Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ανάπτυξη δεδομένων εκμάθησης για συσχετιστικά νευρωνικά δίκτυα: εύρεση πυρήνα συνόλου δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ατσιδάκου Αλεξία

Επιβλέπων: Γιώργος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2019

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Ανάπτυξη δεδομένων εκμάθησης για συσχετιστικά νευρωνικά δίκτυα: εύρεση πυρήνα συνόλου δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ατσιδάκου Αλεξία

Επιβλέπων: Γιώργος Στάμου
Αναπληρωτής Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^{η} Νοεμβρίου 2019.

....	....	....
Γιώργος Στάμου	Ανδρέας-Γεώργιος Σταφυλοπάτης	Δημήτριος Φωτάκης
Αν. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Αν. Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2019

....
Αλεξία Ατσιδάκου
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Περίληψη

Τα νευρωνικά δίκτυα είναι ένα από τα βασικότερα εργαλεία που χρησιμοποιούνται στη μηχανική μάθηση. Η κύρια χρήση τους είναι η εύρεση μοτίβων ή συμπεριφορών που είναι δύσκολο να βρεθούν από τον άνθρωπο, είτε λόγω της μεγάλης περιπλοκότητας τους, είτε λόγω του τεράστιου χώρου στον οποίο πρέπει κανείς να αναζητήσει. Αν και η ιδέα των τεχνητών νευρωνικών δικτύων έχει προταθεί από τη δεκαετία του 1940, τις τελευταίες δεκαετίες έχουν γίνει ιδιαιτέρως δημοφιλή.

Στην παρούσα διπλωματική εργασία μελετάται το πρόβλημα της εύρεσης πυρήνα συνόλου οντολογικών δεδομένων, στα πλαίσια της ανάπτυξης δεδομένων εκμάθησης για συσχετιστικά νευρωνικά δίκτυα. Η παρούσα διπλωματική βρίσκεται στην τομή των νευρωνικών δικτύων με άλλες περιοχές της τεχνητής νημοσύνης, όπως η αναπαράσταση και αποθήκευση γνώσης. Η αναπαράσταση γνώσης είναι ο τομέας που μελετά πώς ένα υποσύνολο του κόσμου, τα γεγονότα, οι ιδέες του και άλλα χαρακτηριστικά του, μπορούν να μοντελοποιηθούν σε ένα σύστημα το οποίο μπορεί να υποστηρίζει αυτοματοποιημένη συλλογιστική. Μεγάλος αριθμός μεθόδων έχει προταθεί για την αναπαράσταση γνώσης, όπως η Λογική Πρώτη Τάξης, οι Λογικές Ανώτερης Τάξης και τα Σημασιολογικά δίκτυα. Ο διάφορες προσεγγίσεις ταλαντεύονται ανάμεσα στο βαθμό εκφραστικότητας της περιγραφής και την υπολογιστική πολυπλοκότητα, καθώς η εκφραστικότερη αναπαράσταση αυξάνει τη δυσκολία εξαγωγής συμπερασμάτων από την αποθηκευμένη γνώση.

Στην παρούσα διατριβή θα μελετηθούν διάφορα προβλήματα εύρεσης ελαχιστοτικών υποσυνόλων μίας βάσης γνώσης. Αρχικά, θα μελετηθεί το πρόβλημα της ελαχιστοποίησης της αποθηκευμένης γνώσης, αλλά και το αντίστοιχο πρόβλημα υπό την παρουσία ερωτημάτων που μπορεί να τεθούν στη γνώση, ενώ στη συνέχεια η μελέτη θα επεκταθεί στην εύρεση όλων των ελαχιστοτικά ισοδύναμων υποσυνόλων της γνώσης. Τα αποτελέσματα μπορούν να χρησιμοποιηθούν από ένα νευρωνικό δίκτυο με σκοπό την εκμάθηση απάντησης ερωτημάτων, ώστε να αποφευχθεί η εκτενής αποθήκευση γνώσης και η χρονοβόρα διαδικασία συλλογιστικής που συνοδεύει την εξαγωγή συμπερασμάτων.

Λέξεις κλειδιά: Νευρωνικά Δίκτυα, Μηχανική Μάθηση, Αναπαράσταση Γνώσης, Συλλογιστική, Ελαχιστοποίηση Βάσης Γνώσης

Abstract

Neural Networks are a popular tool that is used for Machine Learning purposes. They are mainly used for finding patterns and behaviors that cannot be easily found by humans, either because they are very complex or because in order to be found, a huge space needs to be searched. Although Neural Networks have been proposed since the 1940s, they have gained popularity during the last decades.

In this thesis, we examine the problem of finding the core of an Ontology, for the purpose of creating training data for Relational Neural Networks. This study is located on the intersection of Neural Networks with another area of Artificial Intelligence, Knowledge Representation and Reasoning. The field of knowledge representation tries to address the problem of how a subset of the information of the world, the events, the ideas and the other details that it contains, can be modeled in a system that can support automated reasoning. Numerous methods of knowledge representation have been proposed, such as First Order Logic, Higher Order Logics and Semantic Networks. These approaches waver between expressiveness and computational complexity, since increasing the expressiveness can increase the difficulty of drawing conclusions about the knowledge represented.

This thesis examines different problems of knowledge base minimization. Firstly, we investigate the problem of minimizing the knowledge, with or without taking into account the existence of relevant queries that can be posed over the knowledge base. Then, the examination is extended to the problem of finding all minimal subsets of the knowledge base. The results can be used as training data, in order for a neural network to be able to answer ontological questions. This approach avoids storing vast amounts of ontological data, as well as the time-consuming query answering process of the traditional knowledge representation systems.

Keywords: Neural Networks, Machine Learning, Knowledge Representation, Reasoning, Knowledge Base Minimisation

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά όσους συνέβαλαν στην παρούσα διατριβή. Αρχικά, θέλω να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Γιώργο Στάμου, για την εξαιρετική διδασκαλία του, η οποία με ενέπνευσε να ασχοληθώ με τον τομέα της Τεχνητής Νοημοσύνης, αλλά και για τις συμβουλές και την καθοδήγηση που μου παρείχε για την διεκπεραίωση της διπλωματικής μου εργασίας. Επίσης, θα ήθελα να ευχαριστήσω τον Αλέξανδρο Χορταρά για τις συμβουλές του, στο πλαίσιο της διπλωματικής εργασίας.

Παράλληλα, θα ήθελα να ευχαριστήσω όσους στάθηκαν δίπλα μου κατά τη διάρκεια των φοιτητικών μου χρόνων. Αρχικά, θα ήθελα να ευχαριστήσω την οικογένειά μου για την απεριόριστη αγάπη και την υποστήριξή τους, όλα αυτά τα χρόνια. Επιπλέον, θα ήθελα να ευχαριστήσω το Γιάννη, για τις αμέτρητες στιγμές χαράς και την αμέριστη συμπαράσταση του σε στιγμές δυσκολίας. Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια του Γιάννη, για τη φροντίδα και την υποστήριξή τους. Τέλος, θέλω να ευχαριστήσω τους φίλους που έκανα κατά τη διάρκεια των σπουδών μου, με τους οποίους περάσαμε μαζί όλες τις φοιτητικές στιγμές, εύκολες και δύσκολες.

