 'βάση' (π.χ. Hand) και έναν 'ορισμό' ( π.χ. hasLaterality leftLaterality). Η βάση αποτελείται από μια βασική έννοια και ο ορισμός από ένα σύνολο κριτηρίων. Ο ορισμός αποτελείται από μια συσχέτιση (hasLaterality) και μια άλλη βασική έννοια (leftLaterality) . Η συσχέτιση αυτή συνδέει τις κατηγορίες (BodyPart, Laterality) στις οποίες ανήκουν η βάση και η έννοια που υπάρχει στον 'ορισμό' (Hand, leftLaterality). Δηλαδή, για να συνδυαστούν δυο όροι για το σχηματισμό μιας σύνθετης θα πρέπει οι κατηγορίες στις οποίες ανήκουν να συνδέονται με κάποια συσχέτιση. Αν υπάρχουν συσχετίσεις μεταξύ των κατηγοριών των όρων, τότε μόνο αυτές μπορούν να χρησιμοποιηθούν για την παραγωγή της σύνθετης έννοιας.

Αν και η γλώσσα GRAIL λόγω των περιορισμένων δυνατοτήτων που έχει ως προς τους τελεστές που παρέχει δε μπορεί να χρησιμοποιηθεί για να αντιμετωπίσει όλα τα προβλήματα που είχαμε θέσει, μπορεί να μας δώσει ιδέες για να αντιμετωπίσουμε το θέμα της σύνθεσης δυναμικών όρων.

Στο επόμενο κεφάλαιο θα αναπτύξουμε μια θεωρία σχετική με δυναμικούς όρους και θα θέσουμε τις απαιτήσεις μας για το σύστημα που θα κατασκευάσουμε.

---

## Κεφάλαιο 6ο

### Δυναμικοί Όροι

Δυναμικός όρος ονομάζεται ο όρος που χρησιμοποιείται για την αναπαράσταση μιας έννοιας και ο οποίος δεν είναι όρος του θησαυρού, αλλά μπορεί να δημιουργηθεί από το συνδυασμό δυο ή περισσότερων όρων του θησαυρού. Υπάρχουν δυο τρόποι σύνθεσης δυναμικών όρων: Με τη χρήση συσχετίσεων που έχουν οριστεί στο μετα-μοντέλο του συστήματος και με τη χρήση λογικών τελεστών.

#### 6.1 Δημιουργία σύνθετων όρων με χρήση συσχετίσεων

Όταν σε ένα θησαυρό θέλουμε να δημιουργήσουμε έναν δυναμικό όρο, θα πρέπει να ξέρουμε το πώς θα συνδυάσουμε τους όρους του θησαυρού για το σχηματισμό του δυναμικού όρου. Αν δεν θέσουμε στο χρήστη του θησαυρού κάποιους περιορισμούς που θα σχετίζονται με το πώς θα συνδυάσει τους όρους για την παραγωγή του δυναμικού, είναι πολύ πιθανόν να κάνει συνδυασμούς που θα οδηγήσουν το θησαυρό σε ασυνεπή κατάσταση. Για παράδειγμα, κατά τη σύνδεση του MERIMEE με τον AAT, κατά τη δημιουργία δυναμικών όρων, συνδυάζονται οι όροι με τη βοήθεια του τελεστή &, ο οποίος έχουμε πει ότι έχει την ερμηνεία της τομής ή της παράθεσης. Συνεπώς, για να δημιουργηθεί ο δυναμικός όρος με την ερμηνεία "factories that produce wooden toys" χρησιμοποιούνται οι όροι "factories", "wooden toys" συνδυαζόμενοι με το &, παράγοντας τον όρο "wooden toys & factories". Ωστόσο, όπως σχολίασαμε στο 2ο κεφάλαιο, ο τελεστής & δεν είναι η καλύτερη λύση στη σύνδεση όρων για την παραγωγή ενός δυναμικού.

Όπως είναι γνωστό, οι όροι ενός θησαυρού είναι ταξινομημένοι σε ιεραρχίες (hierarchies) και αυτές με τη σειρά τους σε όψεις (facets). Τόσο μεταξύ όψεων, όσο και μεταξύ ιεραρχιών μπορούμε να ορίσουμε συσχετίσεις που θα χρησιμοποιούνται για τη σύνθεση δυναμικών όρων. Μόνο όροι των οποίων οι ιεραρχίες ή οι όψεις συδέονται με συσχετίσεις μπορούν να συνδυαστούν χρησιμοποιώντας τη συγκεκριμένη συσχέτιση. Οι συσχετίσεις αυτές, επειδή θα είναι δηλωμένες στο μετα-μοντέλο του θησαυρού, ονομάζονται μετα-συσχετίσεις. Για τον θησαυρό που χρησιμοποιούμε ως παράδειγμα,

μπορούμε να δούμε στο σχήμα 6.1 ένα υποσύνολο των συσχετίσεων που επιλέξαμε για να συνδέσουμε τις κατηγορίες του.

Error! Not a valid link.

### Σχήμα 6.1 Εισαγωγή συσχετίσεων στις ιεραρχίες του AAT

Ένας δυναμικός όρος που δημιουργείται με αυτό τον τρόπο έχει τρία βασικά χαρακτηριστικά:

- **Βασικός Όρος:** Είναι ο όρος ο οποίος αναπαριστά μια έννοια ευρύτερη από την αντίστοιχη του δυναμικού και ο οποίος συμμετέχει στη σύνθεση του δυναμικού όρου. Για παράδειγμα, η σύνθετη έννοια "factories which are\_built\_for woodworking" έχει βασική έννοια τον όρο "factories". Ο όρος "factories" είναι ο βασικός όρος και του δυναμικού όρου "factories which are\_built\_for producing which has\_result toys".
- **Όρος τροποποίησης:** Είναι ο όρος ο οποίος χρησιμοποιείται στο δυναμικό όρο για να εξειδικεύσει την έννοια που αναπαριστά ο βασικός όρος. Στο προηγούμενο παράδειγμα, ο όρος "woodworking" είναι ο όρος τροποποίησης. Ένας όρος τροποποίησης μπορεί να είναι είτε δυναμικός είτε απλός. Για παράδειγμα, στο δυναμικό όρο "factories which are\_built\_for producing which has\_result toys", ο όρος τροποποίησης είναι ο δυναμικός όρος "producing which has\_result toys", ο οποίος με τη σειρά του έχει βασικό όρο το "producing" και όρο τροποποίησης τον "toys". Επιπλέον είναι δυνατόν ένας δυναμικός όρος να έχει παραπάνω από έναν όρους τροποποίησης οι οποίοι θα συνδυάζονται με τη χρήση των λογικών τελεστών που θα δούμε παρακάτω.
- **Συσχέτιση τροποποίησης:** Είναι η μέτα-συσχέτιση η οποία χρησιμοποιείται για τη δημιουργία του δυναμικού όρου και η οποία θα πρέπει να συσχετίζει κάποια από τις ιεραρχίες ή τις όψεις στις οποίες ανήκουν ο βασικός όρος και ο όρος τροποποίησης, επιτρέποντας με αυτό τον τρόπο το συνδυασμό τους. Στην περίπτωση που ο όρος τροποποίησης είναι δυναμικός, η μέτα-συσχέτιση θα πρέπει να συνδέει τις ιεραρχίες ή τις όψεις στις οποίες ανήκουν ο βασικός όρος του υπό σύνθεση δυναμικού όρου και ο βασικός όρος του όρου τροποποίησης. Στο σύνθετο παράδειγμα που δώσαμε προηγουμένως, η συσχέτιση μετωτροπής για το δυναμικό όρο "factories which are\_built\_for producing which has\_result toys", είναι η "are\_built\_for", η οποία συσχετίζει τις ιεραρχίες "Single Built Works" και "Processes and Techniques" στις οποίες έχουν ταξινομηθεί οι όροι "factories" και

