Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Κατασκευή Ομοσπονδιακού Σχήματος για ΟΣΒΔ από Τοπικά Σχεσιακά Σχήματα

Χριστόδουλος Χαμηλάκης

Μεταπτυχιακή Εργασία

Ηράκλειο, Ιούνιος 1996

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημων
Τμήμα Επιστήμης Υπολογιστών

Κατασκευή Ομοσπονδιακού Σχήματος για ΟΣΒΔ από Τοπικά Σχεσιακά Σχήματα

Εργασία που υποβλήθηκε από τον
Χριστόδουλο Χαμηλάκη
ως μερική εκπλήρωση των απαιτήσεων
για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Χριστόδουλος Χαμηλάκης
Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής, Επόπτης

Χρήστος Νικολάου, Αναπληρωτής Καθηγητής, Μέλος

Γεώργιος Γεωργακόπουλος, Επίκουρος Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος, Αναπληρωτής Καθηγητής
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, Ιούνιος 1996

Κατασκευή Ομοσπονδιακού Σχήματος για ΟΣΒΔ από Τοπικά Σχεσιακά Σχήματα

Χριστόδουλος Χαμηλάκης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

ΠΕΡΙΛΗΨΗ

Ένας από τους ενεργούς τομείς στην περιοχή των Βάσεων Δεδομένων είναι τα Ετερογενή Συστήματα Βάσεων Δεδομένων (ΕΣΒΔ). Τα ΕΣΒΔ αποτελούνται από βάσεις δεδομένων που διαφέρουν ως προς τα συστήματα διαχείρισής τους, το σχήμα τους, τους τύπους δεδομένων τους κ.τ.λ.. Μια κατηγορία ΕΣΒΔ, είναι τα Ομοσπονδιακά Συστήματα Βάσεων Δεδομένων (ΟΣΒΔ), που αποτελούν ομοσπονδίες συνεργαζόμενων, αυτόνομων και ετερογενών συστημάτων βάσεων δεδομένων. Ένα από τα βασικά στοιχεία της αρχιτεκτονικής ενός ΟΣΒΔ είναι το ομοσπονδιακό του σχήμα, που συνήθως ακολουθεί ένα σημασιολογικά πλούσιο μοντέλο δεδομένων.

Το αντικείμενο αυτής της εργασίας είναι το πρόβλημα της κατασκευής του ομοσπονδιακού σχήματος ενός ΟΣΒΔ που αποτελείται από τοπικές σχεσιακές βάσεις δεδομένων. Η προσέγγισή μας περιλαμβάνει την ανάπτυξη μιας μεθόδου μετάφρασης των σχεσιακών σχημάτων προς σχήματα ενός σημασιολογικού μοντέλου δεδομένων, και μιας μεθόδου ενοποίησης των μεταφρασθέντων σχημάτων προς ένα ενοποιημένο ομοσπονδιακό σχήμα.

Οι υποθέσεις που κάνουμε για τα τοπικά προς μετάφραση σχεσιακά σχήματα είναι αρκετά χαλαρές, γεγονός που επιτρέπει στη μέθοδο μετάφρασης να μπορεί να εφαρμοστεί σε ένα μεγάλο αριθμό σχεσιακών σχημάτων σε σχέση με άλλες γνωστές μεθόδους ή αλγορίθμους μετάφρασης. Το σημασιολογικό μοντέλο δεδομένων που ακολουθούν

τα μεταφρασθέντα σχήματα, έχει τα χαρακτηριστικά της εκφραστικότητας και του σημασιολογικού συσχετισμού, που θεωρούνται απαραίτητα για τα κανονικά μοντέλα δεδομένων των ΟΣΒΔ.

Η μέθοδος ενοποίησης αποτελείται από πέντε στάδια : το στάδιο προενοποίησης, το στάδιο σύγκρισης, το στάδιο προσαρμογής, το στάδιο συγχώνευσης, και το στάδιο ανακατασκευής. Ιδιαίτερα για το στάδιο προσαρμογής έχουμε αναπτύξει ένα σύνολο τελεστών μετατροπής, το οποίο παρέχεται στο χρήστη/σχεδιαστή του ενοποιημένου σχήματος προκειμένου να επιλύσει τις ονομαστικές και δομικές συγκρούσεις των συστατικών προς ενοποίηση σχημάτων. Όσο για το στάδιο συγχώνευσης χρησιμοποιήσαμε ένα έτοιμο αλγόριθμο συγχώνευσης σχημάτων.

Τέλος, προτείνουμε την υλοποίηση ενός εργαλείου κατασκευής ομοσπονδιακών σχημάτων για ΟΣΒΔ από τοπικά σχεσιακά σχήματα, το οποίο στηρίζεται στις μεθόδους μετάφρασης και ενοποίησης που έχουμε αναπτύξει.

Επόπτης : Πάνος Κωνσταντόπουλος
Αναπληρωτής Καθηγητής Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Construction of a Federated Schema for a FDBS from Local Relational Schemeta

Christodoulos Hamilakis

Master of Science Thesis

Department of Computer Science
University of Crete

ABSTRACT

Nowadays, an active topic in the area of databases is Heterogeneous Database Systems (HDBSs). HDBSs consist of databases which differ in terms of management systems, schemas, data types etc. One category of HDBSs are Federated Database Systems (FDBSs) which are federations of cooperative, autonomous and heterogeneous database systems. One of the basic components of the architecture of a FDBS is its federated schema, which usually follows a semantically rich data model.

The topic of this thesis is the problem of the construction of a federated schema for a FDBS which consists of local relational databases. Our approach includes the development of a method for translating relational schemas to schemas of a semantic data model, and a method for integrating the translated schemas into a federated schema.

The assumptions for the local relational schemas are quite loose, and so the translation method can be applied on a large number of relational schemas compared to other known methods or algorithms. The semantic data model followed by the translated schemas has the characteristics of expressiveness and semantic relativism, that are considered necessary for canonical data models of the FDBSs.

The integration method comprises five stages : the preintegration stage, the comparison stage, the conforming stage, the merging stage, and the restructuring stage. Especially for the the conforming stage we developed a set of transforming operators, which are offered to the

user/designer of the integrated schema in order to resolve the naming and structural conflicts of the component schemas. As far as the merging stage is concerned we used an already existing algorithm for merging schemas.

Finally, we propose the implementation of a tool for the construction of federated schemas for FDBSs from local relational schemas, which is based on the methods of translation and integration that we have developed.

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

Ευχαριστίες

Στο σημείο αυτό αισθάνομαι την ανάγκη να διατυπώσω τις ευχαριστίες μου σε όλους όσους με τον ένα ή τον άλλο τρόπο με βοήθησαν να φέρω σε πέρας αυτή την εργασία.

Πρώτα απ' όλα ευχαριστώ το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης που με έκανε δεκτό στο μεταπτυχιακό του πρόγραμμα δίνοντας μου την ευκαιρία να αναπτύξω τις ερευνητικές μου προσπάθειες, καθώς και το Ινστιτούτο Πληροφορικής του Ι.Τ.Ε. που με την οικονομική και υλικοτεχνική υποστήριξη που μου παρείχε, στήριξε τις προσπάθειες αυτές.

Ιδιαίτερα ευχαριστώ όλα τα μέλη της ομάδας Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής για τη συνεργασία και τη βοήθεια που απλόχερα μου πρόσφεραν όποτε και για ό,τι την είχα ανάγκη. Ειδικότερα θα ήθελα να ευχαριστήσω τη Χριστίνα Γκριτζάπη και τον Gerd Hillebrand για τις ιδέες τους που μοιράστηκαν μαζί μου, καθώς επίσης το Martin Doerr και τη Χρυσούλα Μπεκιάρη για όλες τις εποικοδομητικές συζητήσεις που είχαμε.

Ένα μεγάλο ευχαριστώ ανήκει στον επόπτη μου, Πάνο Κωνσταντόπουλο, ο οποίος με τις συμβουλές και τις παρατηρήσεις του, όχι μόνο μου "ξεκαθάριζε" το ερευνητικό τοπίο αλλά ταυτόχρονα μου έδινε κουράγιο και δύναμη να συνεχίσω. Πέρα όμως από την επιστημονική μας συνεργασία τον τελευταίο χρόνο μου δόθηκε η ευκαιρία να τον γνωρίσω καλύτερα μέσα από διάφορα προσωπικά γεγονότα και να αναγνωρίσω στο πρόσωπό του έναν άνθρωπο που μου φέρθηκε πραγματικά σα δεύτερος πατέρας. Ελπίζω και στο μέλλον να φανώ αντάξιος της εμπιστοσύνης και της εκτίμησής του.

Για την οικογένεια μου τί να πω. Οι λίγες αυτές γραμμές σε σύγκριση μ' αυτά που μου έχουν προσφέρει είναι σταγόνα στον ωκεανό. Τί να πρωτοαναφέρω. Την ηθική, ψυχολογική, αλλά και οικονομική βοήθεια των γονιών και της γιαγιάς μου, ή την ανοχή και αντοχή της αδελφής μου κατά τα χρόνια της συγκατοίκησής μας. Αυτό όμως που όλοι τους μου χάρισαν αυτά τα χρόνια και θεωρώ αναντικατάστατο, ήταν και είναι η αγάπη και η εμπιστοσύνη τους. Και γιαυτό τους ευχαριστώ περισσότερο απ' όλα.

Το τελευταίο αλλά όχι και μικρότερο ευχαριστώ ανήκει στη Σούλα Αναστασιάδη. Το κορίτσι που τα τελευταία πέντε χρόνια μου συμπαραστέκεται εμψυχώνοντάς με στις δύσκολες ώρες και στηρίζοντάς με ψυχολογικά. Ο άνθρωπος που υπέμεινε με περισσή αντοχή όλα τα άγχη και τις ανασφάλειές μου και που πολλές φορές θυσίασε δικά της ψυχικά αποθέματα για να με κάνει να αισθανθώ καλύτερα.

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

1 Εισαγωγή **1**

2 Ετερογενή Συστήματα Βάσεων Δεδομένων **5**

2.1 Ιστορικά Στοιχεία 5

2.2 Ορισμοί και Χαρακτηριστικά των ΟΣΒΔ 6

2.3 Ταξινομήσεις των ΕΣΒΔ 9

2.4 Αρχιτεκτονική Ομοσπονδιακών Συστημάτων Βάσεων Δεδομένων 11

2.4.1 Βασικά Στοιχεία της Αρχιτεκτονικής 11

2.4.2 Τύποι Επεξεργαστών 12

2.4.3 Τύποι Σχημάτων της Αρχιτεκτονικής 17

2.5 Υπάρχοντα Ομοσπονδιακά Συστήματα Βάσεων Δεδομένων 21

3 Ορισμός του Προβλήματος και Βιβλιογραφική Ανασκόπηση **25**

3.1 Το Πρόβλημα 25

3.2 Μέθοδοι Μετάφρασης Σχημάτων 28

3.3 Μέθοδοι Ενοποίησης Σχημάτων 35

4 Μετάφραση Σχεσιακού Σχήματος σε Σημασιολογικό Δίκτυο **45**

4.1 Σχεσιακό και Σημασιολογικό Μοντέλο Δεδομένων 45

4.2 Μέθοδος Μετάφρασης 49

4.2.1 Υποθέσεις 50

4.2.2 Φάση Αναγνώρισης Λανθανουσών Οντοτήτων 51

4.2.3 Φάση Κατασκευής Τελικού Γράφου Εγκλεισμού 59
4.2.4 Φάση Δημιουργίας Σημασιολογικού Δικτύου 62
4.2.5 Σχέσεις Αντιστοίχισης 65
4.3 Παράδειγμα Εφαρμογής της Μεθόδου Μετάφρασης 67

5 Ενοποίηση Σημασιολογικών Δικτύων 73
5.1 Υποθέσεις 73
5.2 Μέθοδος Ενοποίησης 75
5.2.1 Στάδιο Προενοποίησης 75
5.2.2 Στάδιο Σύγκρισης και Προσαρμογής Συστατικών Σχημάτων 77
5.2.3 Στάδιο Συγχώνευσης Συστατικών Σχημάτων 82
5.2.4 Στάδιο Ανακατασκευής Τελικού Ενοποιημένου Σχήματος 85
5.3 Παράδειγμα Εφαρμογής της Μεθόδου 89
5.3.1 Στάδιο Προενοποίησης 90
5.3.2 Στάδια Σύγκρισης και Προσαρμογής Συστατικών Σχημάτων 90
5.3.3 Στάδιο Συγχώνευσης 92
5.3.4 Στάδιο Ανακατασκευής 92

6 Πρόταση Υλοποίησης 95
6.1 Αδρή Αρχιτεκτονική Εργαλείου 95
6.2 Αρχιτεκτονική και Λειτουργία της Μεταφραστικής Μονάδας 96
6.3 Αρχιτεκτονική και Λειτουργία της Μονάδας Ενοποίησης 101

7 Επίλογος 103
7.1 Συμπεράσματα 103
7.2 Επεκτάσεις - Βελτιώσεις 104

Βιβλιογραφία 105

Κεφάλαιο 1

Εισαγωγή

Ένας από τους ενεργούς τομείς στην περιοχή των βάσεων δεδομένων είναι τα **Ετερογενή Συστήματα Βάσεων Δεδομένων**(ΕΣΒΔ). Τα ΕΣΒΔ είναι η εξέλιξη των κατανεμημένων συστημάτων ΒΔ και αποτελούνται από βάσεις δεδομένων που διαφέρουν ως προς τα συστήματα διαχείρισής τους, το σχήμα τους, τους τύπους δεδομένων τους κ.τ.λ.. Μια κατηγορία ΕΣΒΔ, είναι τα **Ομοσπονδιακά Συστήματα Βάσεων Δεδομένων**(ΟΣΒΔ), που αποτελούν ομοσπονδίες συνεργαζόμενων, αυτόνομων και εν γένει ετερογενών συστημάτων βάσεων δεδομένων. Κάθε σύστημα ΒΔ που συμμετέχει στην ομοσπονδία καλείται **συστατικό σύστημα ΒΔ**.

Μερικά από τα βασικά στοιχεία της αρχιτεκτονικής ενός ΟΣΒΔ είναι τα δεδομένα του, η πληροφορία δηλαδή που διαχειρίζεται, οι βάσεις δεδομένων από τις οποίες αποτελείται, οι αποθήκες δηλαδή που φυλούν τα δεδομένα, τα σχήματα του, οι περιγραφές δηλαδή των δεδομένων σε κάποιο μοντέλο δεδομένων, και οι αντιστοιχίσεις μεταξύ σχημάτων, οι σχέσεις δηλαδή που συσχετίζουν αντικείμενα ενός σχήματος με αντικείμενα ενός άλλου σχήματος. Ειδικότερα, τα σχήματα από τα οποία αποτελείται ένα ΟΣΒΔ δομούνται σε πέντε επίπεδα σχηματίζοντας την **αρχιτεκτονική σχήματος** του. Στο χαμηλότερο επίπεδο της αρχιτεκτονικής βρίσκονται τα **τοπικά σχήματα**, που είναι τα εννοιολογικά σχήματα των συστατικών ΒΔ. Στο αμέσως ψηλότερο επίπεδο βρίσκονται τα **συστατικά σχήματα**, που αποτελούν τις μεταφράσεις των τοπικών σχημάτων σε ένα **κανονικό μοντέλο δεδομένων**. Στο επόμενο επίπεδο βρίσκονται τα **εξαγόμενα σχήματα**, που είναι τα υποσύνολα των συστατικών σχημάτων που θα συμμετέχουν στην ομοσπονδία, και αμέσως μετά τα **ομοσπονδιακά σχήματα**, που προέρχονται από την ενοποίηση των διαφόρων εξαγομένων σχημάτων. Στο τελευταίο επίπεδο βρίσκονται τα **εξωτερικά σχήματα**, που είναι τα υποσύνολα των ομοσπονδιακών σχημάτων με τα οποία έρχονται σε επαφή οι χρήστες ή οι εφαρμογές του ομοσποδιακού συστήματος.

Στην εργασία μας το πρόβλημα που αντιμετωπίσαμε ήταν η κατασκευή ενός ομοσπονδιακού σχήματος από δεδομένα τοπικά σχεσιακά σχήματα. Η προσέγγισή μας αποτελείται από δύο βήματα. Στο πρώτο βήμα γίνεται η μετάφραση των τοπικών σχεσιακών σχημάτων σε σχήματα ενός κανονικού μοντέλου δεδομένων, και στο δεύτερο βήμα τα μεταφρασθέντα σχήματα ενοποιούνται δίνοντας το ομοσπονδιακό σχήμα που επίσης εκφράζεται στο κανονικό μοντέλο δεδομένων. Το κανονικό μοντέλο δεδομένων που χρησιμοποιούμε είναι το σημασιολογικό μοντέλο δεδομένων που ακολουθούν τα σχήματα που παράγονται από τη γλώσσα παράστασης γνώσης Telos (σημασιολογικά δίκτυα). Το μοντέλο αυτό μας παρέχει τα χαρακτηριστικά της εκφραστικότητας και του σημασιολογικού συσχετισμού που ζητάμε να έχουν τα κανονικά μοντέλα δεδομένων.

Για τη μετάφραση των σχεσιακών σχημάτων σε σημασιολογικά δίκτυα αναπτύξαμε μια μέθοδο μετάφρασης. Οι υποθέσεις που θεωρούμε ότι πληρούν τα σχεσιακά σχήματα της εισόδου είναι αρκετά χαλαρές, γεγονός που επιτρέπει την εφαρμογή της μεθόδου σε μεγάλο αριθμό σχεσιακών σχημάτων, αλλά από την άλλη μεριά απαιτεί την επέμβαση του χρήστη για την επίλυση κάποιων ασαφειών σε ορισμένες περιπτώσεις. Ακριβώς αυτός είναι ο λόγος που μιλάμε για ”μέθοδο” και όχι για ”αλγόριθμο” ο οποίος θα επέτρεπε την αυτοματοποίηση της διαδικασίας μετάφρασης. Η μέθοδος περιλαμβάνει τρεις φάσεις. Στην πρώτη φάση, ανακαλύπτονται στο αρχικό σχεσιακό σχήμα ”λανθάνουσες” οντότητες που περιγράφονται από υποσύνολα γνωρισμάτων των σχημάτων-σχέσεων, και δημιουργούνται νέα σχήματα-σχέσεων που τις εκφράζουν. Στη δεύτερη φάση, αφαιρούνται από το ενδιάμεσο σχεσιακό σχήμα που προέκυψε από την προηγούμενη φάση πλεονασμοί που οφείλονται στην κάθετη κατάτμηση πληροφορίας και στη μεταβατικότητα της σχέσης εγκλεισμού. Στην τρίτη και τελευταία φάση, το τελικό σχεσιακό σχήμα που προέκυψε από την προηγούμενη φάση μεταφράζεται βάσει τριών μεταφραστικών κανόνων στο σημασιολογικό δίκτυο, ενώ ταυτόχρονα δηλώνονται και οι σχέσεις αντιστοίχισης μεταξύ των δομών του σχεσιακού σχήματος με τις δομές του σημασιολογικού δικτύου. Οι σχέσεις αυτές περιέχουν όλη την πληροφορία μετάφρασης.

Για την ενοποίηση των σημασιολογικών δικτύων αναπτύξαμε μια μέθοδο ενοποίησης. Η μέθοδος αποτελείται από πέντε στάδια. Στο πρώτο στάδιο, καθορίζεται η στρατηγική δυαδικής ενοποίησης των σημασιολογικών δικτύων, καθώς και ο τρόπος επέμβασης του χρήστη στα επόμενα στάδια της μεθόδου. Στο δεύτερο και τρίτο στάδιο, ανακαλύπτονται και επιλύονται αντίστοιχα οι ονομαστικές και δομικές συγκρούσεις μεταξύ των σημασιολογικών δικτύων της εισόδου. Τα στάδια αυτά λαμβάνουν χώρα επαναληπτικά μέχρις ότου να επιλυθούν όσο το δυνατόν περισσότερες συγκρούσεις.

Ιδιαίτερα για το τρίτο στάδιο έχουμε αναπτύξει ένα σύνολο τελεστών τροποποίησης που επιτρέπει στο χρήστη να επιλύει ορισμένα είδη συγκρούσεων. Στο τέταρτο στάδιο, εφαρμόζεται επαναληπτικά ο αλγόριθμος δυαδικής συγχώνευσης του Kosky ο οποίος παράγει ένα συγχωνευμένο σχήμα που δεν είναι όμως σημασιολογικό δίκτυο. Η εφαρμογή του αλγορίθμου αυτού γίνεται με βάση τη δυαδική στρατηγική συγχώνευσης που έχει καθοριστεί από το πρώτο στάδιο. Τέλος στο πέμπτο στάδιο, το σχήμα που έχει παραχθεί μετατρέπεται στο τελικό ενοποιημένο σημασιολογικό δίκτυο.

Στα πλαίσια επίσης αυτής της εργασίας κάναμε μια πρόταση υλοποίησης ενός εργαλείου κατασκευής ομοσπονδιακών σχημάτων από τοπικά σχεσιακά σχήματα. Το εργαλείο αποτελείται από δύο μονάδες. Η πρώτη μονάδα μεταφράζει τα τοπικά σχεσιακά σχήματα βάσει της μεθόδου που έχουμε αναπτύξει και παράγει τα σημασιολογικά δίκτυα και τις σχέσεις αντιστοίχισης που έχουν δηλωθεί. Η αποθήκευση των σχέσεων αντιστοίχισης γίνεται βάσει του **Μεταφραστικού Μετασχήματος**, ενός σημασιολογικού δικτύου που περιγράφει τα είδη των σχέσεων αντιστοίχισης. Η δεύτερη μονάδα ενοποιεί τα σημασιολογικά δίκτυα της μετάφρασης στο ενοποιημένο σημασιολογικό δίκτυο. Εκτός του ενοποιημένου σημασιολογικού δικτύου το εργαλείο παράγει και τα τροποποιημένα συστατικά σημασιολογικά δίκτυα μαζί με σχέσεις αντιστοίχισής τους.

Στο Κεφάλαιο 2 της εργασίας, παραθέτουμε μερικά γενικά στοιχεία για τα Ετερογενή Συστήματα Βάσεων Δεδομένων.

Στο Κεφάλαιο 3, ορίζουμε το πρόβλημα με το οποίο ασχοληθήκαμε και εξετάζουμε σχετική δουλειά που έχει γίνει και αφορά τη μεθοδολογία επίλυσης που προτείνουμε.

Στο Κεφάλαιο 4, παρουσιάζουμε και αναλύουμε τη μέθοδο που αναπτύξαμε για τη μετάφραση ενός σχεσιακού σχήματος σε ένα σημασιολογικό δίκτυο.

Στο Κεφάλαιο 5, παρουσιάζουμε και αναλύουμε τη μέθοδο που αναπτύξαμε για την ενοποίηση των σημασιολογικών δικτύων προς ένα ενοποιημένο σημασιολογικό δίκτυο.

Στο Κεφάλαιο 6, παρουσιάζουμε την πρόταση υλοποίησης ενός εργαλείου κατασκευής ομοσπονδιακού σχήματος, που χρησιμοποιεί τις μεθόδους μετάφρασης και ενοποίησης, που έχουμε αναπτύξει.

Η εργασία κλείνει με το κεφάλαιο 7, όπου αναφέρονται ορισμένα συμπεράσματα και προτάσεις βελτίωσης.

Κεφάλαιο 2

Ετερογενή Συστήματα Βάσεων Δεδομένων

Στο κεφάλαιο αυτό θα παρουσιάσουμε ορισμένα γενικά στοιχεία για τα Ετερογενή Συστήματα Βάσεων Δεδομένων (ΕΣΒΔ), τα οποία θα βοηθήσουν στο να γίνει κατανοητό το περιβάλλον μέσα στο οποίο αναπτύσσεται αυτή η εργασία. Αρχικά θα δούμε μερικά ιστορικά στοιχεία για τη γέννηση και ανάπτυξη αυτής της ερευνητικής περιοχής και κατόπιν θα αναφέρουμε τους κυριότερους ορισμούς και τα χαρακτηριστικά των ΟΣΒΔ. Θα δούμε επίσης ταξινομήσεις διαφόρων ΕΣΒΔ με βάση κάποια από τα χαρακτηριστικά τους, ενώ θα παρουσιάσουμε και τη γενική αρχιτεκτονική ενός τέτοιου συστήματος. Το κεφάλαιο θα κλείσει με τη σύντομη παρουσίαση υπαρχόντων συστημάτων.

2.1 Ιστορικά Στοιχεία

Κατά τη διάρκεια της δεκαετίας του ’70 τα *κεντρικοποιημένα* Συστήματα Βάσεων Δεδομένων (*centralized DBSs*) ήταν κυρίαρχα. Με τον όρο *Σύστημα Βάσεων Δεδομένων (ΣΒΔ)* εννοούμε τη συνύπαρξη ενός συνόλου δεδομένων που ονομάζεται *Βάση Δεδομένων (ΒΔ)* και ενός λογισμικού τμήματος που διαχειρίζεται τη ΒΔ και ονομάζεται *Σύστημα Διαχείρισης Βάσης Δεδομένων (ΣΔΒΔ).* Η δομή και η οργάνωση των δεδομένων της ΒΔ περιγράφεται από ένα *σχήμα*, το οποίο ακολουθεί ένα *μοντέλο δεδομένων* (σχεσιακό, οντοκεντρικό, κ.α.). Το ΣΔΒΔ επιτρέπει στο χρήστη να οργανώσει κατάλληλα τα δεδομένα του ώστε να έχει μικρούς χρόνους ανάκλησης, δυνατότητα προσπέλασης μέσω *επερωτήσεων* εκφρασμένων σε μια *γλώσσα επερωτήσεων*, συνέπεια και οικονομία χώρου αποφεύγοντας την επανάληψη της πληροφορίας.

Με την ανάπτυξη του υλικού (hardware) και των δικτυακών υπηρεσιών, άρχισε στη δεκαετία του '80, η αποκεντρικοποίηση των ΣΒΔ με την εμφάνιση των πρώτων *κατανεμημένων* ΣΒΔ. Έτσι, ενώ στα κεντρικοποιημένα ΣΒΔ είχαμε τη ΒΔ εγκατεστημένη σε ένα κόμβο, στα κατανεμημένα ΣΒΔ η ΒΔ είναι διάσπαρτη σε διάφορους κόμβους του δικτύου. Τόσο στη μία όσο και στην άλλη περίπτωση, υπάρχει μόνο ένα σχήμα που περιγράφει τη ΒΔ. Σε ένα κατανεμημένο ΣΒΔ λοιπόν, έχουμε *ομοιογένεια* υπό την έννοια ότι όλα τα φυσικά του τμήματα τρέχουν το ίδιο κατανεμημένο ΣΔΒΔ.

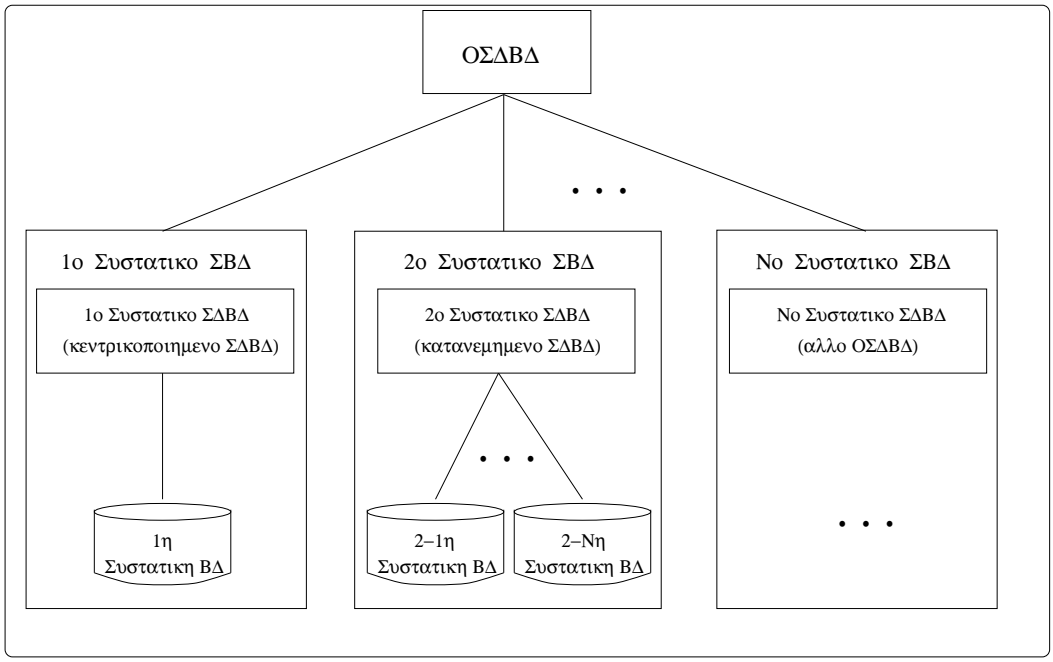
Η εξέλιξη των κατανεμημένων ΣΒΔ ήταν τα *ετερογενή* ΣΒΔ (ΕΣΒΔ). Όπως τα κατανεμημένα ΣΒΔ ήραν τον περιορισμό της κεντρικοποίησης, έτσι τα ετερογενή ΣΒΔ ήραν τον περιορισμό της ομοιογένειας. Στα συστήματα αυτά, τα επιμέρους φυσικά τους τμήματα παρουσιάζουν διαφορές στο μοντέλο δεδομένων, στη γλώσσα επερωτήσεων ή στο σχήμα. Η ιδέα της ανάπτυξης των ΕΣΒΔ προέκυψε τόσο από το γεγονός ότι ήδη μέσα σε ένα πληροφοριακό χώρο υπήρχαν εγκατεστημένες αυτόνομες ΒΔ οι οποίες δεν είχαν δυνατότητα συνεργασίας, όσο και από το ότι για γεωγραφικά απομακρυσμένες ΒΔ με κοινή πληροφορία δεν υπήρχε δυνατότητα επερωτήσεων γι' αυτή. Η βασικότερη διαφορά ενός ΕΣΒΔ από ένα κατανεμημένο ΣΒΔ είναι η δυνατότητα των επιμέρους τμημάτων του πρώτου να λειτουργούν αυτόνομα.

Μέχρι σήμερα η βιβλιογραφία έχει εμπλουτιστεί αρκετά με άρθρα πάνω σε εξειδικευμένα θέματα ΕΣΒΔ, με πρακτικά συνεδρίων, με ειδικές εκδόσεις περιοδικών ([38], [1], [32], [41]) και βιβλίων ([23]) αφιερωμένες στα ΕΣΒΔ, αποδεικνύοντας έτσι το έντονο ενδιαφέρον της επιστημονικής κοινότητας γι' αυτόν τον τομέα.

2.2 Ορισμοί και Χαρακτηριστικά των ΟΣΒΔ

Αν και η ορολογία που χρησιμοποιείται στη διεθνή βιβλιογραφία ([42], [45], [30], [10], [36]) είναι σε πολλές περιπτώσεις αρκετά διαφορετική, εμείς θα στηριχθούμε στους ορισμούς που δίνονται στο [42], θεωρώντας τους σαν περισσότερο αντιπροσωπευτικούς. Έτσι, ένα *Σύστημα Βάσης Δεδομένων* (ΣΒΔ) αποτελείται από ένα τμήμα λογισμικού, που ονομάζεται *Σύστημα Διαχείρισης Βάσης Δεδομένων* (ΣΔΒΔ) και από μία ή περισσότερες βάσεις δεδομένων (ΒΔ) που τις διαχειρίζεται. Όπως αναφέραμε και στην προηγούμενη ενότητα οι *Ετερογενείς Βάσεις Δεδομένων* (ΕΒΔ) είναι ΒΔ που διαφέρουν μεταξύ τους ως προς το σχήμα, τη γλώσσα επερωτήσεων και το μοντέλο δεδομένων. Ένα *Ομοσπονδιακό Σύστημα Βάσεων Δεδομένων* (ΟΣΒΔ) είναι μία συλλογή από συνεργαζόμενα, αυτόνομα και εν γένει ετερογενή ΣΒΔ. Καθένα από τα επιμέρους ΣΒΔ που συμμετέχουν στην ομοσπονδία

ονομάζεται *συστατικό ΣΒΔ*. Το λογισμικό που παρέχει το συντονισμό των συστατικών ΣΒΔ, λέγεται *Ομοσπονδιακό Σύστημα Διαχείρισης Βάσεων Δεδομένων* (ΟΣΔΒΔ). Τα συστατικά ΣΒΔ μπορούν να είναι κεντρικοποιημένα ΣΒΔ, κατανεμημένα ΣΒΔ, ή άλλα ΟΣΒΔ, όπως φαίνεται και στο Σχήμα 2.1.



Σχήμα 2.1: Ομοσπονδιακό Σύστημα Διαχείρισης Βάσεων Δεδομένων

Τα ΟΣΔΒΔ είναι περίπτωση μια γενικότερης κατηγορίας συστημάτων που αποτελούνται από πολλαπλές ΒΔ, τα *Συστήματα ΠολυΒάσεων Δεδομένων* (ΣΠΒΔ). Ένα ΣΠΒΔ υποστηρίζει λειτουργίες σε πολλαπλά συστατικά ΣΒΔ, καθένα από τα οποία έχει το δικό του (εν γένει διαφορετικό) ΣΔΒΔ. Τα ΟΣΔΒΔ λοιπόν είναι εκείνα τα ΣΠΒΔ των οποίων τα συστατικά ΣΒΔ είναι αυτόνομα. Το ΣΠΒΔ ονομάζεται *ομογενές* αν τα ΣΔΒΔ των συστατικών ΣΒΔ είναι τα ίδια, αλλιώς ονομάζεται *ετερογενές*.

Τέλος, ένα σύστημα το οποίο επιτρέπει περιοδική ανταλλαγή δεδομένων ανάμεσα σε πολλές ΒΔ, ονομάζεται *Σύστημα Ανταλλαγής Δεδομένων*, ενώ ένα σύστημα που παρέχει πρόσβαση σε πολλά ΣΔΒΔ, επιτρέποντας τη σύνδεση με ένα από αυτά κάθε φορά, ονομάζεται *Διεπαφή Απομακρυσμένων ΣΔΒΔ*.

Ας δούμε όμως τα κυριότερα χαρακτηριστικά των ΣΠΒΔ. Τέτοια συστήματα μπορούν να χαρακτηριστούν σε τρεις διαστάσεις: την κατανομή, την ετερογένεια και την αυτονομία. Αυτές οι τρεις διαστάσεις μπορούν να αποτελέσουν τομείς σύμφωνα με

τους οποίους, μπορούμε να ταξινομήσουμε τα διάφορα συστήματα όπως θα δούμε και παρακάτω.

- *Κατανομή*

  Τα δεδομένα μπορούν να είναι κατανεμημένα σε πολλές (εν γένει γεωγραφικά απομακρυσμένες) ΒΔ, με διάφορους τρόπους όπως η οριζόντια ή η κάθετη τμηματοποίηση. Πλεονεκτήματα της κατανομής των δεδομένων είναι η αυξημένη διαθεσιμότητα και αξιοπιστία, καθώς και οι βελτιωμένοι χρόνοι προσπέλασης των δεδομένων. Στην περίπτωση των ΟΣΒΔ, η κατανομή των δεδομένων οφείλεται στην ύπαρξη πολλών ΣΒΔ πριν τη δημιουργία του ΟΣΒΔ.

- *Ετερογένεια*

  Οι τύποι της ετερογένειας χωρίζονται σ' αυτούς που οφείλονται σε διαφορές των ΣΔΒΔ και σ' αυτούς που οφείλονται σε διαφορές στη σημασιολογία των δεδομένων. Ετερογένειες που οφείλονται σε διαφορές των ΣΔΒΔ, προέρχονται από διαφορές στα μοντέλα δεδομένων τους (δομικές διαφορές, διαφορές στους περιορισμούς, διαφορές στις γλώσσες επερωτήσεων) και στο επίπεδο συστήματος (διαφορές στις τεχνικές διαχείρισης δοσοληψιών, διαφορές στις απαιτήσεις υλικού και λογισμικού συστήματος, διαφορές στις δυνατότητες επικοινωνίας). Ετερογένειες που οφείλονται σε διαφορές στη σημασιολογία των δεδομένων, προέρχονται από τις διαφωνίες για τη σημασία, την ερμηνεία ή τη χρήση των ίδιων ή σχετικών δεδομένων.

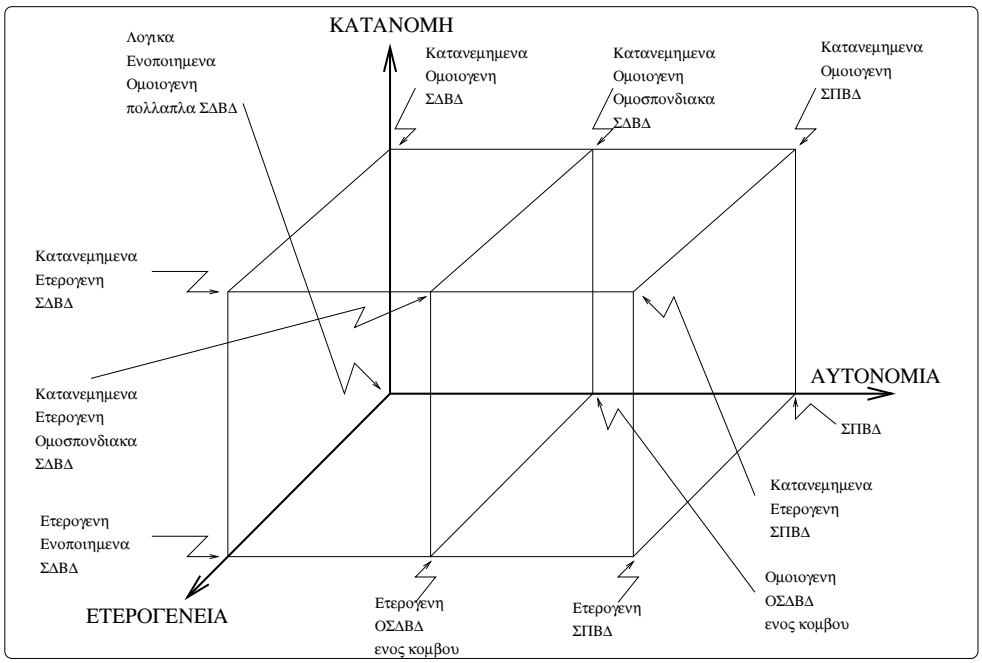
- *Αυτονομία*

  Υπάρχουν τέσσερις τύποι *συστατικής αυτονομίας*, δηλαδή αυτονομίας των συστατικών ΣΒΔ ενός ΟΣΒΔ: αυτονομία σχεδιασμού, αυτονομία επικοινωνίας, αυτονομία εκτέλεσης και αυτονομία συσχέτισης. Η *αυτονομία σχεδιασμού* αναφέρεται στην δυνατότητα ενός συστατικού ΣΒΔ να επιλέξει το δικό του σχεδιασμό σε ότι αφορά τα δεδομένα που διαχειρίζεται, την παράσταση και ονοματολογία τους, τη σημασιολογική ερμηνεία τους, τους περιορισμούς τους, τη λειτουργικότητα του συστήματος, τη συσχέτιση και την κοινωνία με άλλα συστήματα και τέλος την υλοποίηση. Η *αυτονομία επικοινωνίας* αναφέρεται στην δυνατότητα ενός συστατικού ΣΒΔ να αποφασίσει πότε θα επικοινωνεί με άλλα ΣΒΔ. Η *αυτονομία εκτέλεσης* αναφέρεται στην δυνατότητα ενός συστατικού ΣΒΔ να εκτελεί τοπικές λειτουργίες (εντολές ή δοσοληψίες) χωρίς παρεμβολή εξωτερικών λειτουργιών και να αποφασίζει τη σειρά με την οποία θα εκτελεί τις εξωτερικές λειτουργίες. Τέλος, η *αυτονομία συσχέτισης* υπονοεί ότι ένα συστατικό ΣΒΔ έχει τη δυνατότητα να

αποφασίζει για το πόση από τη λειτουργικότητα και τους πόρους του θα συνεισφέρει στην κοινωνία με τα άλλα ΣΒΔ.

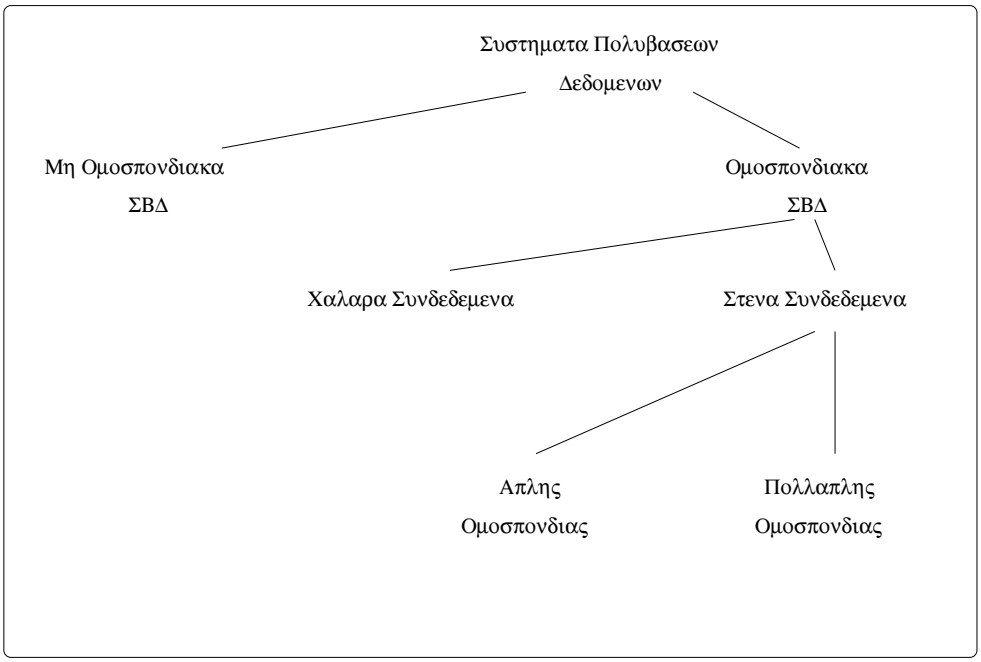
2.3 Ταξινομήσεις των ΕΣΒΔ

Στο [36] υπάρχει μια ταξινόμηση των ΣΔΒΔ για κατανεμημένα περιβάλλοντα, σύμφωνα με τους τρεις άξονες της αυτονομίας, κατανομής και ετερογένειας, όπως φαίνεται και στο Σχήμα 2.2.



Σχήμα 2.2: Ταξινόμηση ΣΔΒΔ για κατανεμημένα περιβάλλοντα

Στον άξονα της αυτονομίας διακρίνουμε: συστήματα τα οποία δεν είναι αυτόνομα, δηλ. παρουσιάζουν μία και μοναδική εικόνα της ΒΔ στους χρήστες, συστήματα ημιαυτόνομα, τα οποία μπορούν να δουλέψουν αυτόνομα αλλά έχουν σχεδιαστεί για μια ομοσπονδία, και τέλος συστήματα πλήρως αυτόνομα, τα οποία μπορούν να δουλέψουν αυτόνομα και δεν ξέρουν τίποτα για τους άλλους που μετέχουν στην ομοσπονδία. Στον άξονα της κατανομής, διακρίνουμε συστήματα όπου τα δεδομένα που διαχειρίζονται είναι πλήρως κατανεμημένα και άλλα που είναι αποθηκευμένα σε ένα μόνο κόμβο. Τέλος στον άξονα της ετερογένειας, όπως αυτή ορίστηκε στην προηγούμενη παράγραφο, διακρίνουμε ομοιογενή και ετερογενή συστήματα.



Σχήμα 2.3: Ταξινόμηση ΣΠΒΔ

Στο [42] έχουμε μια άλλη ταξινόμηση, όπως βλέπουμε και στο Σχήμα 2.3. Ανάλογα με το χαρακτηριστικό της αυτονομίας των συστατικών ΣΒΔ τους, τα ΣΠΒΔ μπορούν να ταξινομηθούν σε : μη-ομοσπονδιακά ΣΒΔ και ομοσπονδιακά ΣΒΔ. Τα *μη-ομοσπονδιακά ΣΒΔ* είναι μια ενοποίηση συστατικών ΣΔΒΔ τα οποία **δεν** είναι αυτόνομα. Έχουν ένα επίπεδο διαχείρισης, όλες οι λειτουργίες εκτελούνται ομοιόμορφα στις συστατικές ΒΔ, ενώ τέλος δε γίνεται διάκριση μεταξύ τοπικών και μη-τοπικών χρηστών. Ένας ειδικός τύπος μη-ομοσπονδιακού ΣΒΔ, όπου όλες οι συστατικές ΒΔ είναι πλήρως ενοποιημένες ώστε να παρέχουν ένα μοναδικό καθολικό σχήμα, ονομάζεται *ενοποιημένο ΣΠΒΔ* και παρουσιάζεται στους χρήστες του όπως ένα κατανεμημένο ΣΒΔ. Τα *ομοσπονδιακά ΣΒΔ* αποτελούνται από συστατικά ΣΒΔ τα οποία αν και αυτόνομα συμμετέχουν στην ομοσπονδία και επιτρέπουν μερική και ελεγχόμενη πρόσβαση στα δεδομένα τους, από άλλα συστήματα της ομοσπονδίας. Μπορούμε να πούμε ότι τα ΟΣΒΔ είναι η μέση οδός ανάμεσα στα πλήρως ενοποιημένα συστήματα όπου η αυτονομία των συστατικών ΣΒΔ είναι ανύπαρκτη, και σ' αυτά που δεν είναι καθόλου ενοποιημένα και στα οποία οι χρήστες καλούνται να έρθουν σ' επαφή σαφώς με πολλαπλά αυτόνομα ΣΒΔ. Τα ΟΣΒΔ μπορούν να ταξινομηθούν επίσης, σύμφωνα με το ποιός διαχειρίζεται την ομοσπονδία και με το πώς ενοποιούνται τα συστατικά ΣΒΔ. Έτσι, λέμε ότι ένα ΟΣΒΔ είναι *χαλαρά συνδεδεμένο*,

αν είναι ευθύνη κάθε χρήστη να δημιουργεί και να διαχειρίζεται την ομοσπονδία. Αντίθετα, αν την ευθύνη αυτή καθώς και τον έλεγχο προσπέλασης στα συστατικά ΣΒΔ, την έχει κάποιος διαχειριστής ΒΔ, λέμε ότι το ΟΣΒΔ είναι *στενά συνδεδεμένο.* Ένα στενά συνδεδεμένο ΟΣΒΔ λέμε ότι έχει *μοναδική ομοσπονδία*, αν επιτρέπει τη δημιουργία και διαχείριση ενός μόνο ομοσπονδιακού σχήματος, αλλιώς λέμε ότι έχει *πολλαπλές ομοσπονδίες*.

Τέλος, στο [10] υπάρχει μια κατηγοριοποίηση των καθολικών συστημάτων κοινωνίας πληροφοριών, καθώς και πίνακες από έργα (projects) που έχουν αναπτυχθεί για κάθε κατηγορία.