Αλεξία

Περιεχόμενα

Λίστα Αλγορίθμων

Λίστα Πινάκων

Κεφάλαιο 1

Εισαγωγή

Οι περιγραφικές λογικές έχουν προταθεί ως μέθοδος αναπαράστασης γνώσης και εξαγωγή συμπερασμάτων λόγω της καλής ισορροπίας εκφραστικότητας, και υπολογιστικής πολυπλοκότητας που παρέχουν. Για παράδειγμα, μία ιδιαιτέρως εκφραστική μέθοδος αναπαράστασης γνώσης είναι η Λογική Πρώτης Τάξης, στην οποία όμως πολλά προβλήματα εξαγωγής συμπερασμάτων είναι μη επιλύσιμα. Μία άλλη μέθοδος αναπαράστασης γνώσης είναι η προτασιακή λογική, όμως αυτή η μέθοδος πάσχει από το μειονέκτημα της πολύ περιορισμένης εκφραστικότητας, δηλαδή των δυνατοτήτων αναπαράστασης που προσφέρει. Οι περιγραφικές λογικές προσφέρουν τη δυνατότητα καταγραφής ιδιοτήτων (εννοιών) των ατόμων του κόσμου, αλλά και διμελών σχέσεων (ρόλων) μεταξύ τους. Επιπλέον, προσφέρουν τη δυνατότητα καταγραφής σχέσεων μεταξύ των εννοιών και των ρόλων, αυτών. Από τις διάφορες περιγραφικές λογικές που έχουν κατά καιρούς προταθεί, στο πλαίσιο της παρούσας εργασίας θα χησιμοποιηθεί η λογική DL-Lite$_{\mathrm{R}}$, τα βασικά δομικά στοιχεία της οποίας θα περιγραφούν στη συνέχεια.

Οι πληροφορίες που είναι αποθηκευμένες σε μία βάση γνώσης δεν είναι απαραίτητα ρητά καταγεγραμμένες, αλλά συχνά μπορούν να υπονοηθούν από τα καταγεγραμμένα στοιχεία της. Συνεπώς, είναι ζητούμενη η εξαγωγή συμπερασμάτων από την αποθηκευμένη γνώση, δηλαδή η εκμαίευση πληροφοριών που δεν είναι ρητά καταγεγραμμένες, και η διαδικασία αυτή ονομάζεται συλλογιστική. Υπάρχουν διάφορα προβλήματα συλλογιστικής, ανάλογα με το είδος των πληροφοριών που είναι επιθυμητό να εξαχθούν. Συχνά προβλήματα συλλογιστικής αποτελούν ο έλεγχος εάν ένα άτομο του κόσμου διαθέτει ένα συγκεκριμένο χαρακτηριστικό, ή εάν ένα χαρακτηριστικό είναι υποσύνολο ενός άλλου.

Ένα βασικό πρόβλημα συλλογιστικής, με το οποίο σχετίζεται σε μεγάλο βαθμό η παρούσα διατριβή, είναι το πρόβλημα της απάντησης ερωτημάτων. Στα πλαίσια αυτού του προβλήματος, ένα ερώτημα τίθεται στη βάση γνώσης, το οποίο συνήθως περιγράφει κάποια επιθυμητά χαρακτηριστικά, και η λύση του περιλαμβάνει την εύρεση όλων των ατόμων που διαθέτουν τα χαρακτηριστικά αυτά, σύμφωνα με τις ρητά, αλλά και τις μη ρητά καταγεγραμμένες πληροφορίες της γνώσης. Μία δημοφιλής προσέγγιση για την εύρεση των απαντήσεων των ερωτημάτων που μπορεί να τεθούν στη γνώση είναι η επαναγραφή ερωτημάτων, στην οποία το αρχικό ερώτημα επαναγράφεται σε ένα σύνολο από ερωτήματα, τα οποία μπορούν να τεθούν στη γνώση χωρίς να απαιτείται επιπλέον συλλογιστική κατά την απάντησή τους, δηλαδή αρκεί να ελεγχθούν οι ρητά καταγεγραμμένες πληροφορίες. Η επαναγραφή ερωτημάτων θα χρησιμοποιηθεί από την παρούσα διατριβή για την επίλυση του προβλήματος της ελαχιστοποίησης της γνώσης υπό την παρουσία ερωτημάτων.

Η ελαχιστοποίηση γνώσης περιλαμβάνει την εύρεση ενός ελαχιστοτικού υποσυνόλου της γνώσης από το οποίο μπορούν να εξαχθούν τα ίδια συμπεράσματα με την αρχική γνώση. Το πρόβλημα αυτό βασίζεται στην ύπαρξη περιττών πληροφοριών που είναι καταγεγραμμένες στη γνώση, των οποίων η αφαίρεση δεν μειώνει πραγματικά την υπονούμενη γνώση. Ο αλγόριθμος ελαχιστοποίησης που προτείνεται στην παρούσα εργασία στοχεύει στην αφαίρεση των περιττών ισχυρισμών. Επιπλέον, ο αλγόριθμος εύρεσης όλων των ελαχιστοτικών ισοδύναμων υποσυνόλων της γνώσης, που προτείνεται στη συνέχεια, βασίζεται στην δυνατότητα εύρεση ενός τέτοιου υποσυνόλου.

Η ελαχιστοποίηση γνώσης υπό την παρουσία ενός συνόλου ερωτημάτων περιλαμβάνει την εύρεση ενός ελαχιστοτικού υποσυνόλου της γνώσης, το οποίο δεν είναι απαραίτητα ιοδυναμο με το αρχικό, όμως δίνει τις ίδιες απαντήσεις εάν τεθούν σε αυτό τα ερωτήματα του συνόλου που αναφέρθηκε. Το πρόβλημα αυτό διαφέρει από την απλή ελαχιστοποίηση της γνώσης, καθώς προστίθενται νέοι λόγοι για τους οποίους κάποιες πληροφορίες της γνώσης μπορεί να είναι περιττές. Είναι εμφανές ότι το πρόβλημα αυτό σχετίζεται με το πρόβλημα της απάντησης ερωτημάτων. Στην προσέγγιση που προτείνεται στην παρούσα διατριβή χρησιμοποιείται η ιδέα της επαναγραφής ερωτημάτων για τον προσδιορισμό όλων των ελαχιστοτικών υποσυνόλων της γνώσης, τα οποία είναι ισοδύναμα με την αρχική γνώση ως προς ένα σύνολο ερωτημάτων.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή

Σε αυτό το κεφάλαιο ο αναγνώστης θα εφοδιαστεί με τα βασικά εργαλεία που απαιτούνται για την κατανόηση των επόμενων κεφαλαίων.

Αρχικά, θα γίνει μία εισαγωγή στις περιγραφικές λογικές, το βασικό φορμαλισμό αναπαράστασης γνώσης που χρησιμοποιείται στη συνέχεια. Θα αναλυθούν τα δομικά στοιχεία των περιγραφικών λογικών καθώς και ο τρόπος χρήσης τους. Επιπλέον, θα παρουσιαστεί ο τρόπος με τον οποίο τα δομικά στοιχεία αυτά συνδυάζονται για την κατασκευή μίας Βάσης Γνώσης, δηλαδή μίας δομής που μπορεί να αποθηκεύσει ρητά και μη ρητά καταγεγγραμμένη γνώση. Τέλος, θα παρουσιαστεί ο τρόπος με τον οποίο τα δομικά στοιχεία της βάσης γνώσης μπορούν να αποκτήσουν ερμηνεία, δηλαδή σαφώς διατυπωμένο νόημα στον κόσμο.

Μία Βάση Γνώσης μπορεί να χρησιμοποιηθεί για την εξαγωγή συμπερασμάτων από την γνώση που είναι αποθηκευμένη σε αυτήν, καθώς και για την απάντηση διαφόρων ερωτημάτων που μπορεί να τεθούν σε αυτή. Με αυτή την ευκαιρία θα γίνει αναφορά στα βασικότερα προβλήματα συλλογιστικής και θα παρουσιαστεί η οικογένεια περιγραφικών λογικών $DL-Lite$, η οποία παρουσιάζει καλή πολυπλοκότητα επίλυσης των προβλημάτων συλλογιστικής, με τη θυσία της πιο περιορισμένης δυνατότητας εκφραστικότητας για την αποθηκευμένη γνώση.