"producing" αντίστοιχα. Ομοίως για το δυναμικό όρο που παίζει το ρόλο του όρου τροποποίησης, η συσχέτιση τροποποίησης είναι η "has\_result", η οποία συσχετίζει τις όψεις "ACTIVITIES" και "OBJECTS", στις οποίες έχουν τοποθετηθεί οι όροι "producing" και "toys" αντίστοιχα. Ένας όρος μπορεί να έχει περισσότερες από μια συσχετίσεις τροποποίησης, η οποίες θα συνδυάζονται με τη χρήση λογικών τελεστών, τους οποίους θα δούμε στη συνέχεια.

Τα τρία αυτά χαρακτηριστικά συνδυάζονται μαζί με τον τελεστή which για τη δημιουργία του δυναμικού όρου. Η γενική μορφή του δυναμικού όρου, χωρίς χρήση λογικών τελεστών, είναι η εξής:

*Βασικός\_Όρος which Συσχέτιση\_τροποποίησης Όρος\_τροποποίησης,*

όπου ο Όρος\_τροποποίησης μπορεί να είναι δυναμικός όρος, όπως έχουμε ήδη σημειώσει.

## 6.2 Η χρήση λογικών τελεστών στη δημιουργία δυναμικών όρων

Όπως είδαμε στην προηγούμενη ενότητα, ένας δυναμικός όρος μπορεί να αποτελείται από έναν βασικό και πολλούς όρους τροποποίησης ή/και πολλές συσχετίσεις τροποποίησης. Στην περίπτωση αυτή θα πρέπει να χρησιμοποιήσουμε τους λογικούς τελεστές για να δηλώσουμε τη συσχέτιση που υπάρχει μεταξύ των πολλαπλών όρων τροποποίησης ή των πολλαπλών συσχετίσεων τροποποίησης. Υπάρχουν συνολικά τρεις τρόποι χρήσης των λογικών τελεστών στο πλαίσιο των δυναμικών όρων:

- Βασικός\_Όρος which Συσχέτιση\_1 (Λογικός\_Τελεστής Όρος1 ... ΌροςN). Σε αυτή την περίπτωση, συνδέουμε έναν βασικό όρο με πολλούς όρους τροποποίησης χρησιμοποιώντας την ίδια συσχέτιση. Για παράδειγμα, "factories which are\_built\_for (or woodworking grinding)".
- Βασικός\_Όρος which (Λογικός\_Τελεστής (Συσχέτιση1 Όρος1) ... (ΣυσχέτισηN Όρος N)). Σε αυτή την περίπτωση συνδέουμε έναν βασικό όρο με πολλούς όρους τροποποίησης, χρησιμοποιώντας όμως διαφορετική συσχέτιση για κάθε όρο. Για παράδειγμα, "factories which (and (are\_built\_for woodworking) (are\_used\_for woodcarving))".
- Όρος1 Λογικός\_Τελεστής ... Λογικός\_Τελεστής ΌροςN. Σε αυτή την περίπτωση, οι λογικοί τελεστές θα πρέπει να συνδέουν ομογενείς όρους, δηλαδή όρους που έχουν ιεραρχική συσχέτιση. Οι όροι αυτοί μπορεί να είναι και δυναμικοί, οι οποίοι όμως αν αναλυθούν μέχρι τέλους αποτελούνται από όρους του θησαυρού. Για παράδειγμα, "factories which

are\_built\_for producing toys which are\_made\_of wood or factories which are\_built\_for woodworking".

Είναι αυτονόητο ότι κάθε μια από τις παραπάνω περιπτώσεις μπορεί να συνδυαστεί με τις υπόλοιπες οδηγώντας σε πολύπλοκες δηλώσεις.

Με τη βοήθεια των δυο παρακάτω θεωρημάτων μπορούμε να μετατρέψουμε τις δυο πρώτες μορφές στην τρίτη.

**Θεώρημα 1** Ένας δυναμικός όρος με μια συσχέτιση τροποποίησης που σχετίζει το βασικό όρο με  $N$  όρους τροποποίησης συνδυαζόμενους με έναν λογικό τελεστή, είναι ισοδύναμος με  $N$  δυναμικούς όρους συνδεόμενους με τον ίδιο λογικό τελεστή και οι οποίοι έχουν τον ίδιο βασικό όρο, την ίδια συσχέτιση τροποποίησης και όρο τροποποίησης έναν από τους  $N$  όρους. Δηλαδή, ο δυναμικός όρος “Βασικός\_Όρος Συσχέτιση\_Τροποποίησης (Λογικός\_Τελεστής Όρος\_Τροποποίησης1 ... Όρος\_Τροποποίησης $N$ )” είναι ισοδύναμος με το δυναμικό όρο “Βασικός\_Όρος Συσχέτιση\_Τροποποίησης Όρος\_Τροποποίησης1 Λογικός\_Τελεστής ... Λογικός\_Τελεστής Βασικός\_Όρος Συσχέτιση\_Τροποποίησης Όρος\_Τροποποίησης $N$ ”.

## ΑΠΟΔΕΙΞΗ ΓΙΑ ΤΟ AND

$$A \text{ R (AND } C \text{ D)} \Leftrightarrow$$

$$\exists (A \cap R.(C \cap D)) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in (C \cap D)^I \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in (C^I \cap D^I) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \wedge c \in D^I \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \wedge \exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in D^I \Leftrightarrow$$

$$\exists (A \cap R.C \cap A \cap R.D) \Leftrightarrow$$

$$A \text{ R } C \text{ AND } A \text{ R } D$$

## ΑΠΟΔΕΙΞΗ ΓΙΑ ΤΟ OR

$$A \text{ R (OR } C \text{ D)} \Leftrightarrow$$

$$\exists (A \cap R.(C \cup D)) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in (C \cup D)^I \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in (C^I \cup D^I) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge (c \in C^I \vee c \in D^I) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \vee \exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in D^I \Leftrightarrow$$

$$\exists (A \wedge R.C \vee A \wedge R.D) \Leftrightarrow$$

$$A \wedge R.C \vee A \wedge R.D$$

**ΘΕΩΡΗΜΑ 2.** Ένας δυναμικός όρος με BN συσχετίσεις τροποποίησης συνδυαζόμενες με έναν λογικό τελεστή είναι ισοδύναμος με N δυναμικούς όρους συνδυαζόμενους με τον ίδιο λογικό τελεστή και οι οποίοι έχουν τον ίδιο βασικό όρο και μια από τις N συσχετίσεις τροποποίησης. Δηλαδή ο δυναμικός όρος “Βασικός\_Όρος (Λογικός\_Τελεστής (Συσχέτιση\_τροποποίησης1 Όρος\_τροποποίησης1) ... (Συσχέτιση\_τροποποίησης N Όρος\_τροποποίησηςN))” είναι ισοδύναμος με το δυναμικό όρο “Βασικός\_Όρος Συσχέτιση\_τροποποίησης1 Όρος\_τροποποίησης1 Λογικός\_τελεστής ... Λογικός\_τελεστής Βασικός\_Όρος Συσχέτιση\_τροποποίησηςN”.