2.4 Αρχιτεκτονική Ομοσπονδιακών Συστημάτων Βάσεων Δεδομένων

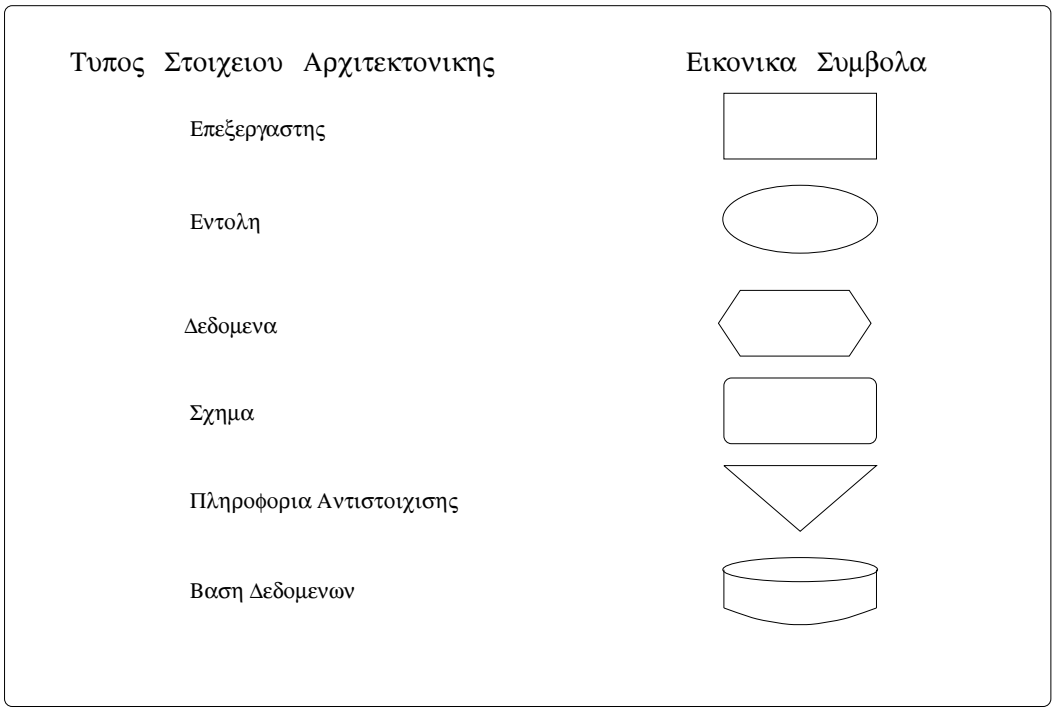
Μια αρχιτεκτονική αναφοράς παρέχει το πλαίσιο μέσα απ' το οποίο μπορούμε να κατανοήσουμε, να κατηγοριοποιήσουμε και να συγκρίνουμε διαφορετικές αρχιτεκτονικές δυνατότητες για την ανάπτυξη των ΟΣΒΔ ([42], [30], [8]).

2.4.1 Βασικά Στοιχεία της Αρχιτεκτονικής

Τα βασικά στοιχεία που αποτελούν ένα ΟΣΒΔ είναι τα δεδομένα, οι ΒΔ, οι εντολές, οι επεξεργαστές, τα σχήματα και οι αντιστοιχίσεις.

- Τα *δεδομένα* είναι τα βασικά γεγονότα και η πληροφορία που διαχειρίζεται ένα ΣΒΔ.
- Η *Βάση Δεδομένων* είναι η αποθήκη (repository) όπου τα δεδομένα αποθηκεύονται σύμφωνα με ένα μοντέλο δεδομένων.
- Οι *εντολές* είναι οι αιτήσεις για συγκεκριμένες ενέργειες, οι οποίες είτε δίνονται από ένα χρήστη είτε παράγονται από ένα επεξεργαστή.
- Οι *επεξεργαστές* είναι κομμάτια λογισμικού που χειρίζονται εντολές και δεδομένα.
- Τα *σχήματα* είναι περιγραφές των δεδομένων. Τα δεδομένα αυτά δύναται να διαχειρίζονται από ένα ή περισσότερα ΣΔΒΔ. Τα σχήματα αποτελούνται από αντικείμενα και από συσχετίσεις μεταξύ αυτών.
- Οι *αντιστοιχίσεις* είναι σχέσεις που συσχετίζουν αντικείμενα ενός σχήματος με αντικείμενα ενός άλλου σχήματος.

Στο Σχήμα 2.4 φαίνονται τα εικονικά σύμβολα που θα χρησιμοποιούμε για καθένα από τα βασικά στοιχεία, σε επόμενα σχήματα.



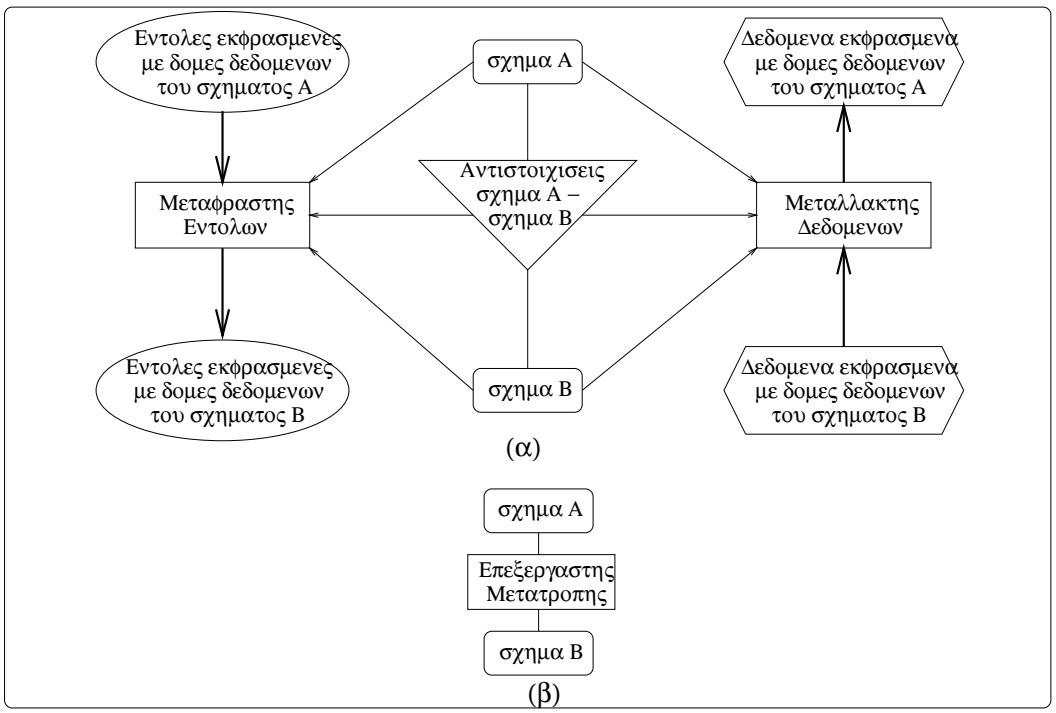
Σχήμα 2.4: Βασικά στοιχεία μιας αρχιτεκτονικής διαχείρισης δεδομένων

Δύο από τα βασικά στοιχεία, οι επεξεργαστές και τα σχήματα, παίζουν ένα ιδιαίτερα σημαντικό ρόλο στον ορισμό και τη διαμόρφωση των διαφόρων αρχιτεκτονικών των ΟΣΒΔ. Οι μεν επεξεργαστές είναι τα τμήματα του λογισμικού ενός ΣΔΒΔ που δεν εξαρτώνται από τις εφαρμογές του ΣΔΒΔ, τα δε σχήματα αντίθετα, είναι τα τμήματα που εξαρτώνται από την εφαρμογή και καθορίζουν τα περιεχόμενα και τη δομή της ΒΔ.

2.4.2 Τύποι Επεξεργαστών

Οι αρχιτεκτονικές της διαχείρισης δεδομένων διαφέρουν ως προς τους τύπους των επεξεργαστών που χρησιμοποιούν, καθώς και τις συσχετίσεις ανάμεσα σε αυτούς. Υπάρχουν τέσσερις τύποι επεξεργαστών, καθένας από τους οποίους εκτελεί διαφορετικές λειτουργίες πάνω σε εντολές χειρισμού δεδομένων και σε προσπελάσιμα δεδομένα: επεξεργαστές μετατροπής, επεξεργαστές φιλτραρίσματος, επεξεργαστές δόμησης και επεξεργαστές προσπέλασης.

Επεξεργαστές Μετατροπής



Σχήμα 2.5: (α) Ζεύγος Επεξεργαστών Μετατροπής (β) Αφηρημένη μορφή επεξεργαστή μετατροπής

Οι *επεξεργαστές μετατροπής* μεταφράζουν εντολές από μια γλώσσα, που καλείται *γλώσσα πηγής*, σε μια άλλη γλώσσα, που καλείται *γλώσσα προορισμού*, ή μετατρέπουν δεδομένα από μια μορφή σε μια άλλη μορφή. Οι επεξεργαστές μετατροπής παρέχουν ένα τύπο ανεξαρτησίας δεδομένων, που καλείται *διαφάνεια μοντέλου δεδομένων*, στον οποίο οι δομές δεδομένων και οι εντολές που χρησιμοποιούνται από ένα επεξεργαστή, είναι κρυμμένες από τους άλλους. Η διαφάνεια μοντέλου δεδομένων κρύβει διαφορές στις γλώσσες επερωτήσεων και στις μορφές δεδομένων.

Είναι δυνατό για μερικούς επεξεργαστές μετατροπής εντολών, να υπάρχουν συνοδευτικοί επεξεργαστές μετατροπής δεδομένων οι οποίοι μετατρέπουν δεδομένα που παρήχθησαν από μετατετραμμένες εντολές, σε δεδομένα συμβατά με τις εντολές πηγής.

στο Σχήμα 2.5 βλέπουμε ένα ζευγάρι επεξεργαστών μετατροπής. Χρησιμοποιώντας πληροφορίες από το σχήμα Α και το σχήμα Β, καθώς και τις αντιστοιχίσεις μεταξύ τους, ο επεξεργαστής μετατροπής εντολών μετατρέπει εντολές εκφρασμένες βάσει του σχήματος Α, σε εντολές εκφρασμένες βάσει του σχήματος Β. Χρησιμοποιώντας τις ίδιες

πληροφορίες, ο συνοδευτικός επεξεργαστής μετατροπής δεδομένων μετατρέπει δεδομένα που περιγράφονται από το σχήμα Β σε δεδομένα που περιγράφονται από το σχήμα Α.

Για την εκτέλεση των μετατροπών, οι επεξεργαστές μετατροπής χρειάζονται τις αντιστοιχίσεις μεταξύ των αντικειμένων των σχημάτων. Η εργασία της *μετάφρασης σχήματος* συνεπάγεται τη μετατροπή ενός σχήματος(σχήμα Α) το οποίο περιγράφει δεδομένα που ακολουθούν ένα μοντέλο δεδομένων, σε ένα ισοδύναμο σχήμα(σχήμα Β) το οποίο περιγράφει δεδομένα που ακολουθούν ένα άλλο μοντέλο δεδομένων. Η εργασία αυτή, παράγει τις αντιστοιχίσεις που συσχετίζουν αντικείμενα του ενός σχήματος(σχήμα Β) με αντικείμενα του άλλου(σχήμα Α). Η εργασία της *μετατροπής εντολών* συνεπάγεται τη μετάφραση εντολών που αφορούν αντικείμενα ενός σχήματος(σχήμα Β), σε εντολές που αφορούν αντικείμενα του άλλου σχήματος(σχήμα Α), χρησιμοποιώντας αυτές τις αντιστοιχίσεις.

Οι αντιστοιχίσεις μπορούν να συσχετιστούν με τον επεξεργαστή μετατροπής με δύο τρόπους. Στην πρώτη περίπτωση, οι αντιστοιχίσεις βρίσκονται μέσα στον κώδικα του επεξεργαστή, κάνοντάς τον έτσι εξειδικευμένο για τα δύο σχήματα. Εναλλακτικά, οι αντιστοιχίσεις μπορούν να αποθηκεύονται σε μια ξεχωριστή δομή δεδομένων που προσπελαύνεται από τον επεξεργαστή όταν μετατρέπει εντολές και δεδομένα. Όπως καταλαβαίνουμε η δεύτερη περίπτωση είναι προτιμότερη, μια και είναι η πιο γενική.

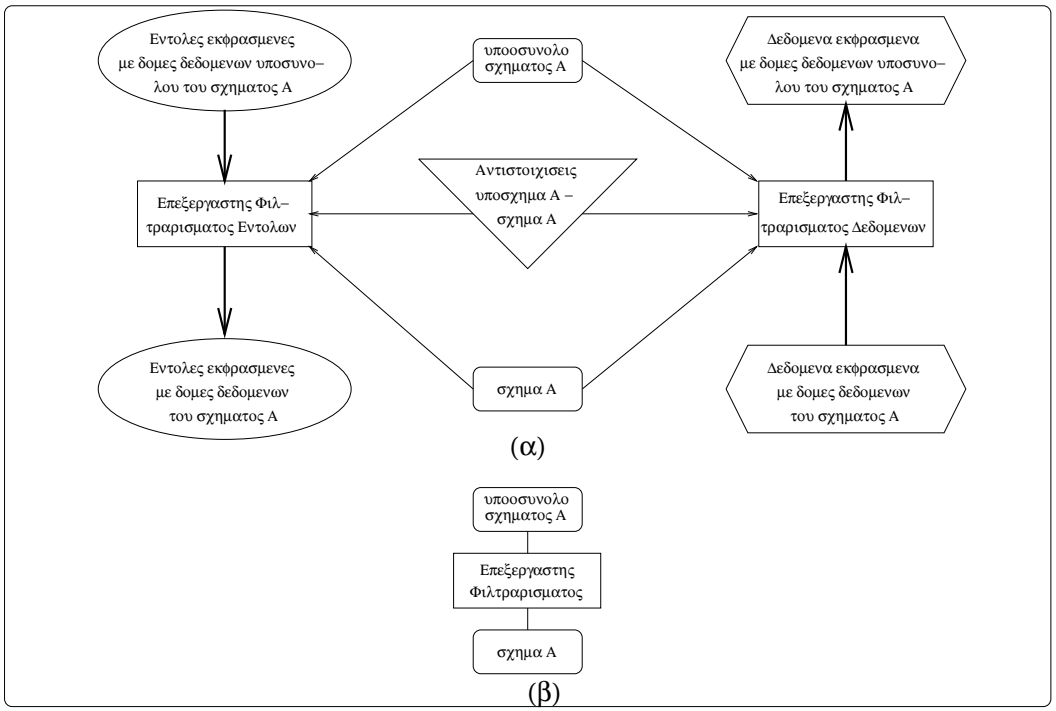
Επεξεργαστές Φιλτραρίσματος

Οι *επεξεργαστές φιλτραρίσματος* περιορίζουν τις εντολές και τα αντίστοιχα δεδομένα που μπορούν να περάσουν σε ένα άλλο επεξεργαστή. Μαζί με κάθε επεξεργαστή φιλτραρίσματος συνδέονται οι αντιστοιχίσεις που περιγράφουν οι περιορισμοί στις εντολές και τα δεδομένα. Αυτοί οι περιορισμοί μπορούν, είτε να είναι εμφυτευμένοι στον κώδικα του επεξεργαστή, είτε να καθορίζονται σε μια ξεχωριστή δομή δεδομένων.

Στο Σχήμα 2.6 βλέπουμε δύο επεξεργαστές φιλτραρίσματος, ένα που ελέγχει τις εντολές και ένα που ελέγχει τα δεδομένα. Μια σημαντική εργασία που μπορεί να διεκπεραιωθεί από ένα επεξεργαστή φιλτραρίσματος, είναι η *ενημέρωση όψεων*. Η εργασία αυτή, εμφανίζεται όταν οι διαφορές στις δομές δεδομένων μεταξύ της όψης και του σχήματος είναι τέτοιες, που μπορεί μια εντολή ενημέρωσης να μεταφραστεί με περισσότερους από ένα τρόπους.

Επεξεργαστές Δόμησης

Οι *επεξεργαστές δόμησης* τεμαχίζουν και/ή δημιουργούν αντίγραφα μιας λειτουργίας που προέρχεται από ένα επεξεργαστή. Τα τεμάχια/αντίγραφα στη συνέχεια, γίνονται

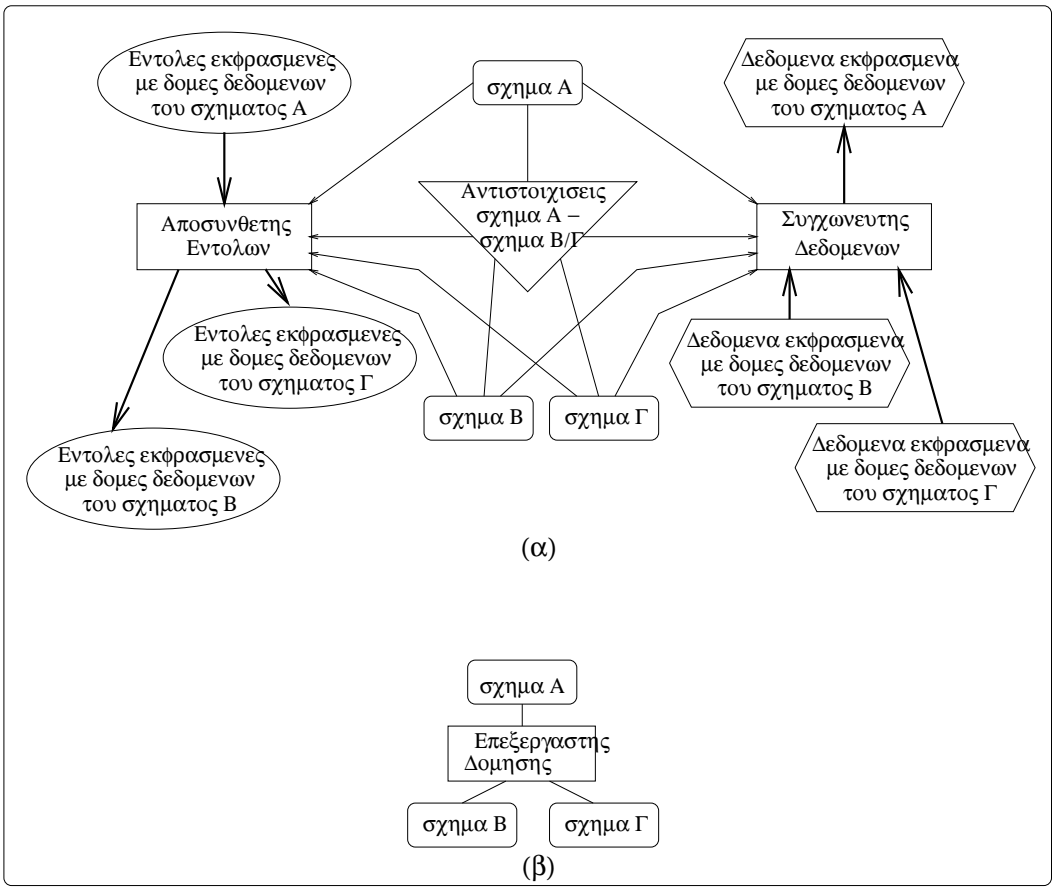


Σχήμα 2.6: (α) Ζεύγος Επεξεργαστών Φιλτραρίσματος (β) Αφηρημένη μορφή επεξεργαστή φιλτραρίσματος

δεκτά από δύο ή περισσότερους άλλους επεξεργαστές. Οι επεξεργαστές δόμησης επίσης, συγχωνεύουν τα δεδομένα που παράγονται από άλλους επεξεργαστές σε ένα σύνολο δεδομένων, το οποίο κατόπιν περνιέται σαν είσοδος σε τρίτο επεξεργαστή. Μπορούν να υποστηρίξουν διαφάνεια θέσης, κατανομής και αντιγράφων, μια και ένας επεξεργαστής που υποβάλλει μια εντολή δεν είναι ανάγκη να γνωρίζει τη θέση, την κατανομή και τον αριθμό των επεξεργαστών που συμμετέχουν στην επεξεργασία της εντολής.

Εργασίες που μπορούν να χειριστεί ένας επεξεργαστής δόμησης είναι :

- **Ενοποίηση Σχημάτων:** Ενοποίηση πολλών σχημάτων σε ένα τελικό σχήμα
- **Διαπραγμάτευση:** Καθορισμός του πρωτοκόλλου που θα χρησιμοποιηθεί από τους ιδιοκτήτες των διαφόρων σχημάτων προς ενοποίηση, ώστε να καθορισθούν τα δεδομένα του ενοποιημένου σχήματος
- **Αποσύνθεση Επερωτήσεων (Εντολών) και Βελτιστοποίηση:** Αποσύνθεση και βελτιστοποίηση μιας επερώτησης (εντολής), που εκφράζεται πάνω στο ενοποιημένο σχήμα

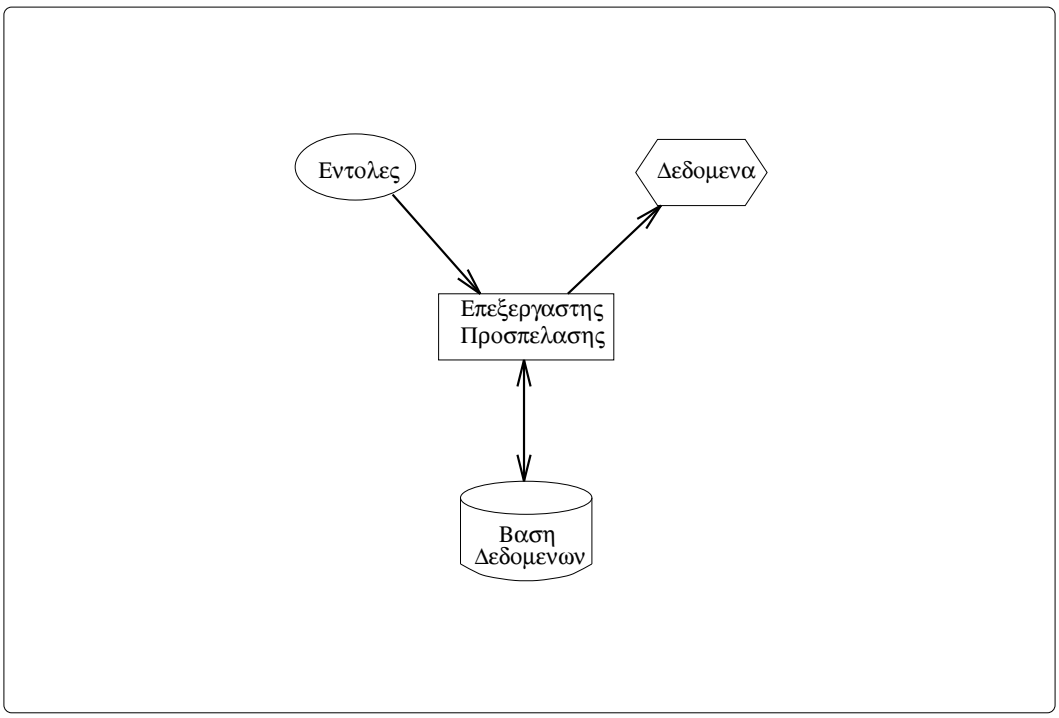


Σχήμα 2.7: (α) Ζεύγος Επεξεργαστών Δόμησης (β) Αφηρημένη μορφή επεξεργαστή δόμησης

- **Διαχείριση Καθολικών Συναλλαγών:** Έλεγχος ατομικότητας και ταυτοχρονισμού

Στο Σχήμα 2.7 βλέπουμε ένα ζευγάρι επεξεργαστών δόμησης. Χρησιμοποιώντας πληροφορίες από το σχήμα Α, το σχήμα Β, το σχήμα Γ, καθώς και τις αντιστοιχίσεις μεταξύ του σχήματος Α και των άλλων δύο, ο αποσυνθετής εντολών παράγει εντολές που χρησιμοποιούν αντικείμενα των σχημάτων Β και Γ, από εντολές που χρησιμοποιούν αντικείμενα του σχήματος Α. Το σχήμα Α είναι ένα *ενοποιημένο σχήμα* που περιέχει περιγραφές όλων, ή μέρους των δεδομένων που περιγράφονται από τα σχήματα Β και Γ. Χρησιμοποιώντας ακριβώς τις ίδιες πληροφορίες, ο συγχωνευτής δεδομένων παράγει δεδομένα στη μορφή των αντικειμένων του σχήματος Α, από δεδομένα της μορφής των αντικειμένων των σχημάτων Β και Γ.

Επεξεργαστές Προσπέλασης



Σχήμα 2.8: Επεξεργαστής Προσπέλασης

Ένας *επεξεργαστής προσπέλασης* δέχεται εντολές και παράγει δεδομένα εκτελώντας τις εντολές πάνω σε μια ΒΔ. Μπορεί να δέχεται εντολές από διαφορετικούς επεξεργαστές και να τις επεξεργάζεται παράλληλα.

Στο Σχήμα 2.8 βλέπουμε ένα επεξεργαστή προσπέλασης, ο οποίος δέχεται εντολές διαχείρισης δεδομένων και χρησιμοποιεί μεθόδους προσπέλασης για να ανακτήσει τα δεδομένα από τη ΒΔ.

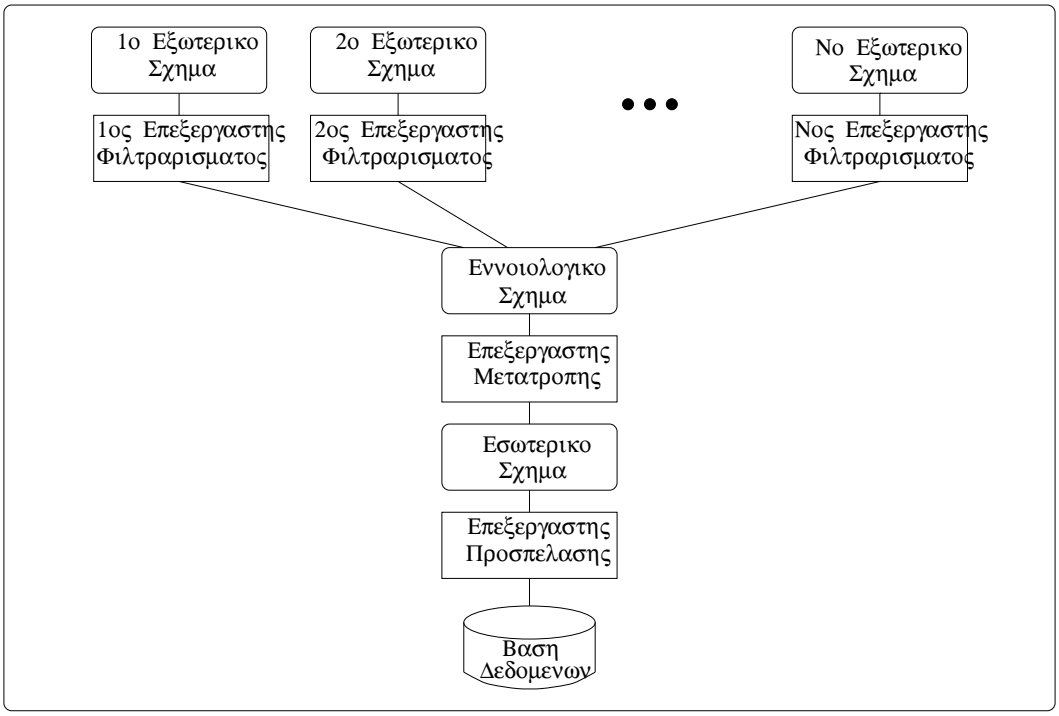
Ζητήματα που αντιμετωπίζονται από τους επεξεργαστές προσπέλασης είναι ο τοπικός έλεγχος ταυτοχρονισμού, η δέσμευση (commitment), η αποθήκευση αντιγράφων (backup), και η ανάνηψη (recovery).

2.4.3 Τύποι Σχημάτων της Αρχιτεκτονικής

Στην ενότητα αυτή θα δούμε αρχικά την πρότυπη αρχιτεκτονική των τριών επιπέδων για σχήματα κεντρικοποιημένων ΣΔΒΔ. Κατόπιν, θα δούμε την επεκτεταμμένη αρχιτεκτονική των πέντε επιπέδων, η οποία αντιμετωπίζει τις απαιτήσεις για κατανομή, αυτονομία και ετερογένεια των ΟΣΒΔ.

ANSI/SPARC Αρχιτεκτονική Σχήματος Τριών Επιπέδων

Η ομάδα μελέτης για Συστήματα Βάσεων Δεδομένων ANSI/X3/SPARC, έδωσε σε γενικές γραμμές την αρχιτεκτονική των τριών επιπέδων για περιγραφή δεδομένων ([16]). Τα τρία επίπεδα για την περιγραφή δεδομένων είναι το εννοιολογικό σχήμα, το εσωτερικό σχήμα και το εξωτερικό σχήμα.



Σχήμα 2.9: Αρχιτεκτονική κεντρικοποιημένου ΣΔΒΔ

Το *εννοιολογικό σχήμα* περιγράφει τις εννοιολογικές και λογικές δομές δεδομένων (δηλ., το σχήμα αποτελείται από αντικείμενα που παρέχουν ένα εννοιολογικό επίπεδο περιγραφής της ΒΔ) και τις συσχετίσεις ανάμεσά τους. Είναι μια προσπάθεια να περιγραφούν όλα τα δεδομένα που ενδιαφέρουν μια επιχείρηση.

Το *εσωτερικό σχήμα* περιγράφει τα φυσικά χαρακτηριστικά των λογικών δομών δεδομένων στο εννοιολογικό σχήμα. Αυτά τα χαρακτηριστικά περιέχουν πληροφορίες για τη θέση των εγγραφών στις φυσικές αποθηκευτικές συσκευές, τη θέση και τον τύπο των ευρετηρίων (indexes) και τη φυσική παράσταση των συσχετίσεων μεταξύ των λογικών εγγραφών.

Το υποσύνολο της ΒΔ που μπορεί να προσπελασθεί από ένα χρήστη, ή μια ομάδα

χρηστών περιγράφεται από ένα *εξωτερικό σχήμα*. Επειδή εν γένει υπάρχουν ομάδες χρηστών που χρειάζονται να προσπελάσουν διαφορετικά κομμάτια της ΒΔ, είναι λογικό να περιμένουμε ότι κάθε ομάδα θα απαιτεί ξεχωριστό εξωτερικό σχήμα.

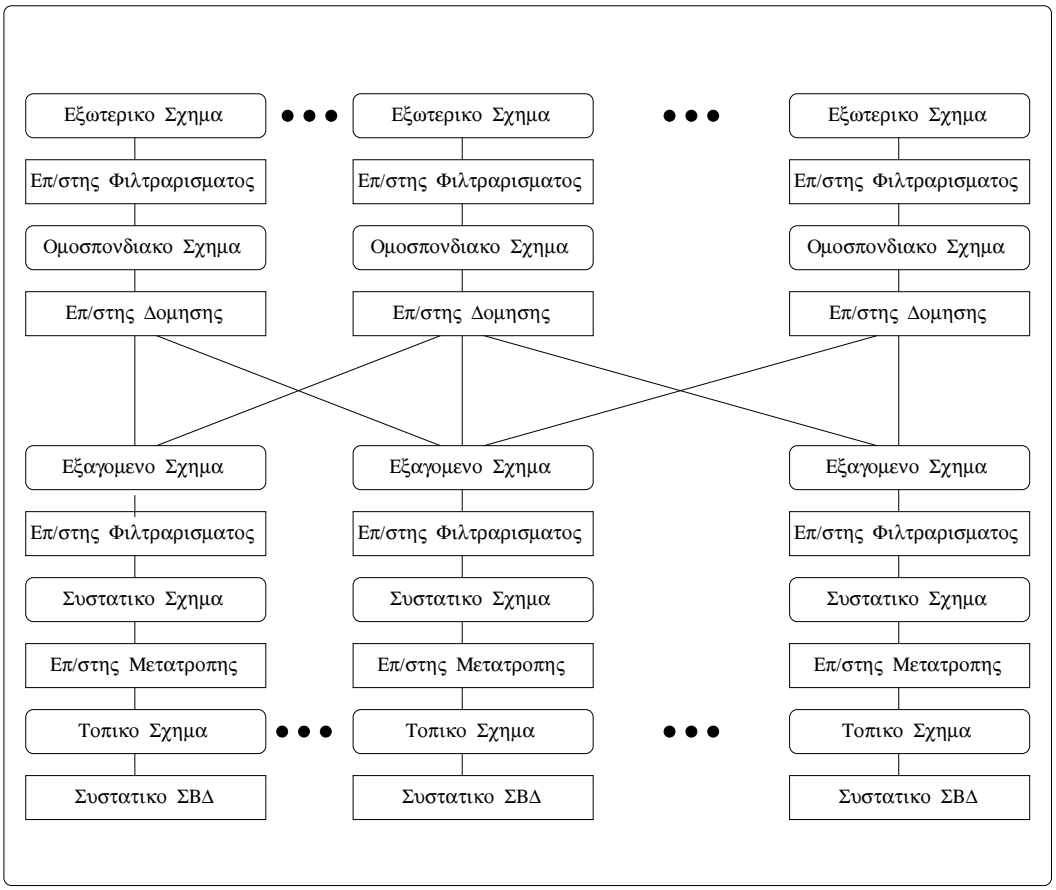
Σε σχέση με τις δομές που αναφέραμε στην προηγούμενη ενότητα, επεξεργαστές φιλτραρίσματος χρησιμοποιούν τις πληροφορίες στα εξωτερικά σχήματα για να ελέγξουν ποιά δεδομένα μπορούν να προσπελασθούν από ποιούς χρήστες. Επίσης, ένας επεξεργαστής μετατροπής μεταφράζει εντολές που εκφράζονται χρησιμοποιώντας αντικείμενα του εννοιολογικού σχήματος, σε εντολές που χρησιμοποιούν τα αντικείμενα του εσωτερικού σχήματος. Τέλος, ένας επεξεργαστής προσπέλασης εκτελεί τις εντολές και ανακτά τα δεδομένα από το φυσικό μέσο. Παραστατικά η αρχιτεκτονική φαίνεται στο Σχήμα 2.9.

Αρχιτεκτονική Σχήματος Πέντε Επιπέδων για ΟΣΒΔ

Η αρχιτεκτονική του σχήματος τριών επιπέδων, είναι κατάλληλη για την περιγραφή αρχιτεκτονικών κεντρικοποιημένων ΣΔΒΔ. Είναι όμως απαραίτητο, αυτή η αρχιτεκτονική να επεκταθεί έτσι ώστε να υποστηρίζει τις τρεις διαστάσεις ενός ΟΣΔΒΔ--κατανομή, ετερογένεια και αυτονομία. Μια δυνατή επέκτασή της είναι αυτή που φαίνεται στο Σχήμα 2.10, η οποία όπως βλέπουμε είναι αρχιτεκτονική σχήματος πέντε επιπέδων. Στο Σχήμα 2.10 βλέπουμε τόσο τα σχήματα, όσο και τους επεξεργαστές που χρησιμοποιούνται από την αρχιτεκτονική. Ας δούμε όμως καλύτερα, ποιά είναι και τί ρόλο παίζουν τα σχήματα κάθε επιπέδου σε μια τέτοια αρχιτεκτονική :

Το **τοπικό σχήμα** είναι το εννοιολογικό σχήμα καθενός από τα επιμέρους ΣΒΔ. Είναι εκφρασμένο στο μοντέλο δεδομένων κάθε ΣΒΔ, άρα διαφορετικά τοπικά σχήματα μπορούν να είναι εκφρασμένα σε διαφορετικά μοντέλα δεδομένων.

Το **συστατικό σχήμα** παράγεται από τη μετάφραση του τοπικού σχήματος σε ένα μοντέλο δεδομένων που ονομάζεται *κανονικό* ή *κοινό μοντέλο δεδομένων* (ΚΜΔ) του ΟΣΒΔ. Με τη χρησιμοποίηση του ΚΜΔ, αφενός πετυχαίνουμε την περιγραφή διαφορετικών σχημάτων χρησιμοποιώντας μια και μόνη παράσταση και αφετέρου έχουμε τη δυνατότητα σημασιολογικά χαρακτηριστικά που λείπουν ή αδυνατούν να εκφραστούν από ένα τοπικό σχήμα να μπορούν να προστεθούν στο αντίστοιχο συστατικό σχήμα. Η διαδικασία της μετάφρασης του τοπικού σχήματος στο συστατικό, παράγει αντιστοιχίσεις μεταξύ αντικειμένων των δύο σχημάτων. Έτσι, οι επεξεργαστές μετατροπής χρησιμοποιούν αυτές τις αντιστοιχίσεις για να μεταφράζουν εντολές για το συστατικό σχήμα σε εντολές για το τοπικό σχήμα. Με τους επεξεργαστές μετατροπής και τα συστατικά σχήματα, υποστηρίζεται το στοιχείο της ετερογένειας του ΟΣΒΔ.



Σχήμα 2.10: Αρχιτεκτονική ΟΣΔΒΔ

Το **εξαγόμενο σχήμα** αντιπροσωπεύει το υποσύνολο του συστατικού σχήματος που είναι διαθέσιμο στο ΟΣΒΔ. Σκοπός του ορισμού εξαγομένων σχημάτων είναι η διευκόλυνση του ελέγχου και της διαχείρισης της αυτονομίας συσχέτισης. Μπορούμε να χρησιμοποιήσουμε έναν επεξεργαστή φιλτραρίσματος για να παρέχουμε έλεγχο προσπέλασης στο εξαγόμενο σχήμα, περιορίζοντας το σύνολο των επιτρεπόμενων λειτουργιών στις οποίες μπορεί να υποβληθεί το αντίστοιχο συστατικό σχήμα. Με αυτούς τους επεξεργαστές φιλτραρίσματος και τα εξαγόμενα σχήματα, υποστηρίζεται το στοιχείο της αυτονομίας του ΟΣΒΔ.

Το **ομοσπονδιακό σχήμα** είναι ένα σχήμα που προκύπτει από την ενοποίηση πολλών εξαγομένων σχημάτων. Οι πληροφορίες για την κατανομή των δεδομένων που παράγονται κατά την ενοποίηση των εξαγομένων σχημάτων, μπορούν είτε να αποθηκευτούν στο ίδιο το ομοσπονδιακό σχήμα, είτε σε κάποιο ξεχωριστό σχήμα που ονομάζεται *σχήμα κατανομής*

ή *σχήμα ανάθεσης*. Μπορούμε να χρησιμοποιήσουμε έναν επεξεργαστή δόμησης για να μετατρέψουμε εντολές για το ομοσπονδιακό σχήμα, σε εντολές για ένα ή περισσότερα εξαγόμενα σχήματα. Με τους επεξεργαστές δόμησης και τα ομοσπονδιακά σχήματα, υποστηρίζεται το στοιχείο της κατανομής του ΟΣΒΔ. Εν γένει μπορούν να υπάρχουν πολλά ομοσπονδιακά σχήματα σε ένα ΟΣΒΔ, ανάλογα με τις ομάδες των ομοσπονδιακών χρηστών (οι οποίοι μπορεί να είναι ένα σύνολο χρηστών και/ή εφαρμογών που εκτελούν συναφείς δραστηριότητες).

Τέλος, το **εξωτερικό σχήμα** είναι το σχήμα με το οποίο έρχεται σε επαφή ένα χρήστης και/ή μια εφαρμογή ή ομάδα χρηστών/εφαρμογών. Λόγοι που δικαιολογούν τη χρησιμοποίηση εξωτερικών σχημάτων σε ΟΣΒΔ είναι ότι πιθανόν το ομοσπονδιακό σχήμα να έχει μεγάλη έκταση και πολυπλοκότητα με αποτέλεσμα ο χρήστης δύσκολα να μπορεί να έρθει σε επαφή με αυτό, ότι πιθανόν να θέλουμε να θέσουμε επιπλέον περιορισμούς ακεραιότητας σε αυτό, ή τέλος ότι θέλουμε να παρέχουμε ένα έλεγχο πρόσβασης στα δεδομένα που διαχειρίζεται το ΟΣΒΔ. Ένας επεξεργαστής φιλτραρίσματος αναλύει τις εντολές στο εξωτερικό σχήμα ώστε να εξασφαλίσει ότι είναι σύμφωνες με τους περιορισμούς ακεραιότητας και τον έλεγχο προσπέλασης του ομοσπονδιακού σχήματος. Δεδομένου ότι υπάρχει περίπτωση το μοντέλο δεδομένων του ομοσπονδιακού και του εξωτερικού σχήματος να είναι διαφορετικά, πιθανόν να χρειάζεται και ένας επεξεργαστής μετατροπής, για να μετατρέπει εντολές για το εξωτερικό σχήμα σε εντολές για το ομοσπονδιακό σχήμα.

Στην αρχιτεκτονική σχήματος πέντε επιπέδων που παρουσιάσαμε, είναι δυνατό να έχουμε διάφορους πλεονασμούς μεταξύ σχημάτων κάποιων εκ των πέντε επιπέδων. Τέτοιοι μπορεί να είναι ο πλεονασμός μεταξύ εξωτερικών και ομοσπονδιακών σχημάτων, ή ο πλεονασμός μεταξύ συστατικών και τοπικών σχημάτων. Οι πλεονασμοί αυτοί αντιμετωπίζονται με την αφαίρεση του πλεονάζοντος σχήματος από την τελική αρχιτεκτονική.

2.5 Υπάρχοντα Ομοσπονδιακά Συστήματα Βάσεων Δεδομένων

Ας δούμε όμως τώρα μια συνοπτική παρουσίαση μερικών συστημάτων ετερογενών και κατανεμημένων βάσεων δεδομένων, που έχουν αναπτυχθεί για εμπορική χρήση ([45], [10]).

Το σύστημα **ADDS** (Amoco Distributed Database System) ([45], [9]), αναπτύχθηκε προς το τέλος του 1983, με σκοπό να ανταποκριθεί στο πρόβλημα της ενοποίησης διαφορετικών βάσεων δεδομένων, που ήταν κατανεμημένες μέσα στο χώρο της εταιρίας Amoco. Παρέχει ομοιόμορφη προσπέλαση σε ετερογενείς κατανεμημένες ΒΔ, βασίζεται στη χρήση του σχεσιακού μοντέλου σαν κανονικό μοντέλο δεδομένων, ενώ υποστηρίζει

ενοποίηση τοπικών ΒΔ με σχεσιακό, δικτυωτό ή ιεραρχικό σχήμα. Χαρακτηρίζεται σαν στενά συνδεδεμένο ΟΣΒΔ που υποστηρίζει την ανάπτυξη πολλαπλών ομοσπονδιακών σχημάτων. Τέλος, διαθέτει χαρακτηριστικά ανεκτικότητας λαθών, καθώς και τη λειτουργία αποσύνδεσης για τους ομοσπονδιακούς χρήστες, η οποία είναι πολύ σημαντική για χρονοβόρες επερωτήσεις.

Το σύστημα **DATAPLEX** ([45], [15]), είναι ένα ετερογενές κατανεμημένο ΣΒΔ που αναπτύχθηκε από τη General Motors Corporation, με σκοπό να αποτελέσει ένα αποδοτικό μέσο για το διαμοιρασμό των δεδομένων, μέσα σ' ένα οργανισμό που υπάρχουν διαφορετικά συστήματα δεδομένων. Και αυτό όπως και το προηγούμενο, είναι ένα στενά συνδεδεμένο ΟΣΒΔ που υποστηρίζει την ανάπτυξη πολλαπλών ομοσπονδιακών σχημάτων. Το μοντέλο του κανονικού μοντέλου δεδομένων που χρησιμοποιεί, είναι και εδώ το σχεσιακό. Τέλος χρησιμοποιεί τη γλώσσα SQL για τη διατύπωση επερωτήσεων, ενώ παρέχει δυνατότητα ενοποίησης τοπικών ΒΔ με σχεσιακό ή ιεραρχικό σχήμα.

Το σύστημα **IMDAS** (Integrated Manufacturing Data Administrator System) ([45]), αναπτύχθηκε με στόχο να παρέχει προσπέλαση από πολλαπλά συστήματα σε πολλαπλές πηγές βιομηχανικών δεδομένων. Ανήκει στην κατηγορία των στενά συνδεδεμένων ΟΣΒΔ και υποστηρίζει ένα και μοναδικό καθολικό σχήμα. Χρησιμοποιεί το σημασιολογικό μοντέλο SAM (Semantic Association Model) σα κανονικό μοντέλο δεδομένων, μια και είναι ικανότερο από σχεσιακό στην παράσταση πολύπλοκων δομών και συσχετίσεων που υπάρχουν σε μια βιομηχανική εταιρία. Τέλος, υποστηρίζει κατανεμημένες ενημερώσεις και ανακλήσεις δεδομένων, χρησιμοποιώντας μια SQL-like γλώσσα επερωτήσεων.

Το σύστημα **Mermaid** ([45], [43]), αναπτύχθηκε από την ανάγκη του Αμερικανικού Υπουργείου Άμυνας, για προσπέλαση και ενοποίηση δεδομένων που ήταν σε αυτόνομες ΒΔ. Είναι ένα στενά συνδεδεμένο ΟΣΒΔ που υποστηρίζει πολλαπλά ομοσπονδιακά σχήματα. Υποστηρίζει αρκετά επίπεδα ετερογένειας όπως: στο υλικό, στο λειτουργικό σύστημα των κόμβων των ΣΔΒΔ, στα μοντέλα δεδομένων (σχεσιακό, δικτυωτό, σειριακό αρχείο), κ.α.. Τέλος, επιτρέπει ανάκληση δεδομένων από πολλαπλές ΒΔ, αλλά ενημέρωση μόνο σε μία. Το εμπορικό προϊόν που είναι απόγονος του Mermaid, λέγεται **InterViso** ([44]) και αναπτύχθηκε από την εταιρία Data Integration Inc..

Τέλος, το σύστημα **MULTIBASE** ([45]), όπως και τα προηγούμενα, σχεδιάστηκε για να δίνει τη δυνατότητα προσπέλασης ενοποιημένων δεδομένων (integrated data) από ήδη υπάρχοντα ΣΒΔ, χωρίς μεταβολή ή αντικατάστασή τους. Χρησιμοποιεί τη γλώσσα ορισμού και διαχείρισης δεδομένων DAPLEX, η οποία βασίζεται στο συναρτησιακό (functional) μοντέλο δεδομένων. Είναι ένα στενά συνδεδεμένο ΟΣΒΔ με πολλαπλά σχήματα ή όψεις.

Υποστηρίζει οριζόντια και κάθετη τμηματοποίηση των δεδομένων, ενώ δεν υποστηρίζει ενημερώσεις. Τέλος, το συναρτησιακό μοντέλο που χρησιμοποιείται σαν κανονικό μοντέλο δεδομένων, επιτρέπει την δυνατότητα στα τοπικά σχήματα των συστατικών ΒΔ να είναι σχεσιακά, ιεραρχικά ή δικτυωτά.

Κεφάλαιο 3

Ορισμός του Προβλήματος και Βιβλιογραφική Ανασκόπηση

Έχοντας παρουσιάσει στο προηγούμενο κεφάλαιο τα πιο σημαντικά στοιχεία των ΟΣΒΔ, είμαστε έτοιμοι να δώσουμε τα πλαίσια μέσα στα οποία κινείται αυτή η εργασία. Έτσι, αρχικά θα ορίσουμε όσο το δυνατόν πληρέστερα το πρόβλημα που μας απασχόλησε, καθώς και τον τρόπο επίλυσης που προτείναμε. Κατόπιν, θα ασχοληθούμε με δουλειά που έχει γίνει διεθνώς και είναι σχετική με τις λύσεις που προτείνουμε, ώστε να μπορούμε αργότερα να εκτιμήσουμε την παρούσα εργασία.