Μεταξύ των προβλημάτων συλλογιστικής, θα γίνει εκτενής αναφορά στο πρόβλημα της απάντησης ερωτημάτων και στις προσεγγίσεις που ακολουθούνται για τη λύση του, καθώς στοιχεία των προσεγγίσεων αυτών θα δανειστούν τα επόμενα κεφάλαια της διατριβής αυτής.

2.2 Περιγραφικές Λογικές

Οι Περιγραφικές Λογικές (ΠΛ) αποτελούν μία οικογένεια γλωσσών αναπαράστασης γνώσης που μπορούν να αναπαραστήσουν την πληροφορία που περιέχεται σε ένα σύστημα με τυπικό και δομημένο τρόπο, ενώ παράλληλα προσφέρουν το πλεονέκτημα της εύκολης κατανόησης της πληροφορίας που αποθηκεύουν. Οι περιγραφικές λογικές διαφέρουν από εναλλακτικές μεθόδους αναπαράστασης γνώσης που έχουν προταθεί, όπως τα σημασιολογικά δίκτυα, στο γεγονός ότι παρέχουν τυπική σημασιολογική αντιστοίχιση με τα διάφορα είδη μαθηματικής λογικής. Πιο συγκεκριμένα, οι περισσότερες ΠΛ μπορούν να αντιστοιχιστούν σε τμήματα της Κατηγορηματικής Λογικής (Πρώτης Τάξης) [7],[21],[3], ενώ ο εφοδιασμός των περιγραφικών γλωσσών με τελεστές που παρέχουν μεγαλύτερη εκφραστικότητα, όπως η δήλωση μεταβατικότητας των ρόλων, μπορεί να τις καταστήσει αντίστοιχες με λογικές ανώτερης τάξης[5].

Ένα μεγάλο πλεονέκτημα των περιγραφικών λογικών σε σύγκριση με διάφορα είδη μαθηματικής λογικής (Προτασιακός Λογισμός, Λογική Πρώτης Τάξης, Λογικές Ανώτερης Τάξης) είναι το ότι προσφέρουν μία καλή ισορροπία ανάμεσα στην εκφραστικότητα για την αναπαράσταση των εκάστοτε πληροφοριών και στην ταχύτητα επίλυσης των διαφόρων προβλημάτων συλλογιστικής. Η Προτασιακή Λογική παρουσιάζει το μειονέκτημα της πολύ περιορισμένης εκφραστικότητας, ενώ η Λογική Πρώτης Τάξης είναι μη αποφασίσιμη. Όπως προαναφέρθηκε, οι Περιγραφικές Λογικές κυμαίνονται από καθαρά υποσύνολα της λογικής πρώτης τάξης, έως υπερσύνολά της, ως προς το βαθμό εκφραστικότητας της γλώσσας, δηλαδή ως προς το είδος των πληροφοριών που μπορεί να αναπαραστήσει. Η επιλογή ΠΛ συγκεκριμένης εκφραστικότητας καθορίζει την ταχύτητα επίλυσης προβλημάτων και την ευκολία εξαγωγής συμπερασμάτων (reasoning) από τη γλώσσα.

Στη συνέχεια, παρουσιάζονται τα βασικά δομικά χαρακτηριστικά των ΠΛ, καθώς και η τυπική σημασία τους.

2.2.1 Δομικά στοιχεία Περιγραφικών Λογικών

Όπως και η Λογική Πρώτης Τάξης, οι ΠΛ περιέχουν δύο κατηγορίες συμβόλων: τα λογικά και τα μη λογικά σύμβολα.

Τα μη λογικά σύμβολα των ΠΛ χωρίζονται σε τρεις κατηγορίες: τα ονόματα ατόμων (individual names - IN), τα ονόματα εννοιών (concept names - CN) και τα ονόματα ρόλων (role names - RN)[8]. Διαισθητικά μπορεί να φαντάζεται κανείς τα ονόματα ατόμων ως προσδιοριστικά για τα αντικείμενα του κόσμου, τις έννοιες ως προσδιοριστικά για τα πιθανά χαρακτηριστικά των ατόμων και τους ρόλους ως προσδιοριστικά για τις πιθανές διμελείς σχέσεις μεταξύ ατόμων.

Παράδειγμα 2.1. Για παράδειγμα, εάν θέλουμε να περιγράψουμε τις σχέσεις των μελών μίας οικογένειας, τότε τα άτομα του κόσμου ίσως υποδηλώνουν ονόματα των ατόμων της οικογένειας, οι έννοιες ίσως υποδηλώνουν ιδιότητες που μπορεί να αποδίδονται σε κάποιον μέσω της συμμετοχής του στην οικογένεια, ενώ οι ρόλοι μπορεί να υποδηλώνουν τις διμελείς συσχετίσεις που μπορεί να έχουν τα άτομα αυτά. Έτσι, ίσως έχουμε:

$$IN = \{\text{Μαρία}, \text{Κωνσταντίνος}, \text{Ελένη}, \text{Ναυσικά}\}$$

$$CN = \{\text{Άντρας}, \text{Γυναίκα}, \text{Πατέρας}, \text{Γονέας}, \text{Μητέρα}\}$$

$$RN = \{\text{παντρεμμένοςΜε}, \text{έχειΠαιδί}, \text{έχειΓονέα}\}$$

Μία σημαντική παρατήρηση είναι ότι η αντιστοιχία ατόμων και αντικειμένων του κόσμου δεν είναι αμφιμονοσήμαντη, αφού οι ΠΛ επιτρέπουν τη σημασιολογική αντιστοίχηση περισσότερων του ενός ονομάτων ατόμων σε κάποιο αντικείμενο του κόσμου. Με άλλα λόγια, στις ΠΛ συνήθως δεν ισχύει η υπόθεση μοναδικού ονόματος (unique name assumption).

Επιπλέον, αξίζει να σημειωθεί ότι, αν και στο παραπάνω παράδειγμα έγινε προσπάθεια να συσχετιστούν τα ονόματα των μη λογικών συμβόλων με διαισθητικές έννοιες, στην πραγματικότητα τα μη λογικά σύμβολα των ΠΛ δεν έχουν προκαθορισμένη ερμηνεία, μόνο συγκεκριμένο τρόπο χρήσης. Αργότερα θα παρουσιαστεί ο τρόπος με τον οποίο τα μη λογικά σύμβολα μπορούν να αποκτήσουν συγκεκριμένη σημασία στον κόσμο. Μετά την απόδοση ερμηνείας στα μη λογικά σύμβολα, τα ονόματα των συμβόλων αυτών είναι δυνατόν να μην αντιπροσωπεύουν το πραγματικό νόημα που τους αποδόθηκε.