## ΑΠΟΔΕΙΞΗ ΓΙΑ ΤΟ AND

$$A \wedge (R.C) \wedge (P.D) \Leftrightarrow$$

$$\exists (A \wedge (R.C \wedge P.D)) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \wedge \exists (a,d) \in P^I \wedge d \in D^I \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \wedge \exists a \in A^I \wedge \exists (a,d) \in P^I \wedge d \in D^I \Leftrightarrow$$

$$\exists (A \wedge R.C \wedge A \wedge P.D) \Leftrightarrow$$

$$A \wedge R.C \wedge A \wedge P.D$$

## ΑΠΟΔΕΙΞΗ ΓΙΑ ΤΟ OR

$$A \vee (R.C) \vee (P.D) \Leftrightarrow$$

$$\exists (A \vee (R.C \vee P.D)) \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \vee \exists (a,d) \in P^I \wedge d \in D^I \Leftrightarrow$$

$$\exists a \in A^I \wedge \exists (a,c) \in R^I \wedge c \in C^I \vee \exists a \in A^I \wedge \exists (a,d) \in P^I \wedge d \in D^I \Leftrightarrow$$

$$\exists (A \vee R.C \vee A \vee P.D) \Leftrightarrow$$

$$A \vee R.C \vee A \vee P.D$$

Με τη βοήθεια των παραπάνω θεωρημάτων, μπορούμε να μετατρέψουμε κάθε δυναμικό όρο που έχει δημιουργηθεί με τη βοήθεια λογικών τελεστών σε μια κοινή μορφή. Η κοινή αυτή μορφή μας δείχνει την βάση όρων από την οποία προέκυψε ο δυναμικός όρος. Μπορούμε έτσι να δημιουργήσουμε ευρύτερους ή στενότερους όρους, απλώς αντικαθιστώντας έναν από τους βασικούς όρους με έναν ευρύτερο ή ειδικότερο. Για παράδειγμα, ο δυναμικός όρος "factories and car assembly plants" έχει δυο βασικούς όρους, τον "factories" και τον "car assembly plants".

### 6.3 Άμεση υπερκλάση

Άμεση υπερκλάση ορίζεται η ειδικότερη από τις υπερκλάσεις ενός όρου και η οποία είτε υπάρχει στο σύστημα είτε μπορεί να δημιουργηθεί ως δυναμικός όρος. Για παράδειγμα, αν έχω το δυναμικό όρο "factories which are\_built\_for producing which has\_results toys which are\_made\_of wood", έχει άμεσες υπερκλάσεις τους όρους:

- "factories which are\_built\_for producing which has\_result <recreational artifacts for noncompetitive activities> which are\_made\_of wood",
- "factories which are\_built\_for producing which has\_result toys which are\_made\_of plant communities" και
- "factories which are\_built\_for <processes and techniques by specific type> which has\_result toys which are\_made\_of wood".

Δηλαδή σχηματίζουμε νέους όρους που προκύπτουν παίρνοντας τον ευρύτερο όρο κάθε όρου μετατροπής.

Στην περίπτωση που ο όρος έχει δημιουργηθεί χρησιμοποιώντας τον λογικό τελεστή AND, τον μετατρέπουμε στην κανονική μορφή και ορίζουμε σαν άμεσες υπερκλάσεις τους βασικούς όρους από τους οποίους αποτελείται. Στο παράδειγμα της προηγούμενης ενότητας, άμεσες υπερκλάσεις του όρου "factories and car assembly plants" είναι οι όροι "factories" και "car assembly plants".

Εαν ο τελεστής που χρησιμοποιείται για τη δημιουργία του δυναμικού όρου είναι το OR, θεωρούμε ότι έχει σαν άμεση υπερκλάση τον πρώτο όρο ο οποίος είναι ευρύτερος όλων των βασικών όρων του. Για παράδειγμα, ο δυναμικός όρος "historians of religion or church historians" έχει σαν άμεση υπερκλάση τον όρο "<people in religion-related occupations>", ο οποίος είναι η ειδικότερη από τις

κοινές υπερκλάσεις των βασικών όρων “historians of religion” και “church historians” που απαρτίζουν το δυναμικό.

## 6.4 Λειτουργίες του συστήματος

Η θεωρία που αναπτύξαμε προηγουμένως σχετικά με τους δυναμικούς όρους, θα μας χρησιμεύσει στη διαχείριση των δυναμικών όρων ενός θησαυρού. Πιο συγκεκριμένα, οι λειτουργίες που θα θέλαμε να εκτελεί ένας θησαυρός και οι οποίες σχετίζονται με τους δυναμικούς όρους είναι οι εξής:

- Κατά τη δήλωση ενός δυναμικού όρου στο σύστημα, να διαπιστώνεται αν το σύστημα παραμένει συνεπές. Να ελέγχεται δηλαδή αν επιτρέπεται ο συνδυασμός των όρων που έχει δηλώσει ο χρήστης για να δημιουργήσει το δυναμικό. Για παράδειγμα, αν δηλώσει ο χρήστης τον δυναμικό όρο "factories which are\_built\_for grinding", χωρίς να έχει δηλωθεί η συσχέτιση are\_built\_for μεταξύ των ιεραρχιών ή των όψεων στις οποίες ανήκουν οι όροι "factories" και "grinding" θα πρέπει το σύστημα να τον ενημερώσει ότι δεν επιτρέπεται τέτοια δήλωση. Η λειτουργία αυτή μπορεί να γίνεται πιο αυτοματοποιημένη, αν το σύστημα αντί να ελέγχει το χρήστη όταν εισάγει δυναμικούς όρους, να τον βοηθάει στο σχηματισμό τους. Πιο συγκεκριμένα, ο χρήστης του συστήματος θα επιλέγει τους όρους που θέλει να συνδυάσει και το σύστημα θα του προτείνει τις επιτρεπόμενες συσχετίσεις που μπορούν να χρησιμοποιηθούν για τη σύνθεση του δυναμικού όρου.
- Επιλέγοντας έναν όρο να μπορεί να εμφανίζεται η άμεση υπερκλάση του. Ένα απλό παράδειγμα είναι στην περίπτωση που δηλώσουμε το δυναμικό όρο "factories which are\_built\_for woodcarving". Ο όρος αυτός έχει άμεση υπερκλάση τον "factories which are\_built\_for woodworking", αφού ο όρος "woodworking" είναι ο ευρύτερος του "woodcarving".
- Ο χρήστης να μπορεί να επιλέξει έναν όρο ως βασικό και μέσω μιας συσχέτισης και μιας ιεραρχίας, να μπορεί να δει την ιεραρχία των δυναμικών όρων που σχηματίζονται. Για να επιλέξει την ιεραρχία των όρων τροποποίησης, αρκεί να επιλέξει τον όρο που βρίσκεται στην κορυφή της ιεραρχίας. Φυσικά, υπάρχει η περίπτωση κάποιος από τους όρους που θα εμφανιστούν στην ιεραρχία να υπάρχει ήδη στο θησαυρό. Για παράδειγμα, αν επιλέξουμε να δούμε τους δυναμικούς όρους που προκύπτουν από τον όρο "factories" συνδυαζόμενο με την ιεραρχία "Processes and Techniques", χρησιμοποιώντας τη συσχέτιση "are\_built\_for", θα δημιουργηθεί κάποια στιγμή και ο όρος "factories which are\_built\_for grinding", αφού ο όρος "grinding" ανήκει στην ιεραρχία "Processes and Techniques". Στον AAT όμως, υπάρχει ο όρος mills, ο οποίος έχει ακριβώς αυτή την ερμηνεία. Συνεπώς, στην ιεραρχία με τους δυναμικούς

όρους που δημιουργούνται, στη θέση του "factories which are\_built\_for grinding" θα πρέπει να εμφανιστεί ο όρος "mills".