3.1 Το Πρόβλημα

Το πρόβλημα που καλούμαστε να αντιμετωπίσουμε, ορίζεται ως εξής:

> *Να κατασκευαστεί ομοσπονδιακό σχήμα από δεδομένα σχεσιακά σχήματα τοπικών βάσεων δεδομένων.*

Το πρόβλημα αυτό, είναι μέρος του μεγαλύτερου προβλήματος της ανάπτυξης ενός ΟΣΒΔ, που θα παρέχει τη δυνατότητα στους ομοσπονδιακούς χρήστες, για προσπέλαση σε δεδομένα των τοπικών (εν γένει γεωγραφικά απομακρυσμένων), ετερογενών και αυτόνομων σχεσιακών βάσεων δεδομένων. Οι προσεγγίσεις που χρησιμοποιούνται για την αντιμετώπιση του γενικότερου προβλήματος προσπέλασης των τοπικών ΒΔ, σ' ένα ΟΣΒΔ, είναι δύο ([10]).

Η πρώτη προσέγγιση είναι η χρησιμοποίηση μιας *γλώσσας πολυβάσεων δεδομένων* ([21], [29]). Οι γλώσσες αυτές χρησιμοποιούνται από χαλαρά συνδεδεμένα ΟΣΒΔ, στα οποία δεν υπάρχει ομοσπονδιακό σχήμα. Επιτρέπουν στους ομοσπονδιακούς χρήστες, τη

διατύπωση επερωτήσεων που περιέχουν αναφορές σε δεδομένα διαφορετικών συστατικών ΒΔ. Είναι φανερό, ότι για να μπορέσει ο χρήστης να τις μεταχειριστεί, πρέπει αφενός να έχει βαθειά γνώση της σημασιολογίας των συστατικών σχημάτων, και αφετέρου να του παρέχεται η δυνατότητα παρουσίασης των διαθέσιμων πληροφοριών, έτσι ώστε να ξέρει τί ακριβώς να ψάξει από το διαθέσιμο σύνολο πληροφοριών.

Η δεύτερη προσέγγιση είναι η χρησιμοποίηση ενός καθολικού σχήματος (ρόλο τον οποίο παίζει το ομοσπονδιακό σχήμα) και η διατύπωση επερωτήσεων πάνω σ' αυτό. Οι επερωτήσεις που διατυπώνονται (π.χ. με τη χρήση μιας γλώσσας επερωτήσεων), μεταφράζονται στη γλώσσα επερωτήσεων των τοπικών ΣΔΒΔ και δρομολογούνται προς αυτά. Μια τέτοια προσέγγιση μπορεί να εφαρμοστεί, τόσο σε στενά, όσο και σε χαλαρά συνδεδεμένα ΟΣΒΔ. Η δυνατότητα επερωτήσεων πάνω στο καθολικό σχήμα παρέχει μια πιο "φιλική" προσπέλαση για το χρήστη, καθώς κρύβει πιθανές διαφορές των συστατικών ΒΔ και παρέχει διαφάνεια αντιγράφων (duplicates) και θέσης. Η χρήση του καθολικού σχήματος, δίνει επίσης τη δυνατότητα στο διαχειριστή του, να ορίσει όψεις (εξωτερικά σχήματα) ανάλογα με τα είδη και την άδεια πρόσβασης των διαφόρων ομάδων ομοσπονδιακών χρηστών. Πιθανά προβλήματα αυτής της προσέγγισης είναι η απαίτηση για το διαχειριστή να έχει βαθειά γνώση της σημασιολογίας των τοπικών σχημάτων, ο χώρος στους δίσκους που απαιτεί το καθολικό σχήμα για να αποθηκευτεί δεδομένου ότι μπορεί να είναι αρκετά μεγάλο, και τέλος η πιθανή ανακατασκευή του καθολικού σχήματος όταν υπάρξουν αλλαγές στα τοπικά σχήματα, ή προσάρτηση νέων.

Βλέπουμε λοιπόν ότι για ΟΣΒΔ, στα οποία δεν υπάρχουν συχνές αλλαγές των τοπικών σχημάτων και στους ομοσπονδιακούς χρήστες επιβάλλονται περιορισμοί προσπέλασης, είναι προτιμότερη η προσέγγιση του καθολικού σχήματος, μια και το περιβάλλον επερωτήσεων που παρέχει στους χρήστες είνα φιλικότερο και ευκολότερο στη χρήση. Για την ανάπτυξη του καθολικού σχήματος ενός ΟΣΒΔ, δύο προσεγγίσεις μπορούν ν' ακολουθηθούν: η *ανιούσα* (bottom-up) και η *κατιούσα* (top-down) προσέγγιση ([42], [14]).

Η πρώτη προσέγγιση, ακολουθείται στην περίπτωση που από δοθέντα τοπικά ΣΒΔ θέλουμε να κατασκευάσουμε ένα ΟΣΔΒΔ, που θα παρέχει πρόσβαση στα δεδομένα τους, διατηρώντας την αυτονομία τους. Έτσι, η διαδικασία που ακολουθείται περιέχει την καταρχήν μετάφραση των τοπικών σχημάτων σ ένα κοινό μοντέλο δεδομένων, τον καθορισμό εξαγομένων σχημάτων από αυτά, και τέλος την ενοποίησή τους σ' ένα καθολικό ομοσπονδιακό σχήμα. Κατόπιν, είμαστε σε θέση, ανάλογα με τις ομάδες χρηστών, να ορίσουμε εξωτερικά σχήματα πάνω από αυτό.

Η δεύτερη προσέγγιση, ακολουθείται στην περίπτωση που ήδη υπάρχει ένα καθολικό σχήμα και θέλουμε να το επεκτείνουμε για να καλύψει νέες ανάγκες χρηστών (π.χ. υποστήριξη νέων εφαρμογών). Η διαδικασία που ακολουθείται σε μια τέτοια περίπτωση περιέχει:

1. **Καθορισμό ή τροποποίηση εξωτερικών σχημάτων:** Συλλέγονται οι απαιτήσεις των ομοσπονδιακών χρηστών και αναλύονται ώστε να ορισθούν νέα εξωτερικά σχήματα ή επεκτάσεις των υπαρχόντων εξωτερικών σχημάτων.

2. **Ανάλυση σχημάτων:** Συγκρίνονται τα σχετικά ομοσπονδιακά με τα εξωτερικά σχήματα για να διαπιστωθούν τα τμήματα των εξωτερικών σχημάτων που ήδη υπάρχουν στο ομοσπονδιακό σχήμα και έτσι υποστηρίζονται από το ΟΣΒΔ. Αν κάποιο τμήμα των εξωτερικών σχημάτων (*προσωρινό σχήμα*) δεν υποστηρίζεται από το ΟΣΒΔ, το ομοσπονδιακό σχήμα πρέπει να επεκταθεί ώστε να περιέχει αυτό το τμήμα. Τελικά λοιπόν, ένα ή περισσότερα συστατικά σχήματα πρέπει να υποστηρίξουν το προσωρινό σχήμα. Αυτό μπορεί να επιτευχθεί με ένα από τους εξής τρεις τρόπους.

   (a) Αν τα δεδομένα που απαιτούνται υπάρχουν σε μία ή περισσότερες συστατικές ΒΔ, γίνεται διαπραγμάτευση με τους διαχειριστές αυτών των ΒΔ έτσι ώστε τα τμήματα των σχημάτων που περιγράφουν αυτά τα δεδομένα να τεθούν στα εξαγόμενα σχήματα με κατάλληλα δικαιώματα προσπέλασης.

   (b) Αν τα δεδομένα που απαιτούνται δεν υπάρχουν σε καμμιά συστατική ΒΔ και κάποιος διαχειριστής συστατικής ΒΔ είναι πρόθυμος να αλλάξει το σχήμα της συστατικής του ΒΔ ώστε να υποστηρίζει αυτά τα δεδομένα, τότε μετατρέπονται το τοπικό, συστατικό και εξαγόμενο σχήμα της ΒΔ αυτής.

   (c) Αν δεν ισχύει τίποτα από τα προηγούμενα, το προσωρινό σχήμα υλοποιείται σα ξεχωριστή ΒΔ ενός ήδη υπάρχοντος ΣΔΒΔ, ή εναλλακτικά ένα νέο ΣΔΒΔ μπορεί να χρησιμοποιηθεί απαιτώντας και ένα νέο επεξεργαστή μετατροπής. Τελικά λοιπόν, το προσωρινό σχήμα γίνεται ένα νέο συστατικό σχήμα.

3. **Ενοποίηση σχημάτων:** Ενοποιείται το προσωρινό σχήμα με το σχετικό ομοσπονδιακό σχήμα.

Δεδομένου ότι στην εργασία μας ζητείται κατασκευή ενός καθολικού σχήματος, από προϋπάρχοντα σχεσιακά σχήματα των τοπικών ΒΔ, είναι φανερό ότι υιοθετήσαμε την πρώτη προσέγγιση για την επίλυση του προβλήματος μας. Η πορεία που ακολουθούμε περιλαμβάνει:

1. Μετάφραση των τοπικών σχεσιακών σχημάτων σε σημασιολογικά δίκτυα, σχήματα δηλαδή του σημασιολογικού μοντέλου δεδομένων.

2. Ενοποίηση των σημασιολογικών δικτύων και παραγωγή του καθολικού σχήματος.

Η χρήση του σημασιολογικού μοντέλου δεδομένων, στο οποίο εκφράζονται τα σχήματα που κατασκευάζονται από τη γλώσσα παράστασης γνώσης TELOS ([7]), θεωρείται επαρκής σύμφωνα με τα κριτήρια της εκφραστικότητας και του σημασιολογικού συσχετισμού, όπως αναφέρονται στο [40]. Για τη μετάφραση των τοπικών σχεσιακών σχημάτων σε σχήματα του σημασιολογικού μοντέλου δεδομένων, αναπτύξαμε μια μέθοδο μετάφρασης η οποία παρουσιάζεται στο κεφάλαιο 4, ενώ για την ενοποίηση των σημασιολογικών μοντέλων αναπτύξαμε μια μέθοδο ενοποίησης που παρουσιάζεται στο κεφάλαιο 5. Παρακάτω, ακολουθεί μια παρουσίαση εργασιών που έχουν γίνει και αφορούν τα δύο αυτά θέματα, της μετάφρασης και της ενοποίησης σχημάτων, έτσι ώστε να μπορέσουμε αργότερα να συγκρίνουμε με αυτές τη δικιά μας δουλειά.

3.2 Μέθοδοι Μετάφρασης Σχημάτων

Το πρόβλημα της μετάφρασης ενός σχήματος που ακολουθεί ένα μοντέλο δεδομένων, προς ένα άλλο σχήμα που ακολουθεί άλλο μοντέλο δεδομένων, είναι αρκετά γνωστό και παλιό στη βιβλιογραφία του τομέα των βάσεων δεδομένων.

Μέχρι παλιότερα μας ενδιέφεραν μόνο μέθοδοι μετάφρασης σχημάτων, που το μοντέλο δεδομένων τους ήταν κάποιο πλούσιο σημασιολογικά (π.χ. Object-Oriented, Entity-Relationship), προς σχήματα που το μοντέλο δεδομένων τους ήταν φτωχότερο σημασιολογικά (π.χ. Σχεσιακό). Ο λόγος γιαυτό είναι ότι, κατά τη σχεδίαση του σχήματος μιας ΒΔ θέλουμε ένα εκφραστικό, περιεκτικό και εύκολο στη χρήση μοντέλο, ενώ κατά την υλοποίηση της ΒΔ θέλουμε ένα μοντέλο που επιτρέπει γρήγορη ανάκτηση δεδομένων, παρέχει οικονομία χώρου και είναι ώριμο τεχνολογικά. Μέθοδοι που πετυχαίνουν την αντίστροφη μετάφραση, άρχισαν να μας ενδιαφέρουν με την εξέλιξη των ΟΣΔΒΔ τα οποία χρησιμοποιούν πλούσια σημασιολογικά μοντέλα ως κανονικά μοντέλα δεδομένων (ΚΜΔ).

Στο πλαίσιο της πρώτης κατηγορίας μεθόδων και εργαλείων μετάφρασης σχημάτων, μια από τις πιο πρόσφατες βιβλιογραφικές αναφορές είναι αυτή του [6].

Στην εργασία αυτή, παρουσιάζεται η αρχιτεκτονική και η λειτουργία ενός εργαλείου, που βοηθά το χρήστη του να υλοποιεί σχεσιακές ΒΔ, σχεδιάζοντάς τις χρησιμοποιώντας το οντοκεντρικό μοντέλο. Το εργαλείο αποτελείται από τρία μέρη : τον *Συντάκτη*

οντοκεντρικών σχημάτων (OMTool), τον *Μεταφραστή οντοκεντρικών σε σχεσιακά σχήματα* (Schemer) και το *Σχεσιακό ΣΔΒΔ* (RDBMS).

Ο Συντάκτης αποθηκεύει δύο παραστάσεις του οντοκεντρικού σχήματος που σχεδιάζει ο χρήστης: τη *γραφική* και τη *λογική*. Στη γραφική παράσταση χρησιμοποιούνται κουτιά, γραμμές και κείμενο, δίνοντας έτσι στο χρήστη μια φιλικότερη παράσταση του σχήματος. Στη λογική παράσταση περιγράφονται οι κλάσεις και οι συσχετίσεις που απορρέουν από το τη λογική αφαίρεση του σχήματος. Ειδικότερα, ο Συντάκτης επιτρέπει στο χρήστη να δημιουργήσει, να "φορτώσει", να συντάξει, να αποθηκεύσει και να εκτυπώσει διαγράμματα αντικειμένων, αυτόματα. Κατά ένα τρόπο λοιπόν, διαχειρίζεται τη λογική διασύνδεση, καθώς ο χρήστης μετακινεί συσχετιζόμενες οντότητες πάνω στην οθόνη.

Ο Μεταφραστής μετατρέπει το αρχείο που περιέχει τη λογική παράσταση του οντοκεντρικού σχήματος που σχεδίασε ο χρήστης, στις απαραίτητες προτάσεις SQL για να δημιουργηθεί το σχεσιακό σχήμα. Για τη λειτουργία του είναι απαραίτητα δύο μεταμοντέλα, το μεταμοντέλο αντικειμένων και το μεταμοντέλο ιδεωδών πινάκων. Καλούνται μεταμοντέλα γιατί είναι μοντέλα μοντέλων. Ειδικότερα, το *μεταμοντέλο αντικειμένων* περιγράφει τη δομή των διαγραμμάτων αντικειμένων που καταλήγουν στο Μεταφραστή από τη λογική αναπαράσταση του Συντάκτη, καθαρά και περιεκτικά. Επίσης, απλουστεύει την υλοποίηση των κανόνων παραγωγής του σχεσιακού σχήματος. Τέλος, για την περίπτωση που διαφορετικά σχεσιακά ΣΔΒΔ έχουν σχήματα που έχουν προέλθει από ένα οντοκεντρικό σχήμα, το μεταμοντέλο αντικειμένων παρέχει ένα κοινό λεξικό δεδομένων για τα διαφορετικά σχεσιακά ΣΔΒΔ. Το *μεταμοντέλο ιδεωδών πινάκων* τώρα, περιγράφει τη δομή των πινάκων που το σχεσιακό σχήμα που θα παραχθεί από το Μεταφραστή, θα περιέχει. Δίνεται έτσι η δυνατότητα στο χρήστη να προσαρμόσει τη δομή των πινάκων, στις ανάγκες του συγκεκριμένου (*ιδεώδους*) σχεσιακού ΣΔΒΔ που χρησιμοποιεί. Τα δύο μεταμοντέλα περιέχουν εκατέρωθεν αναφορές, επιτρέποντας έτσι την πλοήγηση μεταξύ οντοκεντρικών και σχεσιακών δομών. Αυτό είναι χρήσιμο για προχωρημένες εφαρμογές, που πρέπει να εκμεταλλεύονται την οντοκεντρική σημασιολογία των εφαρμογών, αλλά να αποθηκεύουν τα δεδομένα σε σχεσιακούς πίνακες.

Η μετατροπή της λογικής παράστασης του οντοκεντρικού σχήματος, στις προτάσεις SQL του σχεσιακού σχήματος, γίνεται σε τρία στάδια. Στο πρώτο στάδιο, διαβάζεται το αρχείο και προστίθεται πληθυσμός στο μεταμοντέλο αντικειμένων. Στο δεύτερο στάδιο, με τη χρήση κανόνων αντιστοίχισης, αντιστοιχίζονται οντοκεντρικές δομές από το μεταμοντέλο αντικειμένων σε "ιδεώδεις" σχεσιακούς πίνακες, που περιγράφονται από το μεταμοντέλο ιδεωδών πινάκων. Τέλος στο τρίτο στάδιο, αντιστοιχίζονται οι ιδεώδεις

πίνακες στη διάλεκτο της SQL, που υποστηρίζεται από το τελικό σχεσιακό ΣΔΒΔ.

Στην ίδια κατηγορία μεθόδων μετάφρασης εμπίπτει και η δουλειά του [33], όπου αναπτύσσεται αλγόριθμος μετάφρασης σχημάτων εκπεφρασμένων στο Επεκτεταμμένο μοντέλο Οντοτήτων Συσχετίσεων (Extended Entity Relationship), προς σχεσιακά σχήματα BCNF μορφής. Στη δουλειά αυτή η διαδικασία μετάφρασης αποτελείται από τέσσερα στάδια.

Αρχικά, ορίζουν την έννοια της *ορθής* παράστασης ενός EER σχήματος σε σχεσιακό μοντέλο, βάσει ορισμένων κριτηρίων ορθότητας, και αναπτύσσουν μέθοδο μετάφρασης ενός EER σχήματος προς ένα *κανονικό* σχεσιακό σχήμα που ακολουθεί αυτά τα κριτήρια. Αυτή η μέθοδος μετάφρασης παράγει μια κλάση σχεσιακών σχημάτων που είναι όμοια ως προς την ονομασία των γνωρισμάτων τους, και αποτελούνται από σχήματα σχέσεων, συναρτησιακές εξαρτήσεις, εξαρτήσεις εγκλεισμού και περιορισμούς ακυρότητας (null constraints).

Κατόπιν ακολουθούν δύο στάδια που μπορούν να εφαρμοστούν με οποιαδήποτε σειρά, καταλήγοντας στην παραγωγή ενός σχεσιακού σχήματος σε BCNF μορφή. Σ' ένα απ' αυτά τα στάδια γίνεται η απόδοση ονομάτων στα γνωρίσματα του σχεσιακού σχήματος. Έτσι με βάση το πλαίσιο που παρέχεται από τις υποθέσεις της *Σχέσης Σύμπαντος* (Universal Relation assumptions) ([31]) που περιορίζει την απόδοση ονομάτων σε σχεσιακά γνωρίσματα, καθορίζονται συνθήκες που πρέπει να ικανοποιούνται από τα ονόματα των γνωρισμάτων και αναπτύσσεται αλγόριθμος απόδοσης *καθολικών* ονομάτων στα γνωρίσματα, έτσι ώστε ο συνολικός αριθμός των ονομάτων αυτών να είναι ο ελάχιστος δυνατός. Στο άλλο στάδιο γίνεται η μετατροπή του κανονικού σχεσιακού σχήματος σε BCNF κανονική μορφή, βάσει διαδικασίας κανονικοποίησης.

Στο τελευταίο στάδιο αναπτύσσουν μια διαδικασία *συγχώνευσης* σχέσεων του σχήματος, προκειμένου να μειώσουν τον αριθμό των σχέσεων και να ελαττώσουν έτσι τις πράξεις γινομένου (joins), πετυχαίνοντας καλύτερη απόδοση στην απάντηση επερωτήσεων. Με τη διαδικασία αυτή εισάγονται στο νέο σχήμα μερικοί ακόμα περιορισμοί ακυρότητας, ενώ παράλληλα διατηρείται η BCNF κανονική μορφή στην οποία βρίσκεται.

Όπως αναφέρουν οι συγγραφείς του [33] οι διαδικασίες που έχουν αναπτύξει σε κάθε στάδιο είτε μπορούν να ενοποιηθούν αποτελώντας μια ολοκληρωμένη διαδικασία μετάφρασης, είτε μπορούν να εκληφθούν ως ξεχωριστά βήματα σε μια πορεία μετάφρασης. Αυτοί ακολουθώντας μια υβριδική τακτική, ανέπτυξαν ένα εργαλείο *Ορισμού και Μετάφρασης Σχήματος* (Schema Definition and Translation tool) στο οποίο η διαδικασία

μετάφρασης του EER σχήματος σε κανονικό σχεσιακό σχήμα και η διαδικασία κανονικοποίησης έχουν ενοποιηθεί, ενώ η απόδοση ονομάτων και η συγχώνευση αποτελούν ξεχωριστές διαδικασίες. Μ' αυτό τον τρόπο δίνεται η δυνατότητα στο χρήστη του εργαλείου να επιλέξει τους αλγορίθμους απόδοσης ονομάτων σε γνωρίσματα και συγχώνευσης, από μια λίστα αλγορίθμων που διαρκώς μπορεί να επεκτείνεται.

Όπως προαναφέραμε με την εξέλιξη των ΟΣΔΒΔ άρχισαν να μας ενδιαφέρουν μέθοδοι και εργαλεία για την αντίστροφη μετάφραση σχημάτων, και αυτό γιατί πιστεύουμε ότι το κανονικό μοντέλο δεδομένων (ΚΜΔ) που χρησιμοποιείται από το ΟΣΒΔ για το ομοσπονδιακό σχήμα πρέπει να είναι ένα σημασιολογικά πλούσιο μοντέλο το οποίο μπορεί να εκφράσει τις δομές των μοντέλων των τοπικών σχημάτων των ΒΔ. Στην εργασία του [40] αναπτύσσεται ένα πλαίσιο χαρακτηριστικών, βάσει των οποίων ελέγχεται η καταλληλότητα ενός μοντέλου δεδομένων για το ρόλο του κανονικού μοντέλου δεδομένων ενός ΟΣΒΔ. Το πλαίσιο αυτό βασίζεται στους δύο παράγοντες της ικανότητας παράστασης ενός μοντέλου δεδομένων, την *εκφραστικότητα* και το *σημασιολογικό συσχετισμό.*

Με τον όρο εκφραστικότητα ενός μοντέλου δεδομένων, εννοούν το βαθμό κατά τον οποίο ένα μοντέλο μπορεί άμεσα να παραστήσει μια εννοιολογία όσο πολύπλοκη κι αν είναι και όσες έννοιες κι αν περιέχει. Η εκφραστικότητα διακρίνεται σε *δομική εκφραστικότητα* και *εκφραστικότητα συμπεριφοράς.* Η δομική εκφραστικότητα μετρά τη δύναμη των δομών του μοντέλου να παριστάνουν έννοιες, ενώ η εκφραστικότητα συμπεριφοράς τη δύναμη του μοντέλου να παριστάνει συμπεριφορές εννοιών. Ειδικότερα, τα χαρακτηριστικά της δομικής εκφραστικότητας που θεωρούν ότι πρέπει να πληρούνται από το ΚΜΔ, είναι οι μηχανισμοί *ταξινόμησης* (classification/instantiation), *γενίκευσης/εξιδείκευσης* (generalization/specialization) και *σύνθεσης/αποσύνθεσης* (aggregation/decomposition). Ως προς τα χαρακτηριστικά της εκφραστικότητας συμπεριφοράς, θεωρούν ότι το ΚΜΔ πρέπει να υποστηρίζει τον ορισμό νέων λειτουργιών και νέων περιορισμών ακεραιότητας.

Με τον όρο τώρα σημασιολογικός συσχετισμός ενός μοντέλου δεδομένων, εννοούν το βαθμό δυνατότητας που παρέχουν οι λειτουργίες αυτού του μοντέλου στον καθορισμό ομοσπονδιακών και εξωτερικών σχημάτων (όψεων). Έτσι για παράδειγμα, το σχεσιακό μοντέλο δεδομένων έχει υψηλό βαθμό σημασιολογικού συσχετισμού, μια και όλη η δύναμη της σχεσιακής άλγεβρας μπορεί να χρησιμοποιηθεί για τον καθορισμό όψεων. Ειδικότερα, τα χαρακτηριστικά του σημασιολογικού συσχετισμού που θεωρούν απαραίτητα για το ΚΜΔ είναι η υποστήριξη τελεστών ενοποίησης σχημάτων, η δυνατότητα των λειτουργιών του μοντέλου να παράγουν εξωτερικά σχήματα, η υποστήριξη από το μοντέλο μιας μόνο

βασικής δομής ώστε να αποφεύγονται κατά το δυνατό οι περιπτώσεις σημασιολογικών συγκρούσεων, και τέλος η υποστήριξη από το μοντέλο πολλαπλών σημασιών.

Ας περάσουμε όμως τώρα να δούμε, προσπάθειες που έχουν γίνει για τη μετάφραση σχεσιακών σχημάτων προς σχήματα σημασιολογικά πλουσιοτέρων μοντέλων δεδομένων, όπως το μοντέλο οντοτήτων συσχετίσεων και το οντοκεντρικό.

Στην κατηγορία αυτών των αλγορίθμων μετάφρασης σχημάτων ανήκει και η δουλειά του [34], στην οποία αναπτύσσεται μια διαδικασία μετάφρασης ενός σχεσιακού σχήματος που πληρεί κάποιες προϋποθέσεις σε ένα διάγραμμα του Επεκτεταμμένου Μοντέλου Οντοτήτων Συσχετίσεων (EER). Οι προϋποθέσεις αυτές συνίστανται στο ότι το σχεσιακό σχήμα πρέπει να βρίσκεται σε BCNF κανονική μορφή, και στο ότι οι εξαρτήσεις εγκλεισμού που ισχύουν σ' αυτό πρέπει να είναι βασισμένες σε κλειδί. Ένα σχεσιακό σχήμα που πληρεί αυτές τις προϋποθέσεις το ονομάζουν *EER μετατρεπόμενο.* Όσο για το Επεκτεταμένο Μοντέλο Οντοτήτων Συσχετίσεων, είναι το ίδιο με αυτό που έχει χρησιμοποιηθεί στο [33], και επιπλέον του γνωστού μοντέλου Οντοτήτων Συσχετίσεων δίνει τη δυνατότητα για γενικεύσεις/εξειδικεύσεις των οντοτήτων, καθώς και για συσχετίσεις που μέλη τους μπορούν να είναι άλλες συσχετίσεις.

Η διαδικασία που προτείνουν κατ' αρχήν μετατρέπει το σχεσιακό σχήμα σε μια μορφή κατάλληλη για αναγνώριση σ' αυτό, δομών αντικειμένων του EER μοντέλου (οντότητες, συσχετίσεις, ασθενείς οντότητες, κ.τ.λ.). Έτσι, αναπτύσσουν διαδικασίες οι οποίες αφαιρούν πιθανούς κύκλους στο σύνολο των εξαρτήσεων εγκλεισμού του σχεσιακού σχήματος (Fold), και διασπούν σχήματα σχέσεων που θεωρούν ότι είναι συγχώνευση άλλων σχημάτων σχέσεων (Split). Το σχεσιακό σχήμα που παράγεται από την εφαρμογή των παραπάνω διαδικασιών, αποδεικνείουν ότι είναι ισοδύναμο ως προς την χωρητικότητα πληροφορίας με το αρχικό, κάτι που σημαίνει ότι το αρχικό σχήμα μπορεί να εκφράσει ακριβώς την ίδια πληροφορία που μπορεί να εκφράσει και το τελικό. Έχοντας τώρα ένα σχεσιακό σχήμα στο οποίο μπορούν να αναγνωρίσουν EER δομές αντικειμένων, προχωρούν στη αντιστοίχιση κάθε σχήματος σχέσης προς ένα σύνολο αντικειμένων. Ο τύπος του συνόλου αντικειμένων (π.χ. ασθενής οντότητα, οντότητα εξειδίκευσης, κ.τ.λ.) καθώς και ο τύπος των συνδέσεων μεταξύ συνόλων αντικειμένων (π.χ. ρόλος, σύνδεση γενίκευσης), καθορίζονται από τις εξαρτήσεις εγκλεισμού και τη δομή των κλειδιών των σχημάτων σχέσεων. Έτσι, αναπτύσσεται η διαδικασία $Rmap^R$ η οποία χρησιμοποιεί τις δύο προηγούμενες διαδικασίες, και είτε καταλήγει στην παραγωγή ενός καλά ορισμένου EER σχήματος, είτε αποτυγχάνει. Λόγοι για την αποτυχία μπορεί να είναι σχεδιαστικά λάθη, όπως η παράλειψη εξαρτήσεων εγκλεισμού και η απρόσεχτη ονοματολογία των

γνωρισμάτων των σχημάτων σχέσεων του σχεσιακού σχήματος. Αξίζει τέλος να σημειωθεί ότι αν ένα EER σχήμα μεταφραστεί σε σχεσιακό με βάση τον αλγόριθμο του [33], και κατόπιν το σχεσιακό αυτό σχήμα περάσει από την $Rmap^R$ τότε παίρνουμε το αρχικό EER σχήμα. Η εφαρμογή των διαδικασιών με την αντίστροφη σειρά σε ένα σχεσιακό σχήμα, καταλήγει όχι στο ίδιο αλλά σε ένα ισοδύναμο ως προς τη χωρητικότητα πληροφορίας σχεσιακό σχήμα.

Μια άλλη διαδικασία αυτόματης μετάφρασης ενός σχεσιακού σχήματος προς ένα σχήμα του συναρτησιακού μοντέλου δεδομένων, αναπτύσσεται στην εργασία του [3]. Στην εργασία αυτή, η μόνη απαίτηση που θέτουν για το σχεσιακό σχήμα είναι οι εξαρτήσεις εγκλεισμού που θα περιέχει, να είναι εξαρτήσεις κλειδιού ή εξαρτήσεις βασισμένες σε κλειδί. Όσο για το συναρτησιακό μοντέλο δεδομένων που χρησιμοποιούν για να παραστήσουν το μεταφρασθέν σχήμα, αξίζει να σημειωθεί ότι προσομοιάζει αρκετά στο οντοκεντρικό μοντέλο, χρησιμοποιώντας τύπους και συναρτήσεις εκεί που το οντοκεντρικό χρησιμοποιεί κλάσεις και γνωρίσματα (ή μεθόδους). Ο αλγόριθμος μετάφρασης, όπως και όλοι οι άλλοι, βασίζεται στο γεγονός ότι όπως στο οντοκεντρικό παράδειγμα υπάρχουν σχέσεις γενίκευσης μεταξύ κλάσεων και αναφορές (γνωρίσματα) από μια κλάση σε μια άλλη, αντίστοιχα στο σχεσιακό υπάρχουν εξαρτήσεις εγκλεισμού μεταξύ των κυρίων κλειδιών σχημάτων σχέσεων, και εξαρτήσεις εγκλεισμού ξένων κλειδιών.

Σύμφωνα με τον αλγόριθμο, αρχικά σχηματίζουν το γράφο εγκλεισμού από το σύνολο των εξαρτήσεων εγκλεισμού που διαθέτουν και από αυτές που μπορούν να επάγουν. Στο γράφο αυτό οι κορυφές αντιστοιχούν στα σχήματα σχέσεων του σχεσιακού σχήματος, και οι ακμές σε εξαρτήσεις εγκλεισμού. Κατόπιν, βρίσκοντας τις συνεκτικές συνιστώσες αυτού του γράφου κατασκευάζουν το συμπυκνωμένο γράφο, και από αυτόν με μεταβατική μείωση (transitive reduction) παίρνουν το μειωμένο γράφο. Αξίζει εδώ να σημειωθεί ότι όπως αναφέρουν και οι συγγραφείς, με το βήμα της εύρεσης των συνεκτικών συνιστωσών ανακαλύπτουν και εξαλείφουν περιπτώσεις κάθετης κατάτμησης, όπου πληροφορία για τα ίδια αντικείμενα είναι διάσπαρτη σε πλέον της μιας σχέσης, ενώ με το βήμα της μεταβατικής μείωσης ανακαλύπτουν και εξαλείφουν περιπτώσεις πλεονασμών στη δήλωση σχέσεων γενίκευσης, κατά τις οποίες υπάρχουν άμεσα δηλωμένες τέτοιες σχέσεις ενώ έμμεσα μπορούν να εξαχθούν από άλλες. Τέλος, από το μειωμένο αυτό γράφο και βάσει συνθηκών που στηρίζονται στη σημασιολογία των εξαρτήσεων εγκλεισμού του, παράγουν το τελικό συναρτησιακό σχήμα. Παράλληλα με την κατασκευή του σχήματος κρατούν και την πληροφορία της αντιστοίχισης, η οποία θα χρησιμεύσει για τη μετάφραση

επερωτήσεων πάνω από το συναρτησιακό σχήμα, σε επερωτήσεις πάνω από το σχεσιακό σχήμα.

Ο αλγόριθμος αυτός χρησιμοποιείται από το Pegasus, ένα ετερογενές σύστημα πολυβάσεων δεδομένων, το οποίο χρησιμοποιεί το συναρτησιακό μοντέλο σαν ΚΜΔ. Μετά τη μετάφρασή του λοιπόν, ένα σχεσιακό σχήμα μπορεί να ενοποιηθεί με το ήδη υπάρχον ομοσπονδιακό σχήμα του Pegasus.

Η τελευταία εργασία που θα δούμε είναι αυτή του [13], όπου αναπτύσσεται μια μεθοδολογία μετάφρασης ενός σχεσιακού σχήματος προς ένα σχήμα που ακολουθεί ένα επεκτεταμένο οντοκεντρικό μοντέλο δεδομένων. Το μοντέλο αυτό, που το ονομάζουν BLOOM, διαθέτει τους μηχανισμούς ταξινόμησης, τέσσερις μορφές μηχανισμού γενίκευσης/εξειδίκευσης (ξένη, συμπληρωματική, εναλλακτική, γενική), τρεις μορφές μηχανισμού σύνθεσης (απλή, συλλογή, σύνθεση συνδέσμου) και δύο μορφές εξαρτήσεων (ενδιαφέροντος, ύπαρξης). Για το σχεσιακό προς μετάφραση σχήμα, δεν απαιτούν να έχει κάποια συγκεκριμένη μορφή ή να είναι καλά σχεδιασμένο. Γιαυτό το λόγο και η διαδικασία της μετάφρασης δεν είναι αυτόματη, όπως στις προηγούμενες περιπτώσεις, αλλά απαιτεί την παρέμβαση του χρήστη σε περιπτώσεις ασαφειών.

Πριν από τη διαδικασία της μετάφρασης, θεωρούν ότι η σχεσιακή ΒΔ περνά από μια φάση *απόκτησης γνώσης*, κατά την οποία εξάγονται από το σχήμα και τα δεδομένα της ΒΔ, στοιχεία που θα βοηθήσουν στη φάση της μετάφρασης. Συγκεκριμένα, το πρώτο βήμα στη φάση αυτή περιλαμβάνει την ανακάλυψη όλων των κλειδιών των σχημάτων σχέσεων, τόσο δηλαδή αυτών που έχουν δηλωθεί στο σχήμα όσο και αυτών που αν και δεν έχουν δηλωθεί πληρούν την υπόθεση μοναδικότητας. Στο δεύτερο βήμα ανακαλύπτονται όλες οι συναρτησιακές εξαρτήσεις που ισχύουν στα σχήματα σχέσεων. Για το λόγο αυτό έχουν αναπτύξει αλγορίθμους ([12]), οι οποίοι χρησιμοποιώντας τα δεδομένα των σχέσεων μιας ΒΔ εξάγουν συναρτησιακές εξαρτήσεις, κάνοντας τις ελάχιστες προσβάσεις στους δίσκους (όπου βρίσκονται τα δεδομένα). Στο τρίτο βήμα κανονικοποιούν το σχεσιακό σχήμα σε μια μορφή όπου σχήματα σχέσεων εκφράζουν μια μοναδική έννοια. Για το σκοπό αυτό γνωστοί αλγόριθμοι για τη μετατροπή σε τρίτη κανονική μορφή ([25]) μπορούν να χρησιμοποιηθούν. Στο τέταρτο βήμα καθορίζεται ο τύπος των αναγνωριστικών κάθε σχήματος σχέσης. Χαρακτηρίζουν ως *κανονονικό αναγνωριστικό* μιας σχέσης εκείνο το αναγνωριστικό που αντιπροσωπεύει την έννοια της σχέσης, και ως *εξωτερικό αναγνωριστικό* κάθε άλλο αναγνωριστικό της. Στο πέμπτο βήμα με τη χρήση αλγορίθμων (που επίσης περιγράφονται στο [12]) εξάγουν όλες τις εξαρτήσεις εγκλεισμού από τα δεδομένα των σχέσεων του σχήματος, ενώ στο έκτο και τελευταίο βήμα πιστοποιούν

εξαρτήσεις αποκλεισμού (exclusion) και συμπληρωματικότητας (complementariness).

Μετά τη φάση απόκτησης γνώσης, περνούν στη φάση μετατροπής του σχήματος. Η φάση αυτή αποτελείται από τρία στάδια. Στο πρώτο στάδιο, από την ανάλυση των εξαρτήσεων εγκλεισμού ανακαλύπτονται λανθάνουσες οντότητες, και καθορίζεται το είδος των σημασιολογικών αφαιρέσεων για τα σχήματα σχέσεων και τα γνωρίσματά τους που συμμετέχουν στις εξαρτήσεις εγκλεισμού. Στο δεύτερο στάδιο, εξετάζονται τα υπόλοιπα γνωρίσματα των σχημάτων σχέσεων, ενώ στο τρίτο και τελευταίο στάδιο δημιουργείται το οντοκεντρικό σχήμα βάσει της ανάλυσης των δύο προηγουμένων σταδίων.

3.3 Μέθοδοι Ενοποίησης Σχημάτων

Η ενοποίηση σχημάτων είναι η διαδικασία εκείνη κατά την οποία, σχήματα βάσεων δεδομένων συγχωνεύονται σε ένα ενοποιημένο σχήμα. Τα σχήματα αυτά μπορούν να ειδωθούν, είτε σα διαφορετικές όψεις ενός ”προς κατασκευήν” σχήματος, οπότε η διαδικασία ονομάζεται *ενοποίηση όψεων*, είτε σα σχήματα ήδη υπαρχουσών βάσεων δεδομένων, οπότε η διαδικασία ονομάζεται *ενοποίηση βάσεων δεδομένων*. Η ενοποίηση όψεων είναι μια διαδικασία που λαμβάνει χώρα κατά τη διάρκεια σχεδίασης και υλοποίησης ενός κεντρικοποιημένου ΣΒΔ, όπου διαφορετικές ομάδες χρηστών και πληροφοριακές ανάγκες ενός οργανισμού παράγουν διαφορετικές όψεις παράστασης του περιβάλλοντος, οι οποίες πρέπει να ενοποιηθούν για την παραγωγή του σχήματος που θα περιγράφει τον πληροφοριακό κόσμο της βάσης δεδομένων του οργανισμού. Η ενοποίηση ΒΔ είναι μια διαδικασία απαραίτητη για στενά συνδεδεμένα ομοσπονδιακά ΣΒΔ, όπως είδαμε και στο κεφάλαιο 2, στα οποία αυτόνομες ΒΔ παρέχουν τμήμα του σχήματός τους (εξαγόμενα σχήματα) για την κατασκευή ενός καθολικού σχήματος.

Μια αντιπροσωπευτική εργασία για τις μεθοδολογίες ενοποίησης σχημάτων είναι αυτή του [4]. Στόχος της εργασίας αυτής, είναι να παράσχει ένα πλαίσιο για το πρόβλημα της ενοποίησης σχημάτων και βάσει ορισμένων κριτηρίων που απορρέουν από το πλαίσιο αυτό, να συγκρίνει εργασίες που έχουν αντιμετωπίσει αυτό το πρόβλημα. Για το λόγο αυτό έχουν επιλεγεί δώδεκα τέτοιες εργασίες, οι οποίες διακρίνονται σ’ αυτές που ασχολούνται μόνο με το πρόβλημα ενοποίησης όψεων, σ’ αυτές που ασχολούνται μόνο με το πρόβλημα ενοποίησης ΒΔ, και σ’ αυτές που αντιμετωπίζουν και τα δύο προβλήματα.

Για τον καθορισμό του πλαισίου, αρχικά εξετάζουν τα αίτια που προκαλούν διαφοροποιήσεις στα ”προς ενοποίηση” σχήματα και κατόπιν αναγνωρίζουν τις ενέργειες

που συνήθως εκτελούν οι διάφορες διαδικασίες ενοποίησης σχημάτων. Αυτές οι ενέργειες θα χρησιμοποιηθούν σα βάση για τη σύγκριση των δώδεκα εργασιών. Τέλος, εξετάζουν την επιρροή των εννοιολογικών μοντέλων στη συνολική διαδικασία ενοποίησης. Σε ότι αφορά λοιπόν τα αίτια που προκαλούν διαφοροποιήσεις (διαφορετικές παραστάσεις ίδιων εννοιών) στα "προς ενοποίηση" σχήματα, αρχικά διακρίνουν τις *διαφορετικές οπτικές* που μπορούν να έχουν διαφορετικά είδη χρηστών ή σχεδιαστών για μια έννοια του πληροφοριακού κόσμου. Μια άλλη αιτία για διαφοροποίηση είναι η *ισοδυναμία μεταξύ δομών του μοντέλου δεδομένων* που ακολουθούν τα "προς ενοποίηση" σχήματα, η οποία επιτρέπει σε δυο σχήματα την ίδια έννοια να την παριστάνουν με διαφορετικές δομές. Η τελευταία αιτία, κατά τους συγγραφείς, είναι οι *ασύμβατες προδιαγραφές σχεδιασμού* (λανθασμένες επιλογές σε ότι αφορά ονόματα, τύπους, περιορισμούς ακεραιότητας, κ.λ.π.) που μπορεί να υπάρχουν μεταξύ εννοιών των σχημάτων. Τα βασικά βήματα από τα οποία θεωρούν ότι αποτελείται η διαδικασία της ενοποίησης είναι :

- *Προενοποίηση*

  Στο βήμα αυτό γίνεται μια ανάλυση των σχημάτων ώστε να αποφασιστεί μια πολιτική ενοποίησης τους. Η ανάλυση αυτή περιλαμβάνει την επιλογή των σχημάτων, τον καθορισμό της σειράς με την οποία θα ενοποιηθούν, καθώς επίσης και θέματα γενικότερης στρατηγικής ενοποίησης, όπως το μέγεθος της αλληλεπίδρασης του χρήστη στη διαδικασία ενοποίησης, ή ακόμα και το πλήθος των σχημάτων που θα ενοποιούνται τη φορά.

- *Σύγκριση Σχημάτων*

  Στο βήμα αυτό τα σχήματα αναλύονται και συγκρίνονται, ώστε να καθορισθούν οι αντιστοιχίες μεταξύ εννοιών και να διαπιστωθούν πιθανές συγκρούσεις.

- *Προσαρμογή Σχημάτων*

  Στο βήμα αυτό και ενώ έχουν διαπιστωθεί πιθανές συγκρούσεις (διαφορετικές παραστάσεις ίδιας έννοιας) μεταξύ σχημάτων, γίνεται προσπάθεια να επιλυθούν έτσι ώστε να είναι εφικτή η συγχώνευσή τους. Λόγο του ότι η αυτόματη επίλυση συγκρούσεων γενικά δεν είναι εφικτή, είναι απαραίτητη η στενή επαφή χρήστη και σχεδιαστή κατά τη διάρκεια της φάσης αυτής.

- *Συγχώνευση και Ανακατασκευή*

  Στο βήμα αυτό είμαστε έτοιμοι να συγχωνεύσουμε τα σχήματα παίρνοντας ενδιάμεσα ενοποιημένα σχήματα. Τα ενδιάμεσα αυτά σχήματα αναλύονται με

βάση κάποια κριτήρια ποιότητας και αν χρειαστεί ανακατασκευάζονται ώστε να πετύχουν καλύτερες επιδόσεις. Τα κριτήρια αυτά είναι:

-- *Πληρότητα και Ορθότητα.* Το ενοποιημένο σχήμα πρέπει να περιέχει όλες τις έννοιες που περιέχονται στα επιμέρους σχήματα. Το ενοποιημένο σχήμα πρέπει να είναι μια παράσταση της ένωσης των πεδίων εφαρμογής που σχετίζονται με τα επιμέρους σχήματα.

-- *Ελαχιστότητα.* Αν η ίδια έννοια παριστάνεται σε άνω του ενός επιμέρους σχημάτων, τότε πρέπει να παριστάνεται μόνο μια φορά στο ενοποιημένο σχήμα.

-- *Δυνατότητα Κατανόησης.* Το ενοποιημένο σχήμα θα πρέπει να είναι εύκολο στην κατανόηση τόσο από το σχεδιαστή όσο και από τον τελικό χρήστη.

Τέλος, εξετάζοντας την επιρροή των μοντέλων δεδομένων στη διαδικασία ενοποίησης παρατηρούν ότι ένα απλούστερο μοντέλο, ένα μοντέλο δεδομένων δηλαδή με λίγες δομές οργάνωσης δεδομένων, έχει πλεονέκτημα σε ότι αφορά τις ενέργειες της προσαρμογής και της συγχώνευσης, μια και η πιθανότητα ύπαρξης συγκρούσεων είναι μικρότερη, οι λειτουργίες μετατροπής είναι ευκολότερες και η συγχώνευση απαιτεί λιγότερες πρωτογενείς λειτουργίες. Απ' την άλλη μεριά, ένα απλούστερο μοντέλο δεδομένων παρέχει ένα πιο αδύνατο εργαλείο για την ανακάλυψη ομοιοτήτων.

Μετά τον καθορισμό του πλαισίου, περνούν στην ανάλυση και σύγκριση των δώδεκα μεθοδολογιών η οποία γίνεται σε τρία επίπεδα. Στο πρώτο επίπεδο, οι διαδικασίες αντιμετωπίζονται σα "μαύρα κουτιά" και εξετάζονται οι είσοδοι που δέχονται και τα αποτελέσματα που παράγουν. Στο δεύτερο επίπεδο, εξετάζεται η αρχιτεκτονική των μεθοδολογιών, όπου ειδικότερα αναλύουν ποια από τα βήματα που προαναφέραμε υπάρχουν στις διαδικασίες και τί είδη ανακυκλώσεων υπάρχουν μεταξύ τους. Από την εξέταση αυτή όπως αναφέρουν, οι μεθοδολογίες μπορούν να ταξινομηθούν σε τέσσερις κατηγορίες. Υπάρχουν οι μεθοδολογίες που εκτελούν επαναληπτικά σύγκριση, προσαρμογή και συγχώνευση σχημάτων ώστε να μη χρειάζονται ανακατασκευή αργότερα. Υπάρχουν οι μεθοδολογίες που παραλείπουν τα τρία πρώτα βήματα και εκτελούν μόνο συγχώνευση και ανακατασκευή είτε επαναληπτικά, είτε για μια φορά. Υπάρχουν οι μεθοδολογίες που παραλείπουν το πρώτο βήμα και εκτελούν τις τέσσερις υπόλοιπες ενέργειες, ενώ τέλος υπάρχουν και εκείνες που αναφέρουν ρητά τη φάση της προενοποίησης. Στο τρίτο τώρα επίπεδο, περνούν στην εξέταση των μεθοδολογιών με βάση τη συμπεριφορά τους σε καθένα απ' τα βήματα της διαδικασίας ενοποίησης.