Τα λογικά σύμβολα των ΠΛ είναι σύμβολα των οποίων η σημασία είναι προκαθορισμένη και δεν εξαρτάται από την εκάστοτε εφαρμογή. Μερικά από αυτά μπορούν να χρησιμοποιηθούν ως κατασκευαστές (constructors) για τη δημιουργία νέων ρόλων ή εννοιών από υπάρχοντες ρόλους ή έννοιες. Τα πιο συνηθισμένα λογικά σύμβολα είναι τα παρακάτω:

- στίξη: "(", ")", "{", "}"
- θετικοί ακέραιοι: 1,2,3, κ.λ.π
- κατασκευαστές: $\neg, \exists, \forall, \sqcap, \sqcup, \geq, \leq, \circ, ^-$
- σύμβολα σύγκρισης: $\equiv, \sqsubseteq, \sqsupseteq, \approx, \not\approx$
- σταθερά σύμβολα εννοιών: $\top, \bot$

Όπως αναφέρθηκε, η σημασία των λογικών συμβόλων είναι προκαθορισμένη. Παραπάνω μπορεί κανείς να εντοπίσει κάποια συμβολα, όπως οι ποσοδείκτες $\forall, \exists$ και οι θετικοί ακέραιοι $1, 2, 3$, κ.λ.π., τα οποία έχουν γνωστή σημασία, και άλλα που είναι λιγότερο διαδεδομένα, όπως οι λογικές έννοιες $\top, \bot$ που υποδηλώνουν την καθολική και την κενή έννοια, αντίστοιχα. Επιπλέον, τα σύμβολα σύγκρισης διαισθητικά χρησιμοποιούνται για να δηλώσουν ότι μία έννοια είναι υποπερίπτωση ($\sqsubseteq$) ή ισοδύναμη ($\equiv$) με κάποια άλλη έννοια, ή ότι δύο ονόματα ατόμων αναφέρονται στο ίδιο αντικείμενο του κόσμου ($\approx$). Τέλος, κάποια από τα γνωστότερα, ίσως, λογικά σύμβολα είναι η άρνηση ($\neg$), η σύζευξη ($\sqcap$) και διάζευξη ($\sqcup$), ενώ σύμβολα όπως η σύνθεση ρόλων ($\circ$) και η αντιστροφή ρόλου ($^-$) είναι ίσως λιγότερο αναγνωρίσιμα. Η τυπική σημασία των λογικών συμβόλων θα δοθεί στη συνέχεια.

Τα βασικά δομικά στοιχεία των Περιγραφικών Λογικών είναι τα άτομα, οι έννοιες και οι ρόλοι. Τα άτομα προκύπτουν άμεσα από το σύνολο των ονομάτων ατόμων. Οι έννοιες (όπως και οι ρόλοι) μπορεί να είναι είτε ατομικές, δηλαδή να αντιστοιχούν σε ονόματα εννοιών (ρόλων), ή σύνθετες. Σύνθετες έννοιες και ρόλοι μπορεί να κατασκευαστούν από στοιχειώδεις έννοιες και ρόλους, όπως φαίνεται από τους παρακάτω ορισμούς.

Ορισμός 2.1. *Κάθε στοιχείο του συνόλου ονομάτων εννοιών είναι μία έννοια.Έστω δύο έννοιες A και B, r ένα όνομα ρόλου, a ένα όνομα ατόμου και n ένας θετικός ακέραιος. Τότε και οι παρακάτω εκφράσεις είναι έννοιες:*

- $\neg A$
- $\exists r.A$
- $\forall r.A$
- $A \sqcap B$
- $A \sqcup B$

- $\geq nr.A$
- $\leq nr.A$
- $\{a\}$
- $\top$
- $\bot$

Παράδειγμα 2.2. Για παράδειγμα, από τα ονόματα εννοιών και ρόλων του προηγούμενου παραδείγματος μπορεί κανείς να κατασκευάσει μία έννοια που θα υποδηλώνει τις μητέρες μέσω της τομή των εννοιών της γυναίκας και του γονέα

$$\text{Γυναίκα} \sqcap \text{Γονέας}$$

Επιπλέον, η έννοια του γονέα θα μπορούσε εναλλακτικά να περιγραφεί από την σύνθετη έννοια:

$$\text{Πατέρας} \sqcup \text{Μητέρα}$$

Παράλληλα, κάποιος που δεν έχει παιδιά θα μπορούσε να περιγραφεί από την σύνθετη έννοια

$$\neg\exists\text{έχειΠαιδί}.\top$$

ενώ κάποιος που κάθε παιδί του είναι άντρας θα μπορούσε να περιγραφεί από την

$$\forall\text{έχειΠαιδί}.\text{Άντρας}$$

Ορισμός 2.2. *Κάθε στοιχείο του συνόλου ονομάτων ρόλων είναι ρόλος. Έστω r και s ονόματα ρόλων. Τότε και οι παρακάτω εκφράσεις είναι ρόλοι:*

- $r \circ s$
- r^-

Παράδειγμα 2.3. Για παράδειγμα, ο ρόλος έχειΓονέα θα μπορούσε να περιγραφεί εναλλακτικά από τον σύνθετο ρόλο:

$$\text{έχειΠαιδί}^-$$

ενώ ο ρόλος έχειΕγγόνι από τη σύνθεση:

$$\text{έχειΠαιδί} \circ \text{έχειΠαιδί}$$

2.2.2 Βάσεις Γνώσης

Σε αυτή την ενότητα θα περιγραφεί ο τρόπος με τον οποίο τα δομικά στοιχεία που αναφέρθηκαν μπορούν να συνδυαστούν για την αναπαράσταση και αποθήκευση γνώσης. Η αναπαράσταση γνώσης σε μία βάση γνώσης πραγματοποιείται μέσω των αξιωμάτων που περιέχει. Τα αξιώματα αυτά μπορούν να διακριθούν σε δύο είδη, σε Ισχυρισμούς και Ορολογία. Το Σώμα Ορολογίας (TBox) περιέχει όλα τα αξιώματα ορολογίας της γνώσης, ενώ το Σωμα Ισχυρισμών (Abox) περιέχει όλα τα αξιώματα ισχυρισμών. Το TBox περιέχει αξιώματα που αφορούν στις σχέσεις των εννοιών και των ρόλων μεταξύ τους, όπως αυτές ορίζονται στον κόσμο στον οποίο ανήκει η γνώση που επιθυμούμε να αναπαραστήσουμε. Το ABox περιέχει αξιώματα που αφορούν σε συγκεκριμένα άτομα, δηλαδή περιέχει αξιώματα που περιγράφουν ιδιότητες των ατόμων του εν λόγω κόσμου.

TBox

Το TBox περιέχει αξιώματα που περιγράφουν κανόνες που πρέπει να ισχύουν ανάμεσα σε διάφορες έννοιες ή ρόλους. Τα αξιώματα που περιέχει το TBox μπορεί να είναι ορισμοί εννοιών ή ρόλων ή αξιώματα υπαγωγής εννοιών ή ρόλων. Οι ορισμοί χρησιμοποιούνται για να οριστούν ονόματα εννοιών και ρόλων, ενώ τα αξιώματα υπαγωγής χρησιμοποιούνται για να δηλωθεί ότι μία έννοια (ρόλος) είναι υποπερίπτωση μίας άλλης έννοιας (ρόλου).

Οι ορισμοί έχουν τη μορφή:

$$A \equiv B$$

$$r \equiv s$$

όπου A είναι ένα όνομα έννοιας, B είναι μία σύνθετη ή απλή έννοια, r είναι ένα όνομα ρόλου και s είναι ένας σύνθετος ή απλός ρόλος.

Παράδειγμα 2.4. Για παράδειγμα, εάν κάποιος θέλει να ορίσει την ιδιότητα Μητέρα ως κάποιον που είναι Γονιός και Γυναίκα τότε μπορεί να προσθέσει το ακόλουθο αξίωμα ορισμού έννοιας στο TBox:

$$\text{Μητέρα} \equiv \text{Γονιός} \sqcap \text{Γυναίκα}$$

Τα αξιώματα υπαγωγής έχουν τη μορφή:

$$A \sqsubseteq B$$

$$r \sqsubseteq s$$

όπου A και B είναι σύνθετες ή ατομικές έννοιες και r και s είναι σύνθετοι ή ατομικοί ρόλοι.

Παράδειγμα 2.5. Εάν κάποιος θέλει να ορίσει ότι η ιδιότητα της μητέρας είναι υποπερίπτωση της ιδιότητας του γονέα μπορεί να ορίσει το ακόλουθο αξίωμα υπαγωγής έννοιας στο TBox:

$$\text{Γονιός} \sqcap \text{Γυναίκα} \sqsubseteq \exists\text{έχειΠαιδί}.\text{Άνθρωπος}$$

Σε κάποιες ΠΛ μπορεί να περιέχονται επιπλέον είδη αξιωμάτων στο TBox που δηλώνουν ιδιότητες των ρόλων όπως μεταβατικότητα, συμμετρικότητα και ανακλαστικότητα.