## 6.5 Αναπαράσταση των δυναμικών όρων στο σύστημα

Ένα αρκετά σημαντικό θέμα που μας απασχολεί είναι το πώς θα αναπαριστούμε τους δυναμικούς όρους στο σύστημά μας. Η πιο εύκολη προσέγγιση θα ήταν να αναπαριστούμε έναν δυναμικό όρο με μια οντότητα, της οποίας το λογικό όνομα θα είναι η πρόταση που προκύπτει από το συνδυασμό του βασικού όρου με τους όρους τροποποίησης μέσω των επιτρεπόμενων συσχετίσεων. Για παράδειγμα για να αναπαραστήσουμε την έννοια factories built for grinding στο AAT, αρκεί να δημιουργήσουμε μια οντότητα με το όνομα "factories which are\_built\_for grinding". Η οντότητα αυτή θα έχει σαν επιπλέον γνωρίσματα, το βασικό όρο, τους όρους τροποποίησης και τις συσχετίσεις από τις οποίες προέκυψε. Ωστόσο η προσέγγιση αυτή έχει δυο κυρίως μειονεκτήματα:

- Συνήθως υπάρχουν άλλοι όροι που αναπαριστούν την ίδια έννοια και οι οποίοι είναι ευρύτερα χρησιμοποιούμενοι. Στη συγκεκριμένη περίπτωση, ο όρος mills αναπαριστά την ίδια έννοια και είναι προφανώς ευρύτερα χρησιμοποιούμενος από τον όρο "factories which are\_built\_for grinding". Το φαινόμενο είναι περισσότερο εμφανές στην περίπτωση πιο πολύπλοκων δυναμικών όρων, όπως ο "complexes which are composed of factories which are\_built\_for producing which has\_result toys which are\_made\_of wood".
- Στην περίπτωση που ο δυναμικός αυτός όρος έχει σαν "συστατικά" του άλλους δυναμικούς όρους, θα πρέπει να τους δηλώσουμε και αυτούς υποχρεωτικά, ακόμα και αν δε θέλουμε να τους έχουμε στο θησαυρό. Για παράδειγμα, στο δυναμικό όρο "factories which are\_built\_for producing which has\_result toys", ο όρος τροποποίησης του factories είναι ο δυναμικός "producing which has\_result toys", ο οποίος θα πρέπει επίσης να δηλωθεί στο θησαυρό αφού είναι ο όρος\_τροποποίησης του αρχικού δυναμικού όρου. Με αυτό τον τρόπο δημιουργείται μια αλυσίδα από δυναμικούς όρους που δε χρειαζόμαστε, με αποτέλεσμα να επιβαρύνουμε το σύστημα.

Μια λύση στο πρόβλημα θα ήταν να αναλύουμε έναν δυναμικό όρο στους απλούς όρους από τους οποίους αποτελείται και να ορίσουμε γνωρίσματα που θα δείχνουν απευθείας σε αυτούς. Στο προηγούμενο παράδειγμα, θα δηλώναμε ότι ο όρος "factories which are\_built\_for producing which has\_result toys" έχει ως βασικό όρο τον όρο "factories", ως συσχετίσεις τροποποίησης τις "are\_built\_for" και "has\_result" και ως όρους τροποποίησης τους "producing" και "toys". Με αυτό

τον τρόπο, δε χρειάζεται να δημιουργούμε οντότητες που δεν θέλουμε. Το πρόβλημα είναι πώς θα γνωρίζουμε ότι ο όρος "toys" δεν είναι όρος τροποποίησης του "factories", αλλά του όρου "producing". Αυτό λύνεται με το να ορίσουμε ένα επιπλέον γνώρισμα του δυναμικού όρου, με τίτλο "has\_description", το οποίο θα δείχνει σε ένα αλφαριθμητικό που θα έχει την περιγραφή του δυναμικού όρου, δηλαδή το "factories which are\_built\_for producing which has\_result toys". Το όνομα με το οποίο ο δυναμικός όρος αποθηκεύεται στο σύστημα θα είναι προφανώς κάποιο ευρύτερα χρησιμοποιούμενο από την περιγραφή του (has\_description), όπως για παράδειγμα: "toys' factory". Με το γνώρισμα "has\_description" θα μπορούμε να διαπιστώσουμε το πώς συνδυάζονται ή σχετίζονται οι όροι τροποποίησης μεταξύ τους. Στο προηγούμενο παράδειγμα θα είχαμε τα γνωρίσματα: base\_term **factories**, term\_modifier **producing**, role\_modifier **are\_built\_for**, term\_modifier **toys**, role\_modifier **has\_result** και has\_description **factories which are\_built\_for producing which has\_result toys**.

## 6.6 Σχεδιασμός Διεργασιών

Στην ενότητα αυτή περιγράφουμε τα βασικά βήματα των διεργασιών που σχεδιάσαμε για να επιτελέσουμε τις παραπάνω λειτουργίες. Τα παραδείγματα που χρησιμοποιούμε είναι πραγματικά και προέρχονται από την εφαρμογή του συστήματός μας στο θησαυρό όρων AAT.

### 6.6.1 Σύνθεση Δυναμικού Όρου

Για τη σύνθεση ενός δυναμικού όρου, θα πρέπει να δώσει ο χρήστης κάποια στοιχεία σαν είσοδο και να αναλάβει το σύστημα να δημιουργήσει έναν κατάλληλο δυναμικό όρο. Πιο συγκεκριμένα, θα πρέπει ο χρήστης να δώσει το βασικό όρο και τους όρους τροποποίησης, οι οποίοι αν είναι επίσης σύνθετοι, θα πρέπει να ακολουθήσει την ίδια διαδικασία (βασικός όρος του όρου τροποποίησης και όροι τροποποίησης του όρου αυτού). Τα βήματα της διεργασίας είναι τα εξής:

1. Ο χρήστης δίνει στο σύστημα το βασικό όρο.
2. Ο χρήστης δίνει τον 1ο όρο τροποποίησης του βασικού όρου. Αν αυτός είναι δυναμικός, αρκεί να δώσει τον βασικό όρο του όρου τροποποίησης.
3. Το σύστημα εμφανίζει μια λίστα με μετα-συσχετίσεις που συνδέουν τις ιεραρχίες και τις όψεις στις οποίες ανήκουν ο βασικός όρος και ο όρος τροποποίησης. Ο χρήστης επιλέγει κάποια από αυτές για να χρησιμοποιηθεί στη σύνθεση του δυναμικού όρου.
4. Ο χρήστης καλείται να δηλώσει αν ο όρος τροποποίησης έχει δικούς του όρους τροποποίησης.