Στο βήμα της προενοποίησης, με δεδομένο ότι υπάρχουν *n* "προς ενοποίση" (συστατικά)

σχήματα, συγκεντρώνουν την προσοχή τους στη *στρατηγική της επεξεργασίας ενοποίησης* και ταξινομούν τις μεθολογίες ανάλογα μ' αυτή. Διακρίνουν για το λόγο αυτό δύο κύριες κατηγορίες στρατηγικών: τη *δυαδική (binary)* και τη *πλειαδική (n-ary)*. Οι δυαδικές στρατηγικές επιτρέπουν την ενοποίηση μόνο δύο σχημάτων κάθε φορά, αντίθετα με τις πλειαδικές που επιτρέπουν ενοποίηση οσοδήποτε σχημάτων. Στην κατηγορία των δυαδικών στρατηγικών ανήκουν οι *κλιμακωτές στρατηγικές* στις οποίες σε κάθε βήμα ενοποιείται ένα συστατικό σχήμα με ένα ενδιάμεσο ενοποιημένο σχήμα και οι *ισορροπημένες στρατηγικές* στις οποίες τα συστατικά σχήματα αρχικά διαιρούνται σε ζεύγη και ενοποιούνται με συμμετρικό τρόπο. Στην κατηγορία τώρα των πλειαδικών στρατηγικών διακρίνουν τις *στρατηγικές ενός βήματος (one shot strategies)* όπου όλα τα συστατικά σχήματα ενοποιούνται σε ένα βήμα και τις *επαναληπτικές στρατηγικές*.

Στο βήμα της σύγκρισης των συστατικών σχημάτων, εντοπίζουν και αναλύουν τα είδη των συγκρούσεων που αντιμετωπίζονται από τις μεθοδολογίες. Δύο είναι οι κύριες κατηγορίες που εντοπίζουν: *συγκρούσεις ονομάτων* και *δομικές συγκρούσεις*. Στην πρώτη κατηγορία ανήκουν οι περιπτώσεις των *ομωνύμων*, όπου σε δυο σχήματα χρησιμοποιείται το ίδιο όνομα για να δηλώσει διαφορετικές έννοιες, και οι περιπτώσεις των *συνωνύμων*, όπου σε δύο σχήματα χρησιμοποιούνται διαφορετικά ονόματα για να δηλώσουν την ίδια έννοια. Στη δεύτερη κατηγορία, ανήκουν περιπτώσεις συγκρούσεων που οφείλονται στη χρησιμοποίηση διαφορετικών δομών και περιορισμών ακεραιότητας του μοντέλου δεδομένων που ακολουθούν τα συστατικά σχήματα. Υπάρχουν λοιπόν *συγκρούσεις τύπων* που προκύπτουν όταν η ίδια έννοια παριστάνεται από διαφορετικές δομές του μοντέλου δεδομένων στα συστατικά σχήματα, *συγκρούσεις εξαρτήσεων* που προκύπτουν όταν μια ομάδα εννοιών συνδέονται μεταξύ τους με διαφορετικές εξαρτήσεις στα συστατικά σχήματα, *συγκρούσεις κλειδιών* που προκύπτουν όταν η ίδια έννοια έχει διαφορετικά κλειδιά στα συστατικά σχήματα, και τέλος *συγκρούσεις συμπεριφοράς* που προκύπτουν όταν διαφορετικές πολιτικές εισαγωγής/διαγραφής συσχετίζονται με την ίδια ομάδα αντικειμένων στα συστατικά σχήματα.

Στο βήμα της προσαρμογής των συστατικών σχημάτων, εντοπίζουν τα είδη των μετατροπών που υφίστανται τα συστατικά σχήματα ώστε να επιλυθούν τα διάφορα είδη συγκρούσεων που έχουν μεταξύ τους. Τέτοιες μετατροπές είναι κυρίως μετονομασίες (για την επίλυση συγκρούσεων ονομάτων), και μετατροπές τύπων (για επίλυση δομικών συγκρούσεων). Το ενδιαφέρον που παρατηρούν, είναι ότι από τις μεθοδολογίες που εξέτασαν, καμμιά δεν παρείχε κάποια ανάλυση ή απόδειξη πληρότητας, που να δείχνει ότι οι λειτουργίες μετατροπής σχήματος που προτείνονται στις μεθοδολογίες είναι ικανές

να επιλύσουν οποιοδήποτε τύπο σύγκρουσης.

Τέλος στα βήματα της συγχώνευσης και της ανακατασκευής του ενοποιημένου σχήματος, παρατηρούν ότι όλες οι μεθοδολογίες πρώτα συγχωνεύουν τα συστατικά σχήματα μέσω μιας απλής υπέρθεσης των κοινών εννοιών, και κατόπιν εκτελούν λειτουργίες ανακατασκευής στο σχήμα που έχει ήδη προκύψει από τη συγχώνευση. Κάθε τέτοια λειτουργία εκτελείται για να βελτιώσει την ποιότητα του σχήματος ως προς τα κριτήρια της πληρότητας, της ελαχιστότητας και της δυνατότητας κατανόησης.

Τα κυριότερα συμπεράσματα στα οποία καταλήγουν οι συγγραφείς, είναι ότι οι περισσότερες από τις μεθοδολογίες που εξέτασαν δεν παρείχαν κάποιες αλγοριθμικές προδιαγραφές για τις ενέργειες ενοποίησης αλλά απλά και μόνο γενικές κατευθυντήριες γραμμές, και ότι επίσης δεν υπήρχαν κάποιες ενδείξεις για το αν το σύνολο των συγκρούσεων που εξέταζαν και των μετατροπών που παρείχαν ήταν κατά κάποιο τρόπο ”πλήρες”.

Ας περάσουμε όμως τώρα να δούμε από πιο κοντά ορισμένες μεθόδους ενοποίησης βάσεων δεδομένων, που προτείνονται στη βιβλιογραφία.

Η πρώτη εργασία που θα εξετάσουμε είναι αυτή του [17]. Στην εργασία αυτή το πρόβλημα που αντιμετωπίζουν, είναι ο ορισμός ενός καθολικού σχήματος από άλλα συστατικά σχήματα εκπεφρασμένα στο συναρτησιακό μοντέλο δεδομένων έτσι ώστε να επιλύσουν τις διαφορές μεταξύ τους, και η μετατροπή μιας επερώτησης προς το καθολικό σχήμα, σε επερωτήσεις προς τα συστατικά σχήματα. Ειδικότερα, δείχνουν ότι η έννοια της γενίκευσης όταν συνδυαστεί με επεκτεταμμένες διευκολύνσεις για ορισμό όψεων, είναι ένα πολύ δυνατό εργαλείο για την ενοποίηση βάσεων δεδομένων.

Αρχικά λοιπόν, επεκτείνουν το συναρτησιακό μοντέλο που χρησιμοποιούν με την έννοια της γενίκευσης μεταξύ οντοτήτων και συναρτήσεων. Με την επέκταση αυτή, τους παρέχεται η δυνατότητα στο καθολικό σχήμα να εκφράσουν σχέσεις γενίκευσης μεταξύ οντοτήτων και συναρτήσεων των συστατικών σχημάτων. Κατόπιν, θεωρώντας ότι το καθολικό σχήμα αποτελείται από όψεις των συστατικών σχημάτων, προχωρούν στην περιγραφή δομών που θα παρέχουν το καθορισμό αυτών των όψεων από τα συστατικά σχήματα. Οι δομές αυτές δίνονται στον σχεδιαστή του καθολικού σχήματος υπό τη μορφή μιας απλής μη διαδικαστικής γλώσσας επερωτήσεων, η οποία του παρέχει τις εξής δυνατότητες:

- *Εγκλεισμού*

  Ο σχεδιαστής του καθολικού σχήματος μπορεί να ορίσει ότι αυτούσιες οντότητες από τα συστατικά σχήματα, εγκλείονται σ’ αυτό.

- *Ορισμού εικονικών οντοτήτων και συναρτήσεων*
  Ο σχεδιαστής του καθολικού σχήματος μπορεί να ορίσει ότι οντότητες ή συναρτήσεις (εικονικές) του καθολικού σχήματος, προκύπτουν σαν αποτέλεσμα επερωτήσεων πάνω σ' ένα συστατικό σχήμα.

- *Καθορισμού υπερτύπων/υποτύπων και υπερσυναρτήσεων/υποσυναρτήσεων*
  Ο σχεδιαστής του καθολικού σχήματος μπορεί να ορίσει νέες οντότητες (τύπους) ή συναρτήσεις στο καθολικό σχήμα, που αποτελούν γενικεύσεις/εξειδικεύσεις υπαρχόντων τύπων ή συναρτήσεων των συστατικών σχημάτων.

Με τις δομές αυτές ο σχεδιαστής προχωρά στη επίλυση συγκρούσεων μεταξύ των συστατικών σχημάτων, και τη δημιουργία του καθολικού σχήματος. Οι συγκρούσεις που θεωρούν ότι μπορούν να αντιμετωπισθούν με τις δομές αυτές, είναι συγκρούσεις ονομάτων, συγκρούσεις κλίμακας, δομικές συγκρούσεις, και συγκρούσεις στο επίπεδο αφαίρεσης. Με δεδομένο λοιπόν το καθολικό σχήμα, σε δεύτερη φάση περνούν στη διατύπωση ενός αλγορίθμου για τη μετατροπή των επερωτήσεων που μπορεί κάποιος να διατυπώσει πάνω από το καθολικό σχήμα σε επερωτήσεις προς τα συστατικά σχήματα. Συμπερασματικά, μπορούμε να πούμε ότι αν και η κατασκευή του καθολικού σχήματος με τη χρησιμοποίηση των δομών που προτείνουν είναι αρκετά ικανοποιητική, δεν είμαστε σίγουροι για την αποτελεσματικότητα της μεθόδου όταν τα συστατικά σχήματα είναι μεγάλα, μια και τότε απαιτείται ευρεία επέμβαση του χρήστη για τον καθορισμό των όψεων τους που θα συμμετέχουν στο καθολικό σχήμα.

Η δεύτερη εργασία που θα δούμε είναι αυτή του [35]. Στην εργασία αυτή όπως και στην προηγούμενη, στόχος είναι η κατασκευή ενός καθολικού σχήματος, που ο συγγραφέας το ονομάζει *υπερόψη*, από συστατικά σχήματα εκπεφρασμένα στο συναρτησιακό μοντέλο δεδομένων, και η διαχείριση των επερωτήσεων που θα γίνονται πάνω στην υπερόψη. Για την επίτευξη αυτού του στόχου, κατασκευάστηκαν δύο εργαλεία: ο *κατασκευαστής εικονικής ΒΔ* και ο *επεξεργαστής εικονικών επερωτήσεων*.

Με το πρώτο εργαλείο ο σχεδιαστής του καθολικού σχήματος, κατασκευάζει μια *εικονική ΒΔ*, μια βάση δεδομένων δηλαδή η οποία έχει σχήμα (την υπερόψη) αλλά δεν έχει δεδομένα. Αντίθετα, έχει αντιστοιχίσεις από το σχήμα της προς άλλα σχήματα (τα συστατικά) που περιέχουν τα πραγματικά δεδομένα. Μ' άλλα λόγια, ενώ οι πραγματικές ΒΔ είναι ζεύγη της μορφής <σχήμα, δεδομένα>, οι εικονικές ΒΔ είναι ζεύγη της μορφής <σχήμα, αντιστοιχίσεις>. Η κατασκευή λοιπόν της εικονικής ΒΔ, γίνεται χρησιμοποιώντας τα αρχικά σχήματα και ένα σύνολο προτάσεων εκφρασμένων σε μια γλώσσα ενοποίησης σχημάτων που ορίζουν το τελικό σχήμα και τις αντιστοιχίσεις από αυτό προς τα αρχικά

σχήματα. Η γλώσσα ενοποίησης περιέχει ένα σύνολο τελεστών οι οποίοι ενοποιούν ή τροποποιούν σχήματα βάσεων δεδομένων. Οι τελεστές αυτοί είναι οι εξής:

1. *Meet.* Παράγει μια κοινή υπερκλάση δύο κλάσεων.
2. *Join.* Παράγει μια κοινή υποκλάση δύο κλάσεων.
3. *Fold.* Παράγει μια νέα κλάση απορροφώντας από μια παλιά κλάση μια υποκλάση της.
4. *Rename.* Μετονομάζει μια κλάση.
5. *Combine and Connect.* Η Combine συγχωνεύει δύο κλάσεις που έχουν το ίδιο σύνολο γνωρισμάτων σε μία, ενώ η Connect δοθείσης μιας κλάσης Α και μιας κλάσης Β με σύνολο γνωρισμάτων υποσύνολο αυτών της Α, εξαναγκάζει την Α στο να απορροφήσει τη Β.
6. *Aggregate.* Από ένα σύνολο γνωρισμάτων μιας κλάσης Α, δημιουργεί μια καινούργια κλάση Β προσάπτοντας της τα γνωρίσματα αυτά, ενώ ταυτόχρονα προσθέτει και την Β στα γνωρίσματα της κλάσης Α.
7. *Telescope.* Αν μια κλάση Β είναι γνώρισμα μιας κλάσης Α, αφαιρεί τη Β και προσάπτει τα γνωρίσματά της στην Α. Ουσιαστικά είναι η αντίστροφη της Aggregate.
8. *Add.* Προσθέτει μια κλάση.
9. *Delete.* Αφαιρεί μια κλάση.

Καταλαβαίνουμε λοιπόν ότι με μια ακολουθία προτάσεων που στηρίζονται στους τελεστές αυτούς, παίρνουμε την υπερόψη. Στην υπερόψη με κάθε κλάση σχετίζεται μια έκφραση που δηλώνει από πού έχει προέλθει (σε σχέση με κλάσεις των αρχικών βάσεων δεδομένων). Η έκφραση αυτή δημιουργείται αυξητικά κατά τη διάρκεια της φάσης ενοποίησης, καθώς κάθε τελεστής ενημερώνει τις εκφράσεις των κλάσεων που τροποποιεί. Οι εκφράσεις αυτές των κλάσεων της υπερόψης είναι και οι αντιστοιχίσεις τους με τις κλάσεις των αρχικών σχημάτων.

Με το δεύτερο εργαλείο, εξυπηρετείται ο χρήστης που κάνει επερωτήσεις στην υπερόψη. Για την επεξεργασία αυτών των επερωτήσεων χρησιμοποιεί τις αντιστοιχίσεις που έχουν καθορισθεί κατά τη διαδικασία ενοποίησης, και αποσυνθέτει κάθε επερώτηση σε ένα σύνολο επερωτήσεων προς τα σχήματα των πραγματικών (συστατικών) βάσεων δεδομένων. Οι απαντήσεις από τις επερωτήσεις αυτές συνθέτονται για να σχηματίσουν την απάντηση

προς την αρχική επερώτηση. Συνεπώς, από το αποτέλεσμα φαίνεται σαν η υπερόψη να αντιδρά στις επερωτήσεις προς αυτήν όπως μια πραγματική βάση δεδομένων.

Μια τρίτη μεθοδολογία ενοποίησης σχημάτων συστατικών βάσεων δεδομένων, παρουσιάζεται στο [19]. Το πρόβλημα που αντιμετωπίζουν στην εργασία αυτή, είναι η δημιουργία ενός ενοποιημένου σχήματος από τοπικά σχήματα αυτόνομων βάσεων δεδομένων εκπεφρασμένα σε κάποιο σημασιολογικά φτωχό μοντέλο δεδομένων. Το ενοποιημένο σχήμα εκφράζεται σε ένα αρκετά πλούσιο σημασιολογικά οντοκεντρικό μοντέλο δεδομένων. Η μεθοδολογία που αναπτύσσουν αποτελείται από τρεις φάσεις: τη *φάση σημασιολογικού εμπλουτισμού*, τη *φάση αναγνώρισης* και τη *φάση επίλυσης συγκρούσεων*.

Η πρώτη φάση είναι η δουλειά του [13], και έχει περιγραφεί αναλυτικά στην προηγούμενη ενότητα. Να θυμήσουμε μόνο, ότι μετά το πέρας της φάσης αυτής τα τοπικά σχήματα έχουν μεταφραστεί σε σχήματα ενός σημασιολογικά πλούσιου οντοκεντρικού μοντέλου δεδομένων, που παίζει το ρόλο του κανονικού μοντέλου δεδομένων.

Η δεύτερη φάση σα στόχο έχει να αναγνωρίσει σημασιολογικά συσχετιζόμενα αντικείμενα ([18]). Αυτό επιτυγχάνεται μέσω μιας διεργασίας σύγκρισης όπου ανακαλύπτονται ομοιότητες και διαφορές μεταξύ αντικειμένων. Αρχικά λοιπόν, χρειάζεται μια *στρατηγική* για να καθοδηγεί την διεργασία, να αποφασίζει δηλαδή ποιά αντικείμενα θα συγκρίνονται κάθε στιγμή. Κατόπιν, χρειάζεται μια συνάρτηση ομοιότητας που θα βασίζεται σε ορισμένα *κριτήρια*. Στην προσέγγιση τους η στρατηγική λειτουργεί σε δύο επίπεδα: στο πρώτο επίπεδο η στρατηγική ανακύπτει ζεύγη εξειδικεύσεων (ομάδες κλάσεων) προς σύγκριση, βάσει της διάστασης γενίκευσης, ενώ στο δεύτερο επίπεδο ανακαλύπτει ζεύγη κλάσεων προς σύγκριση, βασισμένη στη διάσταση σύνθεσης (aggregation). Κάτω από μια συγκεκριμένη σειρά και με αντίστοιχες "ποινές" μπορούν να εφαρμοσθούν χαλαρώσεις στις αφαιρέσεις (και των δύο διαστάσεων) των κλάσεων ώστε αυτές εικονικά να προσαρμοστούν και να συγκριθούν. Οι ποινές αυτές παίζουν ρόλο στη τελική τιμή της συνάρτησης ομοιότητας.

Η τρίτη φάση στόχο έχει την κατασκευή του ενοποιημένου σχήματος, έχοντας σα δεδομένες από τη δεύτερη φάση τις σημασιολογικές σχέσεις μεταξύ αντικειμένων. Η κύρια πράξη ενοποίησης είναι η *διακριτική γενίκευση*. Η πράξη αυτή είναι παρεμφερής της γενίκευσης και όταν εφαρμόζεται σε δύο σημασιολογικά σχετιζόμενες κλάσεις παράγει μια *εικονική* υπερκλάση τους η οποία κληρονομεί την περιγραφή των υποκλάσεων της (παίρνει το σύνολο των αφαιρέσεών τους). Λέμε εικονική γιατί οι περιπτώσεις αυτής της υπερκλάσης είναι οι περιπτώσεις των υποκλάσεων της. Κάθε αντικείμενο μιας υποκλάσης,

όταν ιδωθεί σα μέλος της υπερκλάσης έχει ένα *γνώρισμα διακριτότητας* που παίρνει την τιμή της βάσης δεδομένων απ' όπου προέρχεται. Στο ομοσπονδιακό επίπεδο λοιπόν, κάθε αντικείμενο έχει μια ετικέτα με το όνομα της βάσης απ' όπου προέρχεται.

Η τελευταία εργασία για ενοποίηση σχημάτων που θα εξετάσουμε, είναι αυτή του [26]. Στην εργασία αυτή, ο συγγραφέας ασχολείται με την ανάπτυξη δύο αλγορίθμων για ενοποίηση όψεων (σχημάτων ίδιας ΒΔ), και για ενοποίηση βάσεων δεδομένων (σχημάτων διαφορετικών ΒΔ). Τα συστατικά σχήματα θεωρούν ότι είναι εκπεφρασμένα σε ένα μοντέλο δεδομένων που προσομοιάζει αρκετά στο συναρτησιακό και ότι πληρούν ορισμένες *απαιτήσεις νομιμότητας*. Και οι δύο αλγόριθμοι έχουν τη μορφή τελεστών, υπό την έννοια ότι δέχονται σαν είσοδο δύο συστατικά νόμιμα σχήματα και παράγουν ένα ενοποιημένο νόμιμο σχήμα. Αυτό τους καθιστά ικανούς να χρησιμοποιηθούν από δυαδικές μεθόδους ενοποίησης. Οι ιδιότητες που επιθυμεί να έχουν οι τελεστές (αλγόριθμοι) που αναπτύσσει, είναι η αντιμεταθετικότητα και η προσεταιριστικότητα. Κιαυτό γιατί, αν ένας τελεστής έχει αυτές τις ιδιότητες και χρησιμοποιηθεί από μια δυαδική μέθοδο ενοποίησης μιας ομάδας σχημάτων, η μέθοδος αυτή δε θα χρειάζεται να ασχολείται με τη σειρά που θα παρέχει τα σχήματα στον τελεστή, μια και το τελικό ενοποιημένο σχήμα θα είναι πάντα το ίδιο. Ας δούμε όμως κάπως ειδικότερα τη δομή των δύο τελεστών.

Για να κατασκευάσει τον πρώτο τελεστή ο συγγραφέας, αρχικά ορίζει ένα είδος *ασθενών σχημάτων*, σχημάτων δηλαδή που δεν πληρούν ακριβώς τις απαιτήσεις νομιμότητας, και δείχνει ότι κάθε νόμιμο σχήμα είναι και ασθενές. Έπειτα, ορίζει μια σχέση υποσυνόλου μεταξύ δύο ασθενών σχημάτων, σύμφωνα με την πληροφορία που αυτά μεταφέρουν. Κατόπιν, κατασκευάζει ένα τελεστή ενοποίησης δύο συστατικών νόμιμων σχημάτων, ο οποίος παράγει ένα ενοποιημένο ασθενές σχήμα που είναι το ελάχιστο άνω φράγμα (σύμφωνα με τη σχέση υποσυνόλου) των συστατικών σχημάτων. Ο τελεστής αυτός έχει και τις δύο ιδιότητες (αντιμεταθετικότητα, προσεταιριστικότητα) που απαιτεί. Τέλος, βάσει ενός αλγορίθμου μετατροπής, από το τελικό ασθενές ενοποιημένο σχήμα κατασκευάζει το τελικό νόμιμο ενοποιημένο σχήμα.

Ανάλογα δουλεύει και στην περίπτωση του δεύτερου τελεστή. Αρχικά, ορίζει ένα **νέο** είδος *ασθενών σχημάτων*, και δείχνει ότι κάθε νόμιμο σχήμα είναι και ασθενές. Έπειτα, στο σύνολο αυτό ορίζει μια **νέα** σχέση υποσυνόλου μεταξύ δύο ασθενών σχημάτων, σύμφωνα με την πληροφορία που αυτά μεταφέρουν. Κατόπιν, κατασκευάζει ένα τελεστή ενοποίησης δύο συστατικών νόμιμων σχημάτων, ο οποίος παράγει ένα ενοποιημένο ασθενές σχήμα που είναι το μέγιστο κάτω φράγμα (σύμφωνα με τη **νέα** σχέση υποσυνόλου) των συστατικών σχημάτων. Τέλος, βάσει ενός **νέου** αλγορίθμου μετατροπής, από το τελικό ασθενές

ενοποιημένο σχήμα κατασκευάζει το τελικό νόμιμο ενοποιημένο σχήμα.

Κεφάλαιο 4

Μετάφραση Σχεσιακού Σχήματος σε Σημασιολογικό Δίκτυο

Στο κεφάλαιο αυτό, θα παρουσιάσουμε τη μέθοδο που αναπτύξαμε, για τη μετάφραση ενός σχεσιακού σχήματος στο αντίστοιχο σημασιολογικό του δίκτυο. Την ονομάζουμε μέθοδο και όχι αλγόριθμο, μια και είναι αναγκαία, για την επίλυση περιπτώσεων ασάφειας, η επέμβαση του χρήστη.

Αρχικά, θα δούμε τις βασικές έννοιες του σχεσιακού και του σημασιολογικού μοντέλου δεδομένων και κατόπιν, θα περάσουμε στην παρουσίαση της μεθόδου. Εκεί, θα δούμε τις υποθέσεις που κάνουμε για την εφαρμογή της, και έπειτα θα δούμε τις τρεις φάσεις της, κατά τις οποίες αναγνωρίζονται λανθάνουσες οντότητες, φτιάχνεται ο τελικός γράφος εγκλεισμού και μεταφράζεται το τελικό σχεσιακό σχήμα στο σημασιολογικό δίκτυο. Μετά θα ορίσουμε τις σχέσεις αντιστοίχισης που χρησιμοποιούνται για να δηλώσουν τις αντιστοιχίσεις μεταξύ δομών του σχεσιακού σχήματος (σχέσεις, εξαρτήσεις εγκλεισμού, γνωρίσματα) με δομές του σημασιολογικού δικτύου (κλάσεις, σχέσεις εξειδίκευσης, σχέσεις γνωρίσματος). Το κεφάλαιο θα κλείσει με την παρουσίαση ενός ολοκληρωμένου παραδείγματος εφαρμογής της μεθόδου.

4.1 Σχεσιακό και Σημασιολογικό Μοντέλο Δεδομένων

Κάθε σχήμα ΒΔ ακολουθεί ένα μοντέλο δεδομένων. Το *μοντέλο δεδομένων* παρέχει τη δυνατότητα περιγραφής των δεδομένων, των σχέσεων μεταξύ τους, της σημασιολογίας τους και των περιορισμών που ισχύουν πάνω τους. Στην κατηγορία των μοντέλων δεδομένων που χρησιμοποιούνται για την παραγωγή του εννοιολογικού σχήματος ενός ΣΔΒΔ, σύμφωνα με

την αρχιτεκτονική τριών επιπέδων, ανήκουν τόσο το σχεσιακό όσο και το σημασιολογικό μοντέλο. Ειδικότερα ([25]), το σχεσιακό μοντέλο ανήκει στην κατηγορία των *λογικών μοντέλων βασισμένων σε εγγραφές*, ενώ το σημασιολογικό μοντέλο στην κατηγορία των *λογικών μοντέλων βασισμένων σε αντικείμενα*. Η πρώτη κατηγορία περιέχει μοντέλα (Δικτυωτό, Σχεσιακό, Ιεραρχικό) σύμφωνα με τα οποία τα δεδομένα δομούνται βάσει ειδικού τύπου εγγραφών με καθορισμένη μορφή. Αντίθετα, τα μοντέλα της δεύτερης κατηγορίας (Οντοτήτων-Συσχετίσεων, Οντοκεντρικό, Σημασιολογικό, Συναρτησιακό, κ.α.) χαρακτηρίζονται από τις δυνατότητες που παρέχουν, για ευέλικτη δόμηση των δεδομένων.

Το *Σχεσιακό Μοντέλο Δεδομένων*, παριστάνει τα δεδομένα και τις συσχετίσεις μεταξύ τους με τη μορφή πινάκων. Λεπτομέρειες για το μοντέλο αυτό, μπορούμε να βρούμε σ' ένα οποιοδήποτε βιβλίο σχετικό με θέματα βάσεων δεδομένων ([46], [25]). Στην ενότητα αυτή θα προσπαθήσουμε να παρουσιάσουμε τις κυριότερες έννοιές του, κατά το δυνατό συντομότερα. Οι βασικές έννοιες του σχεσιακού μοντέλου, είναι το γνώρισμα και η σχέση. Σα *γνώρισμα* ορίζουμε μία ιδιότητα που μπορούν να έχουν τα δεδομένα μας. Ένα γνώρισμα (π.χ. ηλικία) έχει ένα *πεδίο* (Domain) (π.χ. σύνολο ηλικιών) απ' όπου παίρνει τιμές και ένα *τύπο* (type) (π.χ. Real) που δηλώνει τη μορφή που αποθηκεύεται μια τιμή του. Δύο γνωρίσματα λέγονται *συμβατά*, αν έχουν το ίδιο πεδίο (προφανώς και τον ίδιο τύπο). Ένα *σχήμα-σχέσης R* (relation scheme) είναι ένα ονοματισμένο σύνολο γνωρισμάτων, που δύναται να εκφράζει μια οντότητα ή μια συσχέτιση (π.χ. φύλο, ηλικία, αρ.ταυτότητας = Άνθρωπος). Συμβολίζουμε ένα σχήμα-σχέσης με όνομα R_i, και με σύνολο γνωρισμάτων X_i, σαν $R_i(X_i)$. Μια *πλειάδα* (tuple) *t* στο *R* είναι μια συνάρτηση που αντιστοιχεί σε κάθε γνώρισμα του R μια τιμή απ' το πεδίο του ή την *άγνωστη τιμή (null)*. Μια πλειάδα λέγεται *ολική*, όταν σε κανένα γνώρισμα δεν αντιστοιχεί την άγνωστη τιμή. Μια *σχέση* στο *R* είναι ένα σύνολο πλειάδων στο R. Στην πράξη τώρα, πλειάδα t στο $R_i(X_i)$, έχει επικρατήσει να ονομάζεται ένα διάνυσμα τιμών απ' τα πεδία των γνωρισμάτων του X_i, και να συμβολίζεται $t[X_i]$, ενώ μια σχέση του R_i (που συμβολίζεται r_i), έχει επικρατήσει να ονομάζεται ένας πίνακας τιμών, στον οποίο κάθε γραμμή είναι μια πλειάδα του $R_i(X_i)$.

Έστω τώρα, ένα σχήμα-σχέσης $R_i(X_i)$ με αντίστοιχη σχέση r_i, και W υποσύνολο του X_i. Η *προβολή* της r_i στο W, συμβολίζεται με $\pi_W(r_i)$, και ισούται με $\{t[W] \mid t \in r_i\}$. Η *ολική προβολή* της r_i στο W, συμβολίζεται με $\pi \downarrow_W (r_i)$, και ισούται με $\{t[W] \mid t \in r_i$ και είναι ολικές$\}$.

Ας θεωρήσουμε δύο σχήματα-σχέσεων $R_i(X_i)$ και $R_j(X_j)$, με αντίστοιχες σχέσεις r_i και r_j. Μια *εξάρτηση εγκλεισμού (EE)*, είναι μια πρόταση της μορφής $R_i[Y] \subseteq R_j[Z]$, όπου

τα Υ και Ζ είναι υποσύνολα των X_i και X_j αντίστοιχα, που αποτελούνται από συμβατά ανά δύο γνωρίσματα. Λέμε ότι, η εξάρτηση εγκλεισμού $R_i[Y] \subseteq R_j[Z]$ *ικανοποιείται* από τις σχέσεις r_i και r_j, αν $\pi\ \downarrow_Y\ (r_i) \subseteq \pi\ \downarrow_Z\ (r_j)$. Αν το Ζ είναι κλειδί του σχήματος-σχέσης R_j, τότε η εξάρτηση εγκλεισμού $R_i[Y] \subseteq R_j[Z]$ λέγεται εξάρτηση *βασισμένη σε κλειδί* και το Υ λέγεται *ξένο κλειδί* του σχήματος-σχέσης R_i, ενώ αν και το Υ είναι κλειδί του σχήματος-σχέσης R_i, λέγεται εξάρτηση *κλειδιού*. Το σύνολο των εξαρτήσεων εγκλεισμού I, μπορεί να παρασταθεί γραφικά, σαν ένας κατευθυνόμενος γράφος $G_I(V, H)$, όπου οι κορυφές του V, αντιπροσωπεύουν τα σχήματα-σχέσεων του σχεσιακού σχήματος και $R_i \rightarrow R_j \in H$ ανν $R_i[Y] \subseteq R_j[Z] \in I$. Ο γράφος αυτός ονομάζεται *γράφος εγκλεισμού*.

Ας θεωρήσουμε ξανά ένα σχήμα-σχέσης με αντίστοιχη σχέση r_i. Μια *συναρτησιακή εξάρτηση* στο R_i, είναι μια πρόταση της μορφής $R_i : Y \rightarrow Z$, όπου τα Υ και Ζ είναι υποσύνολα του X_i. Λέμε ότι, η συναρτησιακή εξάρτηση $R_i : Y \rightarrow Z$ *ικανοποιείται* από την r_i, αν για οποιεσδήποτε δύο πλειάδες t και t' της r_i, $t[Y] = t'[Y]$ συνεπάγεται $t[Z] = t'[Z]$. Σαν *υπερκλειδί* ενός σχήματος-σχέσης $R_i(X_i)$, ορίζουμε το υποσύνολο K_i του X_i, για το οποίο ισχύει ότι η $R_i : K_i \rightarrow X_i$, ικανοποιείται στη σχέση r_i. Ένα υπερκλειδί για το οποίο οποιοδήποτε αυστηρό υποσύνολο του **δεν** είναι υπερκλειδί, λέγεται *κλειδί* (ή υποψήφιο κλειδί). Είναι δυνατό σε μια σχέση να υπάρχουν περισσότερα του ενός κλειδιά, εκ των οποίων ένα θεωρείται *κύριο*. Στην εργασία μας ορίζουμε, το κλειδί που αντιπροσωπεύει την έννοια που δηλώνει η σχέση στην οποία βρίσκεται, σαν *αντιπροσωπευτικό κλειδί*, ενώ τα υπόλοιπα σαν *μη-αντιπροσωπευτικά*. Έτσι για παράδειγμα, στην παρακάτω σχέση

τμήμα(ΌνομαΤμήματος, Τμηματάρχης, Έσοδα)

με συναρτησιακές εξαρτήσεις

{ΟνομαΤμήματος} → {ΟνομαΤμήματος, Τμηματάρχης, Έσοδα}

{Τμηματάρχης} → {ΟνομαΤμήματος, Τμηματάρχης, Έσοδα}

το αντιπροσωπευτικό κλειδί είναι το ΌνομαΤμήματος, ενώ το Τμηματάρχης είναι μη-αντιπροσωπευτικό. Εν γένει τα κλειδιά μιας σχέσης μπορούν να είναι *απλά* ή *σύνθετα*, ανάλογα με το αν περιέχουν ένα ή περισσότερα γνωρίσματα.

Το *Σημασιολογικό Μοντέλο Δεδομένων*, ανήκει στην κατηγορία των μοντέλων δεδομένων που μπορούν να χρησιμοποιηθούν από ένα οντοκεντρικό ΣΒΔ, μια και διαθέτει ορισμένα χαρακτηριστικά που υποστηρίζονται από όλα τα μοντέλα δεδομένων που δύνανται να παίξουν αυτό το ρόλο. Τα χαρακτηριστικά αυτά είναι:

- **Οντότητα.** Κάθε αντικείμενο του πραγματικού κόσμου παριστάνεται στη βάση σα μια ανεξάρτητη οντότητα.
- **Αναγνωριστικό Οντότητας.** Κάθε οντότητα έχει ένα μοναδικό αναγνωριστικό που

ορίζεται από το σύστημα.

- **Γνωρίσματα Οντοτήτων.** Μια οντότητα μπορεί να συνδέεται με μια άλλη μέσω μιας *σχέσης*, που αποτελεί ένα *γνώρισμα* αυτής της οντότητας.

- **Κλάσεις Οντοτήτων.** Οντότητες με κοινά χαρακτηριστικά μπορούν να ταξινομηθούν/κατηγοριοποιηθούν σε μια *κλάση οντοτήτων*, η οποία με τη σειρά της αποτελεί μια οντότητα. Κάθε οντότητα ταξινομείται κάτω από τουλάχιστον μία κλάση οντοτήτων.

- **Γενικεύσεις/Εξειδικεύσεις Κλάσεων Οντοτήτων.** Μια οντότητα μπορεί να περιγραφεί ως *εξειδίκευση* μιας άλλης από την οποία κληρονομεί ορισμένα χαρακτηριστικά (η δεύτερη οντότητα αποτελεί *γενίκευση* της πρώτης).

Στην εργασία μας χρησιμοποιήσαμε σα κανονικό μοντέλο δεδομένων το σημασιολογικό μοντέλο, το οποίο ακολουθούν τα σχήματα (σημασιολογικά δίκτυα) που παράγονται από μια υλοποίηση[1] της γλώσσας παράστασης γνώσης Telos ([7]). Θεωρούμε λοιπόν σκόπιμο, να παρουσιάσουμε στην ενότητα αυτή, τις ιδιαιτερότητες του συγκεκριμένου μοντέλου και των σημασιολογικών δικτύων που το ακολουθούν.

Στο μοντέλο αυτό κάθε αντικείμενο του πραγματικού κόσμου αποτελεί μια ξεχωριστή *οντότητα (Object)*, η οποία έχει ένα μοναδικό αναγνωριστικό που καθορίζεται από το σύστημα και δύναται να έχει ένα *λογικό όνομα* που καθορίζεται από το χρήστη. Κάθε οντότητα ταξινομείται σε μία ή περισσότερες *κλάσεις οντοτήτων*. Κάθε οντότητα ενός σημασιολογικού δικτύου της Telos, ταξινομείται στην κλάση του συστήματος *Object*. Τα γνωρίσματα των οντοτήτων θεωρούνται και αυτά οντότητες και έτσι ταξινομούνται και αυτά σε κλάσεις γνωρισμάτων που ονομάζονται *κατηγορίες (categories)*. Για τις τιμές ακεραίων, πραγματικών αριθμών και ορμαθών χαρακτήρων χρησιμοποιούνται αντίστοιχα οι κλάσεις του συστήματος *Telos_Integer, Telos_Real, Telos_String*. Οι κλάσεις αυτές δεν έχουν γνωρίσματα και οι τιμές τους δεν μπορούν να δημιουργηθούν αλλά μόνο να γίνει αναφορά σ' αυτές. Οι μηχανισμοί αφαίρεσης του μοντέλου είναι ο *μηχανισμός ταξινόμησης*, ο *μηχανισμός απόδοσης γνωρίσματος* και ο *μηχανισμός γενίκευσης/εξειδίκευσης*.

Με το μηχανισμό ταξινόμησης οντότητες που ορίζονται από το χρήστη με κοινά χαρακτηριστικά ταξινομούνται σε μία ή περισσότερες κλάσεις οντοτήτων. Μια οντότητα που ταξινομείται σε μια κλάση οντοτήτων, λέμε ότι είναι *περίπτωση* αυτής της κλάσης.

[1] Η υλοποίηση αυτή έγινε στα πλαίσια ανάπτυξης του Συστήματος Σημασιολογικού Ευρετηριασμού, από την Ομάδα Πληροφοριακών Συστημάτων και Τεχνολογίας Λογισμικού του Ινστιτούτου Πληροφορικής του Ιδρύματος Τεχνολογίας και Έρευνας

Οι κλάσεις οντοτήτων με τη σειρά τους, θεωρούμενες σαν οντότητες μπορούν να ταξινομηθούν σε κλάσεις που οι περιπτώσεις τους είναι κλάσεις οντοτήτων κ.ο.κ.. Με τον τρόπο αυτό δημιουργείται μια μη φραγμένη ιεραρχία από κλάσεις. Στο πρώτο επίπεδο της ιεραρχίας αυτής βρίσκονται οι *ατομικές οντότητες (Tokens)*, στο δεύτερο οι *απλές κλάσεις (S_Classes)* οντοτήτων που περιέχουν ατομικές οντότητες, στο τρίτο οι *μετακλάσεις (M1_Classes)* που περιέχουν απλές κλάσεις οντοτήτων, κ.ο.κ.. Οι κλάσεις στην Telos χρησιμοποιούνται για την ταξινόμηση *ανεξαρτήτων οντοτήτων (Individuals)*, αλλά και σχέσεων μεταξύ οντοτήτων ή *γνωρισμάτων (Attributes)*.

Με το μηχανισμό γενίκευσης μπορούμε να ορίσουμε μια σχέση υποσυνόλου (σχέση μερικής διάταξης) μεταξύ κλάσεων του ίδιου επιπέδου ταξινόμησης (εκτός του επιπέδου των ατομικών οντοτήτων). Μια κλάση μπορεί να είναι εξειδίκευση (υποσύνολο/υποκλάση) μίας ή περισσοτέρων άλλων κλάσεων, καθώς και γενίκευση (υπερσύνολο/υπερκλάση) μίας ή περισσοτέρων άλλων κλάσεων. Μια ενδιαφέρουσα ιδιότητα της σχέσης γενίκευσης είναι η κληρονομικότητα των γνωρισμάτων. Η υποκλάση λοιπόν κληρονομεί όλα τα γνωρίσματα των υπερκλάσεων της, ενώ ταυτόχρονα δύναται να έχει και δικά της γνωρίσματα.

Με το μηχανισμό απόδοσης γνωρίσματος αποδίδονται γνωρίσματα στις οντότητες. Τα γνωρίσματα θεωρούνται διμελείς σχέσεις μεταξύ οντοτήτων, που έχουν ένα *σύνολο αφετηρίας* (την οντότητα στην οποία αποδίδονται -- *οντότητα-αφετηρία*) και ένα *σύνολο αφίξεως* (την οντότητα στην οποία καταλήγουν -- *οντότητα-προορισμού*). Να σημειωθεί εδώ ότι η οντότητα προορισμού πρέπει να είναι περίπτωση των ανεξαρτήτων οντοτήτων. Δεδομένου ότι και τα γνωρίσματα θεωρούνται οντότητες μπορούν κιαυτά να έχουν δικά τους γνωρίσματα. Τα γνωρίσματα ταξινομούνται σε μία ή περισσότερες κλάσεις (κατηγορίες) γνωρισμάτων των οποίων οι περιπτώσεις βρίσκονται ένα επίπεδο πιο κάτω στην ιεραρχία ταξινόμησης.

4.2 Μέθοδος Μετάφρασης

Η μέθοδος που προτείνουμε συγκεντρώνει στοιχεία, από άλλες μεθόδους ([13]), ή αλγορίθμους ([34], [3]) για τη μετάφραση σχεσιακών σχημάτων, σε σημασιολογικά πλουσιότερα μοντέλα δεδομένων (π.χ. Extended O-O, Functional, Extended E-R). Οι υποθέσεις που κάνουμε είναι αρκετά χαλαρές, ώστε να περιλαμβάνουν, κατά το δυνατό, μεγαλύτερο αριθμό σχεσιακών σχημάτων. Το τίμημα αυτής της χαλαρότητας (ιδίως στη φύση των εξαρτήσεων εγκλεισμού), το πληρώνουμε με την ανάγκη για επέμβαση του

χρήστη, ώστε να επιλυθούν διάφορες ασάφειες στο είδος της σχέσης μεταξύ οντοτήτων, όπως θα δούμε παρακάτω. Μετά την επίλυση των ασαφειών και την ανακάλυψη λανθανουσών οντοτήτων (οι οποίες εκφράζονται με εικονικά σχήματα-σχέσεων, που επαυξάνουν το αρχικό σχεσιακό σχήμα), επεξεργαζόμαστε το γράφο εγκλεισμού που έχει προκύψει (στο επαυξημένο σχεσιακό σχήμα), και απαλείφουμε πιθανούς πλεονασμούς. Από τον τελικό γράφο εγκλεισμού, και με τη χρήση συγκεκριμένων κανόνων μετάφρασης, δημιουργούμε το σημασιολογικό δίκτυο και εντοπίζουμε τις σχέσεις αντιστοίχισης που υπάρχουν μεταξύ αυτού και του (επαυξημένου) σχεσιακού σχήματος.

4.2.1 Υποθέσεις

Οι υποθέσεις που θεωρούμε, για τη φύση του σχεσιακού σχήματος που θα μεταφραστεί, καθώς και για τη φύση των εξαρτήσεων εγκλεισμού, που ισχύουν σ' αυτό, είναι όπως θα δούμε αρκετά χαλαρές, έτσι ώστε να επιτρέπουν την εφαρμογή της μεθόδου, σε μεγάλο πλήθος σχεσιακών σχημάτων. Ας δούμε όμως πρώτα, τί είδους υποθέσεις κάνουν παρόμοιες μέθοδοι, ή αλγόριθμοι μετάφρασης σχεσιακών σχημάτων.

Στο [34], απαιτείται το σχεσιακό σχήμα να βρίσκεται σε Boyce-Codd κανονική μορφή, και οι εξαρτήσεις εγκλεισμού να είναι βασισμένες σε κλειδιά. Όμως η Boyce-Codd κανονική μορφή, είναι μια αρκετά αυστηρή μορφή στην οποία υπάρχουν σχήματα που **δεν** μπορούν ν' αναχθούν. Τί γίνεται όμως όταν υπάρχουν και εξαρτήσεις εγκλεισμού που δεν είναι βασισμένες σε κλειδιά;

Το ερώτημα αυτό δεν το απαντά ούτε ο αλγόριθμος που παρουσιάζεται στο [3]. Εκεί, δεν υποθέτουν ότι το σχεσιακό σχήμα είναι σε Boyce-Codd κανονική μορφή, όμως και αυτοί θεωρούν μόνο εξαρτήσεις εγκλεισμού βασισμένες σε κλειδιά. Υποστηρίζουν μάλιστα, ότι ακόμα κι αν υπάρχουν εξαρτήσεις εγκλεισμού που δεν είναι βασισμένες σε κλειδιά, δεν φαίνεται ότι είναι σχετικές με τη μετάφραση του σχεσιακού σχήματος στο οντοκεντρικό ή σημασιολογικό μοντέλο.

Μόνο στη μέθοδο του [13], [11] βρίσκουμε την αξιοποίηση τέτοιων ειδών εξαρτήσεων, οι οποίες χρησιμεύουν στην ανίχνευση λανθανουσών οντοτήτων, που υπάρχουν λόγω κακού σχεδιασμού του σχεσιακού σχήματος. Στη μέθοδο αυτή, τα μόνα δεδομένα που χρησιμοποιούν είναι το σχεσιακό σχήμα, τα δεδομένα της σχεσιακής ΒΔ, και τις δηλώσεις που έχουν γίνει απ' τους χρήστες για συναρτησιακές εξαρτήσεις, εξαρτήσεις εγκλεισμού και κλειδιά. Κατόπιν, περνούν από μια φάση απόκτησης γνώσης, όπως την ονομάζουν, στην οποία με αλγορίθμους που ανιχνεύουν στα δεδομένα της ΒΔ, αλγορίθμους κανονικοποίησης, καθώς και με ερωτήσεις προς το χρήστη, εντοπίζουν **όλα** τα κλειδιά των

σχέσεων, **όλες** τις συναρτησιακές εξαρτήσεις, φέρνουν το σχήμα σε τρίτη κανονική μορφή με κάθε σχέση να αντιπροσωπεύει μία έννοια, βρίσκουν **όλες** τις εξαρτήσεις εγκλεισμού και τέλος χαρακτηρίζουν (αντιπροσωπευτικά, μη-αντιπροσωπευτικά) τα κλειδιά των σχέσεων ([12]).

Θεωρώντας ότι οι υποθέσεις του [13] είναι οι λιγότερο αυστηρές, μπορούμε στο εξής να θεωρούμε ότι για τα σχεσιακά σχήματα που θα μεταφράσουμε ισχύουν τα εξής:

- βρίσκονται σε τρίτη κανονική μορφή, με κάθε σχήμα-σχέσης να εκφράζει μόνο μία έννοια (συσχέτιση, οντότητα, τμήμα οντότητας)
- τα κλειδιά των σχέσεων έχουν χαρακτηριστεί σε αντιπροσωπευτικά ή μη-αντιπροσωπευτικά
- έχουν εντοπιστεί οποιουδήποτε είδους εξαρτήσεις εγκλεισμού

Είναι φανερό ότι, τα δεδομένα αυτά, είτε θα μας δίνονται εξ αρχής, είτε θα μπορούμε να τα συμπεράνουμε όπως προτείνεται στο [13].