ABox

Τα αξιώματα που περιέχονται στο το ABox αφορούν σε ιδιότητες που έχουν τα άτομα. Με άλλα λόγια, σε αντίθεση με τη γνώση που παρέχεται στο TBox, όπου δίνονται οι κανόνες που ισχύουν στον κόσμο, η γνώση που περιέχεται στο ABox περιέχει συγκεκριμένες πληροφορίες για συγκεκριμένα άτομα του κόσμου.

Τα αξιώματα του ABox μπορεί να είναι ισχυρισμοί ισότητας ή ανισότητας ατόμων, οι οποίοι δηλώνουν ότι τα αντικείμενα του κόσμου στα οποία αναφερονται τα δύο άτομα ταυτίζονται ή είναι διαφορετικά μεταξύ τους, αντίστοιχα:

$$\alpha \approx \beta$$

$$\alpha \not\approx \beta$$

όπου στους παραπάνω ισχυρισμούς τα α,β είναι ονόματα ατόμων.

Επιπλέον, το ABox μπορεί να περιέχει ισχυρισμούς εννοιών ή ρόλων, που υποδηλώνουν την ύπαρξη μίας συγκεκριμένης ιδιότητας σε ένα άτομο ή την συσχέτιση δύο ατόμων μέσω ενός συγκεκριμένου ρόλου, αντίστοιχα:

$$A(a)$$

$$r(a, \beta)$$

όπου στους παραπάνω ισχυρισμούς τα α,β είναι ονόματα ατόμων, το A είναι μία απλή ή σύνθετη έννοια και το r είναι απλός ή σύνθετος ρόλος.[5]

Παράδειγμα 2.6. Το ABox μπορεί να περιέχει τον ισχυρισμό Πατέρας(Κωνσταντίνος), που δηλώνει ότι το άτομο του κόσμου στο οποίο αναφέρεται το όνομα ατόμου Κωνσταντίνος, έχει την ιδιότητα Πατέρας. Επίσης, ο ισχυρισμός Ναυσικά $\not\approx$ Ελένη υποδηλώνει ότι τα δύο αυτά ονόματα ατόμων αναφέρονται σε διαφορετικά αντικείμενα του κόσμου.

Μία σημαντική παρατήρηση είναι ότι στις ΠΛ ισχύει η υπόθεση ανοικτού κόσμου (open world assumption) σύμφωνα με την οποία δε μπορεί να υποτεθεί ότι οι πληροφορίες που δεν είναι αποθηκευμένες στη γνώση δεν ισχύουν. Για παράδειγμα, εάν στη γνώση δεν είναι αποθηκευμένη η πληροφορία ότι το άτομο Ελένη είναι στιγμιότυπο της έννοιας Γυναίκα, δηλαδή εάν δεν υπάρχει ο ισχυρισμός έννοιας Γυναίκα(Ελένη), δεν μπορεί να υποτεθεί ότι $\neg$Γυναίκα(Ελένη).

2.2.3 Σημασιολογία

Όπως προαναφέρθηκε, τα μη λογικά σύμβολα, παρόλο που ίσως έχουν ονόματα που μπορεί να δημιουργούν διαισθητικά κάποιο οικείο νόημα, δεν έχουν εξαρχής ορισμένη σημασία εντός των ΠΛ. Προκειμένου να δοθεί το νόημα των στοιχείων αυτών πρέπει να οριστεί η τυπική ερμηνεία τους [8]. Οι ερμηνείες περιλαμβάνουν ένα σύνολο Δ που αποτελείται από αντικείμενα του κόσμου και μία συνάρτηση I που αντιστοιχίζει κάθε μη λογικό σύμβολο σε κάποια δομή του κόσμου. Τα άτομα αντιστοιχίζονται σε αντικείμενα του κόσμου, οι έννοιες σε υποσύνολα των αντικειμένων του κόσμου, τα οποία έχουν την ιδιότητα που υποδηλώνει η έννοια και οι ρόλοι αντιστοιχίζονται σε ζεύγη αντικειμένων του κόσμου, τα οποία συνδέονται με αυτόν το διμελή ρόλο. Συγκεκριμένα:

Ορισμός 2.3. *Ως ερμηνεία μίας βάσης γνώσης ορίζεται ένα ζεύγος (Δ, I) όπου το Δ είναι ένα σύνολο αντικειμένων του κόσμου, το οποίο ονομάζεται πεδίο της ερμηνείας και I είναι μία απεικόνιση, η οποία ονομάζεται αντιστοίχηση ερμηνείας, από μη λογικά σύμβολα της ΠΛ σε στοιχεία και σχέσεις του Δ, ως εξής:*

- *για κάθε άτομο a, ισχύει $I[a] \in \Delta$*
- *για κάθε απλή έννοια A, ισχύει $I[A] \subseteq \Delta$*
- *για κάθε απλό ρόλο r, ισχύει $I[r] \subseteq \Delta \times \Delta$*
- *για κάθε σύνθετη έννοια, ανάλογα με τη μορφή της ισχύει:*
  - $a \in I[\neg A]$ *ανν* $a \notin I[A]$
  - $a \in I[A \sqcap B]$ *ανν* $a \in I[A]$ *και* $a \in I[B]$.
  - $a \in I[A \sqcup B]$ *ανν* $a \in I[A]$ *ή* $a \in I[B]$.
  - $a \in I[\exists r.A]$ *ανν υπάρχει* $\beta \in \Delta$ *ώστε* $(a, \beta) \in I[r]$ *και* $\beta \in I[A]$.
  - $a \in I[\forall r.A]$ *ανν για κάθε* $\beta \in \Delta$ *με* $(a, \beta) \in I[r]$ *ισχύει* $\beta \in I[A]$.
  - $a \in I[\geq nr.A]$ *ανν υπάρχουν τουλάχιστον* n *διαφορετικά στοιχεία* $\beta_1, ..., \beta_n \in \Delta$ *ώστε* $(a, \beta_i) \in I[r]$ *και* $\beta \in I[A]$.
  - $a \in I[\leq nr.A]$ *ανν υπάρχουν το πολύ* n *διαφορετικά στοιχεία* $\beta_1, ..., \beta_n \in \Delta$ *ώστε* $(a, \beta_i) \in I[r]$ *και* $\beta \in I[A]$.
  - $a \in I[\{a\}]$ *ανν* $a = I[a]$.
  - *για κάθε* $a \in \Delta$ *ισχύει* $a \in I[\top]$.
  - *δεν υπάρχει* $a \in \Delta$ *ώστε* $a \in I[\bot]$.
- *για κάθε σύνθετο ρόλο, ανάλογα με τη μορφή του ισχύει:*
  - $(a, \beta) \in I[r^-]$ *ανν* $(\beta, a) \in I[r]$
  - $(a, \beta) \in I[r \circ s]$ *ανν υπάρχει* $\gamma \in \Delta$ *ώστε* $(a, \gamma) \in I[r]$ *και* $(\gamma, \beta) \in I[s]$

Είναι εμφανές ότι η ερμηνεία είναι υπεύθυνη να δώσει νόημα στα μη λογικά σύμβολα. Επιπλέον, είναι εμφανές από τους παραπάνω κανόνες ότι η σημασία των λογικών συμβόλων είναι σταθερή, ανεξαρτήτως της ερμηνείας που αποδίδεται στα μη λογικά σύμβολα.