5. Επαναλαμβάνονται τα βήματα από το 2-4 μέχρι να δηλώσει ο χρήστης όλους τους όρους τροποποίησης του βασικού.
6. Για όσους όρους έχουν την ίδια συσχέτιση τροποποίησης δηλώνεται ο λογικός τελεστής που θα τους συνδέει.
7. Αν υπάρχουν διαφορετικές συσχετίσεις τροποποίησης, δηλώνεται ο λογικός τελεστής που θα τις συνδέει.
8. Για κάθε όρο τροποποίησης που ο χρήστης είχε δηλώσει ότι έχει δικούς του όρους τροποποίησης, επαναλαμβάνονται τα βήματα 2-7, θεωρώντας ως βασικό όρο το συγκεκριμένο όρο τροποποίησης.
9. Αν ο δυναμικός όρος που προκύπτει έχει λογικούς τελεστές, το σύστημα αναλαμβάνει να τον μετατρέψει στην κανονική μορφή.

**Παράδειγμα** Έστω ότι θέλουμε να παραγάγουμε το δυναμικό όρο που αναπαριστά την έννοια "factories which produce wooden or metal toys". Ο χρήστης δίνει το βασικό όρο factories και στη συνέχεια τον όρο τροποποίησης producing για τον οποίο δηλώνει ότι έχει και αυτός όρους τροποποίησης. Η ιεραρχία Single Built Works στην οποία ανήκει ο όρος factories και η ιεραρχία Processes and Techniques στην οποία ανήκει ο όρος producing, συνδέονται με τις συσχετίσεις are\_built\_for, are\_result\_of και are\_used\_for. Το σύστημα προτείνει αυτές τις συσχετίσεις στο χρήστη και αυτός επιλέγει τη συσχέτιση are\_built\_for. Μέχρι τη φάση αυτή έχει δημιουργηθεί ο όρος "factories which are\_built\_for producing" Για τον όρο τροποποίησης producing, ο χρήστης δηλώνει ότι έχει όρο τροποποίησης ενώ ο όρος factories δεν έχει άλλο όρο τροποποίησης. Συνεπώς επαναλαμβάνεται η ίδια διαδικασία θεωρώντας τον όρο producing ως βασικό. Όρος τροποποίησης του producing δηλώνεται ο όρος toys. Από τις συσχετίσεις που υπάρχουν μεταξύ της ιεραρχίας Processes and Techniques και Recreational Artifacts, στην οποία ανήκει ο όρος toys, ο χρήστης επιλέγει τη συσχέτιση has\_result. Ο όρος που έχει δημιουργηθεί μέχρι τώρα είναι ο "factories which are\_built\_for producing which has\_result toys". Ο όρος toys έχει όρο τροποποίησης, ενώ ο όρος producing δεν έχει άλλο όρο τροποποίησης, οπότε επαναλαμβάνεται η διαδικασία για τον όρο toys. Ο όρος toys έχει επίσης όρους τροποποίησης, οι οποίοι είναι το wood και το metal, οι οποίοι ανήκουν στην ιεραρχία Materials. Από τις προτεινόμενες συσχετίσεις, ο χρήστης επιλέγει και για τους δυο όρους την ίδια συσχέτιση, και πιο συγκεκριμένα, την are\_made\_of. Αφού οι δυο όροι έχουν την ίδια συσχέτιση τροποποίησης, ο χρήστης καλείται να δηλώσει τον τελεστή με τον οποίο αυτοί συνδέονται, και ο οποίος στη συγκεκριμένη περίπτωση είναι ο or. Συνεπώς ο όρος που έχει δημιουργηθεί είναι ο "factories which are\_built\_for producing which has\_result toys which are\_made\_of (or wood metal)", ο οποίος μετατρέπεται στην κανονική μορφή "factories which

are\_built\_for producing which has\_result toys which are\_made\_of wood or factories which are\_built\_for producing which has\_result toys which are\_made\_of metal".

### 6.6.2 Εύρεση άμεσης υπερκλάσης/υποκλάσης

Σε αυτή την περίπτωση θέλουμε να μπορούμε να βρούμε τις άμεσες υπερκλάσεις ή υποκλάσεις ενός δυναμικού όρου. Όπως είχαμε παρατηρήσει σε προηγούμενη ενότητα, αυτό επιτυγχάνεται με το να αντικαταστήσουμε έναν-έναν τους όρους τροποποίησης με τους ευρύτερους ή ειδικότερους τους. Τα βήματα της διεργασίας είναι τα εξής:

1. Ο χρήστης επιλέγει το δυναμικό όρο του οποίου τις υπερκλάσεις/υποκλάσεις θέλει να δει.
2. Το σύστημα βρίσκει τους όρους τροποποίησης του δυναμικού όρου και για καθένα από αυτούς εκτελεί το επόμενο βήμα:
3. Αντικαθιστά τον όρο τροποποίησης με τον ευρύτερό/ειδικότερό του και εμφανίζει ως άμεση υπερκλάση/υποκλάση του, το νέο δυναμικό όρο που προκύπτει. Αν ο δυναμικός όρος που προκύπτει χρησιμοποιείται ως περιγραφή κάποιου άλλου όρου του θησαυρού, τότε ως άμεση υπερκλάση/υποκλάση εμφανίζεται ο όρος αυτός. Αυτό το συμπεραίνουμε αν σε αυτό το δυναμικό όρο δείχνει το γνώρισμα has\_description κάποιου άλλου όρου.

**Παράδειγμα.** Έστω ότι θέλουμε να δούμε τις άμεσες υπερκλάσεις του δυναμικού όρου: “factories which are\_built\_for producing which has\_results toys which are\_made\_of wood”. Οι όροι τροποποίησης του δυναμικού όρου είναι οι producing, toys, wood. Για καθέναν από αυτούς παίρνουμε τους αμέσως ευρύτερους όρους, οι οποίοι είναι αντίστοιχα:

<processes and techniques by specific type>, <recreational artifacts for noncompetitive activities>, και plant communities. Οπότε οι δυναμικοί όροι που προκύπτουν και οι οποίοι είναι άμεσες υπερκλάσεις του αρχικού είναι οι:

“factories which are\_built\_for producing which has\_result <recreational artifacts for noncompetitive activities> which are\_made\_of wood”,

- “factories which are\_built\_for producing which has\_result toys which are\_made\_of plant communities” και
- “factories which are\_built\_for <processes and techniques by specific type> which has\_result toys which are\_made\_of wood”.

Κανέννας από τους παραπάνω όρους δεν έχει ορισθεί ως η περιγραφή κάποιου άλλου όρου του θησαυρού. Αν για παράδειγμα υπήρχε ο όρος “factories producing wooden recreational artifacts”, ο οποίος είχε περιγραφή (has\_description) “factories which are\_built\_for producing which has\_result <recreational artifacts for noncompetitive activities> which are\_made\_of wood”, θα αντικαθιστούσαμε την πρώτη υπερκλάση που βρήκαμε, με τον όρο αυτό.