4.2.2 Φάση Αναγνώρισης Λανθανουσών Οντοτήτων

Όπως αναφέραμε και νωρίτερα, είναι δυνατόν τα σχήματα μας να μην είναι καλά σχεδιασμένα, με αποτέλεσμα οντότητες που θα έπρεπε να υπάρχουν με τη μορφή σχημάτων-σχέσεων, να είναι "λανθάνουσες" σαν υποσύνολα γνωρισμάτων μέσα στα σχήματα-σχέσεων. Σκοπός της φάσης αυτής, ανάλογη της οποίας συναντάμε και στη μέθοδου του [13], είναι να ανακαλυφθούν αυτές οι οντότητες και να δημιουργηθούν τα σχήματα-σχέσεων (*εικονικά σχήματα-σχέσεων*) που τις εκφράζουν. Έτσι, το αρχικό σχήμα θα επαυξηθεί με την προσθήκη των εικονικών σχέσεων, οι οποίες σχηματίζονται από την προβολή σχέσεων του αρχικού σχήματος στα υποσύνολα των γνωρισμάτων τους που αντιστοιχούν στις λανθάνουσες οντότητες. Στη φάση αυτή, θα χρειαστεί και η βοήθεια του χρήστη για να αποφασίσει σε ορισμένες περιπτώσεις, για τη δημιουργία ή όχι, κάποιων εικονικών σχέσεων.

Όπως και στις άλλες προσπάθειες που είδαμε, το κύριο εργαλείο είναι το σύνολο των εξαρτήσεων εγκλεισμού που θα μας βοηθήσει τόσο στην ανίχνευση των λανθανουσών οντοτήτων, όσο και στη δημιουργία του τελικού σημασιολογικού δικτύου. Από τις υποθέσεις που κάναμε, προκύπτουν οι παρακάτω πιθανές μορφές εξαρτήσεων εγκλεισμού ($R(a) \subseteq S(b)$, όπου a, b σύνολο συμβατών γνωρισμάτων των σχημάτων-σχέσεων R και S).

S(b)	**Κλειδί**	**Τμήμα ΣΚ**	**Όχι Κλειδί**
R(a)	**(Κ)**	**(τ.ΣΚ)**	**(ΟΚ)**
Κλειδί (Κ)	Κ-Κ	Κ-τ.ΣΚ	Κ-ΟΚ
Κάθε τ.ΣΚ	κ.τ.ΣΚ-Κ	κ.τ.ΣΚ-τ.ΣΚ	κ.τ.ΣΚ-ΟΚ
Μόνο τ.ΣΚ	μ.τ.ΣΚ-Κ	μ.τ.ΣΚ-τ.ΣΚ	μ.τ.ΣΚ-ΟΚ
Όχι Κλ. (ΟΚ)	ΟΚ-Κ	ΟΚ-τ.ΣΚ	ΟΚ-ΟΚ

Πίνακας 1

Οι περιπτώσεις που το αριστερό μέλος της ΕΕ είναι της μορφής **Κάθε τ.ΣΚ** ή **Μόνο τ.ΣΚ**, σημαίνουν ότι κάθε τμήμα ενός σύνθετου κλειδιού συμμετέχει σε μια ΕΕ ή ότι μόνο κάποιο τμήμα ενός σύνθετου κλειδιού συμμετέχει σε μια ΕΕ.

Τα βήματα που θα ακολουθήσουμε για την ανακάλυψη των λανθανουσών οντοτήτων είναι τα παρακάτω:

- **Βήμα 1**

  Εξετάζουμε τις εξαρτήσεις εγκλεισμού που το δεξί τους μέλος είναι τμήμα κλειδιού ή όχι κλειδί με την εξής σειρά (όπου Χ, οποιοδήποτε από τα στοιχεία της στήλης R(a) του Πίνακα 1):

  Στήλη 2: Περιπτώσεις Χ ⊆ τ.ΣΚ *(Πρώτη Μορφή ΕΕ)*
  Στήλη 3: Περιπτώσεις Χ ⊆ ΟΚ *(Δεύτερη Μορφή ΕΕ)*

  Σ' αυτό το βήμα, οι λανθάνουσες οντότητες αντιστοιχούν στα υποσύνολα γνωρισμάτων των σχημάτων σχέσεων που συμμετέχουν στο δεξί μέλος των ΕΕ (δηλαδή στο υποσύνολο b του σχήματος σχέσης S).

- **Βήμα 2**

  Εξετάζουμε τις εξαρτήσεις εγκλεισμού σε μη-αντιπροσωπευτικά κλειδιά με την εξής σειρά (όπου Χ, οποιοδήποτε από τα στοιχεία της στήλης R(a) του Πίνακα 1 για την τρίτη μορφή ΕΕ, και οποιοδήποτε από τα στοιχεία της γραμμής S(b) του Πίνακα 1 για την τέταρτη μορφή ΕΕ):

  Περιπτώσεις Χ ⊆ μ.α.κ *(Τρίτη Μορφή ΕΕ)*
  Περιπτώσεις μ.α.κ ⊆ Χ *(Τέταρτη Μορφή ΕΕ)*

  Στην τρίτη μορφή ΕΕ, η λανθάνουσα οντότητα αντιστοιχεί στο υποσύνολο γνωρισμάτων b του σχήματος σχέσης S, ενώ στην τέταρτη μορφή χρειάζεται η

άποψη του χρήστη για να διευκρινίσει αν υπάρχει ή όχι, μια λανθάνουσα οντότητα που αντιστοιχεί στο υποσύνολο γνωρισμάτων a του σχήματος σχέσης R.

- **Βήμα 3**
  Εξετάζουμε κάθε τμήμα ενός σύνθετου κλειδιού ενός σχήματος σχέσης που **δεν** έχει αναμιχθεί σε εξάρτηση εγκλεισμού, και ζητάμε από το χρήστη να διευκρινίσει αν υπάρχει ή όχι, λανθάνουσα οντότητα που να αντιστοιχεί στο τμήμα (υποσύνολο γνωρισμάτων) αυτό.

Βλέπουμε λοιπόν ότι αρχικά ανακαλύπτουμε **αυτόματα** όλες τις λανθάνουσες οντότητες που μπορούμε (περιπτώσεις του πρώτου βήματος και πρώτη περίπτωση δεύτερου βήματος), ενώ κατόπιν ζητάμε την άποψη του χρήστη για την ανακάλυψη μιας λανθάνουσας οντότητας (δεύτερη περίπτωση δεύτερου βήματος και περιπτώσεις τρίτου βήματος).

Στη διάρκεια καθεμιάς περίπτωσης από τα παραπάνω βήματα, μόλις ανακαλυφθεί μια λανθάνουσα οντότητα (υποσύνολο γνωρισμάτων ενός σχήματος-σχέσης) εκτελούμε με τη σειρά τις πιο κάτω ενέργειες:

- *Ενέργεια 1*
  Δημιουργούμε ένα εικονικό σχήμα-σχέσης για τη λανθάνουσα οντότητα που ανακαλύψαμε, που έχει σα σύνολο γνωρισμάτων τα γνωρίσματα της σχέσης απ' όπου προήλθε που αντιστοιχούν στη λανθάνουσα οντότητα, και σαν όνομα τη συγχώνευση των ονομάτων των γνωρισμάτων της. Αν λοιπόν, ένα εικονικό σχήμα σχέσης προέρχεται από ένα σύνολο γνωρισμάτων *b* ενός σχήματος σχέσης *S*, συμβολίζεται με S_b, έχει σαν όνομα τη συγχώνευση των ονομάτων των γνωρισμάτων του συνόλου *b* και σα σύνολο γνωρισμάτων το σύνολο *b*.

- *Ενέργεια 2*
  Θέτουμε σαν αντιπροσωπευτικό κλειδί (αν.κ) της νέας εικονικής σχέσης το σύνολο των γνωρισμάτων της, ενώ καθένα από αυτά αλλάζει το όνομά του ανάλογα με τον τύπο του. Αν ο τύπος ενός γνωρίσματος ήταν ορμαθός χαρακτήρων (String), προσθέτουμε στο παλιό του όνομα το επίθημα $_όνομα$, ενώ σε κάθε άλλη περίπτωση προσθέτουμε το επίθημα $_νούμερο$. Συμβολίζουμε με b', το αν.κ. της εικονικής σχέσης S_b.

- *Ενέργεια 3*
  Όλες οι ΕΕ που το δεξί τους μέλος είναι το γνώρισμα (ή τα γνωρίσματα) του σχήματος

σχέσης απ' όπου προήλθε η εικονική σχέση, μεταφέρονται σ' αυτήν, ως εξής (όπου Χ, οποιοδήποτε από τα στοιχεία της στήλης R(a) του Πίνακα 1):

Χ $\subseteq$ απλό τ.ΣΚ $\longrightarrow$ Χ $\subseteq$ Απλό Κ

Χ $\subseteq$ σύνθετο τ.ΣΚ $\longrightarrow$ Χ $\subseteq$ Σύνθετο Κ

Χ $\subseteq$ απλό ΟΚ $\longrightarrow$ Χ $\subseteq$ Απλό Κ

Χ $\subseteq$ σύνθετο ΟΚ $\longrightarrow$ Χ $\subseteq$ Σύνθετο Κ

Χ $\subseteq$ μ.α.κ $\longrightarrow$ Χ $\subseteq$ αν.κ

Αν λοιπόν S_b είναι η νέα εικονική σχέση, όλες οι ΕΕ της μορφής $K(x) \subseteq S(b)$ με x και b σύνολα συμβατών γνωρισμάτων, μετατρέπονται σε ΕΕ της μορφής $K(x) \subseteq S_b(b')$. Με τον τρόπο αυτό, οι εξαρτήσεις εγκλεισμού που μεταφέρονται στην εικονική σχέση, μετατρέπονται σε ΕΕ κλειδιού αν το x είναι αν.κ της K, ή ΕΕ βασισμένες σε κλειδί αν το x δεν είναι αν.κ της K.

- *Ενέργεια 4*

  Όλες οι ΕΕ που το αριστερό τους μέλος είναι το γνώρισμα (ή τα γνωρίσματα) του σχήματος σχέσης απ' όπου προήλθε η εικονική σχέση, μεταφέρονται σ' αυτήν. Αν λοιπόν S_b είναι η νέα εικονική σχέση, όλες οι ΕΕ της μορφής $S(b) \subseteq K(x)$ με x και b σύνολα συμβατών γνωρισμάτων, μετατρέπονται σε ΕΕ της μορφής $S_{b'} \subseteq K(x)$. Αυτές οι ΕΕ, θα μετατραπούν σε ΕΕ που το αριστερό τους μέλος θα είναι απλό ή σύνθετο αντιπροσωπευτικό κλειδί, ανάλογα με το αν το b είναι μονοσύνολο ή όχι.

- *Ενέργεια 5*

  Δημιουργείται μια νέα ΕΕ. Αν είναι R το αρχικό σχήμα-σχέσης (στο οποίο περικλείεται η λανθάνουσα οντότητα), R_A το εικονικό (το οποίο αντιστοιχεί στη λανθάνουσα οντότητα), και Α το σύνολο των γνωρισμάτων του R απ' τα οποία προέκυψε το R_A, η ΕΕ έχει τη μορφή $R(A) \subseteq R_A(A')$ (A'αν.κ.R_A). Δημιουργείται επίσης και η αντίστροφή της ΕΕ, $R_A(A') \subseteq R(A)$, μια και το εικονικό σχήμα-σχέσης R_A αντιπροσωπεύει ακριβώς τις οντότητες που συμμετέχουν στην R, δηλ. το R(A) και το $R_A(A')$ είναι τα ίδια.

Κοιτάζοντας τον Πίνακα 1 παρατηρούμε ότι οι εξαρτήσεις εγκλεισμού που δεν είναι βασισμένες σε κλειδί, ανήκουν σε κάποια από τις περιπτώσεις των δύο τελευταίων στηλών του. Μετά την επεξεργασία όμως των περιπτώσεων του πρώτου βήματος αυτές οι εξαρτήσεις εγκλεισμού έχουν μετατραπεί σε ΕΕ βασισμένες σε κλειδιά (των εικονικών σχέσεων), ενώ η επεξεργασία των περιπτώσεων των άλλων δύο βημάτων είναι προφανές ότι εισάγει μόνο ΕΕ βασισμένες σε κλειδιά (των εικονικών σχέσεων). Αυτό λοιπόν σημαίνει

ότι, εκτός από τις ΕΕ των εικονικών σχέσεων προς τις σχέσεις από τις οποίες προήλθαν, όλες οι άλλες ΕΕ είναι βασισμένες σε κλειδιά. Επειδή ο μοναδικός ρόλος των πρώτων ΕΕ, είναι να εξασφαλίσουν ότι οι εικονικές σχέσεις περιέχουν την ίδια ακριβώς πληροφορία με την προβολή των σχέσεων από τις οποίες προήλθαν πάνω στο σύνολο των γνωρισμάτων από τα οποία προήλθαν, μπορούμε να τις παραλείψουμε από το γράφο εγκλεισμού, μια και δεν πρόκειται να μας βοηθήσουν στη μετάφραση του σχεσιακού σχήματος. Αυτό λοιπόν που πετυχαίνουμε με τη διαδικασία που μόλις περιγράψαμε, είναι να μετατρέψουμε όλες τις εξαρτήσεις εγκλεισμού, σε ΕΕ κλειδιών ή σε ΕΕ βασισμένων σε κλειδιά, άροντας έτσι κάποια λάθη σχεδιασμού του αρχικού σχεσιακού σχήματος. Κάτι τέτοιο μας επιτρέπει από εδώ και πέρα να χειριστούμε τη μετάφραση του σχεσιακού σχήματος στο οποίο έχουμε καταλήξει (με την εισαγωγή των εικονικών σχέσεων), με τρόπο ανάλογο με αυτούς που προτείνονται στις εργασίες [3] και [34].

Πριν προχωρήσουμε όμως σε μια πιο τυπική παρουσίαση της διαδικασίας, ας δούμε κάποια παραδείγματα που θα βοηθήσουν στην κατανόηση της. (Τα αν.κ. είναι διπλά υπογραμμισμένα, ενώ τα μ.αν.κ. είναι απλά υπογραμμισμένα)

Παράδειγμα 1: (πρώτη περίπτωση πρώτου βήματος, Χ ⊂ τ.ΣΚ) Ας θεωρήσουμε το εξής σχεσιακό σχήμα :

σχολή(πανεπιστήμιο, σχ_όνομα, τύπος)
ιδιωτικό_πανεπ(όνομα, κατηγορία, ιδιοκτήτης)
δημόσιο_πανεπ(όνομα, κατηγορία, προϋπολογισμός)
κτίριο_πανεπ(πανεπ, κτρ, διεύθυνση, χωρητικότητα)

με εξαρτήσεις εγκλεισμού

ιδιωτικό_πανεπ(όνομα) ⊂ σχολή(πανεπιστήμιο)
δημόσιο_πανεπ(όνομα) ⊂ σχολή(πανεπιστήμιο)
κτίριο_πανεπ(πανεπ) ⊂ σχολή(πανεπιστήμιο)

απ' όπου παίρνουμε :

--εικονική σχέση: πανεπιστήμιο(πανεπιστήμιο_όνομα)
--μεταφορά ΕΕ προς/από σχολή(πανεπιστήμιο), σε ΕΕ προς/από πανεπιστήμιο(πανεπιστήμιο_όνομα) :
ιδιωτικό_πανεπ(όνομα) ⊂ πανεπιστήμιο(πανεπιστήμιο_όνομα)
δημόσιο_πανεπ(όνομα) ⊂ πανεπιστήμιο(πανεπιστήμιο_όνομα)
κτίριο_πανεπ(πανεπ) ⊂ πανεπιστήμιο(πανεπιστήμιο_όνομα)
--νέες ΕΕ: σχολή(πανεπιστήμιο) ⊂ πανεπιστήμιο(πανεπιστήμιο_όνομα)
πανεπιστήμιο(πανεπιστήμιο_όνομα) ⊂ σχολή(πανεπιστήμιο)

□

Παράδειγμα 2: (πρώτη περίπτωση δεύτερου βήματος, Χ ⊂ μ.αν.κ) Ας θεωρήσουμε το εξής σχεσιακό σχήμα :
τμήμα(τμ#, τμ_όνομα, διευθυντής)
έργο(έρ#, έρ_όνομα, υπεύθυνος)

με εξαρτήσεις εγκλεισμού
έργο(υπεύθυνος) ⊂ τμήμα(διευθυντής)
απ' όπου παίρνουμε :

--εικονική σχέση: διευθυντής(διευθυντής_όνομα)
--μεταφορά ΕΕ προς/από τμήμα(διευθυντής), σε ΕΕ προς/από διευθυντής(διευθυντής_όνομα) :
έργο(υπεύθυνος) ⊂ διευθυντής(διευθυντής_όνομα)
--νέες ΕΕ: τμήμα(διευθυντής) ⊂ διευθυντής(διευθυντής_όνομα)
διευθυντής(διευθυντής_όνομα) ⊂ τμήμα(διευθυντής)

Στο σημείο αυτό επειδή η ΕΕ
έργο(υπεύθυνος) ⊂ διευθυντής(διευθυντής_όνομα)
είναι της τέταρτης μορφής, ερωτάται ο χρήστης αν θέλει ή όχι να δημιουργηθεί και νέα εικονική σχέση. Αν αυτός απαντήσει καταφατικά, τότε παίρνουμε :

--εικονική σχέση: υπεύθυνος(υπεύθυνος_όνομα)
--μεταφορά ΕΕ προς/από έργο(υπεύθυνος), σε ΕΕ προς/από υπεύθυνος(υπεύθυνος_όνομα) :
υπεύθυνος(υπεύθυνος_όνομα) ⊂ διευθυντής(διευθυντής_όνομα)
--νέες ΕΕ: έργο(υπεύθυνος) ⊂ υπεύθυνος(υπεύθυνος_όνομα)
υπεύθυνος(υπεύθυνος_όνομα) ⊂ έργο(υπεύθυνος)

□

Πιο τυπικά η διαδικασία που περιγράψαμε μπορεί να παρουσιαστεί με μορφή ψευδοκώδικα σε μια Pascal-like γλώσσα. Θα θεωρήσουμε ότι το σύνολο των ΕΕ *I*, περιγράφεται από ονοματισμένο γράφο εγκλεισμού. Αν συμβολίσουμε $G = (V, E)$ το γράφο αυτό τότε:
V = σύνολο σχημάτων-σχέσεων του αρχικού σχεσιακού σχήματος
Επίσης, αν $R(a) \subseteq S(b) \in I$, τότε $(R, S, (a, b)) \in E$.

Μια ακμή λοιπόν του Ε είναι μια τριάδα , όπου το R είναι ο κόμβος πηγής, το S ο κόμβος προορισμού και το (a,b) (ζεύγος συνόλων συμβατών γνωρισμάτων) το όνομα της ακμής.

Η διαδικασία αναγνώρισης λανθανουσών οντοτήτων, ονομάζεται *ΕυρεσηΛανθανουσωνΟντοτητων(V, E)*. Δέχεται σαν αναφορικά ορίσματα ένα σύνολο κορυφών *V* και ακμών *E* (σχήματα-σχέσεων και ΕΕ, αντίστοιχα), τα οποία αρχικά περιέχουν τις κορυφές και ακμές του αρχικού γράφου, ενώ τελικά θα περιέχουν επιπλέον τις κορυφές και ακμές που έχουν προέλθει από τις εικονικές σχέσεις και τη δημιουργία/μεταφορά των ΕΕ αντίστοιχα. Νωρίτερα είδαμε ότι για τις τρεις πρώτες μορφές εξαρτήσεων εγκλεισμού η λανθάνουσα οντότητα (εικονική σχέση) βρίσκεται αυτόματα (χωρίς επέμβαση του χρήστη). Στον κορμό της λοιπόν, η διαδικασία *ΕυρεσηΛανθανουσωνΟντοτητων(V, E)* καλεί διαδοχικά για καθεμιά από τις πρώτες τρεις μορφές ΕΕ, τη διαδικασία *ΑυτοματηΔημιουργιαΕικονικηςΣχεσης(V, E, Μορφή ΕΕ)* η οποία ανάλογα με τη μορφή ΕΕ δημιουργεί τις εικονικές σχέσεις και μεταφέρει τις ΕΕ. Για την έκτη μορφή ΕΕ και για τις περιπτώσεις τμημάτων σύνθετων κλειδιών σχημάτων σχέσεων που δε συμμετέχουν σε ΕΕ, καλούνται διαδοχικά οι διαδικασίες *ΕΕσεΜηΑντιπροωπευτικαΚλειδια(V, E)* και *ΤμηματαΣΚχωρισΕΕ(V, E)*, όπου με απόφαση του χρήστη δημιουργούνται ή όχι νέες εικονικές σχέσεις και μεταφέρονται οι ΕΕ.

```
procedure ΕυρεσηΛανθανουσωνΟντοτητων(var V, E);
        { Επεξεργασία πρώτης περίπτωσης πρώτου βήματος }
        ΑυτοματηΔημιουργιαΕικονικηςΣχεσης(V, E, Πρώτη Μορφή ΕΕ);
        { Επεξεργασία δεύτερης περίπτωσης πρώτου βήματος }
        ΑυτοματηΔημιουργιαΕικονικηςΣχεσης(V, E, Δεύτερη Μορφή ΕΕ);
        { Επεξεργασία πρώτης περίπτωσης δεύτερου βήματος }
        ΑυτοματηΔημιουργιαΕικονικηςΣχεσης(V, E, Τρίτη Μορφή ΕΕ);
        { Επεξεργασία δεύτερης περίπτωσης δεύτερου βήματος }
        ΕΕσεΜηΑντιπροωπευτικαΚλειδια(V, E);
        { Επεξεργασία τρίτου βήματος }
        ΤμηματαΣΚχωρισΕΕ(V, E);
end procedure;

procedure ΑυτοματηΔημιουργιαΕικονικηςΣχεσης(var V, E, ΜορφηΕΕ);
        { Έλεγχος για την ύπαρξη ακμής (ΕΕ) που δεν έχουμε ξαναεξετάσει,
        απ' όπου μπορεί να προκύψει εικονική σχέση }
        while στο E υπάρχει ακμή (R, S, (a, b)) της μορφής ΜορφηΕΕ ΚΑΙ
                ((S, R, (b, a)) με a αν.κ του R) ∉ E do
        begin
                { Δημιουργία εικονικής σχέσης S_b
```

```
                και καθορισμός αν. κ. (Ενέργειες 1, 2)}
                V := V ∪ {S_b};
                Θέσε b' αν.κ της S_b;
                { Μεταφορά ΕΕ προς τη νέα εικονική σχέση (Ενέργεια 3)}
                E := E ∪ {(K, S_b, (x, b')) | (K, S, (x, b)) ∈ E} \ {(K, S, (x, b)) ∈ E};
                { Μεταφορά ΕΕ από τη νέα εικονική σχέση (Ενέργεια 4)}
                E := E ∪ {(S_b, K, (b', x)) | (S, K, (b, x)) ∈ E} \ {(S, K, (b, x)) ∈ E};
                { Δημιουργία ΕΕ εικονικής σχέσης
                με τη σχέση απ' όπου προήλθε (Ενέργεια 5)}
                E := E ∪ {(S, S_b, (b, b')), (S_b, S, (b', b))};
        end;
end procedure;

procedure ΕΕσεΜηΑντιπροωπευτικαΚλειδια(var V, E);
        { Έλεγχος για την ύπαρξη ακμής (ΕΕ) που δεν έχουμε ξαναεξετάσει,
        απ' όπου μπορεί να προκύψει εικονική σχέση }
        while στο Ε υπάρχει ακμή (R, S, (a, b)) που δεν έχει εξεταστεί
                ΚΑΙ είναι της τέταρτης μορφής
                ΚΑΙ ((S, R, (b, a)) με b αν.κ του S) ∉ E do
        begin
                Ρώτα το χρήστη αν επιθυμεί δημιουργία εικονικής
                σχέσης με γνωρίσματα a;
                if (απάντηση = 'ΝΑΙ') then begin
                   { Δημιουργία εικονικής σχέσης R_a
                   και καθορισμός αν. κ. (Ενέργειες 1, 2)}
                   V := V ∪ {R_a};
                   Θέσε a' αν.κ της R_a;
                   { Μεταφορά ΕΕ προς τη νέα εικονική σχέση (Ενέργεια 3)}
                   E := E ∪ {(K, R_a, (x, a')) | (K, R, (x, a)) ∈ E} \ {(K, R, (x, a)) ∈ E};
                   { Μεταφορά ΕΕ από τη νέα εικονική σχέση (Ενέργεια 4)}
                   E := E ∪ {(R_a, K, (a', x)) | (R, K, (a, x)) ∈ E} \ {(R, K, (a, x)) ∈ E};
                   { Δημιουργία ΕΕ εικονικής σχέσης
                   με τη σχέση απ' όπου προήλθε (Ενέργεια 5)}
                   E := E ∪ {(R, R_a, (a, a')), (R_a, R, (a', a))};
                end;
```

```
            end;
end procedure;

procedure ΤμηματαΣΚχωριςΕΕ(var V, E);
            { Έλεγχος για την ύπαρξη τμημάτων ΣΚ σχημάτων σχέσεων που δε
            συμμετέχουν σε ΕΕ και δεν έχουμε ξαναεξετάσει, απ' όπου μπορούν
            να προκύψουν εικονικές σχέσεις }
            while υπάρχουν τμήματα ΣΚ που δε συμμετέχουν σε ΕΕ
                  ΚΑΙ δεν έχουν εξεταστεί do
            begin
                  Αν a το σύνολο των γνωρισμάτων ενός τμήματος του ΣΚ
                  ενός σχήματος σχέσης R που δε συμμετέχει σε ΕΕ,
                  ρώτα το χρήστη αν θέλει να δημιουργηθεί μια εικονική
                  σχέση από το τμήμα αυτό;
                  if (απάντηση = 'ΝΑΙ') then begin
                     {Δημιουργία εικονικής σχέσης R_a
                     και καθορισμός αν. κ. (Ενέργειες 1, 2)}
                     V := V ∪ {R_a};
                     Θέσε a' αν.κ της R_a;
                     { Δεδομένου ότι το σύνολο γνωρισμάτων a της σχέσης R
                     απ' όπου προέρχεται η εικονική σχέση δε συμμετείχε
                     σε κάποια ΕΕ, δε χρειάζονται οι ενέργειες 3 και 4
                     και προχωρούμε στη δημιουργία ΕΕ εικονικής σχέσης
                     με τη σχέση απ' όπου προήλθε (Ενέργεια 5)}
                     E := E ∪ {(R, R_a, (a, a')), (R_a, R, (a', a))};
                  end;
            end;
end procedure;
```

4.2.3 Φάση Κατασκευής Τελικού Γράφου Εγκλεισμού

Στόχος της φάσης αυτής, είναι η κατασκευή ενός γράφου εγκλεισμού, πάνω στον οποίο θα εφαρμοστούν οι κανόνες μετάφρασης που θα δώσουν το σημασιολογικό δίκτυο. Ο γράφος αυτός θα πρέπει κατά το δυνατόν να είναι απαλλαγμένος από πλεονασμούς, έτσι ώστε το σημασιολογικό δίκτυο να περιέχει μόνο την απαραίτητη πληροφορία. Δύο λοιπόν είναι τα είδη των πλεονασμών που θα απαλείψουμε σ' αυτή τη φάση : πλεονασμοί που προέρχονται

από την κάθετη κατάτμηση σχημάτων σχέσεων, και πλεονασμοί που προέρχονται από τη μεταβατικότητα της σχέσης εγκλεισμού.

Υπάρχουν περιπτώσεις σχεδίασης ενός σχεσιακού σχήματος, όπως αναφέρεται και στο [3], στις οποίες πληροφορίες για μια οντότητα, είναι διάσπαρτες σε πλέον της μίας σχέσεις (κάθετη κατάτμηση). Κάτι τέτοιο έχει σα συνέπεια την ύπαρξη κύκλων με ακμές που αντιστοιχούν σε ΕΕ αντιπροσωπευτικών κλειδιών στο γράφο εγκλεισμού. Ας θεωρήσουμε για παράδειγμα το σχήμα :

Μαθητής(α.μ., διεύθυνση)

ΑρχείαΜαθητών(α.μ., μέσοςόρος, επιβλέπωνκαθηγητής)

Καθηγητής(όνομα, διεύθυνση)

στο οποίο ισχύουν οι εξής ΕΕ βασισμένες σε κλειδιά

Μαθητής(α.μ.) $\subseteq$ *ΑρχείαΜαθητών(α.μ.)*

ΑρχείαΜαθητών(α.μ.) $\subseteq$ *Μαθητής(α.μ.)*

ΑρχείαΜαθητών(επιβλέπωνκαθηγητής) $\subseteq$ *Καθηγητής(όνομα)*

Στο παράδειγμα αυτό βλέπουμε ότι πληροφορία για την οντότητα Μαθητής, είναι διάσπαρτη σε δύο πίνακες, που συνδέονται μεταξύ τους με ΕΕ αντιπροσωπευτικού κλειδιού σχηματίζοντας κύκλο στο αντίστοιχο γράφο εγκλεισμού. Άρα, το πρώτο βήμα για την κατασκευή ενός τελικού γράφου είναι η εξάλειψη τέτοιων κύκλων από τον αρχικό γράφο εγκλεισμού. Ας δούμε όμως πώς μπορούμε να αντιμετωπίσουμε το πρόβλημα αυτό.

Αρχικά, ας θεωρήσουμε σαν G=(V, E) το γράφο εγκλεισμού που έχουμε κατασκευάσει μέχρι στιγμής, και σαν G'=(V', E') τον υπογράφο του G με ακμές μόνο εκείνες που αντιστοιχούν σε ΕΕ αντιπροσωπευτικού κλειδιού, δηλαδή

V' = V,

E' = { (R, S, (a, b)) | (R, S, (a, b)) $\in$ E ΚΑΙ a αν.κ της R
ΚΑΙ b αν.κ της S}

Κατόπιν, ας ορίσουμε μια διμελή σχέση ισοδυναμίας ($\equiv$) στο V' ως εξής :

R $\equiv$ S $\Leftrightarrow$ στον G' υπάρχει κατευθυνόμενο μονοπάτι από την R στην S ΚΑΙ
στον G' υπάρχει κατευθυνόμενο μονοπάτι από την S στην R

Η διμελής αυτή σχέση ισοδυναμίας διαμερίζει το V' σε κλάσεις ισοδυναμίας. Αν V_i' μια τέτοια κλάση ισοδυναμίας και E_i' το σύνολο των ακμών του G' που ενώνουν τις κορυφές του V_i', τότε ο υπογράφος G_i'=(V_i', E_i') του καλείται *συνεκτική συνιστώσα* του G'. Οι κορυφές που βρίσκονται στην ίδια κλάση ισοδυναμίας, είναι προφανές ότι αντιστοιχούν σε σχήματα σχέσεων που έχουν πληροφορία για την ίδια οντότητα, μια και για οποιεσδήποτε δύο από

αυτές υπάρχει κύκλο που τις συνδέει. Κάθε κλάση ισοδυναμίας δηλαδή αντιστοιχεί σε ένα σχήμα σχέσης που έχει κατατμηθεί κάθετα σε τμήματα που αντιστοιχούν στις κορυφές (σχήματα σχέσεων) της κλάσης. Στόχος μας λοιπόν για την εξάλειψη του πλεονασμού, είναι η εύρεση όλων των κλάσεων ισοδυναμίας του V', ή αλλιώς όλων των συνεκτικών συνιστωσών του G'. Το πρόβλημα αυτό είναι γνωστό στη βιβλιογραφία και αλγόριθμος για την επίλυση του υπάρχει στο [2].

Μ' αυτό τον τρόπο μπορώ να φτιάξω ένα *μειωμένο* γράφο G''=(V'', E''), στον οποίο υπάρχει μία κορυφή για κάθε κλάση ισοδυναμίας του V'. Κάθε κορυφή του μειωμένου γράφου αντιστοιχεί σε ένα σχήμα-σχέσης με αν.κ, το αν.κ όλων των σχημάτων-σχέσεων της κλάσης ισοδυναμίας. Στα γνωρίσματα του αν.κ του νέου σχήματος σχέσης δίνονται τα πιο κατάλληλα κατά το χρήστη ονόματα, ενώ τα ονόματα των υπολοίπων γνωρισμάτων του, παραμένουν όπως παλιά. Για τα τελευταία, σε περιπτώσεις που μπορούν να εμφανιστούν ομώνυμα, τα ονόματά τους προθηματίζονται με το όνομα του σχήματος-σχέσης από το οποίο προήλθαν. Το σύνολο E'', των ακμών του μειωμένου γράφου σχηματίζεται ως εξής:

E'' = $\{(V_i, V_j, (a, b)) \mid V_i \in V''$ ΚΑΙ $V_j \in V''$ ΚΑΙ
υπάρχει ακμή (R, S, (a, b)) ∈ G με R $\in V_i$ ΚΑΙ S $\in V_j$

Στο προηγούμενο παράδειγμα, μετά την κατασκευή του μειωμένου γράφου εγκλεισμού το σχήμα έχει γίνει :

ΜαθητήςΑρχείαΜαθητών(α.μ., διεύθυνση, μέσοςόρος, επιβλέπωνκαθηγητής)
Καθηγητής(όνομα, διεύθυνση)

με τις εξής ΕΕ βασισμένες σε κλειδιά

ΜαθητήςΑρχείαΜαθητών(επιβλέπωνκαθηγητής) $\subseteq$ *Καθηγητής(όνομα)*

Ο δεύτερος πλεονασμός που θα αφαιρέσουμε από το μειωμένο γράφο στον οποίο έχουμε καταλήξει, είναι ο πλεονασμός ακμών του που αντιστοιχούν σε ΕΕ αντιπροσωπευτικού κλειδιού και οι οποίες μπορούν να εξαχθούν από άλλες, βάσει της μεταβατικότητας της σχέσης εγκλεισμού. Αν για παράδειγμα στο μειωμένο γράφο υπάρχουν οι ακμές $(V_1, V_2, (a, b))$, $(V_2, V_3, (b, c))$ και $(V_1, V_3, (a, c))$, με a, b, c αντιπροσωπευτικά κλειδιά των V_1, V_2, V_3 αντίστοιχα, είναι προφανές ότι η ακμή $(V_1, V_3, (a, c))$ είναι περιττή, μια και αντιστοιχεί σε μια ΕΕ που μπορεί, λόγω της μεταβατικότητας, να εξαχθεί από τις ΕΕ που αντιστοιχούν στις δύο πρώτες ακμές. Το πρόβλημα αυτό στη βιβλιογραφία ([2]) αναφέρεται ως το πρόβλημα εύρεσης της *μεταβατικής μείωσης* ενός γράφου και ορίζεται ως εξής :

> Η μεταβατική μείωση ενός κατευθυνόμενου γράφου G=(V, E) ορίζεται να είναι ένας γράφος G'=(V, E') με κατά το δυνατό λιγότερες ακμές, έτσι ώστε η

μεταβατική κλειστότητα του G' να ισούται με τη μεταβατική κλειστότητα του G.

Αν λοιπόν ο G_k=(V'', E_k) υπογράφος του G'', με E_k τις ακμές του G'' που αντιστοιχούν σε ΕΕ αντιπροσωπευτικού κλειδιού, αυτό που μας ενδιαφέρει, είναι η εύρεση της μεταβατικής μείωσης G'_k=(V'', E'_k) του G_k. Για το πρόβλημα αυτό, στο [22] υπάρχει αλγόριθμος που το επιλύει. Ο τελικός γράφος $G_f = (V_f, E_f)$, θα παρθεί αν από τον G'' αφαιρέσουμε τις ακμές E_k και προσθέσθουμε τις ακμές E'_k της μεταβατικής μείωσης του G_k, δηλαδή $G_f = (V_f, E_f)$, με V_f = V'', και E_f = E'' $\setminus E_k \cup E'_k$.

4.2.4 Φάση Δημιουργίας Σημασιολογικού Δικτύου

Πριν προχωρήσουμε στην ανάλυση της φάσης δημιουργίας του σημασιολογικού δικτύου, τη φάση δηλαδή της μετάφρασης του τελικού σχεσιακού σχήματος, ας συνοψίσουμε λίγο αυτά που έχουμε κάνει μέχρι στιγμής. Αρχικά, θεωρήσαμε σα δεδομένο ένα σχεσιακό σχήμα $RS_{src} = (\mathcal{R}_{src}, \mathcal{F}_{src}, \mathcal{I}_{src})$, με σύνολο σχημάτων σχέσεων $\mathcal{R}_{src}$, σύνολο συναρτησιακών εξαρτήσεων $\mathcal{F}_{src}$, και σύνολο εξαρτήσεων εγκλεισμού $\mathcal{I}_{src}$, στο οποίο αντιστοιχεί ένας αρχικός γράφος εγκλεισμού $G_{src} = (V_{src}, E_{src})$. Με τη ανάλυση των εξαρτήσεων εγκλεισμού και τη διαδικασία ανακάλυψης λανθανουσών οντοτήτων, προστέθηκαν νέα (εικονικά) σχήματα σχέσεων στο αρχικό σχήμα που προέρχονταν από προβολές σε σχήματα σχέσεων του $\mathcal{R}_{src}$, ενώ ταυτόχρονα μεταφέρθηκαν και προστέθηκαν νέες εξαρτήσεις εγκλεισμού στο $\mathcal{I}_{src}$. Με τον τρόπο αυτό πήραμε ένα ενδιάμεσο σχεσιακό σχήμα $RS_{int} = (\mathcal{R}_{int}, \mathcal{F}_{int}, \mathcal{I}_{int})$, στο οποίο αντιστοιχεί ένας ενδιάμεσος γράφος εγκλεισμού $G_{int} = (V_{int}, E_{int})$. Τελικά μετά τη φάση αφαίρεσης των πλεονασμών, η οποία περιείχε συγχωνεύσεις σχημάτων σχέσεων του $\mathcal{R}_{int}$ προς νέα (εικονικά) σχήματα σχέσεων και αφαιρέσεις περιττών εξαρτήσεων εγκλεισμού από το $\mathcal{I}_{int}$, καταλήξαμε στο τελικό σχεσιακό σχήμα $RS_{fin} = (\mathcal{R}_{fin}, \mathcal{F}_{fin}, \mathcal{I}_{fin})$, και ένα τελικό γράφο εγκλεισμού $G_{fin} = (V_{fin}, E_{fin})$. Από το τελικό αυτό σχήμα και με τη βοήθεια μεταφραστικών κανόνων, θα πάρουμε το σημασιολογικό δίκτυο που θα αντιστοιχεί στο αρχικό σχεσιακό σχήμα. Στο σημασιολογικό αυτό δίκτυο θα υπάρχουν κλάσεις που θα ενώνονται μεταξύ τους με σχέσεις γενίκευσης, κατ' αναλογία με τα σχήματα σχέσεων του σχεσιακού σχήματος που συνδέονται μεταξύ τους με εξαρτήσεις εγκλεισμού αντιπροσωπευτικών κλειδιών. Επίσης, κλάσεις του σημασιολογικού δικτύου θα συνδέονται με σχέσεις γνωρίσματος, όπως τα σχήματα σχέσεων στο σχεσιακό σχήμα συνδέονται με εξαρτήσεις (εγκλεισμού) ξένων κλειδιών. Τέλος, κλάσεις του σημασιολογικού δικτύου θα έχουν γνωρίσματα που θα παίρνουν τιμές συγκεκριμένου τύπου, όπως τα σχήματα σχέσεων του σχεσιακού σχήματος

έχουν γνωρίσματα που παίρνουν τιμές συγκεκριμένων τύπων. Για τεχνικούς λόγους, δεν θα χρησιμοποιηθούν οι κλάσεις συστήματος *Telos_Integer, Telos_String* και *Telos_Real*, για να εκφράσουν τους τύπους τιμών των γνωρισμάτων αυτών, αλλά θα ορίσουμε νέες κλάσεις που θα αντιστοιχούν στους τύπους τιμών που χρησιμοποιεί το σχεσιακό μοντέλο.

Ας περάσουμε όμως τώρα στην ανάλυση της φάσης μετάφρασης (κατασκευής σημασιολογικού δικτύου). Η φάση αποτελείται από τρία βήματα σε καθένα από τα οποία εφαρμόζονται ορισμένοι μεταφραστικοί κανόνες. Τα βήματα αυτά είναι τα εξής:

1. Επεξεργασία των σχημάτων σχέσεων $\mathcal{R}_{fin}$ του τελικού σχήματος RS_{fin}.
2. Επεξεργασία των εξαρτήσεων εγκλεισμού $\mathcal{I}_{fin}$ του τελικού σχήματος RS_{fin}.
3. Επεξεργασία των γνωρισμάτων των σχημάτων σχέσεων του RS_{fin}, τα οποία είτε συμμετέχουν μόνο στο δεξί μέλος εξαρτήσεων εγκλεισμού του $\mathcal{I}_{fin}$, είτε δε συμμετέχουν καθόλου σε εξαρτήσεις εγκλεισμού του $\mathcal{I}_{fin}$.

Στο πρώτο βήμα εφαρμόζουμε τον εξής μεταφραστικό κανόνα :

Μεταφραστικός Κανόνας 1:
Για κάθε σχήμα σχέσης του $\mathcal{R}_{fin}$, δημιούργησε μια απλή κλάση (S_Class) στο σημασιολογικό δίκτυο, με λογικό όνομα που παράγεται από το όνομα του σχήματος σχέσης.

Στο δεύτερο βήμα, ο μεταφραστικός κανόνας θα καθορίσει τις σχέσεις (γενίκευσης και γνωρίσματος) μεταξύ των απλών κλάσεων του σημασιολογικού δικτύου που έχουν δημιουργηθεί στο προηγούμενο βήμα, βασιζόμενος στο είδος των εξαρτήσεων εγκλεισμού του $\mathcal{I}_{fin}$. Δοθέντος ότι οι εξαρτήσεις εγκλεισμού του $\mathcal{I}_{fin}$ είναι όλες βασισμένες σε (αντιπροσωπευτικό) κλειδί, μπορούν να ταξινομηθούν σε μια από τις περιπτώσεις της πρώτης στήλης, του πίνακα εγκλεισμών (Πίνακας 1). Ο μεταφραστικός κανόνας λοιπόν, τίθεται ως εξής:

Μεταφραστικός Κανόνας 2:
Θεωρώντας μια εξάρτηση εγκλεισμού $R(a) \subseteq S(b)$ του $\mathcal{I}_{fin}$, διέκρινε σε ποια περίπτωση του Πίνακα 1 βρίσκεται και πράξε ανάλογα.

- *Περιπτώσεις K-K ή OK-K*
  Αν το a είναι αντιπροσωπευτικό κλειδί στο R, τότε δημιούργησε μια σχέση εξειδίκευσης μεταξύ των απλών κλάσεων R και S του σημασιολογικού δικτύου, έτσι ώστε η κλάση S να είναι γενίκευση της κλάσης R.

Αν το a είναι μη αντιπροσωπευτικό κλειδί ή όχι κλειδί για το R, τότε δημιούργησε μια σχέση γνωρίσματος από την απλή κλάση R προς την απλή κλάση S. Το λογικό όνομα της σχέσης γνωρίσματος μπορεί να είναι το όνομα του γνωρίσματος a, αν το a είναι μονοσύνολο, ή η συγχώνευση των ονομάτων των γνωρισμάτων του a, αν αυτό έχει πλέον του ενός στοιχεία.

- *Περιπτώσεις κ.τ.ΣΚ-Κ ή μ.τ.ΣΚ-Κ*
  Στις περιπτώσεις αυτές, τα τμήματα[2] $a_1, ..., a_m$ του a που είναι σύνθετο κλειδί του σχήματος σχέσης R, συμμετέχουν στο αριστερό μέλος εξαρτήσεων εγκλεισμού με αντιπροσωπευτικά κλειδιά $b_1, ..., b_m$ άλλων σχημάτων σχέσεων $S_1, ..., S_m$. Τότε, για κάθε εξάρτηση εγκλεισμού $R(a_i) \subseteq S_i(b_i)$, δημιούργησε μια σχέση γνωρίσματος από την απλή κλάση R προς την απλή κλάση S_i[3]. Το λογικό όνομα της σχέσης γνωρίσματος μπορεί να είναι το όνομα του γνωρίσματος a_i, αν το a_i είναι μονοσύνολο, ή η συγχώνευση των ονομάτων των γνωρισμάτων του a_i, αν αυτό έχει πλέον του ενός στοιχεία.

Στο τελευταίο βήμα, πριν εφαρμόσουμε τον τρίτο μεταφραστικό κανόνα δημιουργούμε απλές κλάσεις (S_Classes) στο σημασιολογικό δίκτυο που αντιστοιχούν στους τύπους τιμών των γνωρισμάτων του σχεσιακού μοντέλου δεδομένων, στο οποίο είναι εκφρασμένο το αρχικό σχεσιακό σχήμα. Ας αναφερόμαστε στο εξής στις κλάσεις αυτές, με τον όρο *ψευδοτύποι*. Ο τρίτος μεταφραστικός κανόνας υπαγορεύει τα εξής:

Μεταφραστικός Κανόνας 3:
Για κάθε σχήμα σχέσης R του $\mathcal{R}_{fin}$, αν a ένα γνώρισμα του R για το οποίο ισχύει ότι:
συμμετέχει μόνο στο δεξί μέλος εξαρτήσεων εγκλεισμού του $\mathcal{I}_{fin}$,
ή
δε συμμετέχει σε καμμία εξάρτηση εγκλεισμού του $\mathcal{I}_{fin}$,
τότε δημιούργησε μια σχέση γνωρίσματος με λογικό όνομα το όνομα του γνωρίσματος a, από την κλάση R προς τον ψευδοτύπο που αντιστοιχεί στον τύπο του a.

[2]Για τις περιπτώσεις κ.τ.ΣΚ-Κ τα $a_1, ..., a_m$ αποτελούν διαμέριση του a.

[3]Στην περίπτωση κ.τ.ΣΚ-Κ, η κλάση R αντιπροσωπεύει μια συσχέτιση, και οι σχέσεις γνωρίσματος τους ρόλους των κλάσεων $S_1, ..., S_m$. Στην περίπτωση μ.τ.ΣΚ-Κ, η κλάση R αντιπροσωπεύει μια ασθενή οντότητα, και οι σχέσεις γνωρίσματος τις σχέσεις ταυτότητας με τις ισχυρές οντότητες $S_1, ..., S_m$.

Ας δούμε μια παρατήρηση για τους μεταφραστικούς κανόνες 2 και 3. Ας θεωρήσουμε την εξής περίπτωση: Μια απλή κλάση *S* είναι υποκλάση μιας απλής κλάσης *R* η οποία έχει ένα γνώρισμα *a* με λογικό όνομα *name_a* προς μια κλάση (απλή ή ψευδοτύπο) R_a. Αν ο μεταφραστικός κανόνας προσπαθήσει να δημιουργήσει μια σχέση γνωρίσματος με λογικό όνομα *name_a* μεταξύ της απλής κλάσης *S* και μιας κλάσης (απλής ή ψευδοτύπου) S_a, η κλάση S_a θα πρέπει να είναι υποκλάση της R_a. Αλλιώς, η δημιουργηθείσα σχέση γνωρίσματος θα προθέσει στο λογικό της όνομα το λογικό όνομα της κλάσης *S*. Ο περιορισμός αυτός τίθεται έτσι ώστε τα σημασιολογικά δίκτυα που παράγονται να είναι σύμφωνα με τους κανόνες ορθότητας της γλώσσας Telos.