Παράδειγμα 2.7. Έστω η ερμηνεία (Δ, I), όπου $\Delta = a, b, c, d$, και

$$I(\text{Μαρία}) = a,\ I(\text{Κωνσταντίνος}) = b,\ I(\text{Ελένη}) = c,\ I(\text{Ναυσικά}) = d$$

Επίσης, για τις έννοιες: Μητέρα, Πατέρας, Γονέας, μπορεί να έχουμε:

$$I(\text{Μητέρα}) = \{a\},\ I(\text{Πατέρας}) = \{b\},\ I(\text{Γονέας}) = \{a, b\}$$

Τέλος, για τους ρόλους: έχειΠαιδί, έχειΠαιδί $\circ$ έχειΠαιδί η συγκεκριμένη ερμηνεία μπορεί να δίνει:

$$I(\text{έχειΠαιδί}) = \{(a, c)(a, d)(b, c)(b, d)\}$$

$$I(\text{έχειΠαιδί} \circ \text{έχειΠαιδί}) = \{\}$$

Μία ερμηνεία δεν είναι απαραίτητο να συμβαδίζει με τη γνώση που έχει αναπαρασταθεί στη βάση γνώσης. Για παράδειγμα, είναι δυνατόν να υπάρχει ο ισχυρισμός $A(x)$ στο ABox, αλλά το αντικείμενο του κόσμου στο οποίο η ερμηνεία έχει αντιστοιχίσει το όνομα x να μην ανήκει στο σύνολο αντικειμένων που η ερμηνεία έχει αντιστοιχίσει με την έννοια A, δηλαδή στο σύνολο αντικειμένων που, σύμφωνα με την ερμηνεία, αποτελούν στιγμιότυπα της έννοιας αυτής. Επομένως, πρέπει να διευκρινιστεί πότε μία ερμηνεία συμβαδίζει με την αποθηκευμένη γνώση και πότε όχι.

Ορισμός 2.4. *Η ερηνεία (Δ, I) της γνώσης ικανοποιεί:*

- *έναν ισχυρισμό έννοιας $C(a)$ ανν $I(a) \in I(C)$,*
- *έναν ισχυρισμό ρόλου $r(a, b)$ ανν $(I(a), I(b)) \in I(r)$,*
- *έναν ισχυρισμό ισότητας ατόμων $a \approx b$ ανν $I(a) = I(b)$,*
- *έναν ισχυρισμό ανισότητας ατόμων $a \not\approx b$ ανν $I(a) \neq I(b)$.*

Μια ερμηνεία λέγεται μοντέλο του ABox εάν ικανοποιεί όλους τους ισχυρισμούς του.

Ορισμός 2.5. *Η ερηνεία (Δ, I) της γνώσης ικανοποιεί:*

- *ένα αξίωμα υπαγωγής εννοιών $A \sqsubseteq B$ ανν $I(A) \subseteq I(B)$,*
- *ένα αξίωμα ισοδυναμίας εννοιών $A \equiv B$ ανν $I(A) = I(B)$,*
- *έναν αξίωμα υπαγωγής ρόλων $r \sqsubseteq s$ ανν $I(r) \subseteq I(s)$,*
- *έναν αξίωμα ισοδυναμίας ρόλων $r \equiv s$ ανν $I(r) = I(s)$.*

Μια ερμηνεία λέγεται μοντέλο του TBox εάν ικανοποιεί όλα τα αξιώματά του.

Ορισμός 2.6. *Μια ερμηνεία λέγεται μοντέλο της γνώσης εάν είναι μοντελο του TBox και του ABox.*

Επομένως, είναι εμφανές ότι για να αποτελεί μία ερμηνεία μοντέλο της γνώσης, πρέπει να αναπαριστά έναν πιθανό κόσμο από αυτούς που περιγράφει η γνώση. Συνεπώς, πρέπει να ικανοποιεί όλα τα αξιώματα και τους ισχυρισμούς της γνώσης.

2.2.4 Συλλογιστική

Πέραν της αποθήκευσης γνώσης, σε μία βάση γνώσης έχει αξία και η εξαγωγή συμπερασμάτων από αυτήν. Το σύνολο των προβλημάτων εξαγωγής συμπερασμάτων από μία βάση γνώσης ονομάζονται Προβλήματα Συλλογιστικής (Reasoning Problems). Μερικά από τα πιο βασικά προβλήματα συλλογιστικής παρουσιάζονται στη συνέχεια.

Υπαγωγή εννοιών (Concept Subsumption) - Δεδομένων δύο εννοιών A, B να αποφασιστεί εάν $A \sqsubseteq B$.

Ικανοποιησιμότητα έννοιας (Concept Satisfiability) - Δεδομένης μίας έννοιας A να αποφασιστεί εάν η A είναι ικανοποιήσιμη, δηλαδή εάν υπάρχει ερμηνεία $J = (\Delta, I)$ ώστε $I(A) \neq \emptyset$.

Τα δύο προβλήματα σχετίζονται μεταξύ τους, αφού ισχύουν οι εξής ισοδυναμίες:

$$A \text{ μη ικανοποιήσιμη} \Leftrightarrow A \sqsubseteq \bot$$

$$A \sqsubseteq B \Leftrightarrow A \sqcap \neg B \text{ μη ικανοποιήσιμη}$$

Τα παραπάνω προβλήματα συλλογιστικής εμπλέκουν μόνο το TBox, όμως υπάρχουν προβλήματα που αφορούν και το ABox:

Έλεγχος ισχυρισμού έννοιας (Instance Checking) - Δεδομένου του ισχυρισμού $A(a)$, όπου A έννοια, a όνομα ατόμου, να ελεγχθεί εάν αυτός ο ισχυρισμός συνεπάγεται από τη βάση γνώσης, δηλαδή εάν για κάθε ερμηνεία $J = (\Delta, I)$ τέτοια ώστε $J \models K$ ισχύει $I(a) \in I(A)$.

Μία βάση γνώσης δεν είναι απαραίτητο να περιέχει 'σωστές' πληροφορίες, δηλαδή πληροφορίες που δεν οδηγούν σε αντιφασκούμενα συμπεράσματα. Για παράδειγμα, στο $ABoxA = \{\text{Άνθρωπος}(a), \neg\text{Άνθρωπος}(a)\}$, λαμβάνουμε την αντιφάσκουσα πληροφορία ότι το άτομο a είναι στιγμιότυπο της έννοιας Άνρθωπος και παράληλα, ότι δεν είναι. Έτσι, μία σημαντική ανάγκη που προκύπτει συχνά είναι ο έλεγχος της βάσης γνώσης για τυχόν αντιφάσεις:

Ικανοποιησιμότητα Βάσης Γνώσης (KB Satisfiability) - Δεδομένης μίας βάσης γνώσης K να ελεχθεί εάν είναι ικανοποιήσιμη, δηλαδή εάν υπάρχει ερμηνεία $J = (\Delta, I)$ ώστε $J \models K$. Διαφορετικά, η βάση γνώσης ονομάζεται μη ικανοποιήσιμη.

Το πρόβλημα του ελέγχου ισχυρισμού έννοιας μπορεί να αναχθεί στο πρόβλημα ικανοποιησιμότητας βάσης γνώσης. Συγκεκριμένα, εάν $K = (A, T)$:

$$B(a) \Leftrightarrow (A \cup \{\neg B(a)\}, T) \text{ μη ικανοποιήσιμο}$$

Ομοίως, ο έλεγχος ικανοποιησιμότητας έννοιας μπορεί να αναχθεί στην ικανοποιησιμότητα βάσης γνώσης ως εξής:

$$B \text{ μη ικανοποιήσιμη} \Leftrightarrow (A \cup \{B(a)\}, T) \text{ μη ικανοποιήσιμο}$$

όπου a στην παραπάνω ισοδυναμία είναι ένα νέο όνομα ατόμου που δεν υπάρχει ήδη στη βάση γνώσης.