Προφανώς, είναι δυνατόν οι άμεσες υπερκλάσεις του αρχικού δυναμικού όρου να μην έχουν κάποια λογική ερμηνεία, δηλ. να μην έχουν κάποιο νόημα για το θησαυρό (π.χ. “factories which are\_built\_for producing which has\_result toys which are\_made\_of plant communities”). Είναι στην ευχέρεια του χρήστη να επιλέξει αν αυτοί οι όροι που προκύπτουν έχει νόημα να αποθηκευθούν στο σύστημα ή όχι.

### 6.6.2 Δημιουργία Ιεραρχίας Δυναμικών Όρων

Μια πολύ σημαντική λειτουργία που θέλουμε να εξυπηρετεί το σύστημά μας είναι να μπορούμε να αναπαριστάνουμε μια ιεραρχία δυναμικών όρων που θα δημιουργούμε αυτόματα. Θέλουμε δηλαδή να μπορεί ο χρήστης να επιλέγει έναν όρο ο οποίος θα είναι ο βασικός του δυναμικού, μια μετασυσχέτιση και μια ιεραρχία όρων που θα χρησιμοποιήσει ως όρους τροποποίησης. Όπως έχουμε παρατηρήσει, για να επιλέξει την ιεραρχία αρκεί να επιλέξει τον όρο που βρίσκεται στην κορυφή της ιεραρχίας. Τα βήματα της διεργασίας είναι τα εξής:

1. Ο χρήστης δίνει το βασικό όρο. Από τον ορισμό του βασικού όρου προκύπτει ότι κάθε δυναμικός όρος που δημιουργείται είναι ειδικότερος από τον βασικό όρο. Θα πρέπει να σημειωθεί ότι ο βασικός όρος μπορεί να είναι και αυτός δυναμικός. Επειδή πρόκειται να δημιουργηθεί μια ιεραρχία δυναμικών όρων, ο χρήστης καλείται να δώσει έναν όρο οδηγό κάτω από τον οποίο θα δημιουργηθεί η ιεραρχία.
2. Ο χρήστης δίνει τον όρο τροποποίησης του οποίου την ιεραρχία ειδικότερων όρων θα χρησιμοποιήσουμε για να δημιουργήσουμε την ιεραρχία δυναμικών όρων.
3. Ο χρήστης επιλέγει κάποια από τις υπάρχουσες μετασυσχετίσεις τροποποίησης.
4. Διατρέχοντας την ιεραρχία του όρου τροποποίησης δημιουργούμε μια αντίστοιχη ιεραρχία δυναμικών όρων, συνδέοντας το βασικό όρο, τη συσχέτιση τροποποίησης και τον τρέχοντα όρο της ιεραρχίας. Για κάθε δυναμικό όρο, ελέγχουμε αν το όνομά του

είναι ίδιο με την περιγραφή (has\_description) κάποιου από τους όρους του θησαυρού και αν είναι, τον αντικαθιστούμε με τον όρο αυτό.

**Παράδειγμα.** Έστω ότι θέλουμε να παραστήσουμε στο θησαυρό τις βίλλες, οι ιδιοκτήτες των οποίων είναι ταυτόχρονα και ιδιοκτήτες κάποιου εργοστασίου. Η έννοια αυτή παριστάνεται στο θησαυρό από το δυναμικό όρο “villas which have\_owner owners which are\_owners\_of factories”, χρησιμοποιώντας τις επιτρεπόμενες συσχετίσεις μεταξύ των όρων “villas”, “owners” και “factories”. Αν θέλουμε όμως να δούμε όλους τους ιδιοκτήτες κάποιας βίλλας με βάση το είδος εργοστασίου που έχουν, χρησιμοποιούμε τον παραπάνω αλγόριθμο, θέτοντας ως σταθερό τμήμα το “villas which have\_owner owners which are\_owners\_of” και ως μεταβλητό τμήμα τον όρο “factories”. Στη συνέχεια, διατρέχουμε την ιεραρχία που έχει σαν κορυφή της τον όρο “factories” και δημιουργούμε τους νέους δυναμικούς όρους. Στο σχήμα 6.1 φαίνεται ένα τμήμα της ιεραρχίας καθώς και η αντίστοιχη ιεραρχία δυναμικών όρων που προκύπτει.

Error! Not a valid link.

Error! Not a valid link.

**Σχήμα 6.1** Ένα τμήμα της ιεραρχίας του όρου factories και η αντίστοιχη ιεραρχία δυναμικών όρων που προκύπτει

## 6.7 Αποτελέσματα

Όπως είδαμε και από τα παραδείγματα, οι διεργασίες που σχεδιάσαμε επιλύουν τα προβλήματα που προκύπτουν κατά τη σύνθεση ενός δυναμικού όρου και επιπλέον βοηθούν το χρήστη στη δημιουργία νέων ιεραρχιών όρων καθώς επίσης και στην εύρεση και εισαγωγή στο σύστημα δυναμικών όρων που είναι ευρύτεροι άλλων.

Εκτός από τα διάφορα πειράματα που κάναμε, δοκιμάσαμε να το εφαρμόσουμε στο παράδειγμα εκείνο που μας έδωσε το ερέθισμα για την εκπόνηση της συγκεκριμένης εργασίας. Πιο

συγκεκριμένα, όπως είχαμε πει κατά τη σύνδεση του γαλλικού θησαυρού όρων MERIMÉE με τον αντίστοιχο αμερικανικό θησαυρό AAT, υπήρχαν γαλλικοί όροι που δεν είχαν κάποιο αντίστοιχο ισοδύναμο στους όρους του AAT. Η λύση που δόθηκε ήταν να δημιουργηθούν δυναμικοί όροι στο AAT ονομαζόμενοι «Συλλεκτικές Έννοιες», οι οποίοι ήταν ουσιαστικά ένας συνδυασμός όρων του AAT. Έτσι για παράδειγμα, ο όρος του MERIMÉE “VOIRRIE” είχε σαν ισοδύναμο τον δυναμικό όρο “cities & streets”. Και οι δυο όροι αναπαριστούν την έννοια “δρόμοι μιας πόλης”. Ωστόσο στην περίπτωση του AAT με τον όρο “cities & streets” δεν είναι καθόλου ξεκάθαρο το ποια έννοια ακριβώς αναπαριστά. Αντίθετα, αν χρησιμοποιήσουμε το σύστημά μας, θα προκύψει ο δυναμικός όρος “streets which are\_located\_in cities”, δίνοντας το ακριβές νόημα που θέλαμε. Ομοίως ο δυναμικός όρος “factories & owners & villas” που περιγράφει τους ιδιοκτήτες βίλλας που είναι ταυτόχρονα και ιδιοκτήτες εργοστασίων, τον αντικαθιστούμε με τον όρο “factories and villas owners”, ο οποίος έχει σαν περιγραφή (has\_description) τον όρο “factories which have\_owners owners which are\_owners\_of villas”. στους σύνθετους όρους που δημιούργησε ο θησαυρός όρων MERIMÉE στον AAT. Συνολικά 2000 δυναμικοί όροι αντικαταστάθηκαν από άλλους που προέκυψαν από το σύστημά μας.