Τέλος, ας παρατηρήσουμε ότι από τον τρόπο κατασκευής του σημασιολογικού δικτύου, είναι φανερό ότι αυτό είναι "επίπεδο", υπό την έννοια ότι περιέχει μόνο απλές κλάσεις οντοτήτων, που σχετίζονται μεταξύ τους με σχέσεις γενίκευσης και γνωρίσματος.

4.2.5 Σχέσεις Αντιστοίχισης

Σύμφωνα, με το [27], η *μετατροπή βάσης δεδομένων* είναι ένα σύνολο από αντιστοιχίσεις των περιπτώσεων ενός σχήματος πηγής, σε περιπτώσεις ενός σχήματος στόχου. Υπάρχουν δύο προσεγγίσεις για την υλοποίηση, των μετατροπών ΒΔ. Σύμφωνα με την πρώτη, τα δεδομένα της ΒΔ πηγής, μετατρέπονται στη μορφή των δεδομένων της ΒΔ στόχου, και αποθηκεύονται εκεί. Μια τέτοια προσέγγιση, μπορεί να χαρακτηριστεί σαν *άπαξ μαζική μετατροπή*. Στη δεύτερη προσέγγιση, τα δεδομένα της ΒΔ πηγής δε μεταφέρονται στη ΒΔ στόχου, αλλά αυτό που γίνεται είναι επερωτήσεις προς το σχήμα της ΒΔ στόχου να μεταφράζονται σε επερωτήσεις προς το σχήμα της ΒΔ πηγής. Σε μια τέτοια προσέγγιση γίνεται μια *μετατροπή κατά παραγγελία*.

Στην πρώτη περίπτωση, ανήκει η δουλειά του [20] όπου αναπτύσσεται ένα σύστημα μεταφοράς δεδομένων από μια σχεσιακή προς μια οντοκεντρική βάση δεδομένων. Στην εργασία αυτή, ορίζεται ένα σύνολο *τελεστών απεικόνισης* οι οποίοι καθορίζουν τον τρόπο απεικόνισης στην οντοκεντρική βάση των δεδομένων που μεταφέρονται από τη σχεσιακή.

Στην περίπτωσή μας, που τα δεδομένα δε μεταφέρονται από τη ΒΔ πηγής (σχεσιακή ΒΔ) προς τη ΒΔ στόχου (σημασιολογικό δίκτυο), αυτό που χρειαζόμαστε είναι μια προσέγγιση που θα κάνει μετατροπή κατά παραγγελία στα δεδομένα της ΒΔ πηγής. Για το λόγο αυτό, ορίζουμε ένα σύνολο σχέσεων αντιστοίχισης, οι οποίες δηλώνουν τις αντιστοιχίσεις μεταξύ των δομών (σχήματα-σχέσεων, εξαρτήσεις εγκλεισμού, γνωρίσματα) του σχεσιακού σχήματος πηγής με δομές (απλές κλάσεις, σχέσεις γενίκευσης, σχέσεις γνωρίσματος) του σημασιολογικού σχήματος στόχου. Οι σχέσεις αντιστοίχισης

αποκαθίστανται κατά τη διάρκεια της φάσης δημιουργίας του σημασιολογικού δικτύου, ταυτόχρονα με την εφαρμογή των μεταφραστικών κανόνων.

- Την πρώτη σχέση αντιστοίχισης, την ονομάσαμε **BECOMES_CLASS**, και δηλώνει αντιστοίχιση ενός σχήματος-σχέσης R του σχεσιακού σχήματος, με μια απλή C του σημασιολογικού δικτύου. Ο συμβολισμός της σχέσης αυτής είναι:

  RELATION-SCHEME *R* **BECOMES_CLASS** S_CLASS *C*

  Σημασιολογικά, η σχέση αυτή υπαγορεύει ότι πλειάδες της σχέσης r (που αντιστοιχεί στο σχήμα-σχέσης R), μετατρέπονται σε ατομικές οντότητες της απλής κλάσης C. Οι σχέσεις αντιστοίχισης αυτού του είδους αποκαθίστανται με την εφαρμογή του πρώτου μεταφραστικού κανόνα.

- Τη δεύτερη σχέση αντιστοίχισης, την ονομάσαμε **BECOMES_SPEC_REL**, και δηλώνει την αντιστοίχιση μιας ΕΕ κλειδιού $R(a) \subseteq S(b)$ του σχεσιακού σχήματος, με μια σχέση εξειδίκευσης μεταξύ των απλών κλάσεων C1 και C2 του σημασιολογικού δικτύου. Προϋπόθεση για να δηλωθεί μια τέτοια σχέση αντιστοίχισης, είναι να έχει δηλωθεί ότι
  RELATION-SCHEME *R* **BECOMES_CLASS** S_CLASS *C1*
  και
  RELATION-SCHEME *S* **BECOMES_CLASS** S_CLASS *C2*
  Ο συμβολισμός της σχέσης αυτής είναι :

  IND *R(a) - S(b)* **BECOMES_SPEC_REL** BETWEEN S_CLASSES *C1* AND *C2*

  Σημασιολογικά, η σχέση αυτή υπαγορεύει ότι η σχέση υποσυνόλου μεταξύ των πλειάδων της $\pi_a(r)$ (r η σχέση που αντιστοιχεί στο σχήμα-σχέσης R) και των πλειάδων της $\pi_b(s)$ (s η σχέση που αντιστοιχεί στο σχήμα-σχέσης S), μετατρέπεται σε σχέση υποσυνόλου μεταξύ του συνόλου των ατομικών οντοτήτων της απλής κλάσης C1 και του συνόλου των ατομικών οντοτήτων της απλής κλάσης C2. Οι σχέσεις αντιστοίχισης αυτού του είδους, αποκαθίστανται με την εφαρμογή του δεύτερου μεταφραστικού κανόνα σε περιπτώσεις ΕΕ της μορφής αν.Κ-Κ.

- Την τρίτη σχέση αντιστοίχισης, την ονομάσαμε **BECOMES_ATTR_REL**, και δηλώνει την αντιστοίχιση μιας ΕΕ βασισμένης σε κλειδί $R(a) \subseteq S(b)$ του σχεσιακού σχήματος, με μια σχέση γνωρίσματος a μεταξύ των απλών κλάσεων C1 και C2 του σημασιολογικού δικτύου. Προϋπόθεση για να δηλωθεί μια τέτοια σχέση αντιστοίχισης, είναι να έχει δηλωθεί ότι

RELATION-SCHEME *R* **BECOMES_CLASS** S_CLASS *C1*

και

RELATION-SCHEME *S* **BECOMES_CLASS** S_CLASS *C2*

Ο συμβολισμός της σχέσης αυτής είναι :

IND *R(a)* - *S(b)* **BECOMES_ATTR_REL** a BETWEEN S_CLASSES *C1* AND *C2*

Σημασιολογικά, η σχέση αυτή υπαγορεύει ότι πλειάδες της $\pi_a(r)$ (r η σχέση που αντιστοιχεί στο σχήμα-σχέσης R), μετατρέπονται σε ατομικές οντότητες του γνωρίσματος a, που συνδέουν ατομικές οντότητες των απλών κλάσεων C1 και C2. Οι σχέσεις αντιστοίχισης αυτού του είδους, αποκαθίστανται με την εφαρμογή του δεύτερου μεταφραστικού κανόνα σε περιπτώσεις ΕΕ της μορφής μ.αν.Κ-Κ και ΟΚ-Κ.

- Τέλος, την τέταρτη σχέση αντιστοίχισης, την ονομάσαμε **BECOMES_ATTR**, και δηλώνει την αντιστοίχιση ενός γνωρίσματος Α ενός σχήματος-σχέσης R με τύπο τιμών Τ, με ένα γνώρισμα a μιας απλής κλάσης C προς τον αντίστοιχο ψευδοτύπο Τ'. Προϋπόθεση για να δηλωθεί μια τέτοια σχέση αντιστοίχισης είναι να έχει δηλωθεί ότι

  RELATION-SCHEME *R* **BECOMES_CLASS** S_CLASS *C*

  Ο συμβολισμός της σχέσης αυτής είναι :

  ATTRIBUTE *A* OF RELATION-SCHEME *R* BECOMES_ATTR *a* OF S_CLASS *C*

  Σημασιολογικά, η σχέση αυτή υπαγορεύει ότι τιμές τύπου Τ του γνωρίσματος Α σε μια πλειάδα της σχέσης r (που αντιστοιχεί στο σχήμα-σχέσης R), μετατρέπονται σε ατομικές οντότητες του γνωρίσματος a, που συνδέουν ατομικές οντότητες της απλής κλάσης C με τιμές του ψευδοτύπου Τ'. Οι σχέσεις αντιστοίχισης αυτού του είδους αποκαθίστανται με την εφαρμογή του τρίτου μεταφραστικού κανόνα.

4.3 Παράδειγμα Εφαρμογής της Μεθόδου Μετάφρασης

Στις προηγούμενες ενότητες, αναλύσαμε τη διαδικασία μετάφρασης και τις σχέσεις αντιστοίχισης που αποκαθίστανται κατά την τελευταία φάση αυτής της διαδικασίας. Ας περάσουμε τώρα να δούμε ένα ολοκληρωμένο παράδειγμα μετάφρασης ενός σχεσιακού σχήματος.

Ας θεωρήσουμε το πιο κάτω "αρχικό" σχεσιακό σχήμα :

Στ_ταυτ_ατόμου(α.τ, όνομα)

Αλλα_στ_ατόμου(α.τ, πλήθος_αυτοκ, τηλ, λέσχη)

Υπάλληλος(α.τ, είδος_εργ, τμήμα)
Αρχεία_υπάλ(α.τ, μισθός, ημερομ_προσλ)
Παιδιά(υπάλ, όνομα, ημερομ_γεν, φύλο)
Τμήμα(όνομα, διευθυντής, προϋπολογισμός)
Αυτοκίνητο(αρ.κυκλ, μοντέλο, ιδιοκτήτης)
Λέσχη(όνομα, τόπος_στέγ, ημερομ_ίδρ)
Ανάθεση_έργου(υπάλ, έργο, ώρες, ημερομ)

Με το εξής ”αρχικό” σύνολο εξαρτήσεων εγκλεισμού :

Στ_ταυτ_ατόμου(α.τ) ⊆ Αλλα_στ_ατόμου(α.τ)
Αλλα_στ_ατόμου(α.τ) ⊆ Στ_ταυτ_ατόμου(α.τ)
Αλλα_στ_ατόμου(λέσχη) ⊆ Λέσχη(όνομα)
Υπάλληλος(α.τ) ⊆ Αρχεία_υπάλ(α.τ)
Αρχεία_υπάλ(α.τ) ⊆ Υπάλληλος(α.τ)
Υπάλληλος(α.τ) ⊆ Στ_ταυτ_ατόμου(α.τ)
Υπάλληλος(τμήμα) ⊆ Τμήμα(όνομα)
Παιδιά(υπάλ) ⊆ Υπάλληλος(α.τ)
Τμήμα(διευθυντής) ⊆ Υπάλληλος(α.τ) *Τέταρτη Μορφή ΕΕ*
Αυτοκίνητο(ιδιοκτήτης) ⊆ Αλλα_στ_ατόμου(α.τ)
Ανάθεση_έργου(υπάλ) ⊆ Υπάλληλος(α.τ)

Μετά τη φάση ανακάλυψης λανθανουσών οντοτήτων, στο ”ενδιάμεσο” σχήμα έχουν προστεθεί οι εξής εικονικές σχέσεις :

Διευθυντής(διευθυντής_όνομα)
Εργο(έργο_όνομα)

ενώ στο ”ενδιάμεσο” σύνολο εξαρτήσεων εγκλεισμού έχει μετατραπεί η ΕΕ

Τμήμα(διευθυντής) ⊆ Υπάλληλος(α.τ) **σε**
Διευθυντής(διευθυντής_όνομα) ⊆ Υπάλληλος(α.τ)

και έχουν προστεθεί οι ΕΕ :

Τμήμα(διευθυντής) ⊆ Διευθυντής(διευθυντής_όνομα)
Διευθυντής(διευθυντής_όνομα) ⊆ Τμήμα(διευθυντής)
Ανάθεση_έργου(έργο) ⊆ Εργο(έργο_όνομα)
Εργο(έργο_όνομα) ⊆ Ανάθεση_έργου(έργο)

Στο σημείο αυτό μπορούμε να αφαιρέσουμε από το ”ενδιάμεσο” σύνολο ΕΕ, τις ΕΕ

Διευθυντής(διευθυντής_όνομα) ⊆ Υπάλληλος(α.τ)

και

Εργο(έργο_όνομα) ⊆ Ανάθεση_έργου(έργο)

μια και όπως έχουμε προαναφέρει, δεν πρόκειται να παίξουν ρόλο στην περαιτέρω επεξεργασία του σχήματος. Προχωρούμε λοιπόν στην τελική φάση αφαίρεσης των πλεονασμών, και παίρνουμε το ”τελικό” σχήμα :

Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(α.τ, όνομα, πλήθος_αυτοκ, τηλ, λέσχη)

ΥπάλληλοςΑρχεία_υπάλ(α.τ, είδος_εργ, τμήμα, μισθός, ημερομ_προσλ)

Παιδιά(υπάλ, όνομα, ημερομ_γεν, φύλο)

Τμήμα(όνομα, διευθυντής, προϋπολογισμός)

Αυτοκίνητο(αρ.κυκλ, μοντέλο, ιδιοκτήτης)

Λέσχη(όνομα, τόπος_στέγ, ημερομ_ίδρ)

Ανάθεση_έργου(υπάλ, έργο, ώρες, ημερομ)

Διευθυντής(διευθυντής_όνομα)

Εργο(έργο_όνομα)

και το ”τελικό” σύνολο εξαρτήσεων εγκλεισμού :

Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(λέσχη) ⊆ Λέσχη(όνομα) *ΟΚ-αν.Κ*

ΥπάλληλοςΑρχεία_υπάλ(α.τ) ⊆ Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(α.τ) *αν.Κ-αν.Κ*

ΥπάλληλοςΑρχεία_υπάλ(τμήμα) ⊆ Τμήμα(όνομα) *ΟΚ-αν.Κ*

Παιδιά(υπάλ) ⊆ ΥπάλληλοςΑρχεία_υπάλ(α.τ) *μ.τ.ΣΚ-αν.Κ*

Αυτοκίνητο(ιδιοκτήτης) ⊆ Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(α.τ) *ΟΚ-αν.Κ*

Διευθυντής(διευθυντής_όνομα) ⊆ ΥπάλληλοςΑρχεία_υπάλ(α.τ) *αν.Κ-αν.Κ*

Τμήμα(διευθυντής) ⊆ Διευθυντής(διευθυντής_όνομα) *μ.αν.Κ-αν.Κ*

Ανάθεση_έργου(υπάλ) ⊆ ΥπάλληλοςΑρχεία_υπάλ(α.τ) *κ.τ.ΣΚ-αν.Κ*

Ανάθεση_έργου(έργο) ⊆ Εργο(έργο_όνομα) *κ.τ.ΣΚ-αν.Κ*

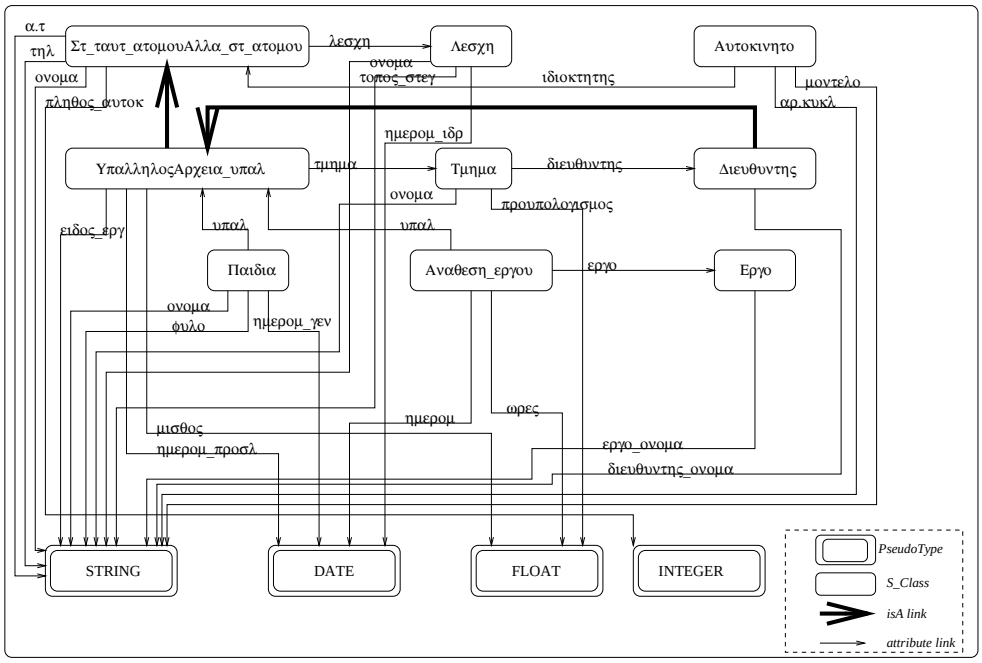
Από το τελικό αυτό σχήμα και το τελικό σύνολο ΕΕ με εφαρμογή των μεταφραστικών κανόνων, κατασκευάζεται το σημασιολογικό δίκτυο που φαίνεται στο Σχήμα 4.1, και δηλώνονται με τη σειρά τις παρακάτω σχέσεις αντιστοίχισης.

(Από την εφαρμογή του **πρώτου** μεταφραστικού κανόνα)

RELATION-SCHEME *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* **BECOMES_CLASS** S_CLASS *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*

RELATION-SCHEME *ΥπάλληλοςΑρχεία_υπάλ* **BECOMES_CLASS** S_CLASS *ΥπάλληλοςΑρχεία_υπάλ*

RELATION-SCHEME *Παιδιά* **BECOMES_CLASS** S_CLASS *Παιδιά*



Σχήμα 4.1: Τελικό Σημασιολογικό Δίκτυο

RELATION-SCHEME *Τμήμα* **BECOMES_CLASS** S_CLASS *Τμήμα*

RELATION-SCHEME *Αυτοκίνητο* **BECOMES_CLASS** S_CLASS *Αυτοκίνητο*

RELATION-SCHEME *Λέσχη* **BECOMES_CLASS** S_CLASS *Λέσχη*

RELATION-SCHEME *Ανάθεση_έργου* **BECOMES_CLASS** S_CLASS *Ανάθεση_έργου*

RELATION-SCHEME *Διευθυντής* **BECOMES_CLASS** S_CLASS *Διευθυντής*

RELATION-SCHEME *Εργο* **BECOMES_CLASS** S_CLASS *Εργο*

(Από την εφαρμογή του **δεύτερου** μεταφραστικού κανόνα)

IND *ΥπάλληλοςΑρχεία_υπάλ(α.τ)* - *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(α.τ)* **BECOMES_SPEC_REL** BETWEEN S_CLASSES *ΥπάλληλοςΑρχεία_υπάλ* AND *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*

IND *Διευθυντής(διευθυντής_όνομα)* - *ΥπάλληλοςΑρχεία_υπάλ(α.τ)* **BECOMES_SPEC_REL** BETWEEN S_CLASSES *Διευθυντής* AND *ΥπάλληλοςΑρχεία_υπάλ*

και

IND *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(λέσχη)* - *Λέσχη(όνομα)* **BECOMES_ATTR_REL** *λέσχη* BETWEEN S_CLASSES *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* AND *Λέσχη*

IND *ΥπάλληλοςΑρχεία_υπάλ(τμήμα)* - *Τμήμα(όνομα)* **BECOMES_ATTR_REL** *τμήμα* BETWEEN S_CLASSES *ΥπάλληλοςΑρχεία_υπάλ* AND *Τμήμα*

IND *Παιδιά(υπάλ)* - *ΥπάλληλοςΑρχεία_υπάλ(α.τ)* **BECOMES_ATTR_REL** *υπάλ* BETWEEN S_CLASSES *Παιδιά*

AND *ΥπάλληλοςΑρχεία_υπάλ*

IND *Αυτοκίνητο(ιδιοκτήτης) - Στ_ταυτ_ατόμουΑλλα_στ_ατόμου(α.τ)* **BECOMES_ATTR_REL** *ιδιοκτήτης* BETWEEN S_CLASSES *Αυτοκίνητο* AND *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*

IND *Τμήμα(διευθυντής) - Διευθυντής(διευθυντής_όνομα)* **BECOMES_ATTR_REL** *διευθυντής* BETWEEN S_CLASSES *Τμήμα* AND *Διευθυντής*

IND *Ανάθεση_έργου(υπάλ) - ΥπάλληλοςΑρχεία_υπάλ(α.τ)* **BECOMES_ATTR_REL** *υπάλ* BETWEEN S_CLASSES *Ανάθεση_έργου* AND *ΥπάλληλοςΑρχεία_υπάλ*

IND *Ανάθεση_έργου(έργο) - Εργο(έργο_όνομα)* **BECOMES_ATTR_REL** *έργο* BETWEEN S_CLASSES *Ανάθεση_έργου* AND *Εργο*

(Τέλος, από την εφαρμογή του **τρίτου** μεταφραστικού κανόνα)

ATTRIBUTE *α.τ* OF RELATION-SCHEME *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* **BECOMES_ATTR** *α.τ* OF S_CLASS *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*
ATTRIBUTE *όνομα* OF RELATION-SCHEME *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* **BECOMES_ATTR** *όνομα* OF S_CLASS *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*
ATTRIBUTE *πλήθος_αυτοκ* OF RELATION-SCHEME *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* **BECOMES_ATTR** *πλήθος_αυτοκ* OF S_CLASS *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*
ATTRIBUTE *τηλ* OF RELATION-SCHEME *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* **BECOMES_ATTR** *τηλ* OF S_CLASS *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου*

ATTRIBUTE *είδος_εργ* OF RELATION-SCHEME *ΥπάλληλοςΑρχεία_υπάλ* **BECOMES_ATTR** *είδος_εργ* OF S_CLASS *ΥπάλληλοςΑρχεία_υπάλ*
ATTRIBUTE *μισθός* OF RELATION-SCHEME *ΥπάλληλοςΑρχεία_υπάλ* **BECOMES_ATTR** *μισθός* OF S_CLASS *ΥπάλληλοςΑρχεία_υπάλ*
ATTRIBUTE *ημερομ_προσλ* OF RELATION-SCHEME *ΥπάλληλοςΑρχεία_υπάλ* **BECOMES_ATTR** *ημερομ_προσλ* OF S_CLASS *ΥπάλληλοςΑρχεία_υπάλ*

ATTRIBUTE *όνομα* OF RELATION-SCHEME *Παιδιά* **BECOMES_ATTR** *όνομα* OF S_CLASS *Παιδιά*
ATTRIBUTE *ημερομ_γεν* OF RELATION-SCHEME *Παιδιά* **BECOMES_ATTR** *ημερομ_γεν* OF S_CLASS *Παιδιά*
ATTRIBUTE *φύλο* OF RELATION-SCHEME *Παιδιά* **BECOMES_ATTR** *φύλο* OF S_CLASS *Παιδιά*

ATTRIBUTE *όνομα* OF RELATION-SCHEME *Τμήμα* **BECOMES_ATTR** *όνομα* OF S_CLASS *Τμήμα*
ATTRIBUTE *προϋπολογισμός* OF RELATION-SCHEME *Τμήμα* **BECOMES_ATTR** *προϋπολογισμός* OF S_CLASS *Τμήμα*

ATTRIBUTE *αρ.κυκλ* OF RELATION-SCHEME *Αυτοκίνητο* **BECOMES_ATTR** *αρ.κυκλ* OF S_CLASS *Αυτοκίνητο*
ATTRIBUTE *μοντέλο* OF RELATION-SCHEME *Αυτοκίνητο* **BECOMES_ATTR** *μοντέλο* OF S_CLASS *Αυτοκίνητο*

ATTRIBUTE *όνομα* OF RELATION-SCHEME *Λέσχη* **BECOMES_ATTR** *όνομα* OF S_CLASS *Λέσχη*

ATTRIBUTE *τόπος_στέγ* OF RELATION-SCHEME *Λέσχη* **BECOMES_ATTR** *τόπος_στέγ* OF S_CLASS *Λέσχη*
ATTRIBUTE *ημερομ_ίδρ* OF RELATION-SCHEME *Λέσχη* **BECOMES_ATTR** *ημερομ_ίδρ* OF S_CLASS *Λέσχη*

ATTRIBUTE *ώρες* OF RELATION-SCHEME *Ανάθεση_έργου* **BECOMES_ATTR** *ώρες* OF S_CLASS *Ανάθεση_έργου*
ATTRIBUTE *ημερομ* OF RELATION-SCHEME *Ανάθεση_έργου* **BECOMES_ATTR** *ημερομ* OF S_CLASS *Ανάθεση_έργου*

ATTRIBUTE *διευθυντής_όνομα* OF RELATION-SCHEME *Διευθυντής* **BECOMES_ATTR** *διευθυντής_όνομα* OF S_CLASS *Διευθυντής*

ATTRIBUTE *έργο_όνομα* OF RELATION-SCHEME *Εργο* **BECOMES_ATTR** *έργο_όνομα* OF S_CLASS *Εργο*

Κεφάλαιο 5

Ενοποίηση Σημασιολογικών Δικτύων

Στο κεφάλαιο αυτό θα παρουσιάσουμε τη μέθοδο που αναπτύξαμε για ενοποίηση σημασιολογικών δικτύων. Αρχικά, θα δούμε ορισμένες υποθέσεις για το είδος των σημασιολογικών δικτύων που μπορούν να ενοποιηθούν από τη μέθοδό μας. Κατόπιν, θα παρουσιάσουμε την αρχιτεκτονική της μεθόδου με βάση το πλαίσιο που είδαμε στο [4], και θα αναλύσουμε τα στάδιά της. Το κεφάλαιο θα κλείσει με ένα αναλυτικό παράδειγμα εφαρμογής της μεθόδου.

5.1 Υποθέσεις

Στο προηγούμενο κεφάλαιο είδαμε ότι τα σημασιολογικά δίκτυα που παράγονται από τη μέθοδο μετάφρασης, έχουν μια αρκετά καθορισμένη μορφή. Συγκεκριμένα, από τον τρόπο κατασκευής τους έχουν τις παρακάτω ιδιότητες :

1. Οι κλάσεις οντοτήτων που περιέχουν, είναι **όλες** απλές.

2. Δεν περιέχουν τις κλάσεις τιμών *Telos_String*, *Telos_Integer*, *Telos_Real*, αλλά στη θέση τους χρησιμοποιούν *ψευδοτύπους* (που είναι απλές κλάσεις οντοτήτων).

3. Οι απλές κλάσεις είναι οργανωμένες στη διάσταση των γνωρισμάτων κατά τρόπο ώστε, μια κλάση να **μην** έχει δύο διαφορετικά γνωρίσματα με το ίδιο λογικό όνομα.

4. Αν δύο απλές κλάσεις συνδέονται με σχέση εξειδίκευσης τότε, αφενός η εξειδικευμένη κλάση κληρονομεί τα γνωρίσματα της γενικευμένης, αφεταίρου αν οι κλάσεις έχουν γνωρίσματα με το ίδιο λογικό όνομα τα γνωρίσματα αυτά προσδιορίζουν (καταλήγουν σε) κλάσεις που συνδέονται με σχέση εξειδίκευσης.

Η υπόθεση λοιπόν που κάνουμε για τα οποιαδήποτε συστατικά σχήματα πρόκειται να ενοποιηθούν βάσει της μεθόδου που προτείνουμε, είναι η τήρηση των παραπάνω τεσσάρων απαιτήσεων.

Στο [26] περιγράφεται ένας αλγόριθμος συγχώνευσης σχημάτων ΒΔ, ο οποίος θεωρεί ότι τα προς συγχώνευση σχήματα, είναι **νόμιμα σχήματα** σύμφωνα με τον παρακάτω ορισμό:

> Ας θεωρήσουμε ένα σύνολο $\mathcal{N}$ **ονομάτων κλάσεων**, και ένα σύνολο $\mathcal{L}$ **ονομάτων γνωρισμάτων**. Τότε, ένα **σχήμα** πάνω από το $\mathcal{N}$ και το $\mathcal{L}$, είναι μια τριάδα συνόλων, $\mathcal{G} = (C, E, S)$, για τα οποία ισχύουν :
> $C \subseteq \mathcal{N}$,
> $E \subseteq C \times \mathcal{L} \times C$, και
> η S είναι μια σχέση μερικής διάταξης (μεταβατική, ανακλαστική, αντισυμμετρική) στο C.
>
> Αν $p, q \in C, a \in \mathcal{L}$ τ.ω. $(p, a, q) \in E$, τότε γράφουμε $p \xrightarrow{a}_G q$, ενώ αν $(p, q) \in S$, τότε γράφουμε $p \Longrightarrow q$.
>
> Τέλος, ένα σχήμα $\mathcal{G} = (\mathcal{C}, \mathcal{E}, \mathcal{S})$ πάνω από το $\mathcal{N}$ και το $\mathcal{L}$, λέγεται **νόμιμο σχήμα** αν ισχύουν οι εξής δύο ιδιότητες νομιμότητας:
>
> 1. $(p \xrightarrow{a}_G q \wedge p \xrightarrow{a}_G r)$ **συνεπάγεται** $r \equiv q$
> 2. $(p \Longrightarrow q \wedge q \xrightarrow{a}_G r)$ **συνεπάγεται** $(\exists s \in C \cdot p \xrightarrow{a}_G s \wedge s \Longrightarrow r)$

Παρατηρούμε λοιπόν ότι, αν αντί για τα σύνολα $\mathcal{N}$ και $\mathcal{L}$ του [26], θεωρήσουμε τα σύνολα των λογικών ονομάτων των απλών κλάσεων οντοτήτων και γνωρισμάτων της Telos, οι ιδιότητες που απαιτούμε να έχουν τα συστατικά μας σχήματα είναι ακριβώς αυτές που έχουν τα νόμιμα σχήματα του [26]. Με άλλα λόγια, αν θεωρήσουμε το συστατικό μας σχήμα (σημασιολογικό δίκτυο) σα μια τριάδα συνόλων $\mathcal{G} = (C, E, S)$, με
C = σύνολο λογικών ονομάτων απλών κλάσεων (εννοείται και ψευδοτύπων) του συστατικού σχήματος,
E = σύνολο ονοματισμένων γνωρισμάτων μεταξύ των απλών κλάσεων του συστατικού σχήματος, και
S = σύνολο των ζευγών απλών κλάσεων που συνδέονται με σχέση εξειδίκευσης στο συστατικό σχήμα
τότε το $\mathcal{G} = (C, E, S)$ είναι ένα νόμιμο σχήμα. Η παρατήρηση αυτή μας οδήγησε στην υιοθέτηση **αυτούσιου** του αλγορίθμου συγχώνευσης του [26], για το στάδιο συγχώνευσης της μεθόδου μας.

5.2 Μέθοδος Ενοποίησης

Στην εργασία [4] που παρουσιάσαμε στο Κεφάλαιο 3, είδαμε ένα πλαίσιο περιγραφής και ανάλυσης μεθόδων ενοποίησης σχημάτων. Στην ενότητα αυτή θα αναλύσουμε την αρχιτεκτονική της μεθόδου μας με βάση εκείνο το πλαίσιο.

Κατ' αρχήν, αν δούμε τη μέθοδό μας σαν ένα μαύρο κουτί στο οποίο εξετάζουμε τις εισόδους και την έξοδο του, τότε μπορούμε να πούμε ότι στην είσοδο του δέχεται ένα πλήθος N νόμιμων σχημάτων καθώς και δηλώσεις επίλυσης συγκρούσεων μεταξύ τους, ενώ στην έξοδο παράγει ένα ενοποιημένο νόμιμο σχήμα.

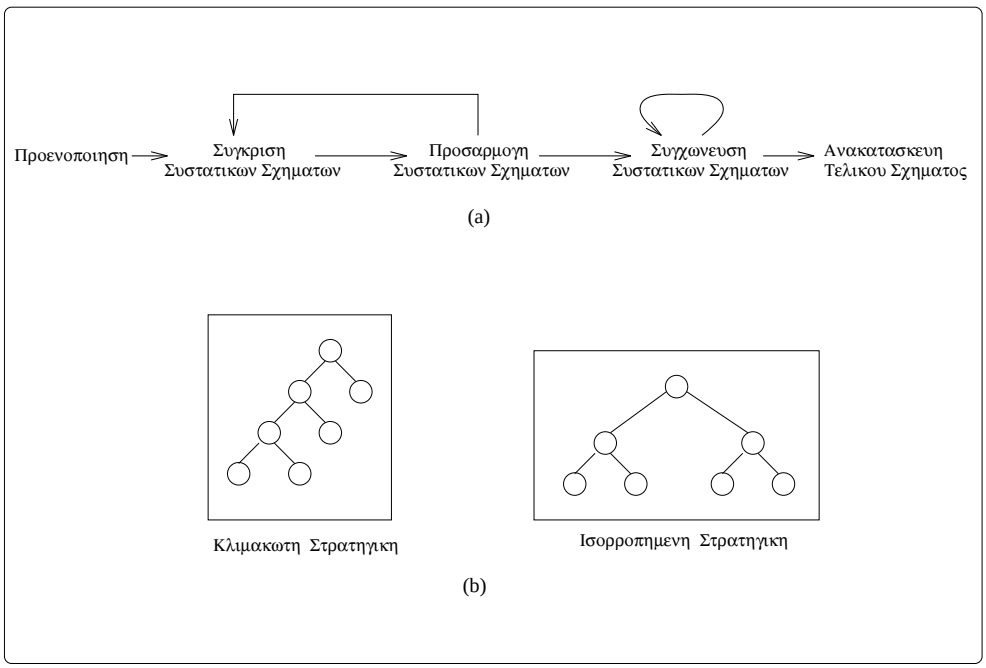
Αναλυτικότερα, η μέθοδος αποτελείται από πέντε στάδια. Σύμφωνα με την ορολογία του [4], είναι το *στάδιο προενοποίησης*, το *στάδιο σύγκρισης συστατικών σχημάτων*, το *στάδιο προσαρμογής συστατικών σχημάτων*, το *στάδιο συγχώνευσης συστατικών σχημάτων*, και το *στάδιο ανακατασκευής του τελικού ενοποιημένου σχήματος*. Η σειρά με την οποία λαμβάνουν χώρα τα πέντε στάδια φαίνεται στο Σχήμα 5.1(a). Στο σχήμα αυτό βλέπουμε ότι αφού περάσουμε το στάδιο της προενοποίησης, που υπαγορεύει ότι τα N συστατικά σχήματα θα ενοποιηθούν με μια δυαδική στρατηγική, κατόπιν εκτελούνται επαναληπτικά τα στάδια σύγκρισης και προσαρμογής μέχρις ότου ανακαλυφθούν και επιλυθούν όλες οι συγκρούσεις (ονομαστικές και δομικές) των συστατικών σχημάτων. Έπειτα, εφαρμόζεται επαναληπτικά ο αλγόριθμος συγχώνευσης όπως ορίζει η δυαδική στρατηγική και παίρνουμε ένα τελικό ενοποιημένο σχήμα που **δεν** είναι όμως νόμιμο. Για το λόγο αυτό, εφαρμόζεται ο αλγόριθμος ανακατασκευής στο σχήμα αυτό, ο οποίος δίνει και το τελικό **νόμιμο** ενοποιημένο σχήμα.

Ας περάσουμε όμως τώρα στην αναλυτική παρουσίαση των σταδίων της μεθόδου.

5.2.1 Στάδιο Προενοποίησης

Το πρώτο στάδιο της μεθόδου είναι το στάδιο της προενοποίησης. Στο στάδιο αυτό καθορίζεται η στρατηγική ενοποίησης των N συστατικών σχημάτων της εισόδου, καθώς και ο τρόπος επέμβασης του χρήστη στα επόμενα στάδια της μεθόδου.

Ξεκινώντας από το δεύτερο έχουμε να πούμε ότι επέμβαση του χρήστη απαιτείται μόνο στο δεύτερο και τρίτο στάδιο της μεθόδου, στα στάδια δηλαδή που ανακαλύπτονται και επιλύονται οι διάφορες συγκρούσεις μεταξύ των συστατικών σχημάτων. Κάτι τέτοιο είναι απόλυτα φυσιολογικό και αναμενόμενο, μια και ο σχεδιαστής του ενοποιημένου σχήματος είναι ο μόνος που γνωρίζει **επακριβώς** πώς εκφράζεται μια έννοια σε καθένα από τα συστατικά σχήματα. Είναι λοιπόν μάλλον μάταιο να προσπαθήσουμε να



Σχήμα 5.1: (a)Στάδια της Μεθόδου, (b)Τύποι Δυαδικών Στρατηγικών

αυτοματοποιήσουμε **πλήρως** τις δύο αυτές διαδικασίες. Όπως είδαμε και στο Κεφάλαιο 3, η καλύτερη λύση που έχει δοθεί ώστε να "βοηθείται" ο σχεδιαστής κατά τη διάρκεια των διαδικασιών αυτών προτείνεται στο [19], με την ανάπτυξη ενός εργαλείου που ανευρίσκει ομοιότητες μεταξύ των συστατικών σχημάτων.

Σχετικά τώρα με τη στρατηγική ενοποίησης, καθορίζεται να εφαρμοστεί μια οποιαδήποτε δυαδική στρατηγική. Θυμίζουμε ότι το χαρακτηριστικό των δυαδικών στρατηγικών είναι ότι το στάδιο συγχώνευσής τους δέχεται σαν είσοδο δύο σχήματα κάθε φορά και παράγει ένα συγχωνευμένο σχήμα στην έξοδο. Υπάρχουν δύο τύποι δυαδικών στρατηγικών που φαίνονται με μορφή δένδρου στο Σχήμα 5.1(b) : η *κλιμακωτή* και η *ισορροπημένη* στρατηγική. Καθένα από τα φύλλα αυτών των δένδρων αντιπροσωπεύει ένα συστατικό σχήμα, η ρίζα αντιπροσωπεύει το τελικό ενοποιημένο σχήμα, ενώ οι ενδιάμεσοι κόμβοι αντιπροσωπεύουν ενδιάμεσα συγχωνευμένα σχήματα.

Δεδομένου ότι για το στάδιο συγχώνευσης χρησιμοποιούμε έναν αλγόριθμο που δέχεται δύο σχήματα και παράγει ένα άλλο, ο αλγόριθμος αυτός μπορεί να θεωρηθεί ως ένας δυαδικός τελεστής σχημάτων. Όπως θα δούμε στην ανάλυση της φάσης συγχώνευσης, ο τελεστής αυτός αφενός είναι **αντιμεταθετικός**, αφετέρου αν τα σχήματα εισόδου του έχουν μια ορισμένη μορφή, είναι και **προσεταιριστικός**. Αυτό σημαίνει ότι η σειρά με την

οποία περνάμε τα συστατικά σχήματα στον αλγόριθμο συγχώνευσης δεν παίζει ρόλο, μια και το τελικό συγχωνευμένο σχήμα είναι πάντα το ίδιο. Πάντως στη γενική περίπτωση που τα συστατικά σχήματα δεν έχουν αυτή την μορφή, θεωρούμε ότι δεν γίνεται κάποια ιδιαίτερη υπόθεση για τη σειρά συγχώνευσης των συστατικών σχημάτων.

Τα επόμενα δύο στάδια της διαδικασίας λαμβάνουν χώρα επαναληπτικά, γιαυτό θα τα παρουσιάσουμε μαζί.

5.2.2 Στάδιο Σύγκρισης και Προσαρμογής Συστατικών Σχημάτων

Στα δύο αυτά στάδια, όπως έχουμε προαναφέρει, γίνεται η αναγνώριση και επίλυση των συγκρούσεων των συστατικών σχημάτων. Στη βιβλιογραφία υπάρχουν αρκετές αναφορές για τα είδη των συγκρούσεων που μπορεί να υπάρξουν μεταξύ δύο σχημάτων ([24], [39], [37], [28], [5]). Θα επιλέξουμε την ορολογία του [37] για να παραθέσουμε μερικά είδη συγκρούσεων σχημάτων. Έχουμε λοιπόν :

- **Σημασιολογικές συγκρούσεις**, οι οποίες εκδηλώνονται όταν σε δύο σχήματα υπάρχουν έννοιες που εκφράζουν αλληλοεπικαλυπτόμενα σύνολα αντικειμένων του πραγματικού κόσμου. Για παράδειγμα, μια κλάση οντοτήτων ”Φοιτητής” στο ένα σχήμα και μια κλάση οντοτήτων ”Φοιτητής-ΗΥ” στο άλλο σχήμα.

- **Περιγραφικές συγκρούσεις**, που περιλαμβάνουν τις :

  -- *Ονομαστικές συγκρούσεις*, περιπτώσεις ομωνύμων και συνωνύμων.

  -- *Συγκρούσεις πεδίου τιμών γνωρισμάτων*, όπου το γνώρισμα μιας κλάσης έχει διαφορετικά πεδία τιμών σε δύο σχήματα.

  -- *Συγκρούσεις κλίμακας τιμών γνωρισμάτων*, όπου το γνώρισμα μιας κλάσης έχει διαφορετικές μονάδες μέτρησης σε δύο σχήματα.

- **Συγκρούσεις ετερογένειας**, που προκαλούνται από τη χρήση διαφορετικών μοντέλων δεδομένων για την παράσταση μιας έννοιας του πραγματικού κόσμου.

- **Δομικές συγκρούσεις**, που εκδηλώνονται όταν μια έννοια του πραγματικού κόσμου παριστάνεται με χρήση διαφορετικών δομών σε δύο σχήματα που ακολουθούν το ίδιο μοντέλο δεδομένων.

Εμείς θα ασχοληθούμε με το χειρισμό ονομαστικών και δομικών συγκρούσεων, μια και αυτές συναντώνται συχνότερα στα σχήματα που έχουν σχηματιστεί μετά τη διαδικασία μετάφρασης. Στο σημείο αυτό να παρατηρήσουμε ότι όπως θα δούμε παρακάτω, κατά το στάδιο ανακατασκευής επιλύονται και συγκρούσεις πεδίου τιμών γνωρισμάτων.

Η αναγνώριση των ονομαστικών και δομικών συγκρούσεων, θεωρούμε ότι γίνεται αποκλειστικά από το χρήστη/σχεδιαστή, μια και αυτός είναι ο μόνος που γνωρίζει τη σημασιολογία των εννοιών που περιγράφονται από τα συστατικά σχήματα. Η εξάλειψη τώρα των συγκρούσεων αυτών από τα συστατικά σχήματα, γίνεται με την εφαρμογή σ' αυτά ορισμένων τελεστών τροποποίησης. Μετά το τέλος των σταδίων αυτών τα συστατικά σχήματα έχουν έρθει σε ένα είδος *κανονικής μορφής*, σύμφωνα με την οποία το λογικό όνομα μια συγκεκριμένης κλάσης οντοτήτων ή μιας σχέσης γνωρίσματος είναι το ίδιο σε όλα τα τροποποιημένα συστατικά σχήματα.

Ας δούμε όμως τώρα αναλυτικά τούς τελεστές τροποποίησης που αναπτύξαμε, οι οποίοι χωρίζονται σε *τελεστές ονομαστικής τροποποίησης* και *τελεστές δομικής τροποποίησης*.