Τέλος, ένα σημαντικό πρόβλημα συλλογιστικής είναι το πρόβλημα της **Απάντησης Ερωτημάτων (Query Answering)**, κατά το οποίο διάφορα είδη ερωτημάτων μπορεί να τεθούν στη βάση γνώσης. Τα ερωτήματα αυτά συνήθως αναμένουν ως απαντήσεις τα ονόματα ατόμων που ικανοποιούν μία σειρά από προυποθέσεις, όμως υπάρχουν και ερωτήματα που δέχονται ως απαντήσεις ¨ΝΑΙ' ή ¨ΟΧΙ¨. Το πρόβλημα της απάντησης ερωτημάτων θα μας απασχολήσει στη συνέχεια της διατριβής και αναλύεται με λεπτομέρεια στην αντίστοιχη ενότητα.

2.2.5 Οικογένεια Λογικών DL-Lite

Η οικογένεια DL-Lite αποτελεί μία ενδιαφέρουσα οικογένεια περιγραφικών λογικών λόγω της σχετικά χαμηλής υπολογιστικής πολυπλοκότητας τους[4]. Σε αυτή την ενότητα θα μελετηθεί η οικογένεια λογικών DL-Lite ως προς τις εκφραστικές της δυνατότητες και την πολυπλοκότητα επίλυσης προβλημάτων συλλογιστικής. Στοιχεία από διάφορες διαλέκτους της οικογένειας DL-Lite θα αποτελέσουν τη βάση για τις λογικές που θα χρησιμοποιηθούν σε επόμενα κεφάλαια.

Δομικά στοιχεία της οικογένειας DL-Lite

Μια από τις εκφραστικότερες διαλέκτους της DL-Lite είναι η $DL - Lite^{HN}_{bool}$. Όπως και οι υπόλοιπες ΠΛ, η $DL - Lite^{HN}_{bool}$ περιέχει ονόματα εννοιών, ονόματα ρόλων και ονόματα ατόμων. Σύνθετες έννοιες και ρόλοι ορίζονται όπως φαίνεται στη συνέχεια:

$$R \ ::= \ P_k \mid P_k^-$$

$$B \ ::= \ A_k \mid \ \geq n.R \mid \bot$$

$$C \ ::= \ B \mid \neg C \mid C_1 \sqcap C_2$$

όπου P_k όνομα ρόλου, A_k όνομα έννοιας, B, C, C_1, C_2 σύνθετες έννοιες, R σύνθετος ρόλος και n θετικός ακέραιος. Οι έννοιες της μορφής B θα ονομάζονται βασικές έννοιες. Το $DL - Lite^{HN}_{bool}$ TBox μπορεί να περιέχει αξιώματα υπαγωγής ρόλων και εννοιών:

$$C_1 \sqsubseteq C_2$$

$$R_1 \sqsubseteq R_2$$

όπου R_1, R_2, C_1, C_2 ρόλοι και έννοιες αντίστοιχα.

Το $DL - Lite^{HN}_{bool}$ ABox μπορεί να περιέχει αξιώματα της μορφής:

$$A_k(a), \quad \neg A_k(a), \quad P_k(a_i, a_j), \quad \neg P_k(a_i, a_j).$$

Τα σύμβολα που χρησιμοποιεί η $DL - Lite^{HN}_{bool}$ ακολουθούν τους ίδιους κανόνες ως προς ερμηνεία τους, με αυτούς που αναφέρθηκαν σε προηγούμενη ενότητα. Επιπλέον, ως προς τη μορφή των αξιωμάτων του TBox, σε διάφορες διαλέκτους της οικογένειας DL-Lite μπορεί να εισάγονται επιπλέον περιορισμοί. Πέραν της μορφής *bool* για το TBox της διαλέκτου, η οποία παρουσιάσηκε παραπάνω, μία πιο περιοριστική μορφή του είναι η $Krom$ στην οποία τα αξιώματα υπαγωγής εννοιών επιτρέπεται να έχουν τη μορφή:

$$B_1 \sqsubseteq B_2, \quad \neg B_1 \sqsubseteq B_2, \quad B_1 \sqsubseteq \neg B_2 \qquad (Krom)$$

όπου όλες οι έννοιες παραπάνω είναι βασικές. Επιπλέον, υπάρχουν τα TBox της μορφής $Horn$, στα οποία τα αξιώματα υπαγωγής εννοιών επιτρέπεται να έχουν τη μορφή:

$$\sqcap B_k \sqsubseteq B \qquad (Horn)$$

όπου οι έννοιες παραπάνω είναι βασικές. Τέλος, σε ένα TBox της μορφής *core* τα αξιώματα υπαγωγής εννοιών περιορίζονται στη μορφή:

$$B_1 \sqsubseteq B_2, \quad B_1 \sqsubseteq \neg B_2 \qquad (core)$$

όπου οι έννοιες παραπάνω είναι βασικές.

Ο εκθέτης του ονόματος της DL-Lite υποδεικνύει την εκφραστικότητα της διαλέκτου ως προς τις σύνθετες έννοιες και ρόλους που μπορεί να αναπαραστήσει.

Συγκεκριμένα, στις διαλέκτους με εκθέτη H, δηλαδή στις $DL-Lite^{H}_{bool}$, $DL-Lite^{H}_{krom}$, $DL-Lite^{H}_{horn}$, $DL-Lite^{H}_{core}$ επιτρέπονται όλα τα σύμβολα για την κατασκευή σύνθετων εννοιών και ρόλων που επιτρέπονται στην $DL-Lite^{HN}$ με αντίστοιχο TBox, εκτός των αριθμητικών περιορισμών, δηλαδή εκφράσεων της μορφής $\geq n.R$, για $n > 1$.

Στις διαλέκτους με εκθέτη HF επιτρέπονται τα σύμβολα για την κατασκευή σύνθετων εννοιών και ρόλων των αντίστοιχων $DL-Lite^{HN}$ με τη διαφορά ότι οι αριθμητικοί περιορισμοί $\geq nR$ επιτρέπονται γενικά μόνο για $n = 1$, ενώ για $n = 2$ επιτρέπονται μόνο για να εκφράσουν περιορισμούς της μορφής $\geq 2R \sqsubseteq \perp$.

Στις διαλέκτους με εκθέτη N επιτρέπονται τα στοιχεία της $DL-Lite^{HN}_{bool}$ εκτός των αξιωμάτων υπαγωγών ρόλων. Στις διαλέκτους με εκθέτη F επιτρέπονται τα στοιχεία της $DL-Lite^{HF}_{bool}$ εκτός των αξιωμάτων υπαγωγών ρόλων.

Τέλος, στις διαλέκτους χωρίς εκθέτη επιτρέπονται τα στοιχεία της $DL-Lite^{HF}_{bool}$ εκτός των αξιωμάτων υπαγωγών ρόλων και εκτός των αξιωμάτων της μορφής $\geq 2R \sqsubseteq \perp$ που προαναφέρθηκαν, δηλαδή οι αριθμητικοί περιορισμοί της μορφής $\geq nR$ επιτρέπονται μόνο για $n = 1$.

Μία παραλλαγή της χρήσης εκθέτη είναι οι εκθέτες εντός παρενθέσεων. Για παράδειγμα, η γλώσσες με εκθέτη (HN) παρέχουν επιπλέον εκφραστικές δυνατότητες για τους ρόλους, όπως για παράδειγμα δηλώσεις συμμετρικών, ανακλαστικών ρόλων. Επιπλέον, επιτρέπουν την δήλωση εκφράσεων της μορφής $\geq nR.C$, υπό περιορισμούς. Παράλληλα, εισάγουν μία σειρά περιορισμών στη μορφή του TBox ώστε να μειώσουν την πολυπλοκότητα των προβλημάτων συλλογιστικής για τη συγκεκριμένη ομάδα γλωσσών.