---

## Κεφάλαιο 7ο

### Επίλογος

#### 7.1 Συμπεράσματα

Το θέμα της εργασίας αυτής ήταν η οικονομία και η συνέπεια σε θησαυρούς όρων. Αφού κάναμε μια σύντομη εισαγωγή στα κυριότερα χαρακτηριστικά των θησαυρών όρων, παρουσιάσαμε το πρόβλημα προσπαθώντας να το αναλύσουμε στις σημαντικότερες συνισταμένες του.

Στη συνέχεια, αφού διαπιστώσαμε ότι υπάρχουν άλλοι τομείς έρευνας και συγκεκριμένα τα συστήματα περιγραφικής λογικής που ασχολούνται με παρόμοια ζητήματα, μεταφέραμε το θησαυρό και τις περιπτώσεις που θέλουμε να αντιμετωπίσουμε στα συστήματα αυτά. Διαπιστώσαμε ότι τα συστήματα περιγραφικής λογικής για να επιλύσουν τα θέματα που μας απασχολούν θέτουν πολλούς περιορισμούς μη αποδεκτούς στην περίπτωση μας και επιπλέον υποθέτουν ότι τα δεδομένα τα οποία χειρίζονται είναι περιορισμένου μεγέθους. Συνεπώς δεν είναι δυνατή η ανίχνευση ασυνεπειών σε ένα θησαυρό απλώς με τη μεταφορά του σε ένα σύστημα περιγραφικής λογικής. Αυτό όμως που μπορούμε να χρησιμοποιήσουμε από τα συστήματα περιγραφικής λογικής είναι ο αλγόριθμος που χρησιμοποιούν για την ανίχνευση ασυνεπειών, αρκεί να καταργήσουμε τους περιορισμούς που θέτει και να εισάγουμε άλλους που αφορούν τους θησαυρούς όρων, έτσι ώστε η εφαρμογή του αλγορίθμου να είναι ικανοποιητική τόσο από άποψη χρόνου όσο και από το ποσοστό ανίχνευσης των ασυνεπειών.

Το άλλο πρόβλημα που αντιμετωπίσαμε είναι αυτό της σύνθεσης δυναμικών όρων. Ορίσαμε συσχετίσεις μεταξύ κατηγοριών όρων οι οποίες δίνουν τη δυνατότητα σε όρους από συγκεκριμένες κατηγορίες να συνδυαστούν για τη δημιουργία δυναμικών όρων. Οι νέοι δυναμικοί όροι που δημιουργούνται έχουν σαφή ερμηνεία και μπορούμε άμεσα να συμπεράνουμε το πώς έχουν δημιουργηθεί. Τέλος εφαρμόσαμε το σύστημα δημιουργίας δυναμικών όρων στο θησαυρό AAT και τα αποτελέσματα που λάβαμε ήταν απολύτως ικανοποιητικά.

## 7.2 Μελλοντικές κατευθύνσεις

Για την πλήρη ολοκλήρωση της παρούσας εργασίας υπάρχει ένας αριθμός από εργασίες οι οποίες θα μπορούσαν να γίνουν στο μέλλον.

Το πρώτο και βασικότερο βήμα είναι να διατυπωθεί και να υλοποιηθεί πλήρως ένας αλγόριθμος παρόμοιος με αυτόν της Περιγραφικής Λογικής για την αυτόματη ταξινόμηση εννοιών, ο οποίος μπορεί να αντιμετωπίζει περιπτώσεις ασυνέπειας και έλλειψης οικονομίας σε θησαυρούς όρων. Η έλλειψη ερευνητικής εργασίας σε αυτό τον τομέα, μας οδήγησε στην ανάπτυξη κυρίως του θεωρητικού υπόβαθρου του προβλήματος, τα χρονικά όμως πλαίσια της εργασίας δεν αρκούσαν για να αναπτύξουμε και να υλοποιήσουμε έναν τέτοιο αλγόριθμο. Παρόλα αυτά, ερευνήσαμε το πώς άλλα συστήματα αντιμετωπίζουν το πρόβλημα αυτό και προτείναμε κάποιες κατευθυντήριες γραμμές που θα βοηθήσουν την μελλοντική έρευνα στον τομέα αυτό.

Αφού πραγματοποιηθεί το προηγούμενο βήμα θα είναι δυνατή η δημιουργία ενός ολοκληρωμένου συστήματος, το οποίο θα μπορεί να ανιχνεύει όλες τις περιπτώσεις ασυνέπειας στο θησαυρό και να

αποτελέσει ένα πολύτιμο εργαλείο στα χέρια των ερευνητών που αναλαμβάνουν τη δημιουργία αλλά και τη συνένωση θησαυρών. Θα μπορεί να χρησιμοποιηθεί παράλληλα με συστήματα συνένωσης θησαυρών αλλά και συστήματα που εντοπίζουν ισοδύναμες έννοιες σε διαφορετικούς θησαυρούς, ώστε να εξοικονομείται πολύτιμος χρόνος από τον έλεγχο της ορθότητας των συσχετίσεων που δημιουργούνται.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [AAT] Art and Architecture Thesaurus. [http://www.gii.getty.edu/aat\\_browser](http://www.gii.getty.edu/aat_browser)
- [BDNS94] M. Buchheit, F. Donini, W. Nutt, and A. Schaerf. Terminological Systems Revisited: Terminology = Schema + Views. In Proceedings of KI'94 WorkShop in KRDB'94, Saarbrücken, Germany, September 20-22, 1994.
- [BM95] U. Badertscher and R. Marti. Integrating Terminological and Deductive Reasoning. Technical Report 1995TR-229, Swiss Federal Institute of Technology, Zurich, March, 1995.
- [BMS<sup>+</sup>91] R. Brachman, D. McGuinness, P. P. Schneider, L. A. Resnick and A. Borgida. Living with CLASSIC: When and How to Use a KL-ONE-like Language. In John F. Sowa, editor, *Principles of Semantic Networks -- Explorations in the Representation of Knowledge*, pages 401-456. Morgan Kaufmann, 1991.
- [Bor95] A. Borgida. Description Logics in Data Management. *IEEE Transactions on Knowledge and Data Engineering*, 7(5):671-682, October 1995.
- [Bor96] A. Borgida. On the Relative Expressiveness of Description Logics and Predicate Logics. *AIJ: Artificial Intelligence*, 82(1-2):353 - 367, April 1996.
- [BMT96] Giorgio Brajnik, Stefano Mizzaro, Carlo Tasso: Evaluating User Interfaces to Information Retrieval Systems: A Case Study on User Support. SIGIR 1996: 128-136
- [Bre95a] P. Bresciani. Querying Databases from Description Logics. In *Proceedings of 2nd Workshop KRDB'95*, Bielefeld, Germany, September 11-12, 1995.
- [BS85] R.J. Brachman and J.G. Schmolze. An Overview of the KL-ONE knowledge Representation System. *Cognitive Science*, 9(2):171-216, March 1985.
- [BP94] A. Borgida and Patel-Schneider, P.F. "A Semantics and Complete Algorithm for Subsumption in the CLASSIC Description Logic", Volume 1, pages 277-308. March 1994.