Τελεστές Ονομαστικής Τροποποίησης

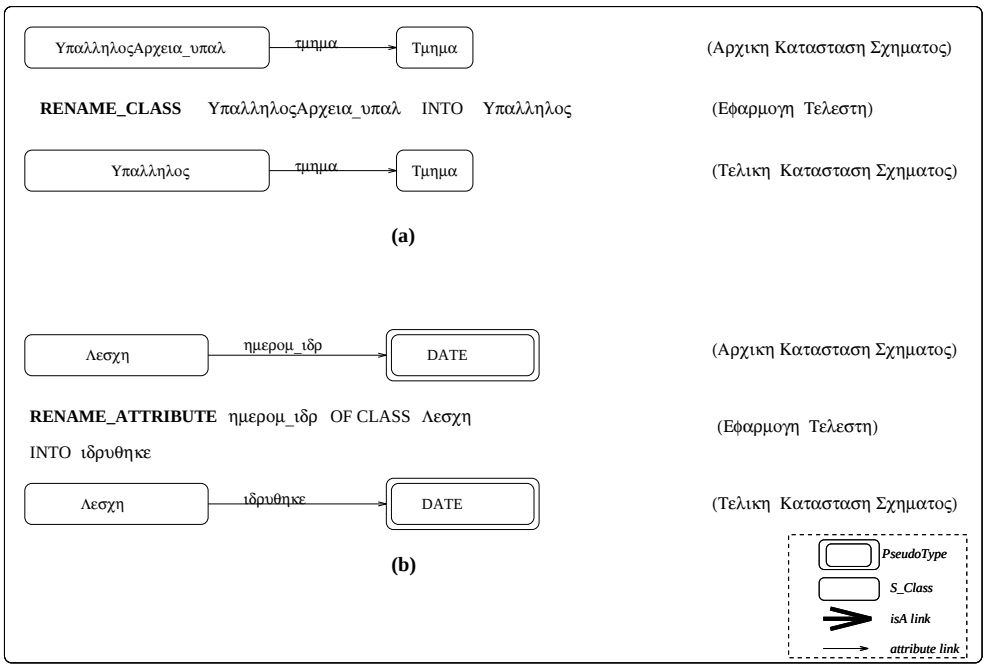
Οι τελεστές ονομαστικής τροποποίησης, χρησιμοποιούνται για την επίλυση ονομαστικών συγκρούσεων μεταξύ των συστατικών σχημάτων, περιπτώσεων δηλαδή *ομωνύμων* και *συνωνύμων*. Στις περιπτώσεις ομωνύμων, διαφορετικές έννοιες του πραγματικού κόσμου εμφανίζονται σε δύο διαφορετικά σχήματα με το ίδιο όνομα, ενώ στις περιπτώσεις συνωνύμων η ίδια έννοια του πραγματικού κόσμου εμφανίζεται σε δύο διαφορετικά σχήματα με διαφορετικό όνομα. Οι τελεστές ονομαστικής τροποποίησης λοιπόν είναι οι εξής :

- Ο τελεστής *μετονομασίας κλάσης* (**RENAME_CLASS**). Ο τελεστής αυτός μετατρέπει το λογικό όνομα <*old_class_name*> μιας απλής κλάσης ενός νομίμου σχήματος, στο λογικό όνομα <*new_class_name*>. Ο συμβολισμός του τελεστή αυτού είναι :
  RENAME_CLASS <*old_class_name*> INTO <*new_class_name*>
  Ο τελεστής αυτός μπορεί να χρησιμοποιηθεί μόνο αν το νέο λογικό όνομα <*new_class_name*> της απλής κλάσης δεν ταυτίζεται με το λογικό όνομα κάποιας άλλης ήδη υπάρχουσας απλής κλάσης του σχήματος. Ένα παράδειγμα εφαρμογής του τελεστή φαίνεται στο Σχήμα 5.2(a)

- Ο τελεστής *μετονομασίας γνωρίσματος* (**RENAME_ATTRIBUTE**). Ο τελεστής αυτός μετατρέπει το λογικό όνομα <*old_attribute_name*> ενός γνωρίσματος μιας απλής κλάσης <*class_name*> ενός νομίμου σχήματος, στο λογικό όνομα <*new_attribute_name*> της απλής κλάσης <*class_name*>. Ο συμβολισμός του τελεστή αυτού είναι :
  RENAME_ATTRIBUTE <*old_attribute_name*> OF CLASS <*class_name*>

INTO <*new_attribute_name*>

Ο τελεστής αυτός μπορεί να χρησιμοποιηθεί μόνο αν το νέο λογικό όνομα <*new_attribute_name*> του γνωρίσματος της απλής κλάσης <*class_name*> δεν ταυτίζεται με λογικό όνομα κάποιου άλλου ήδη υπάρχοντος γνωρίσματος της ίδιας κλάσης. Ένα παράδειγμα εφαρμογής του τελεστή φαίνεται στο Σχήμα 5.2(b)



Σχήμα 5.2: Παραδείγματα Εφαρμογής Τελεστών Ονομαστικής Τροποποίησης

Τελεστές Δομικής Τροποποίησης

Οι τελεστές δομικής τροποποίησης, χρησιμοποιούνται για την επίλυση δομικών συγκρούσεων μεταξύ των συστατικών σχημάτων. Οι δομικές συγκρούσεις προκαλούνται όταν σε δύο σχήματα χρησιμοποιούνται διαφορετικές δομές του μοντέλου δεδομένων, για να παρασταθεί η ίδια έννοια του πραγματικού κόσμου. Οι τελεστές δομικής τροποποίησης λοιπόν είναι οι εξής :

- Ο τελεστής *αναστροφής γνωρίσματος* (**REVERSE_ATTRIBUTE**). Ο τελεστής αυτός μετατρέπει το γνώρισμα με λογικό όνομα <*old_attribute_name*> μιας απλής κλάσης <*class_name_A*> προς μια απλή κλάση <*class_name_B*>, σε ένα γνώρισμα με λογικό όνομα <*new_attribute_name*> της απλής κλάσης <*class_name_B*> προς την απλή

κλάση <*class_name_A*>. Ο συμβολισμός του τελεστή αυτού είναι :

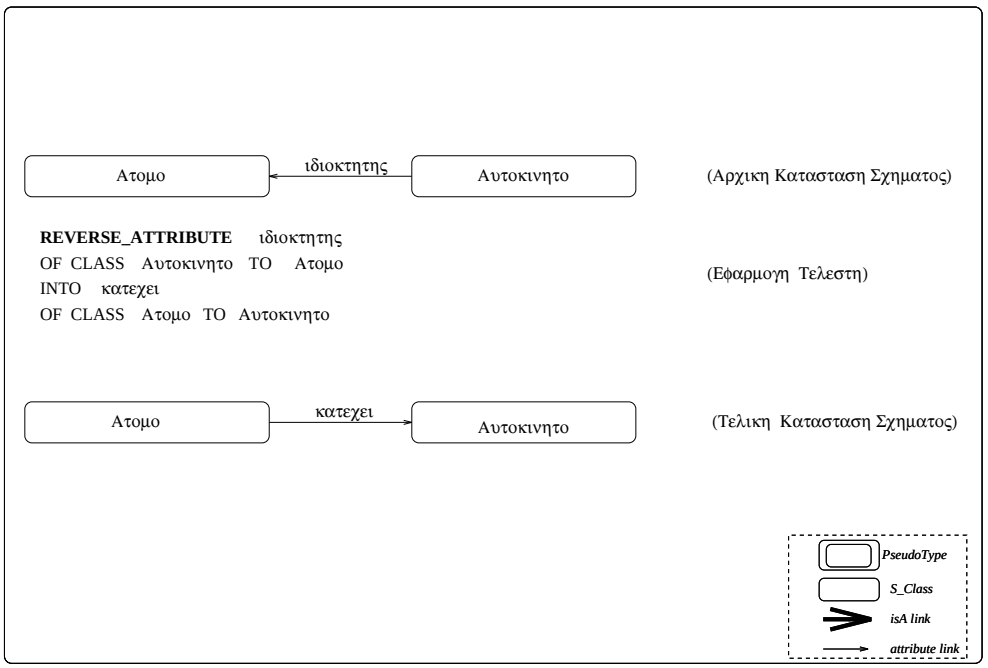
REVERSE_ATTRIBUTE <*old_attribute_name*>

OF CLASS <*class_name_A*> TO <*class_name_B*>

INTO <*new_attribute_name*>

OF CLASS <*class_name_B*> TO <*class_name_A*>

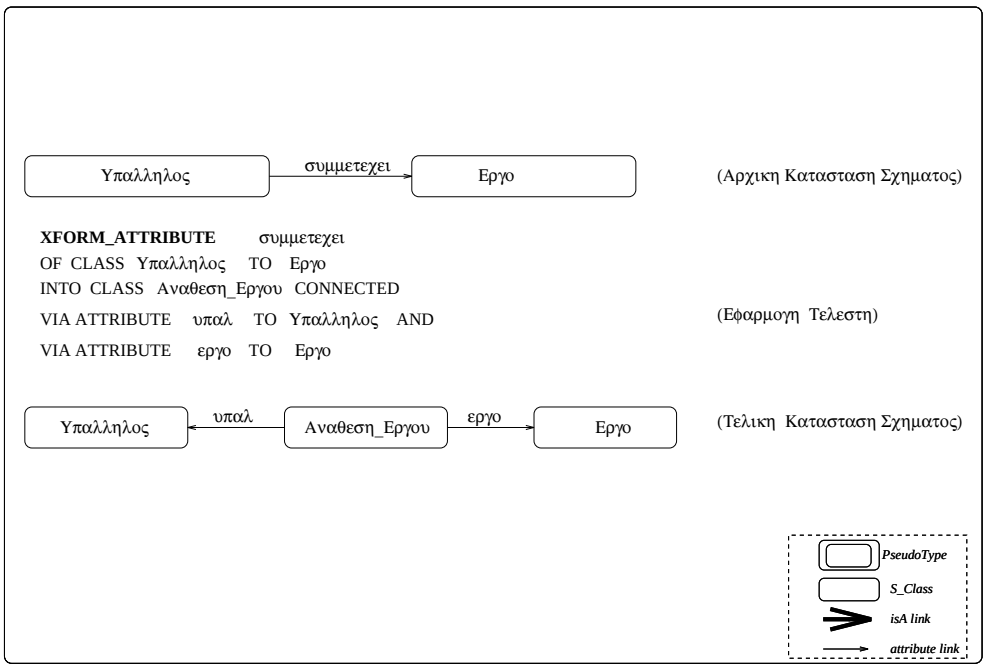
Ο τελεστής αυτός μπορεί να χρησιμοποιηθεί μόνο αν το νέο λογικό όνομα <*new_attribute_name*> του γνωρίσματος της απλής κλάσης <*class_name_B*> δεν ταυτίζεται με λογικό όνομα κάποιου άλλου ήδη υπάρχοντος γνωρίσματος της ίδιας κλάσης. Ένα παράδειγμα εφαρμογής του τελεστή φαίνεται στο Σχήμα 5.3.



Σχήμα 5.3: Παράδειγμα Εφαρμογής Τελεστή Αναστροφής Γνωρίσματος

- Ο τελεστής *μετατροπής γνωρίσματος* (**XFORM_ATTRIBUTE**). Ο τελεστής αυτός μετατρέπει το γνώρισμα με λογικό όνομα <*old_attribute_name*> μιας απλής κλάσης <*class_name_A*> προς μια απλή κλάση <*class_name_B*>, σε μια απλή κλάση με λογικό όνομα <*new_class_name*> η οποία συνδέεται μέσω των γνωρισμάτων με λογικά ονόματα <*new_attribute_name_A*> και <*new_attribute_name_B*>, με τις απλές κλάσεις <*class_name_A*> και <*class_name_B*> αντίστοιχα. Ο συμβολισμός του τελεστή αυτού είναι :

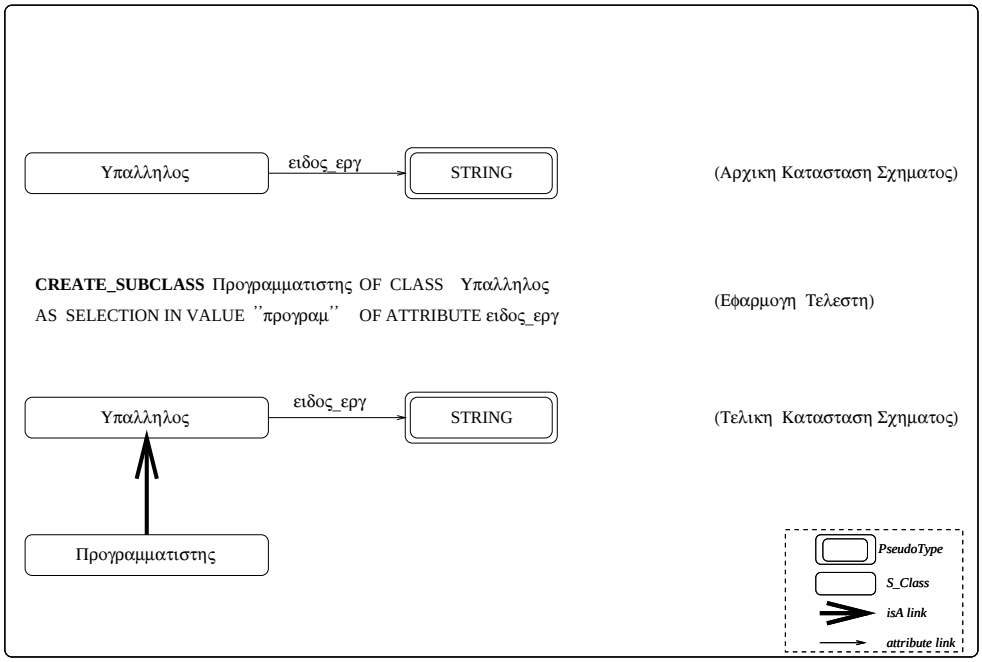
XFORM_ATTRIBUTE <*old_attribute_name*>
OF CLASS <*class_name_A*> TO <*class_name_B*>
INTO CLASS <*new_class_name*> CONNECTED
VIA ATTRIBUTE <*new_attribute_name_A*> TO <*class_name_A*> AND
VIA ATTRIBUTE <*new_attribute_name_B*> TO <*class_name_B*>
Ο τελεστής αυτός μπορεί να χρησιμοποιηθεί μόνο αν το νέο λογικό όνομα <*new_class_name*> της δημιουργηθείσας απλής κλάσης, δεν ταυτίζεται με το λογικό όνομα κάποιας άλλης ήδη υπάρχουσας απλής κλάσης του σχήματος. Ένα παράδειγμα εφαρμογής του τελεστή φαίνεται στο Σχήμα 5.4.



Σχήμα 5.4: Παράδειγμα Εφαρμογής Τελεστή Μετατροπής Γνωρίσματος

- Ο τελεστής *δημιουργίας υποκλάσης* (**CREATE_SUBCLASS**). Ο τελεστής αυτός δημιουργεί μια υποκλάση με λογικό όνομα <*new_class_name*> που περιέχει τις οντότητες της απλής κλάσης με λογικό όνομα <*class_name*> οι οποίες έχουν την τιμή <*value*> για το γνώρισμα <*attribute_name*>. Ο συμβολισμός του τελεστή αυτού είναι :
  CREATE_SUBCLASS <*new_class_name*> OF CLASS <*class_name*>
  AS SELECTION IN VALUE <*value*> OF ATTRIBUTE <*attribute_name*>

Ο τελεστής αυτός μπορεί να χρησιμοποιηθεί μόνο αν το νέο λογικό όνομα *<new_class_name>* της δημιουργηθείσας απλής κλάσης, δεν ταυτίζεται με το λογικό όνομα κάποιας άλλης ήδη υπάρχουσας απλής κλάσης του σχήματος. Ένα παράδειγμα εφαρμογής του τελεστή φαίνεται στο Σχήμα 5.5.



Σχήμα 5.5: Παράδειγμα Εφαρμογής Τελεστή Δημιουργίας Υποκλάσης

Κλείνοντας την παρουσίαση των τελεστών τροποποίησης και πριν περάσουμε στην ανάλυση του σταδίου συγχώνευσης, ας παρατηρήσουμε ότι οι τελεστές ''διατηρούν τη νομιμότητα'' των συστατικών σχημάτων, υπό την έννοια ότι μετά την εφαρμογή τους σε ένα νόμιμο συστατικό σχήμα δίνουν ένα νέο **νόμιμο** σχήμα.

5.2.3 Στάδιο Συγχώνευσης Συστατικών Σχημάτων

Στο στάδιο αυτό σύμφωνα με στρατηγική συγχώνευσης, εφαρμόζουμε επαναληπτικά ένα αλγόριθμο συγχώνευσης πάνω σε δύο σχήματα (συστατικά ή ενδιάμεσα) κάθε φορά, και παίρνουμε το τελικό ενοποιημένο σχήμα. Στην ενότητα 5.1, είδαμε ότι το είδος των σημασιολογικών δικτύων που θέλουμε να ενοποιήσουμε μας επιτρέπει να χρησιμοποιήσουμε αυτούσιο τον αλγόριθμο συγχώνευσης σχημάτων ΒΔ που παρουσιάζεται στο [26]. Στο υπόλοιπο αυτής της ενότητας λοιπόν, θα παρουσιάσουμε τον

τρόπο λειτουργίας αυτού του αλγορίθμου.

Αλγόριθμος Συγχώνευσης Σχημάτων

Στην ενότητα 5.1 είδαμε τον ορισμό των νομίμων σχημάτων. Χαλαρώνοντας την πρώτη ιδιότητα νομιμότητας των σχημάτων αυτών, παίρνουμε μια κατηγορία σχημάτων που ονομάζονται **ασθενή** και ορίζονται ως εξής :

> Ένα **ασθενές σχήμα** πάνω από τα ονόματα κλάσεων $\mathcal{N}$ και τα ονόματα γνωρισμάτων $\mathcal{L}$, είναι μια τριάδα (C, E, S), όπου $C \subseteq \mathcal{N}$, $E \subseteq C \times \mathcal{L} \times C$, και $S \subseteq C \times C$, τέτοια ώστε
>
> 1. Αν $p, q, r \in C$ και $a \in \mathcal{L}$ είναι τέτοια ώστε $p \xrightarrow{a} q$ και $p \xrightarrow{a} r$ και $q \Longrightarrow r$, τότε $q = r$.
> 2. Η S είναι σχέση μερικής διάταξης στο C.
> 3. Για κάθε $p, q, r \in C$ και $a \in \mathcal{L}$, αν $p \Longrightarrow q$ και $p \xrightarrow{a} r$ και υπάρχει ένα $s \in C$ τέτοιο ώστε $q \xrightarrow{a} s$, τότε υπάρχει ένα $t \in C$ τέτοιο ώστε $q \xrightarrow{a} t$ και $r \Longrightarrow t$.

Προφανώς κάθε νόμιμο σχήμα είναι και ασθενές. Ο αλγόριθμος του [26] τώρα δέχεται σαν είσοδο δύο ασθενή σχήματα, και παράγει στην έξοδο ένα συγχωνευμένο ασθενές σχήμα. Το σχήμα αυτό θα δούμε ότι είναι το *μέγιστο κάτω φράγμα* των δύο σχημάτων εισόδου, κάτω από μια σχέση *πληροφοριακής διάταξης*.

> Ορίζουμε τη σχέση **πληροφοριακής διάταξης** ($\sqsubseteq$) στο σύνολο των ασθενών σχημάτων έτσι ώστε, αν τα $\mathcal{G}_1$ και $\mathcal{G}_2$ είναι ασθενή σχήματα πάνω από το $\mathcal{N}$ και το $\mathcal{L}$, τότε $\mathcal{G}_1 \sqsubseteq \mathcal{G}_2$ αν και μόνο αν
>
> 1. $C_1 \supseteq C_2$.
> 2. Αν $p \Longrightarrow_1 q$ και $p \in C_2$ ή $q \in C_2$, τότε $p \Longrightarrow_2 q$.
> 3. Αν $p \xrightarrow{a}_2 q$, τότε υπάρχει ένα $r \in C_1$ τέτοιο ώστε $p \xrightarrow{a}_1 r$ και $q \Longrightarrow_1 r$.

Ας προσέξουμε ότι η δεύτερη συνθήκη του προηγούμενου ορισμού, εννοεί ότι για να περιέχει ένα ασθενές σχήμα G_1, λιγότερη πληροφορία από ένα δεύτερο σχήμα G_2, το G_1 δεν θα πρέπει να περιέχει **νέες** σχέσεις εξειδίκευσης που θα συμμετέχουν κλάσεις του G_2, ενώ μπορεί να περιέχει **νέες** σχέσεις εξειδίκευσης που θα συμμετέχουν κλάσεις που δεν περιέχονται στο G_2. Κατόπιν, ο συγγραφέας παρατηρεί ότι η σχέση πληροφοριακής διάταξης είναι ανακλαστική, και αντισυμμετρική, όμως στη γενική περίπτωση δεν είναι

μεταβατική. Θα δούμε παρακάτω, υπό ποιές προυποθέσεις η σχέση αυτή είναι και μεταβατική.

Κατόπιν, ορίζει ένα δυαδικό τελεστή "$\sqcap$", μεταξύ δύο ασθενών σχημάτων $\mathcal{G}_1$ και $\mathcal{G}_2$, και αποδεικνύει ότι το αποτέλεσμα του τελεστή αυτού είναι το **μέγιστο κάτω φράγμα** των $\mathcal{G}_1$ και $\mathcal{G}_2$, υπό την έννοια ότι αν $\mathcal{G} = \mathcal{G}_1 \sqcap \mathcal{G}_2$ τότε

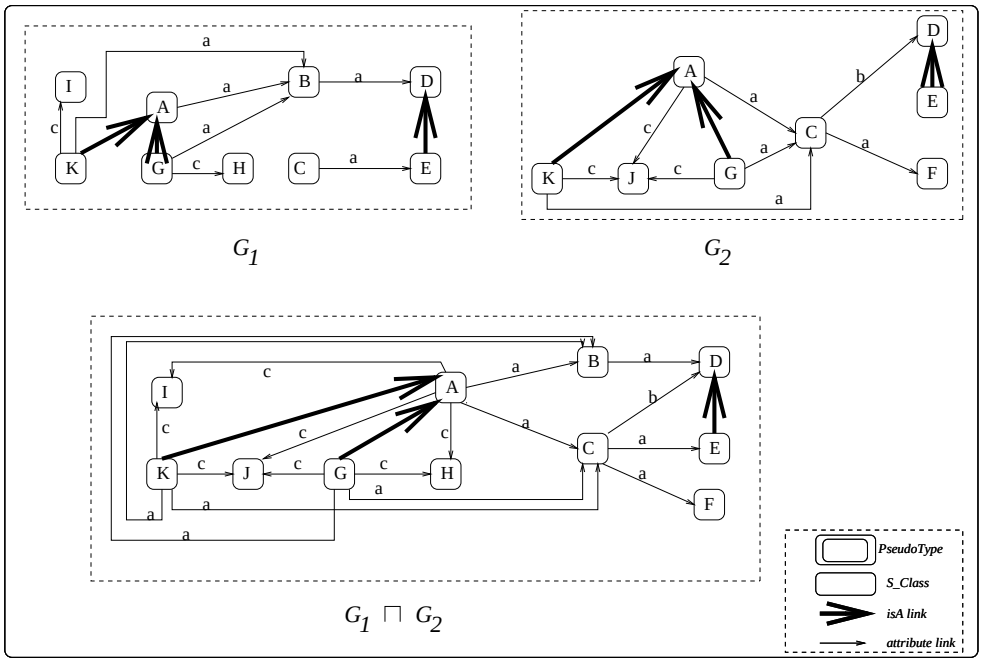
- το $\mathcal{G}$ είναι ένα ασθενές σχήμα,
- $\mathcal{G} \sqsubseteq \mathcal{G}_1$ και $\mathcal{G} \sqsubseteq \mathcal{G}_2$, και
- για κάθε $\mathcal{G}_3$ τέτοιο ώστε $\mathcal{G}_3 \sqsubseteq \mathcal{G}_1$ και $\mathcal{G}_3 \sqsubseteq \mathcal{G}_2$, ισχύει ότι $\mathcal{G}_3 \sqsubseteq \mathcal{G}$

Ο τρόπος υπολογισμού του τελεστή "$\sqcap$" αποτελεί τον αλγόριθμο συγχώνευσης των ασθενών σχημάτων $\mathcal{G}_1$ και $\mathcal{G}_2$. Έχουμε λοιπόν τον πιο κάτω ορισμό :

> Αν $\mathcal{G}_1 = (C_1, E_1, S_1)$ και $\mathcal{G}_2 = (C_2, E_2, S_2)$ δύο ασθενή σχήματα, ορίζουμε το σχήμα $\mathcal{G} = \mathcal{G}_1 \sqcap \mathcal{G}_2$, με $\mathcal{G} = (C, E, S)$ ως εξής :
>
> $$
> \begin{aligned}
> C =& C_1 \cup C_2 \\
> S =& \{(p,q) \in S_1 \cup S_2 \mid ((p,q) \in S_1 \cap S_2) \vee \\
> & \quad (p \notin C_1 \wedge q \notin C_1) \vee (p \notin C_2 \wedge q \notin C_2)\} \\
> E =& \{(p,a,q) \mid (\exists s \in C \cdot (p,a,s) \in E_1 \cup E_2) \wedge \\
> & \quad (\exists r \in C \cdot r \Longrightarrow p \wedge (r,a,q) \in E_1 \cup E_2) \wedge \\
> & \quad (\forall r,s \in C \cdot r \Longrightarrow p \wedge q \Longrightarrow s \wedge (r,a,s) \in E_1 \cup E_2 \text{ συνεπάγεται } q \equiv s)\}
> \end{aligned}
> $$

Στο Σχήμα 5.6 φαίνεται ένα μικρό παράδειγμα εφαρμογής του τελεστή πάνω σε δύο ασθενή σχήματα. Στο παράδειγμα αυτό βλέπουμε ότι το τελικό ασθενές σχήμα παραβιάζει την πρώτη ιδιότητα νομιμότητας.

Είναι προφανές από τον τρόπο ορισμού του τελεστή, ότι αυτός είναι αντιμεταθετικός. Δεδομένου όμως ότι η σχέση πληροφοριακής διάταξης δεν είναι πάντα μεταβατική, ο συγγραφέας υποστηρίζει ότι ο τελεστής μέγιστου κάτω φράγματος δεν είναι πάντα προσεταιριστικός. Αυτό σημαίνει ότι αν επαναληπτικά εφαρμόζεται σε μια ομάδα ασθενών σχημάτων κάθε φορά με διαφορετική σειρά, τα τελικά ασθενή σχήματα που θα παράγει θα είναι διαφορετικά. Ισχυρίζεται όμως, ότι για μια κατηγορία *καλά-συμπεριφερόμενων* ασθενών σχημάτων η σχέση πληροφοριακής διάταξης είναι μεταβατική, και άρα ο τελεστής μέγιστου άνω φράγματος γιαυτά τα σχήματα είναι προσεταιριστικός. Ορίζει λοιπόν την κατηγορία των σχημάτων αυτών ως εξής :



Σχήμα 5.6: Παράδειγμα Εφαρμογής Τελεστή Συγχώνευσης

Μια συλλογή ασθενών σχημάτων, $\mathcal{G}_1, ..., \mathcal{G}_n$, λέγεται ότι είναι **καλά-συμπεριφερόμενη** αν και μόνο αν, για οποιοδήποτε $\mathcal{G}_i$ και $\mathcal{G}_j$ και οποιεσδήποτε κλάσεις p και q, αν $p \Longrightarrow_i q$ και $p \Longrightarrow_j q$ τότε, για οποιαδήποτε r αν $r \Longrightarrow_i p$ και $r \in C_j$ τότε $r \Longrightarrow_j p$, και για οποιαδήποτε s αν $q \Longrightarrow_i s$ και $s \in C_j$ τότε $q \Longrightarrow_j s$.

5.2.4 Στάδιο Ανακατασκευής Τελικού Ενοποιημένου Σχήματος

Όπως είδαμε στο προηγούμενο στάδιο, μετά τη επαναληπτική εφαρμογή του τελεστή όπως ορίζει η στρατηγική συγχώνευσης, καταλήγουμε σε ένα ασθενές συγχωνευμένο σχήμα. Στόχος του σταδίου ανακατασκευής του τελικού ενοποιημένου σχήματος είναι η μετατροπή του τελικού ασθενούς σχήματος σε ένα νόμιμο σχήμα. Για το λόγο αυτό ακολουθούμε τη διαδικασία που επίσης αναφέρεται στο [26]. Για την περιγραφή της διαδικασίας, αρχικά είναι απαραίτητοι ορισμένοι συμβολισμοί.

Για κάθε $p \in C$ και κάθε $a \in \mathcal{L}$ γράφουμε $E(p, a)$ για να δηλώσουμε το σύνολο

$$E(p, a) = \{q \in C \mid p \xrightarrow{a} q\}$$

. Ακόμα, για κάθε σύνολο $X \subseteq C$, γράφουμε $E(X, a)$ για να δηλώσουμε το σύνολο

$$E(X, a) = Max_S(\{q \in \mathcal{N} \mid \forall p \in X \cdot p \xrightarrow{a} q\})$$

όπου η συνάρτηση $Max_S : \mathcal{P}(\mathcal{N}) \rightarrow \mathcal{P}(\mathcal{N})$ ορίζεται έτσι ώστε, για κάθε σύνολο $X \subseteq \mathcal{N}$, το $Max_S(X)$ να είναι το σύνολο των μέγιστων στοιχείων του X σύμφωνα με τη σχέση μερικής διάταξης S. Δηλαδή

$$Max_S(X) = \{p \in X \mid \forall q \in X \cdot p \Longrightarrow q \text{συνεπάγεται} q = p\}$$

Προχωρούμε λοιπόν τώρα στην ανάλυση της διαδικασίας παραγωγής του νόμιμου σχήματος $\mathcal{G}' = (C', E', S')$ από ένα ασθενές $\mathcal{G}$ ως ακολούθως :

1. Αρχικά για να άρουμε την παραβίαση της πρώτης ιδιότητας νομιμότητας, θα εισάγουμε στο σχήμα μας κάποιες νέες *υπονοούμενες* κλάσεις. Οι κλάσεις αυτές αν και μπορεί να μην εκφράζουν κάποια έννοια του πραγματικού κόσμου, θα εισαχθούν όπως θα δούμε και παρακάτω ως υπερκλάσεις ήδη υπαρχόντων κλάσεων και θα αντιπροσωπεύουν την ένωση των πεδίων τιμών ορισμένων γνωρισμάτων. Η εισαγωγή λοιπόν αυτών των κλάσεων στο τελικό νόμιμο σχήμα, θα άρει τις περιπτώσεις *συγκρούσεων πεδίου τιμών γνωρισμάτων* μεταξύ των συστατικών σχημάτων. Ξεκινάμε δημιουργώντας ένα σύνολο, $Imp \subseteq \mathcal{P}(\mathcal{C})$, από σύνολα κλάσεων, που αντιστοιχεί στο σύνολο των υπονοουμένων κλάσεων. Το σύνολο αυτό θα κατασκευαστεί αναδρομικά ως εξής:

$$I^0 = \{\{p\} \mid p \in C\}$$
$$I^{n+1} = \{E(X, a) \mid X \in I^n, a \in \mathcal{L}\}$$
$$I^\infty = \bigcup_{n=1}^{\infty} I^n$$

και τελικά

$$Imp = \{X \in I^\infty \mid |X| > 1\}$$

Ας παρατηρήσουμε ότι από τη στιγμή που υπάρχουν πεπερασμένα υποσύνολα του του $\mathcal{C}$, θα βρούμε τελικά ένα I_n τέτοιο ώστε $I_n \subseteq \bigcup_{i=1}^{n-1}(I_i)$, οπότε και θα τερματίσουμε την αναδρομή. Για το παράδειγμα του σχήματος 5.6 λοιπόν παίρνουμε τις εξής υπονούμενες κλάσεις :

$I^0 = \{\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{G\}, \{H\}, \{I\}, \{J\}, \{K\}\}$
$I^1 = \{\{B, C\}, \{J, H, I\}, \{D\}, \{E, F\}, \{H, J\}, \{I, J\}\}$
$I^2 = \{\emptyset\}$
$I^\infty = I^0 \cup I^1$
$Imp = \{\{B, C\}, \{J, H, I\}, \{E, F\}, \{H, J\}, \{I, J\}\}$

2. Ορίζουμε το σύνολο των κλάσεων C' παίρνοντας αρχικά το C και προσθέτοντας του μια νέα κλάση X' για κάθε $X \in Imp$. Έτσι έχουμε

$$C' = C \cup \{X' \mid X \in Imp\}$$

Το λογικό όνομα μιας κλάσης X', προκύπτει από τη συγχώνευση των λογικών ονομάτων των κλάσεων που συμμετέχουν στο αντίστοιχο σύνολο κλάσεων $X \in Imp$.

Συνεχίζοντας για το παράδειγμα του σχήματος 5.6

$$C' = \{A, B, C, D, E, F, G, H, I, J, K, \{B, C\}', \{E, F\}', \{H, J\}', \{I, J\}', \{H, I, J\}'\}$$

3. Ορίζουμε το σύνολο των γνωρισμάτων E' έτσι ώστε αν υπάρχει ένα μοναδικό q τέτοιο ώστε $p \xrightarrow{a} q$ τότε $(p, a, q) \in E'$, ενώ αν $p \xrightarrow{a} q$ για κάθε $q \in X$ τότε $(p, a, X') \in E'$. Δηλαδή, η υπονοούμενη κλάση X′ αντιπροσωπεύει το νέο πεδίο τιμών του γνωρίσματος a της κλάσης p. Τυπικά :

$$\begin{aligned} E' = \{\ (x, a, q) \mid\ & x \in C', a \in \mathcal{L}, E(x, a) = \{q\}\} \\ & \cup \{(x, a, Y') \mid x \in C', a \in \mathcal{L}, E(x, a) = Y \wedge |Y| > 1\} \end{aligned}$$

όπου $E(X', a) = E(X, a)$ για κάθε $X \in Imp$.
Για το παράδειγμά μας

$$\begin{aligned} E' = \{\ & (A, a, \{B, C\}'), (B, a, D), (C, b, D), (C, a, \{E, F\}'), (K, a, \{B, C\}'), \\ & (G, a, \{B, C\}'), (A, c, \{H, I, J\}'), (K, c, \{I, J\}'), (G, c, \{H, J\}')\} \end{aligned}$$

4. Ορίζουμε τη σχέση $\overline{\in}$ μεταξύ των συνόλων κλάσεων C και C' έτσι ώστε, για κάθε $p \in C$ και $x \in C'$, έχουμε $p \overline{\in} x$ αν και μόνο αν είτε $x \equiv p$, είτε $x = X'$ για κάποιο $X \in Imp$ και $p \in X$.

Ορίζουμε τώρα το S' παίρνοντας αρχικά το S και προσθέτοντας :

- κάθε ζεύγος (X', Y') που είναι τέτοιο ώστε κάθε κλάση του X είναι εξειδίκευση κάποιας κλάσης του Y,
- κάθε ζεύγος (X', p) όπου κάθε κλάση του X είναι εξειδίκευση της p,
- κάθε ζεύγος (p, X') όπου η p είναι εξειδίκευση κάποιας κλάσης του X.

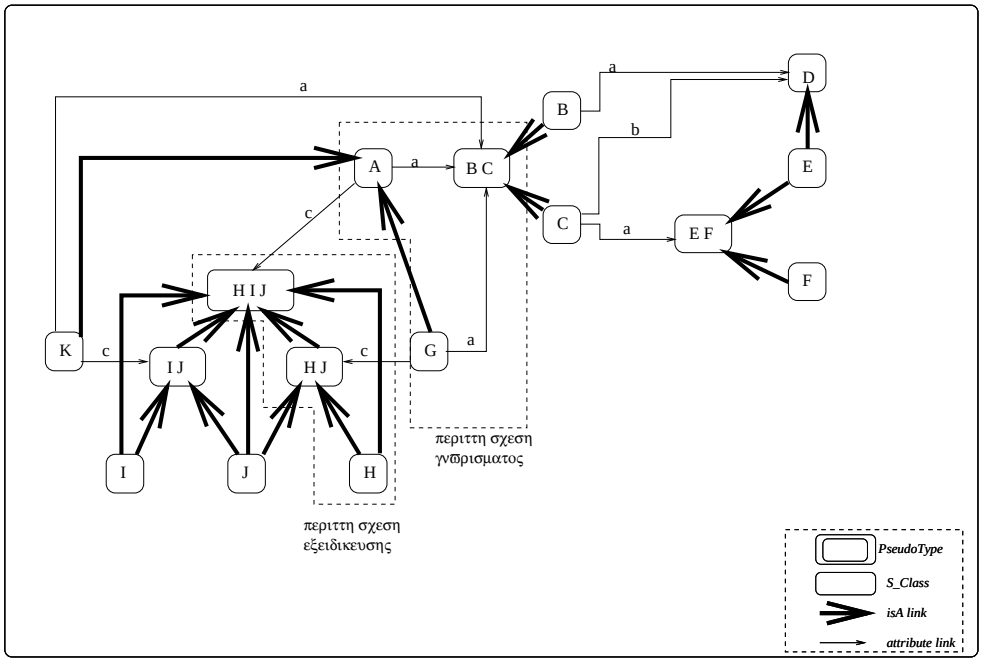
Με τον τρόπο αυτό δηλώνουμε ότι οι υπονοούμενες κλάσεις είναι η ένωση των κλάσεων από τις οποίες προήλθαν. Τυπικά,

$$S' = \{(x, y) \mid x, y \in C', \forall p \overline{\in} x \cdot \exists q \overline{\in} y \cdot (p, q) \in S\}$$

Για το παράδειγμά μας

$$\begin{aligned} S' = \{ & (E, D), (C, \{B, C\}'), (B, \{B, C\}'), (F, \{E, F\}'), (E, \{E, F\}'), (K, A), (G, A), \\ & (\{I, J\}', \{H, I, J\}'), (\{H, J\}', \{H, I, J\}'), (I, \{I, J\}'), (J, \{I, J\}'), (J, \{H, J\}'), \\ & (H, \{H, J\}'), (I, \{H, I, J\}'), (J, \{H, I, J\}'), (H, \{H, I, J\}')\} \end{aligned}$$

Στο Σχήμα 5.7 βλέπουμε το τελικό σχήμα που προέκυψε εφαρμόζοντας τη διαδικασία ανακατασκευής στο ασθενές σχήμα του παραδείγματος (Σχήμα 5.6).



Σχήμα 5.7: Τελικό Νόμιμο Σχήμα

Παρατηρούμε όμως ότι το τελικό νόμιμο σχήμα (όπως αυτό του παραδείγματος) που προκύπτει από τη διαδικασία ανακατασκευής περιέχει δύο είδη πλεονασμών που χρειάζεται να εξαλειφθούν προκειμένου να πάρουμε το τελικό Telos σημασιολογικό δίκτυο που αντιπροσωπεύει το ενοποιημένο σχήμα. Οι πλεονασμοί αυτοί είναι οι περιττές σχέσεις εξειδίκευσης που οφείλονται στη μεταβατικότητα της σχέσης εξειδίκευσης, και τα περιττά

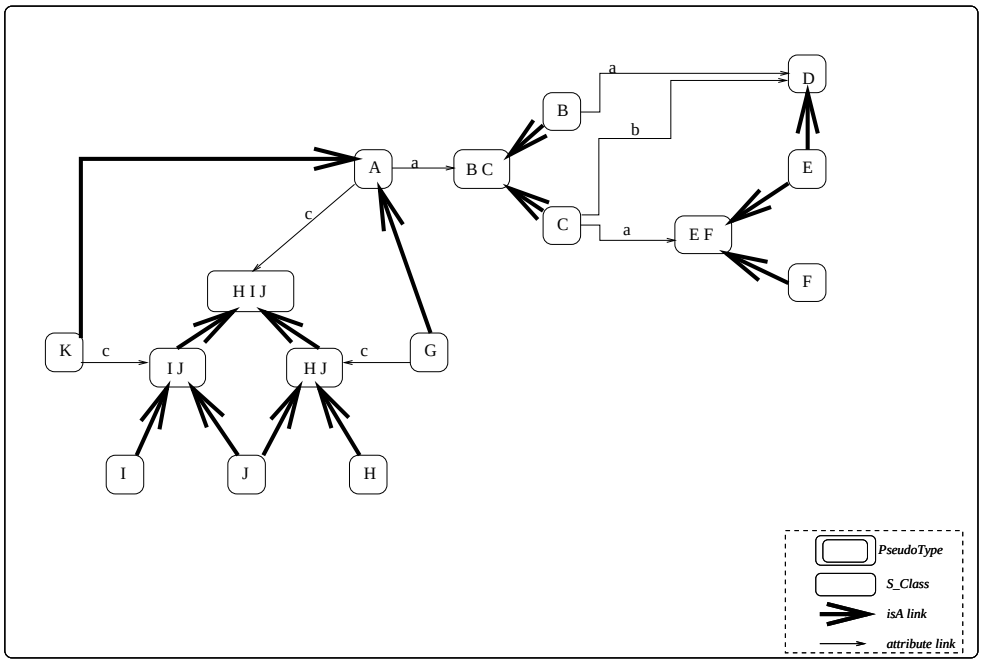
γνωρίσματα ορισμένων κλάσεων που οφείλονται στην ιδιότητα της κληρονομικότητας των γνωρισμάτων μεταξύ κλάσεων που συνδέονται με τη σχέση εξειδίκευσης. Αν λοιπόν $\mathcal{G}'= (C', S', E')$ είναι το τελικό νόμιμο σχήμα, τότε το Telos σημασιολογικό δίκτυο θα έχει τη μορφή $\mathcal{G}'_{fin} = (C'_{fin}, S'_{fin}, E'_{fin})$ με

$$C'_{fin} = C'$$

$$S'_{fin} = S' \setminus \{(p,q) \in S' \mid \exists r \in C' \cdot r \neq p \wedge r \neq q \wedge (p,r) \in S' \wedge (r,q) \in S'\}$$

$$E'_{fin} = E' \setminus \{(p,a,q) \in E' \mid \exists r \in C' \cdot r \neq p \wedge p \Longrightarrow r \wedge (r,a,q) \in E'\}$$

Μετά την εξάλειψη των πλεονασμών από το τελικό νόμιμο σχήμα του παραδείγματος, το Telos σημασιολογικό δίκτυο που προκύπτει φαίνεται στο Σχήμα 5.8.



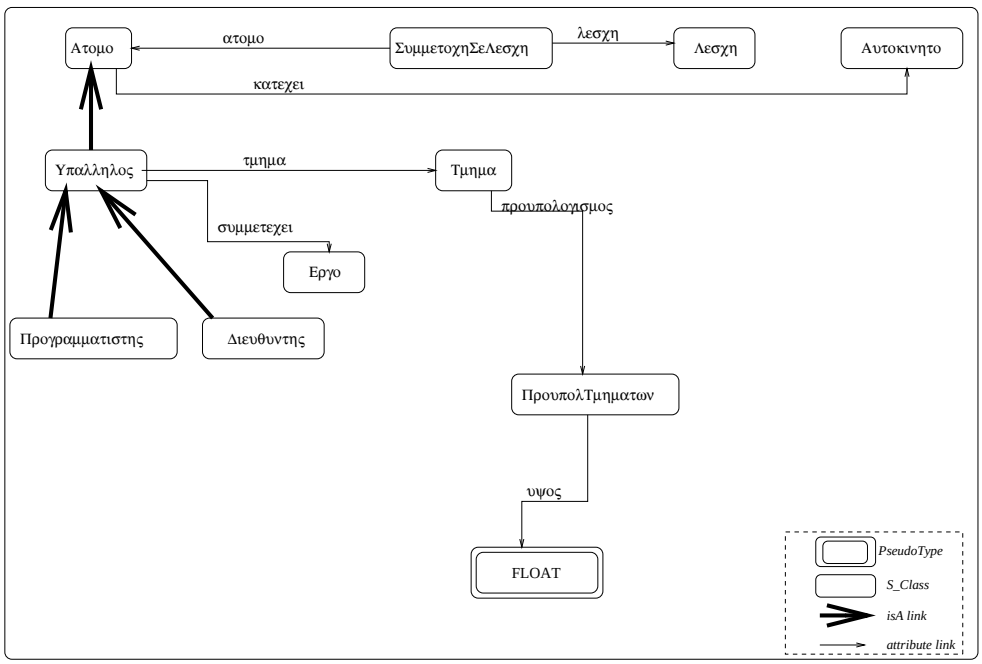
Σχήμα 5.8: Τελικό Ενοποιημένο Σημασιολογικό Δίκτυο

Παρακάτω ακολουθεί ένα αναλυτικό παράδειγμα εφαρμογής της μεθόδου.

5.3 Παράδειγμα Εφαρμογής της Μεθόδου

Το κεφάλαιο θα κλείσει με ένα αναλυτικό παράδειγμα εφαρμογής της μεθόδου ενοποίησης σχημάτων που παρουσιάσαμε. Για λόγους απλότητας και οικονομίας χώρου θα

θεωρήσουμε μόνο δύο συστατικά σχήματα (σημασιολογικά δίκτυα). Το πρώτο συστατικό σχήμα φαίνεται στο Σχήμα 4.1, και το δεύτερο φαίνεται στο Σχήμα 5.9.



Σχήμα 5.9: Συστατικό Σημασιολογικό Δίκτυο

Για λόγους επίσης οικονομίας χώρου, στα δύο αυτά συστατικά σχήματα δεν απεικονίζονται τα γνωρίσματα ορισμένων κλάσεων που κληρονομούνται από τις υπερκλάσεις τους, καθώς και οι σχέσεις εξειδίκευσης που προκύπτουν από τη μεταβατικότητα της σχέσης εξειδίκευσης.

5.3.1 Στάδιο Προενοποίησης

Δεδομένου ότι τα συστατικά σχήματα είναι μόνο δύο, στο στάδιο προενοποίησης απλά καθορίζουμε ότι ο τελεστής συγχώνευσης θα εφαρμοστεί μόνο μία φορά.

5.3.2 Στάδια Σύγκρισης και Προσαρμογής Συστατικών Σχημάτων

Κατά τη διάρκεια αυτών των σταδίων εφαρμόζουμε στα συστατικά σχήματα τους τελεστές τροποποίησης για να άρουμε τις ονομαστικές και δομικές συγκρούσεις τους. Για το πρώτο συστατικό σχήμα λοιπόν δηλώνουμε :

- **Επίλυση Ονομαστικών Συγκρούσεων**

-- **RENAME_CLASS** *ΥπάλληλοςΑρχεία_υπάλ* INTO *Υπάλληλος*

-- **RENAME_CLASS** *Στ_ταυτ_ατόμουΑλλα_στ_ατόμου* INTO *Άτομο*

Τα αποτελέσματα εφαρμογής του πρώτου τελεστή φαίνονται στο Σχήμα 5.2(a), ενώ ανάλογα είναι και τα αποτελέσματα εφαρμογής του δεύτερου τελεστή.

- **Επίλυση Δομικών Συγκρούσεων**

-- **REVERSE_ATTRIBUTE** *ιδιοκτήτης*
OF CLASS *Αυτοκίνητο* TO *Άτομο*
INTO *κατέχει*
OF CLASS *Άτομο* TO *Αυτοκίνητο*

-- **XFORM_ATTRIBUTE** *λέσχη*
OF CLASS *Άτομο* TO *Λέσχη*
INTO CLASS *ΣυμμετοχήΣεΛέσχη* CONNECTED
VIA ATTRIBUTE *άτομο* TO *Άτομο* AND
VIA ATTRIBUTE *λέσχη* TO *Λέσχη*

-- **CREATE_SUBCLASS** *Προγραμματιστής* OF CLASS *Υπάλληλος*
AS SELECTION IN VALUE *"προγραμ"* OF ATTRIBUTE *είδος_εργ*

Τα αποτελέσματα εφαρμογής του πρώτου και του τρίτου τελεστή φαίνονται στα Σχήματα 5.3 και 5.5 αντίστοιχα, ενώ τα αποτελέσματα εφαρμογής του δεύτερου τελεστή είναι ανάλογα αυτών που φαίνονται στο Σχήμα 5.4.

Για το δεύτερο συστατικό σχήμα δηλώνουμε :

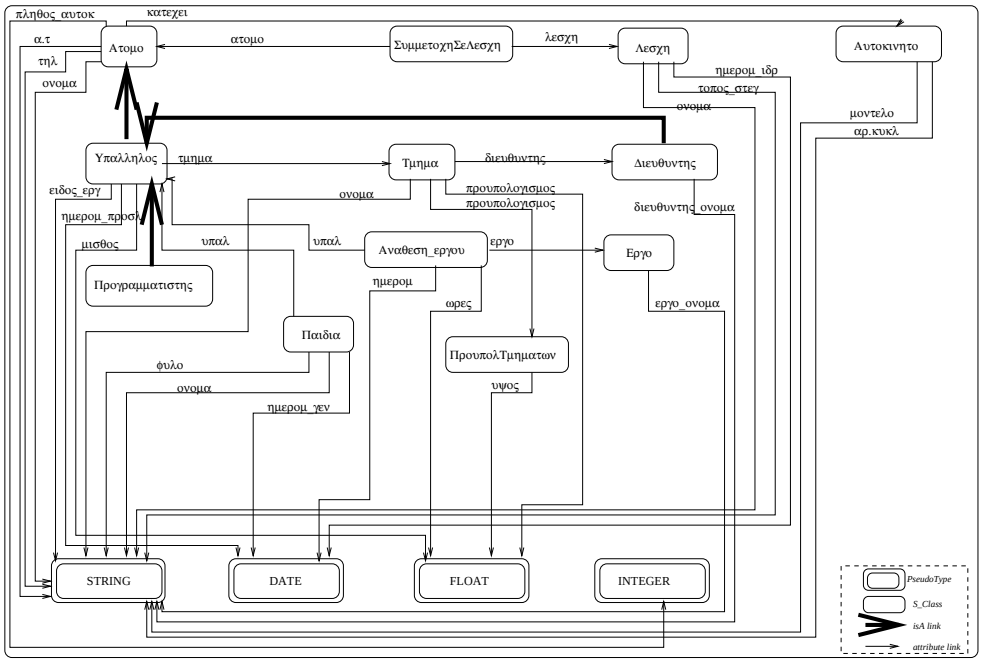
- **Επίλυση Δομικών Συγκρούσεων**

-- **XFORM_ATTRIBUTE** *συμμετέχει*
OF CLASS *Υπάλληλος* TO *Έργο*
INTO CLASS *Ανάθεση_Έργου* CONNECTED
VIA ATTRIBUTE *υπάλ* TO *Υπάλληλος* AND
VIA ATTRIBUTE *έργο* TO *Έργο*

Τα αποτελέσματα εφαρμογής του τελεστή φαίνονται στο Σχήμα 5.4.

5.3.3 Στάδιο Συγχώνευσης

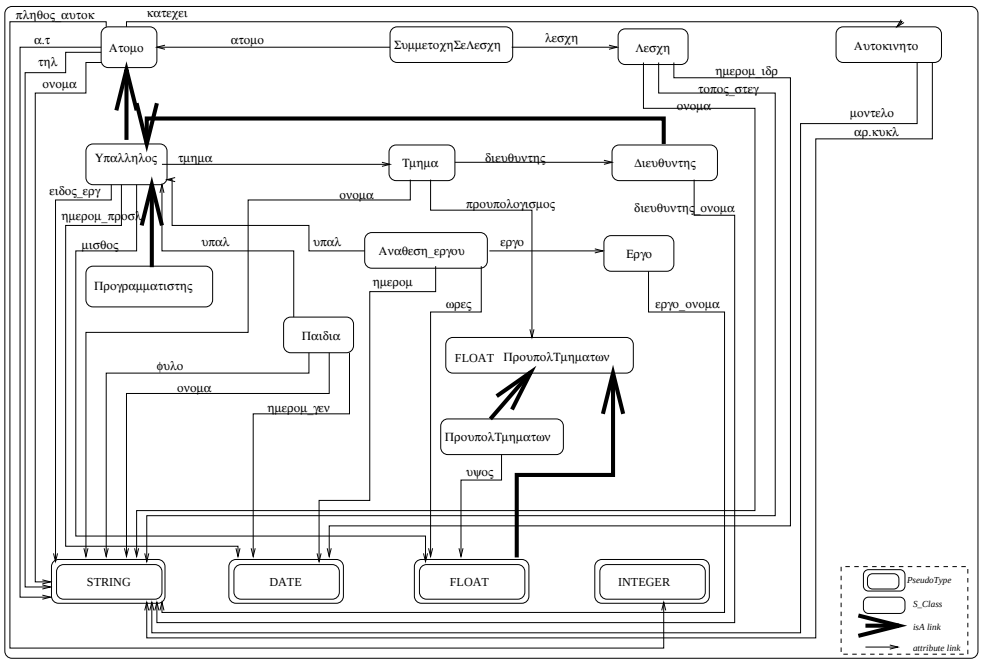
Στο στάδιο αυτό εφαρμόζεται ο τελεστής συγχώνευσης στα τροποποιημένα συστατικά σχήματα που προέκυψαν από το προηγούμενο στάδιο. Το αποτέλεσμα του σταδίου αυτού είναι το ασθενές σχήμα που φαίνεται στο Σχήμα 5.10.



Σχήμα 5.10: Ασθενές Συγχωνευμένο Σχήμα

5.3.4 Στάδιο Ανακατασκευής

Στο συγχωνευμένο ασθενές σχήμα του προηγούμενου σταδίου εφαρμόζεται η διαδικασία παραγωγής του τελικού νόμιμου σχήματος και αφαίρεσης των πλεονασμών που δίνει το τελικό ενοποιημένο Telos σημασιολογικό δίκτυο που φαίνεται στο Σχήμα 5.11.



Σχήμα 5.11: Τελικό Ενοποιημένο Σημασιολογικό Δίκτυο

Κεφάλαιο 6

Πρόταση Υλοποίησης

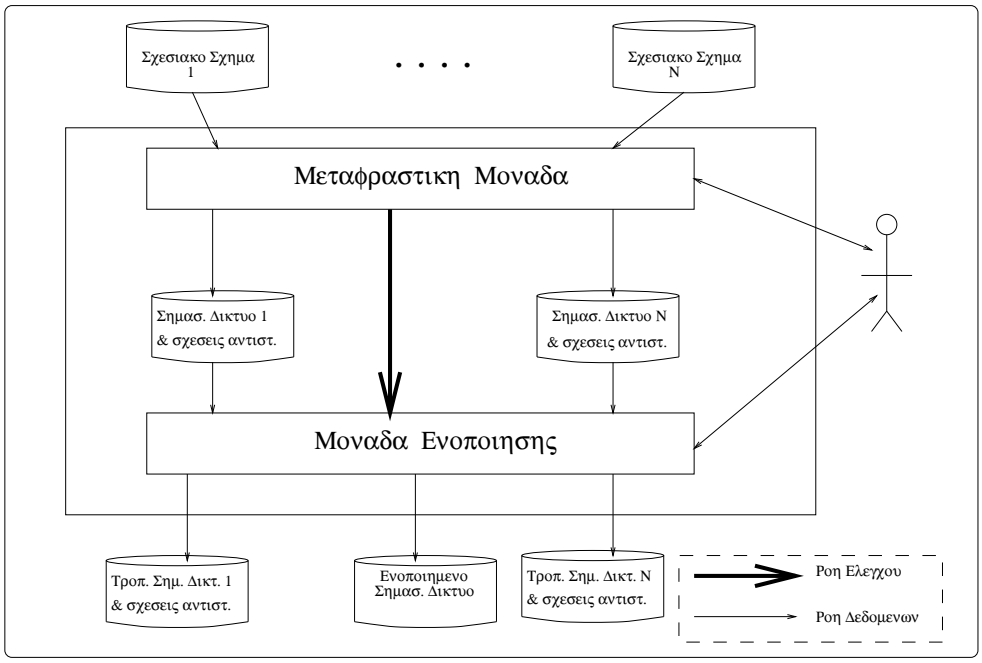
Στο κεφάλαιο αυτό κάνουμε μία πρόταση υλοποίησης ενός εργαλείου κατασκευής ομοσπονδιακού σχήματος που ακολουθεί το σημασιολογικό μοντέλο δεδομένων, από τοπικά σχεσιακά σχήματα αυτόνομων ΒΔ. Αρχικά θα αναλύσουμε την αδρή αρχιτεκτονική του εργαλείου, ενώ κατόπιν θα περάσουμε στην ανάλυση της αρχιτεκτονικής και λειτουργίας των τμημάτων του.

6.1 Αδρή Αρχιτεκτονική Εργαλείου

Στο Σχήμα 6.1 βλέπουμε την αδρή αρχιτεκτονική του εργαλείου που προτείνουμε. Το εργαλείο δέχεται σαν είσοδο ένα πλήθος Ν σχεσιακών σχημάτων και παράγει στην έξοδο του ένα σημασιολογικό δίκτυο που παίζει το ρόλο του ομοσπονδιακού σχήματος, καθώς και τις αντιστοιχίσεις δομών των τοπικών σχεσιακών σχημάτων με δομές του ομοσπονδιακού σημασιολογικού δικτύου. Εσωτερικά αποτελείται από δύο τμήματα: τη **Μεταφραστική Μονάδα** και τη **Μονάδα Ενοποίησης**.

Η Μεταφραστική Μονάδα δέχεται σαν είσοδο τα Ν τοπικά σχεσιακά σχήματα και για καθένα από αυτά παράγει στην έξοδο ένα συστατικό σημασιολογικό δίκτυο (που είναι η μετάφραση του τοπικού σχεσιακού σχήματος), καθώς και αντιστοιχίσεις δομών του τοπικού σχεσιακού σχήματος με δομές του δικτύου αυτού. Κατά τη διάρκεια της διαδικασίας μετάφρασης, που στηρίζεται στη μέθοδο μετάφρασης που αναπτύξαμε στο Κεφάλαιο 4, υπάρχει αλληλεπίδραση της Μονάδας Μετάφρασης με το χρήστη/σχεδιαστή για την επίλυση ασαφειών παράστασης.

Μετά το πέρας της διαδικασίας μετάφρασης των τοπικών σχημάτων, ο έλεγχος περνά στη Μονάδα Ενοποίησης η οποία δέχεται στην είσοδο της τα συστατικά σημασιολογικά δίκτυα με τις αντιστοιχίσεις και παράγει στην έξοδο ένα ενοποιημένο σημασιολογικό



Σχήμα 6.1: Αρχιτεκτονική Εργαλείου

δίκτυο (που παίζει το ρόλο του ομοσπονδιακού σχήματος), καθώς και τις αντιστοιχίσεις δομών των τοπικών σχεσιακών σχημάτων με δομές των τροποποιημένων σημασιολογικών δικτύων. Κατά τη διάρκεια της διαδικασίας ενοποίησης, υπάρχει αλληλεπίδραση με το χρήστη/σχεδιαστή για την αναγνώριση και επίλυση δομικών και ονομαστικών συγκρούσεων μεταξύ των συστατικών σημασιολογικών δικτύων.

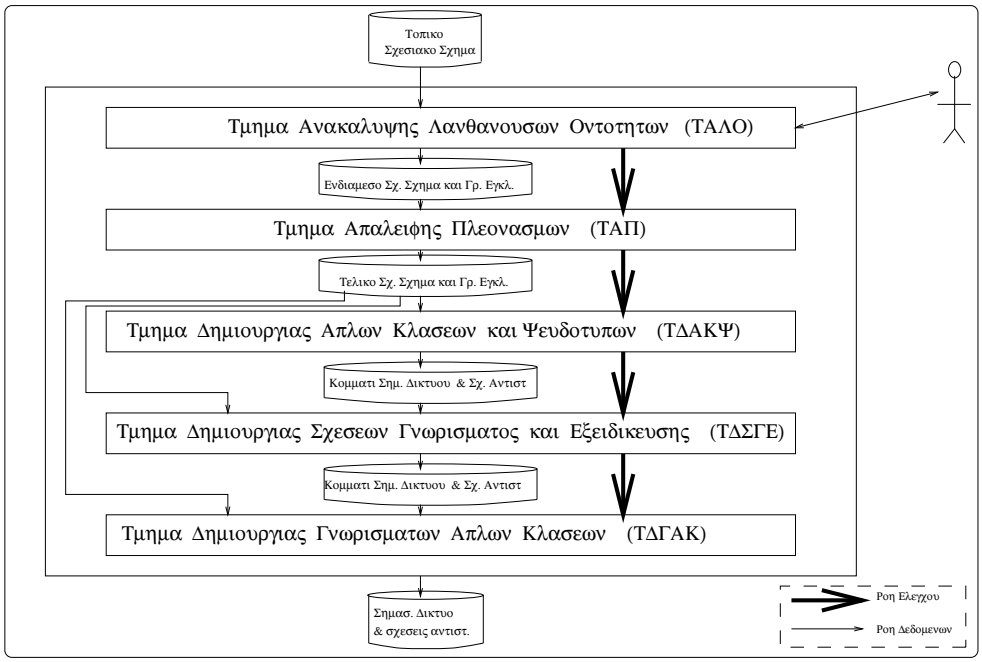
Ας δούμε όμως τώρα αναλυτικότερα την αρχιτεκτονική και τη λειτουργία των δύο μονάδων του εργαλείου.

6.2 Αρχιτεκτονική και Λειτουργία της Μεταφραστικής Μονάδας

Η αρχιτεκτονική της Μεταφραστικής Μονάδας φαίνεται στο Σχήμα 6.2.

Η λειτουργία της Μεταφραστικής Μονάδας στηρίζεται στη μέθοδο που παρουσιάσαμε στο Κεφάλαιο 4. Η είσοδος της Μεταφραστικής Μονάδας είναι ένα τοπικό σχεσιακό σχήμα σε 3NF, στο οποίο έχουν αναγνωρισθεί αντιπροσωπευτικά και μη-αντιπροσωπευτικά κλειδιά και έχουν εντοπιστεί οι ΕΕ για τα σχήματα σχέσεών του.

Η είσοδος αυτή περνά από το Τμήμα Ανακάλυψης Λανθανουσών Οντοτήτων (ΤΑΛΟ), όπου με τη βοήθεια του χρήστη κατασκευάζεται ένα ενδιάμεσο σχεσιακό σχήμα και ένας



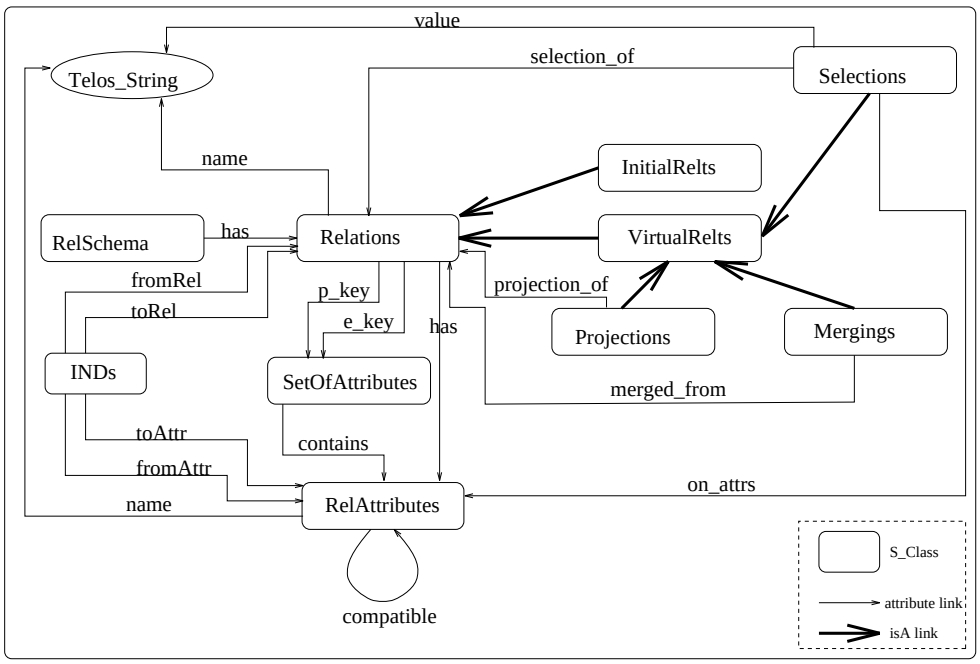
Σχήμα 6.2: Αρχιτεκτονική Μεταφραστικής Μονάδας

ενδιάμεσος γράφος εγκλεισμού. Στο σχήμα αυτό υπάρχουν τα αρχικά σχήματα σχέσεων, καθώς και εικονικά σχήματα σχέσεων που αντιστοιχούν στις λανθάνουσες οντότητες.

Κατόπιν, ο έλεγχος περνά στο Τμήμα Απαλειφής Πλεονασμών (ΤΑΠ), απ' όπου παράγεται το τελικό σχεσιακό σχήμα και ο τελικός γράφος εγκλεισμού. Στο σχήμα αυτό υπάρχουν αρχικά σχήματα σχέσεων, εικονικά σχήματα σχέσεων του ενδιάμεσου σχήματος και νέα εικονικά σχήματα σχέσεων που προέρχονται από συγχωνεύσεις σχεσιακών σχημάτων του ενδιάμεσου σχήματος.

Τέλος, ο έλεγχος περνά διαδοχικά από το Τμήμα Δημιουργίας Απλών Κλάσεων και Ψευδοτύπων (ΤΔΑΚΨ), το Τμήμα Δημιουργίας Σχέσεων Γνωρίσματος και Εξειδίκευσης (ΤΔΣΓΕ) και το Τμήμα Δημιουργίας Γνωρισμάτων Απλών Κλάσεων (ΤΔΓΑΚ), όπου με είσοδο το τελικό σχεσιακό σχήμα και τον τελικό γράφο εγκλεισμού παράγουν διαδοχικά τις απλές κλάσεις και ψευδοτύπους του σημασιολογικού δικτύου, τις σχέσεις γνωρίσματος και εξειδίκευσης μεταξύ απλών κλάσεων και τις σχέσεις γνωρίσματος απλών κλάσεων με ψευδοτύπους. Καθένα απ' τα τρία αυτά τμήματα, ταυτόχρονα με την παραγωγή κάποιου μέρους του σημασιολογικού δικτύου (απλές κλάσεις, σχέσεις γνωρίσματος, κλπ), αποθηκεύει τις σχέσεις αντιστοίχισης μεταξύ των δομών του τελικού σχεσιακού σχήματος και του σημασιολογικού δικτύου.

Τα δεδομένα που ανταλλάσσονται μεταξύ των τμημάτων της μονάδας, αποθηκεύονται σε μία Telos Βάση Δεδομένων. Έτσι, το τοπικό σχεσιακό σχήμα, το ενδιάμεσο σχεσιακό σχήμα, και το τελικό σχεσιακό σχήμα, αποθηκεύονται στο κομμάτι της Telos Βάσης Δεδομένων, που ακολουθεί το σημασιολογικό δίκτυο του Σχήματος 6.3.

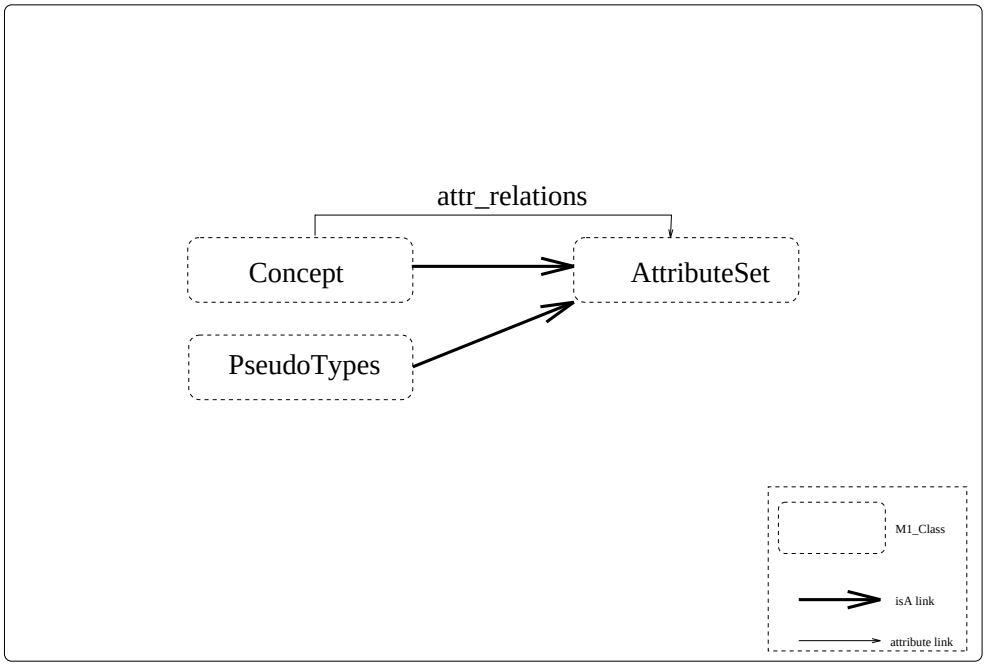


Σχήμα 6.3: Παράσταση σχεσιακού σχήματος στο Μεταφραστικό Μετασχήμα

Συγκεκριμένα, για το τοπικό σχεσιακό σχήμα, τα σχήματα σχέσεών του ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση InitialRelts, τα γνωρίσματα κάθε σχεσιακού σχήματος ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση RelAttributes, και οι ΕΕ μεταξύ των σχεσιακών σχημάτων ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση INDs. Για το ενδιάμεσο σχεσιακό σχήμα, τα εικονικά σχήματα σχέσεων που έχουν προκύψει, ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση Projections, ενώ ταυτόχρονα συνδέονται, μέσω της κατηγορίας projection_of, με το σχήμα σχέσης από το οποίο προήλθαν (με προβολή). Τα γνωρίσματα κάθε νέου σχήματος σχέσης ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση RelAttributes, ενώ ταυτόχρονα συνδέονται, μέσω της κατηγορίας compatible, με τα συμβατά γνωρίσματα του σχήματος σχέσης από το οποίο προήλθαν. Οι νέες ΕΕ που έχουν δημιουργηθεί, ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση INDs. Τέλος, για το τελικό σχεσιακό σχήμα, τα νέα εικονικά σχήματα σχέσεων που έχουν

προκύψει, ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση Mergings, ενώ ταυτόχρονα συνδέονται, μέσω της κατηγορίας merged_from, με τα σχήματα σχέσεων από τα οποία έχουν προέλθει (με συγχώνευση). Τα γνωρίσματα κάθε νέου σχήματος σχέσης ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση RelAttributes, ενώ ταυτόχρονα συνδέονται, μέσω της κατηγορίας compatible, με τα συμβατά γνωρίσματα των σχημάτων σχέσεων από τα οποία προήλθαν. Οι νέες ΕΕ που έχουν δημιουργηθεί, ταξινομούνται ως ατομικές οντότητες κάτω από την απλή κλάση INDs.

Τα κομμάτια του τελικού σημασιολογικού δικτύου, που παράγονται από τα τμήματα ΤΔΑΚΨ, ΤΔΣΓΕ και ΤΔΓΑΚ της Μεταφραστικής Μονάδας, αποθηκεύονται στο κομμάτι της Telos Βάσης Δεδομένων που ακολουθεί το σημασιολογικό δίκτυο του Σχήματος 6.4.



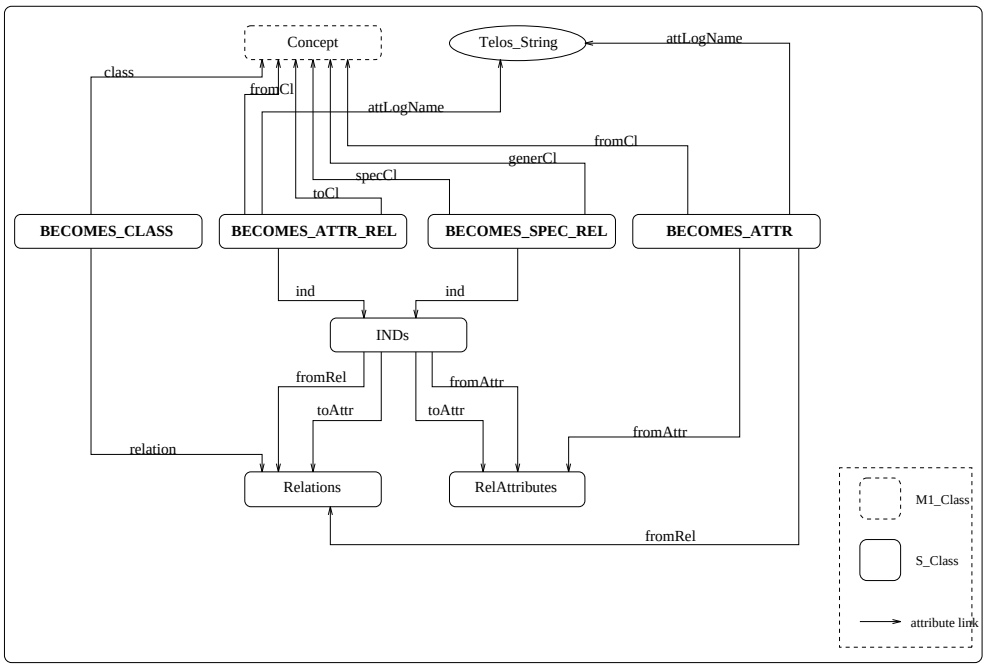
Σχήμα 6.4: Παράσταση σημασιολογικού δικτύου στο Μεταφραστικό Μετασχήμα

Για το τμήμα ΤΔΑΚΨ που επεξεργάζεται τα σχήματα σχέσεων του τελικού σχεσιακού σχήματος και εντοπίζει τους τύπους τιμών των γνωρισμάτων των σχεσιακών σχημάτων, οι απλές κλάσεις που παράγει ταξινομούνται κάτω από την μετακλάση Concept, ενώ οι ψευδοτύποι που παράγει ταξινομούνται κάτω από την μετακλάση PseudoTypes.

Για το τμήμα ΤΔΣΓΕ, που επεξεργάζεται τις ΕΕ του τελικού σχεσιακού σχήματος, οι σχέσεις γνωρίσματος που παράγει, ταξινομούνται κάτω από την μετακατηγορία attr_relations.

Για το τμήμα ΤΔΓΑΚ, που επεξεργάζεται τα γνωρίσματα των σχημάτων σχέσεων του τελικού σχεσιακού σχήματος, που είτε συμμετέχουν στο δεξί μόνο μέλος των ΕΕ, είτε δεν συμμετέχουν σε καμία ΕΕ, οι σχέσεις γνωρίσματος που παράγει, ταξινομούνται ομοίως κάτω από την μετακατηγορία attr_relations.

Τέλος, οι σχέσεις αντιστοίχισης, που παράγονται από τα τμήματα ΤΔΑΚΨ, ΤΔΣΓΕ και ΤΔΓΑΚ της Μεταφραστικής Μονάδας, αποθηκεύονται στο κομμάτι της ΒΔ, που ακολουθεί το σημασιολογικό δίκτυο του Σχήματος 6.5. Κάθε σχέση αντιστοίχισης που παράγεται από τα τρία αυτά τμήματα, ταξινομείται ως ατομική οντότητα κάτω από την απλή κλάση που αντιστοιχεί στο είδος της.



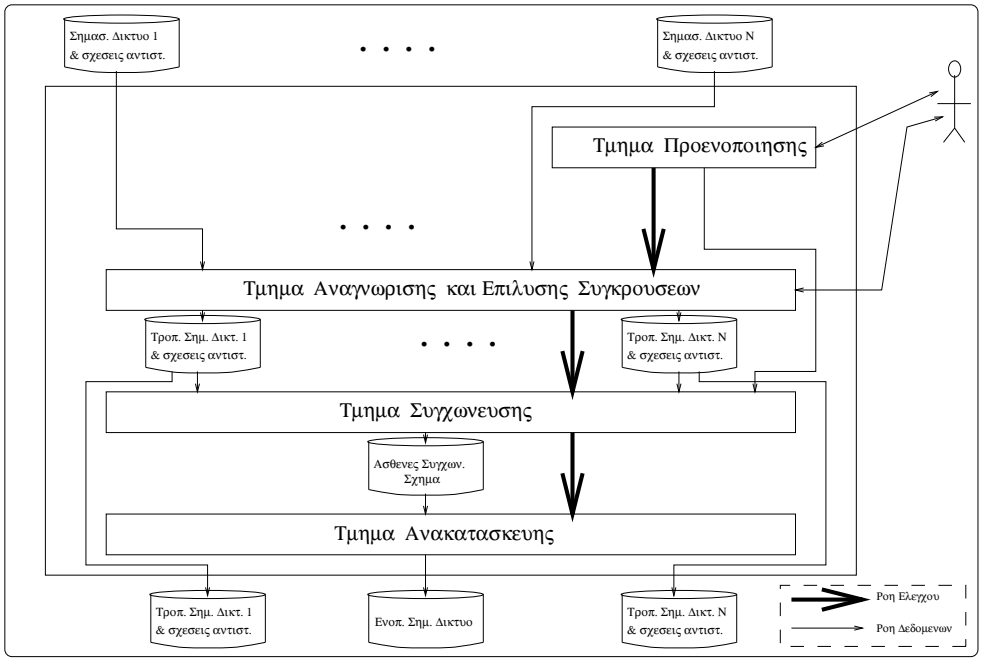
Σχήμα 6.5: Παράσταση σχέσεων αντιστοίχισης στο Μεταφραστικό Μετασχήμα

Έτσι τελικά, η Telos Βάση Δεδομένων που το σχήμα της περιγράφεται στα Σχήματα 6.3, 6.4 και 6.5, περιέχει το αρχικό, ενδιάμεσο και τελικό σχεσιακό σχήμα, το σημασιολογικό δίκτυο, και τις σχέσεις αντιστοίχισης που παρήγαγε η Μεταφραστική Μονάδα. Στο εξής θα ονομάζουμε το σχήμα αυτής της Telos Βάσης Δεδομένων, *Μεταφραστικό Μετασχήμα*.

Κλείνοντας, ας παρατηρήσουμε ότι το Μεταφραστικό Μετασχήμα που αναπτύξαμε, μοιάζει πολύ σαν ιδέα με τα Μεταμοντέλα Αντικειμένων και Ιδεωδών Πινάκων που χρησιμοποιεί το εργαλείο μετάφρασης σχημάτων OMTool ([6]), τα οποία (όπως αναφέραμε στο Κεφάλαιο 3) περιγράφουν τη δομή των σχημάτων εισόδου και εξόδου του εργαλείου.

6.3 Αρχιτεκτονική και Λειτουργία της Μονάδας Ενοποίησης

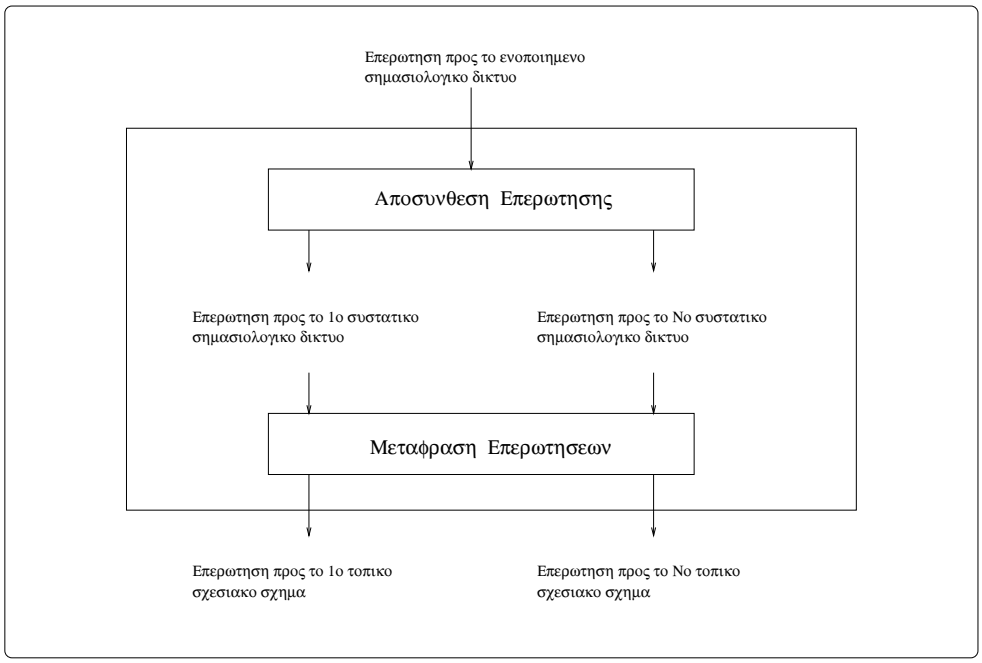
Η αρχιτεκτονική της Μονάδας Ενοποίησης φαίνεται στο Σχήμα 6.6.



Σχήμα 6.6: Αρχιτεκτονική Μονάδας Ενοποίησης

Η λειτουργία της Μονάδας Ενοποίησης στηρίζεται στη μέθοδο ενοποίησης, που περιγράψαμε στο Κεφάλαιο 5. Αρχικά, ο χρήστης/σχεδιαστής επικοινωνεί με το τμήμα προενοποίησης, για να καθορίσει το είδος της δυαδικής στρατηγικής που θα ακολουθήσει το τμήμα συγχώνευσης. Κατόπιν, ο έλεγχος περνά στο Τμήμα Αναγνώρισης και Επίλυσης Συγκρούσεων, στο οποίο ο χρήστης ανακαλύπτει τις ονομαστικές και δομικές συγκρούσεις των συστατικών σχημάτων της εισόδου και τις επιλύει με την εφαρμογή των τελεστών ονομαστικής και δομικής τροποποίησης. Είναι φανερό, ότι με την τροποποίηση των συστατικών σχημάτων τροποποιούνται και οι σχέσεις αντιστοίχισης, που τα συνοδεύουν. Μετά την επίλυση των συγκρούσεων, τα συστατικά σχήματα περνούν στο Τμήμα Συγχώνευσης όπου συγχωνεύονται σύμφωνα με τη δυαδική στρατηγική που έχει καθοριστεί από το τμήμα Προενοποίησης. Το αποτέλεσμα είναι ένα συγχωνευμένο ασθενές σχήμα που μετατρέπεται στο ενοποιημένο σημασιολογικό δίκτυο από το Τμήμα Ανακατασκευής. Το ενοποιημένο σημασιολογικό δίκτυο, τα τροποποιημένα συστατικά σημασιολογικά δίκτυα και οι σχέσεις αντιστοίχισής τους παρέχονται στην έξοδο της

Μονάδας Ενοποίησης προκειμένου να αξιοποιηθούν από *Εργαλεία Επερωτήσεων* του ενοποιημένου σημασιολογικού δικτύου. Τα εργαλεία αυτά λειτουργούν σε δύο φάσεις, όπως φαίνεται στο Σχήμα 6.7.



Σχήμα 6.7: Φάσεις Λειτουργίας Εργαλείων Επερωτήσεων

Αρχικά αποσυνθέτουν τις επερωτήσεις που γίνονται προς το ενοποιημένο σημασιολογικό δίκτυο σε επερωτήσεις προς τα συστατικά σημασιολογικά δίκτυα, και κατόπιν με τη βοήθεια των σχέσεων αντιστοίχισης τις μεταφράζουν προς τα αρχικά τοπικά σχεσιακά σχήματα.

Κεφάλαιο 7

Επίλογος

Κλείνοντας αυτή την εργασία, ας δούμε τα συμπεράσματα στα οποία καταλήξαμε και τις βελτιώσεις/επεκτάσεις που προτείνουμε.

7.1 Συμπεράσματα

Στην εργασία αυτή, αντιμετωπίσαμε το πρόβλημα της κατασκευής ενός ομοσπονδιακού σχήματος από τοπικά σχεσιακά σχήματα αυτόνομων Βάσεων Δεδομένων. Για την επίλυση του χρησιμοποιήσαμε αρχικά ως κανονικό μοντέλο δεδομένων, το σημασιολογικό μοντέλο στο οποίο εκφράζονται τα σχήματα των Βάσεων Δεδομένων που παράγονται από τη γλώσσα παράστασης γνώσης Telos. Κατόπιν, αναπτύξαμε μία μέθοδο μετάφρασης σχεσιακών σχημάτων σε σημασιολογικά δίκτυα και μία μέθοδο ενοποίησης σημασιολογικών δικτύων. Τέλος, προτείναμε την υλοποίηση ενός εργαλείου κατασκευής ομοσπονδιακού σχήματος που αποτελείται από ένα μεταφραστή μετατροπής και ένα επεξεργαστή δόμησης, η λειτουργία των οποίων στηρίζεται στις μεθόδους μετάφρασης και ενοποίησης.

Ως προς την επιλογή του κανονικού μοντέλου, το σημασιολογικό μοντέλο που διαλέξαμε να χρησιμοποιήσουμε πληρεί τα περισσότερα χαρακτηριστικά των ιδιοτήτων εκφραστικότητας και σημασιολογικού συσχετισμού που παρουσιάζονται στο [40]. Είναι λοιπόν αρκετά κατάλληλο για να εκφράσει τα συστατικά και το ομοσπονδιακό σχήμα.

Η Μέθοδος Μετάφρασης που αναπτύξαμε στο Κεφάλαιο 4, κάνει τις λιγότερες υποθέσεις για το είδος των σχεσιακών σχημάτων σε σχέση με όλες τις άλλες γνωστές μεθόδους ή αλγορίθμους μετάφρασης σχεσιακών σχημάτων. Από την άλλη μεριά όμως, υστερεί έναντι των αλγορίθμων μετάφρασης αφού απαιτεί την επέμβαση του χρήστη σε κάποια σημεία της.

Η Μέθοδος Ενοποίησης που προτείνουμε στο Κεφάλαιο 5, αποτελείται και από τα πέντε στάδια που προτείνονται στο πλαίσιο του [4]. Ο αλγόριθμος συγχώνευσης που χρησιμοποιεί το στάδιο συγχώνευσης, κάνει αναγκαία την επέμβαση του χρήστη **μόνο** στα στάδια ανακάλυψης και επίλυσης συγκρούσεων. Τέλος, το ενοποιημένο σχήμα που παράγει ικανοποιεί τα κριτήρια πληρότητας και ελαχιστότητας.

Το εργαλείο κατασκευής ομοσπονδιακού σχήματος που προτείναμε στο Κεφάλαιο 6, παράγει από ένα πλήθος σχεσιακών σχημάτων ένα ενοποιημένο σημασιολογικό δίκτυο που παίζει το ρόλο του ομοσπονδιακού σχήματος, καθώς και ένα σύνολο σχέσεων αντιστοίχισης μεταξύ δομών των σχεσιακών σχημάτων με δομές των συστατικών σημασιολογικών δικτύων. Οι σχέσεις αυτές αποθηκεύονται μαζί με τα συστατικά σημασιολογικά δίκτυα και τα σχεσιακά σχήματα σε Telos Βάσεις Δεδομένων που ακολουθούν το Μεταφραστικό Μετασχήμα. Δεδομένου ότι το σύνολο των σχέσεων αντιστοίχισης που είναι αποθηκευμένο στις Telos Βάση Δεδομένων περιέχει όλη την πληροφορία της αντιστοίχισης, δίνεται η δυνατότητα ανάπτυξης μεταφραστών επερωτήσεων, οι οποίοι από μία επερώτηση προς το ομοσπονδιακό σχήμα θα παράγουν επερωτήσεις προς τα τοπικά σχήματα βάσει της πληροφορίας αντιστοίχισης.

7.2 Επεκτάσεις - Βελτιώσεις

Στη Μέθοδο Ενοποίησης, θα μπορούσε να επεκταθεί το σύνολο των τελεστών τροποποίησης ώστε να επιλύει και άλλες περιπτώσεις συγκρούσεων. Πάντως, μέχρι στιγμής δεν έχουμε ανακαλύψει στη βιβλιογραφία κάποια δουλειά που να υποστηρίζει την *πληρότητα* κάποιου συνόλου τελεστών τροποποίησης. Πληρότητα με την έννοια ότι καλύπτουν όλες τις δυνατές περιπτώσεις συγκρούσεων.

Ακόμα, πιστεύουμε ότι είναι σχεδόν απαραίτητη η ανάπτυξη κάποιας μεθόδου ανακάλυψης συγκρούσεων η οποία θα καθοδηγεί το χρήστη να βρίσκει τις συγκρούσεις μεταξύ των σχημάτων όταν αυτά είναι πολύ μεγάλα.

Στη Μέθοδο Μετάφρασης μία πιθανή επέκταση είναι η ανάπτυξη μεθόδων μετάφρασης σχημάτων όχι μόνο του σχεσιακού αλλά και άλλων μοντέλων δεδομένων προς σημασιολογικά δίκτυα.

Βιβλιογραφία

[1] VLDB Journal. Issue is dedicated to the topic of "Federated Databases", 1(1), July 1992.

[2] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The Design and Analysis of Computer Algorithms.* Computer Science and Information Processing. Addison-Wesley, 1974.

[3] Joseph Albert, Rafi Ahmed, Mohammed Ketabchi, William Kent, and Ming-Chien Shan. Automatic Importation of Relational Schemas in Pegasus. In: Proceedings, 3rd International Workshop on Research Issues in Data Engineering : Interoperability in Multidatabase Systems, Vienna, Austria, pp. 105-113, 1993.

[4] C. Batini, M. Lenzerini, and S. B. Navathe. A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*, 18(4):pp. 322--364, December 1986.

[5] H. Bhargava, S. Kimbrough, and R. Krishnan. Unique Names Violations, a Problem for Model Integration or You Say Tomato, I say tomahto. *ORSA Journal on Computing*, 3(2):pp. 107--120, Spring 1991.

[6] Michael Blaha, William Premerlani, and Hwa Shen. Converting OO-Models into RDBMS Schema. *IEEE Software*, 11(3):pp. 28--39, May 1994.

[7] Alex Borgida, Matthias Jarke, Manolis Koubarakis, and John Mylopoulos. Telos: Representing Knowledge About Information Systems. *ACM Transactions on Information Systems*, 8(4):pp. 352--362, October 1990.

[8] Ralf Bottger, Yaron Engel, Gerd Kachel, Silvia Kolmschlag, Dietmar Nolte, and Elke Radeke. Enhancing the Data Openness of Frameworks by Database Federation Services. Cadlab--Joint Venture of University of Paderborn and Siemens Nixdorf Informationssysteme, 1995.

[9] Y. Breitbart, P.L. Olson, and G.R. Thompson. Database Integration in a Distributed Heterogeneous Database System. In *Proceedings Second International Conference on Data Engineering*, pages pp. 301--310, 1986.

[10] M.W. Bright, A.R. Hurson, and Simin H. Pakzad. A Taxonomy and Current Issues in Multidabase Systems. *IEEE Computer*, 25(3):pp. 50--59, March 1992.

[11] Malu Castellanos. A Methodology for Semantically Enriching Interoperable Databases. In *Proceedings, 11th British National Conference on Databases*, pages pp.58--75, 1993.

[12] Malu Castellanos and Felix Saltor. Extraction of Data Dependencies. In: Jaakkola, Kangassalo, Kitahashi and Markus (eds), Information Modelling and Knowledge Bases V (3rd European-Japanese Seminar on Information Modelling and Knowledge Bases , Budapest, 1993). IOS Press, Amsterdam,pp. 400-420,1994.

[13] Malu Castellanos, Felix Saltor, and Manuel Garcia-Solaco. Semantically Enriching Relational Databases into an Object-Oriented Semantic Model . In: D. Karagiannis (ed.): Database and Expert Systems Applications (5th International Conference DEXA'94, Athens, 1994). Springer Verlag, 1994, pp. 125-134.

[14] Stefano Ceri, Barbara Pernici, and Gio Wiederhold. Distributed Database Design Methodologies. *Proceedings of the IEEE*, 75(5):pp. 533--546, May 1987.

[15] Chin-Wan Chung. DATAPLEX : An Access to Heterogeneous Distributed Databases. *Communications of the ACM*, 33(1):pp.70--80, January 1990.

[16] D. Tsichritzis and F. Lochovsky. Eds. 1978. The ANSI/X3/SPARC DBMS framework. Inf. Syst. 3, 4.

[17] Umeshwar Dayal and Hai-Yann Hwang. View Definition and Generalization for Database Integration in a Multidatabase System. *IEEE Transactions on Software Engineering*, 10:pp. 628--645, November 1984.

[18] Manuel Garcia-Solaco, Malu Castellanos, and Felix Saltor. Discovering Interdatabase Resemblance of Classes for Interoperable Databases. In: Schek, Sheth and Czejdo (eds) Proceedings, 3rd International Workshop on Research Issues in Data Engineering: Interoperability in Multidatabase Systems, Vienna, Austria, 1993, pp.26-33.

[19] Manuel Garcia-Solaco, Felix Saltor, and Malu Castellanos. A Structured Based Schema Integration Methodology. Accepted for the 11th International Conference on Data Engineering, Taipei, March 1995.

[20] Χριστίνα Γκριτζάπη. Μεταφορά δεδομένων από μια σχεσιακή σε μια οντοκεντρική βάση δεδομένων. Μεταπτυχιακή Εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, Νοέμβριος 1995.

[21] John Grant, Witold Litwin, Nick Roussopoulos, and Timos Sellis. Query Languages for Relational Multidatabases. *VLDB Journal*, 2(2):pp. 153--171, 1993.

[22] Rajiv Gupta and Ellis Horowitz, editors. *Object-Oriented Databases with Applications to CASE, Networks, and VLSI CAD*, pages pp. 45--46. Data and Knowledge Base Systems. Prentice Hall, 1991.

[23] A.R. Hurson, M.W. Bright, and S. Pakzad, editors. *Multidatabase Systems: An Advanced Solution for Global Information Sharing*. IEEE Computer Society Press, 1994.

[24] Won Kim and Jungyun Seo. Classifying Schematic and Data Heterogeneity in Multidatabase Systems. *IEEE Computer*, 24(12):pp. 12--18, December 1991.

[25] Henry F. Korth and Abraham Silberschatz. *Database System Concepts*. Computer Science. McGraw-Hill Inc., second edition, 1991.

[26] Anthony Kosky. Modeling and Merging Database Schemas. Technical Report MS-CIS-91-65, Dept. of Computer and Information Sciences, University of Pennsylvania, September 1991.

[27] Anthony Kosky, Susan Davidson, and Peter Buneman. Semantics of Database Transformations. Technical Report MS-CIS-95-25, Dept. of Computer and Information Sciences, University of Pennsylvania, July 1995.

[28] Ravi Krishnamurthy, Witold Litwin, and William Kent. Interoperability of Heterogeneous Databases with Schematic Discrepancies. In *Proceedings, First International Workshop on Interoperability in Multidatabase Systems*, pages pp. 144--151, 1991.

[29] Witold Litwin and Abdelaziz Abdellatif. An Overview of the Multi-Database Manipulation Language MDSL. *Proceedings of the IEEE*, 75(5):pp. 621--631, May 1987.

[30] Witold Litwin, Leo Mark, and Nick Roussopoulos. Interoperability of Multiple Autonomous Database. *ACM Computing Surveys*, 22(3):pp. 267--293, September 1990.

[31] D. Maier, D. Rozenshtein, and J. Stein, editors. *Advances in Computing Research*, volume 3, chapter Window Functions, pages pp. 213--246. JAI Press, 1986.

[32] Salvatore T. March. ACM Computing Surveys. Special Issue : Heterogeneous Databases, 22(3), September 1990. ACM Press.

[33] Victor Markowitz and Arie Shoshani. Representing Extended Entity-Relationship Structures in Relational Databases: A Modular Approach. *ACM Transactions on Database Systems*, 17(3):pp. 423--464, September 1992.

[34] Victor M. Markowitz and Johann A. Makowsky. Identifying Extended Entity-Relationship Structures in Relational Schemas. *IEEE Transactions On Software Engineering*, 16(8):pp. 77--90, August 1990.

[35] Amihai Motro. Superviews : Virtual Integration of Multiple Databases. *IEEE Transactions On Software Engineering*, 13(7):pp. 785--798, July 1987.

[36] Tamer M. Ozsu and Patrick Valduriez. Distributed Database Systems : Where Are We Now ? *IEEE Computer*, 24(8):pp. 68--78, August 1991.

[37] Christine Parent, Stefano Spaccapietra, and Yahu Dupont. Model Independent Assertions for Integration of Heterogeneous Schemas. *VLDB Journal*, pages pp. 81--126, July 1992.

[38] Sudha Ram. IEEE Computer. Special Issue : Heterogeneous Distributed Database Systems, 24(12), December 1991. IEEE Computer Society Press.

[39] Felix Saltor, Malu Castellanos, and Manuel Garcia-Solaco. Overcoming Schematic Discrepancies in Interoperable Databases. In: Hsiao, Neuhold and Sacks-Davis (eds) Interoperable Database Systems (Proceedings, IFIP WG2.6 Database Semantics Conference on Interoperable Database Systems , Lorne, Victoria, Australia, 1992), 1993, pp. 191-205.

[40] Felix Saltor, Malu Castellanos, and Manuel Garcia-Solaco. Suitability of Data Models as Canonical Models for Federated Databases. *ACM SIGMOD Record*, 20(4):pp. 44--48, December 1991.

[41] Amit Sheth. ACM SIGMOD RECORD. Special Issue : Semantic Issues in Multidatabase Systems, 20(4), December 1991. ACM Press.

[42] Amit P. Sheth and James A. Larson. Federated Database Systems for Managing Distributed, Heterogeneous and Autonomous Databases. *ACM Computing Surveys*, 22(3):pp.183--236, September 1990.

[43] Marjorie Templeton, David Brill, Son K. Dao, Eric Lund, Patricia Ward, Arbee L.P. Chen, and Robert McGregor. Mermaid--A Front-End to Distributed Heterogeneous Databases. In *Proceedings IEEE*, pages pp. 695--708, May 1987.

[44] Marjorie Templeton, Herbert Henley, Edward Maros, and Darrel J. Van Buer. InterViso: Dealing with the Complexity of Federated Database Access. Data Integration Inc., 1994.

[45] Gomer Thomas, Glenn R. Thompson, Chin-Wan Chung, Edward Barkmeyer, Fred Carter, Marjorie Templeton, Stephen Fox, and Berl Hartman. Heterogeneous Distributed Database Systems for Production Use. *ACM Computing Surveys*, 22(3):pp. 237--266, September 1990.

[46] Jeffrey D. Ullman. *Principles of Database And Knowledge-Base Systems, Volume I : Classical Database Systems.* Computer Science Press, 1988.