- [CD95] P. Constantopoulos and M. Doerr. Component Classification in the Software Information Base. In Nierstrasz O. and Tsichritzis D. editors, *Object-Oriented Software Composition*. Prentice Hall, 1995.
- [CLBD93] H. Chen, K. Lynch, K. Basu and T. Dorbin. Generating, Integrating and Activating Thesauri for Concept-Based Document Retrieval. *IEEE Expert*, April 1993.
- [DDC] Dewey Decimal Classification Web Site. <http://www.oclc.org/fp/>
- [DF98a] M. Doerr and I. Fundulaki, "SIS - TMS: A Thesaurus Management System for Distributed Digital Collections" Proc. 2nd European Conference, ECDL'98, September 1998, Heraklion, Crete, Greece.
- [DF98b] M. Doerr and I. Fundulaki. A proposal on extended interthesaurus links semantics. Technical Report FORTH-ICS/TR-215, 9 pages, March 1998.
- [DKT95] M. Doerr, P. Klimathianakis and M. Theodorakis. *SIS Data Entry Language User's Manual*. Institute of Computer Science, Foundation of Research and Technology -- Hellas, Herklion, Crete, Greece, 1996.
- [DLNN94] F.M. Donini, M. Lenzerini, D. Nardi and W. Nutt. Queries, Rules and Definitions as Epistemic Sentences in Concept Languages. *Lecture Notes in Computer Science*, 810:113,1994.
- [DLNN95] F. M. Donini, M. Lenzerini, D. Nardi and W. Nutt. The Complexity of Concept Languages. Research Report RR-95-07, Deutsches Forschungszentrum fur Kunstliche Intelligenz, Kaiserslautern, Germany, 1995.
- [DLNS96] F. D. Donini, M. Lenzerini, D. Nardi and A. Schaerf. Reasoning in Description Logics. In G. Brewka, editor, *Principles of Knowledge Representation and Reasoning*, Studies in Logic, Language and Information, pages 193 - 238. CLSI Publications, 1996.
- [ISO85] ISO. Documentation-Guidelines for the Development and Establishment of Multilingual Thesauri, February 1985. International Standard ISO 5964.
- [ISO86] ISO. Documentation-Guidelines for the Development and Establishment of Monolingual Thesauri, November 1986. International Standard ISO 2788.

- [Hor98a] I. Horrocks. The FaCT System. Automated Reasoning with Analytic Tableaux and Related Methods. *International Conference Tableaux'98*, May, 1998, pp307-312.
- [Hor98b] I. Horrocks. Using an Expressive Description Logic: FaCT or Fiction?. A. G. Cohn, L. K. Schubert, and S. C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference (KR'98)*, Morgan Kaufmann Publishers, San Francisco, CA, 1998.
- [HS99] I. Horrocks and P. F. Patel-Schneider. Optimising Description Logic Subsumption. *Journal of Logic and Computation*, 9(3):267-293, 1999.
- [Hus94] U. Hustadt. Do we need the closed world assumption in knowledge representation? In *Proceedings of KI'94 Workshop KRDB'94*, Saarbrocken, Germany, September 20-22, 1994.
- [Kri94] J. Kristensen. Expanding User Queries Using a Search-Aid Thesaurus. *Information Processing and Management*, 29(6):733-744, 1994.
- [Kris93] J. Kristensen. Expanding User Queries Using a Search-Aid Thesaurus. *Information Processing and Management*, 29(6):733-744, 1993.
- [KS93] T. Z. Kalaboukis and M. M. Syntichakis. Suffix Stripping with Modern Greek. Department of Informatics, Athens University of Economics, 1993.
- [LKL94] J. H. Lee, M. H. Kim and Y. J. Lee. Ranking Documents in Thesaurus-Based Boolean Retrieval Systems. *Information Processing and Management*, 30(1):79-91, 1994.
- [MBJK90] J. Mylopoulos, A. Borgida, M. Jarke and M. Koubarakis. Telos: Representing Knowledge About Information Systems. *ACM Transactions on Information Systems*, 8(4):325-362, 1990.
- [MERIMEE] TheMerimeeDatabase. <http://www.culture.fr/documentation/merimee/accueil.htm>
- [Mil91] J. Milstead. Specifications for Thesaurus Software. *Information Processign and Management*, 27(2):165-175, 1991.

- [MR88] H. Mili and R. Rada. Merging Thesauri. Principles and Evaluation. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 10(2):204-220, 1988.
- [MSH] Medical Subject Headings. <http://www.nlm.nih.gov/mesh/meshhome.html>
- [Neb91] B. Nebel. Terminological Cycles: Semantics and Computational Properties. In John F. Sowa, editor, *Principles of Semantic Networks*, pages 331-362. Morgan Kaufmann, San Mateo, 1991.
- [Pai91] C. Paice. A Thesaural Model of Information Retrieval. *Information Processing and Management*, 27(5):433-447, 1991.
- [RBG<sup>+</sup>97] A. Rector S. Bechhofer, C. Goble, I. Horrocks, W. Nowlan, W. Solomon. The GRAIL Concept Modelling Language for Medical Terminology. THE GALEN Project, Medical Informatics Group, Department of Computer Science, University of Manchester, England, 1997.
- [RGG<sup>+</sup>94] A. Rector, A. Gangemi, E. Galeazzi, A. Glowinski, A. Rossi-Mori. The GALEN CORE Model Scheme\ata for Anatomy: Towards a Re-usable Application-Independent Model of Medical Concepts. *Proceedings of Medical Informatics Europe*, MIE 94, Lisbon, pp 186-189.
- [RSNR94] A. Rector, W. Solomon, W. Nowlan, T. Rush. A Terminology Server for Medical Information Systems, 1994.
- [RRP96] A. Rector, J. Rogers, P. Pole. The GALEN High Level Ontology. *Medical Informatics in Europe (MIE)*, Copenhagen, 1996.
- [Sal89] G. Salton. Automatic Text Processing. Addison Wessley, 1989.
- [Sch91] Klaus Schild: From Terminological Logics to Modal Logics. pages 101-104, 1991.
- [Sch94] A. Shaerf. *Query Answering in Concept-Based Knowledge Representation Systems: Algorithms, Complexity, and Semantic Issues*. Ph.D thesis Dipartimento di Informatica e Sistemistica Univerita di Roma: La Sapienza, 1994.

- [Smea96] Alan F. Smeaton: Information Retrieval: Still Butting Heads with Natural Language Processing? SCIE 1996: 115-138
- [Soe95] D. Soergel. The Arts and Architecture Thesaurus (AAT): A Critical Appraisal. *Visual Resources*, 10(4):369-400, 1995.
- [Spin94] Amanda Spink: Term Relevance Feedback and Query Expansion: Relation to Design. SIGIR 1994. 81-90.
- [Spin96] Elicitations During Information Retrieval: Implications for IR System Design. SIGIR 1996, 120-127.
- [Sve89] E. Svenonius. Design of Controlled Vocabularies. In *Encyclopedia of Library and Information Science*, pages 82-109. Marcek Dekker, 1989.
- [UMLS] UMLS Knowledge Source Server.  
<http://www.nlm.nih.gov/pubs/factsheets/umlskss.html>
- [Wil95] H. P. Wilhelmus. *Selecting Knowledge Representation Systems*. PhD thesis, Universiteit Twente Enschede, 1995.
- [ZHF+97] P.E. Zanstra, E. van der Haring, F. Flier, J. Rogers, W Solomon. Using the GRAIL Language for Classification Management. 1997