Μία τελευταία παραλλαγή αποτελούν οι γλώσσες με εκθέτη της μορφής $(x)+$, όπου το x μπορεί να είναι κάποιο εκ των $\{HF, HN\}$, οι οποίες επιτρέπουν όσα οι γλώσσες με τον αντίστοιχο εκθέτη x εντός παρένθεσης, με την προσθήκη της ικανότητας δήλωσης μεταβατικών ρόλων.

Σε αυτή τη διατριβή θα μας απασχολήσει η γλώσσα $DL-Lite_{R}$ που είναι διαφορετικό όνομα για τον χαρακτηρισμό της διαλέκτου $DL-Lite^{H}_{core}$, συχνά με την προσθήκη της δυνατότητας καταγραφής αξιωμάτων υπαγωγής ρόλων, όπως θα δούμε σε επόμενο κεφάλαιο. Συχνά, η γλώσσα αυτή επεκτείνεται με την δυνατότητα προσθήκης εκφράσεων της μορφής $\exists R.C$ στο δεξί μέλος των αξιωμάτων υπαγωγής εννοιών. Όμως, αυτή η κατασκευή μπορεί να υπαινιχθεί από άλλες δομές της γλώσσας [12], επομένως στην υπόλοιπη διατριβή δεν θα θεωρηθεί ότι ανήκει ρητά στη γλώσσα.

Πολυπλοκότητα συλλογιστικής στην DL-Lite

Όπως αναφέρθηκε προηγουμένως, οι διάφορες μορφές της DL-Lite προτιμώνται για τη σχετικά χαμηλή υπολογιστική πολυπλοκότητα που προκύπτει για τα διάφορα προβλήματα συλλογιστικής σε αυτές. Για τα προβλήματα ικανοποιησιμότητας και ελέγχου ισχυρισμού έννοιας οι παράμετροι ενδιαφέροντος μπορεί να είναι το μέγεθος του TBox T και του ABox A, που μπορεί να συμβολίζονται με $|T|$ και $|A|$, αντίστοιχα. Για την απάντηση ερωτημάτων, σημασία έχει και το μέγεθος του ερωτήματος (query), το οποίο συνήθως θεωρείται φραγμένο από κάποια σταθερά [4].

Συμπληρωματικά με τα παραπάνω, μπορούν να διακριθούν δύο συνηθισμένοι τρόποι μελέτης της πολυπλοκότητας των προβλημάτων συλλογιστικής. Ο πρώτος είναι να θεωρηθεί ως είσοδος το $|T|+|A|$, επομένως μιλάμε για συνδυασμένη πολυπλοκότητα (combined complexity). Εναλλακτικά, μπορεί να θεωρηθεί ως είσοδος μόνο το μέγεθος του ABox, επομένως μιλάμε για πολυπλοκότητα των δεδομένων (data complexity) [36].

Ο πίνακας 2.1 παρουσιάζει μερικά συγκεντρωτικά αποτελέσματα της πολυπλοκότητας επίλυσης προβλημάτων συλλογιστικής στην οικογένεια γλωσσών $DL - Lite$, όπως αυτά αναλύονται στο [4]. Στον πίνακα αυτό συγκρίνονται οι πολυπλοκότητες των προβλημάτων ικανοποιησιμότητας ισχυρισμού έννοιας και απάντησης ερωτημάτων σε διάφορες διαλέκτους DL-Lite. Αξίζει να σημειωθεί ότι το πρόβλημα της απάντησης ερωτημάτων είναι γενίκευση του προβλήματος ικανοποιησιμότητας ισχυρισμού έννοιας(instance checking) [25], επομένως είναι ένα από τα δυσκολότερα προβλήματα συλλογιστικής στις βάσεις γνώσης.

Γλώσσα	Data Complexity ικανοποιησιμότητα ισχυρισμού έννοιας	Data Complexity απάντησης ερωτημάτων
$DL - Lite_{core}^{[\ \vert H]}$	AC^0	AC^0
$DL - Lite_{horn}^{[\ \vert H]}$	AC^0	AC^0
$DL - Lite_{krom}^{[\ \vert H]}$	AC^0	$coNP$
$DL - Lite_{bool}^{[\ \vert H]}$	AC^0	$coNP$
$DL - Lite_{core}^{[F \vert N \vert (HF) \vert (HN)]}$	AC^0	AC^0
$DL - Lite_{horn}^{[F \vert N \vert (HF) \vert (HN)]}$	AC^0	AC^0
$DL - Lite_{krom}^{[F \vert N \vert (HF) \vert (HN)]}$	AC^0	$coNP$
$DL - Lite_{bool}^{[F \vert N \vert (HF) \vert (HN)]}$	AC^0	$coNP$
$DL - Lite_{core/horn}^{HF}$	P	P
$DL - Lite_{krom/bool}^{HF}$	$coNP$	$coNP$
$DL - Lite_{core/horn}^{HN}$	$coNP$	$coNP$
$DL - Lite_{krom/bool}^{HN}$	$coNP$	$coNP$

Table 2.1: Πολυπλοκότητα οικογένειας λογικών $DL - Lite$.

2.3 Ανάκτηση Σημασιολογικών Δεδομένων

Όπως είδαμε, οι Βάσεις Γνώσης, δηλαδή ο συνδυασμός ενός ABox και ενός TBox, μπορούν να αποθηκεύσουν αποτελεσματικά ρητή ή υπονοούμενη γνώση του κόσμου ενδιαφέροντος. Συχνά, έχει νόημα να μπορεί κάποιος να θέτει ερωτήματα, τα οποία απαντώνται βάσει της γνώσης που έχει αποθηκευτεί για τον κόσμο. Τα ερωτήματα αυτά συνήθως συνθέτουν μία γενική περιγραφή των ατόμων, απαιτώντας τα άτομα να διαθέτουν συγκεκριμένα χαρακτηριστικά ή σχέσεις με άλλα άτομα, τα οποία μπορεί επίσης να διαθέτουν συγκεκριμένα χαρακτηριστικά, κ.ο.κ. Ένα σύστημα απάντησης των ερωτημάτων αυτών θα πρέπει να επιστρέψει όλα τα άτομα που τηρούν τις προδιαγραφές του ερωτήματος, συμβουλευόμενο τη βάση γνώσης.

Προφανώς, για να έχει πρακτικό νόημα η υλοποίηση ενός τέτοιου συστήματος, πρέπει να υπάρχει μεγάλος όγκος αποθηκευμένης γνώσης πάνω στην οποία έχει νόημα να τεθούν τέτοιου είδους ερωτήματα. Ομολογουμένως, μέχρι σήμερα μεγάλος όγκος πληροφορίας αποθηκεύεται σε σχεσιακές βάσεις δεδομένων. Έτσι, έχει προκύψει η ανάγκη ορισμού μία αντιστοίχησης των δεδομένων που είναι αποθηκευμένα σε σχεσιακές βάσεις με στοιχεία οντολογιών. Μια τέτοιου είδους αντιστοίχηση επιτρέπει την πραγματοποίηση διαδικασιών συλλογιστικής, όπως η απάντηση ερωτημάτων, χρησιμοποιώντας τη γνώση που μπορεί να εξαχθεί από μία σχεσιακή βάση δεδομένων.

Για την επίλυση του προβλήματος που αναφέρθηκε υπάρχουν δύο πιθανές προσεγγίσεις. Η πρώτη προσέγγιση είναι η μετάφραση όλης της πληροφορίας που είναι διατυπωμένη στη σχεσιακή βάση σε όρους οντολογίας για την κατασκευή ενός ABox που θα περιέχει όλη την αποθηκευμένη στην σχεσιακή βάση γνώση. Η μέθοδος αυτή παρουσιάζει το μειονέκτημα ότι, σε περίπτωση που διατηρούνται ταυτόχρονα η σχεσιακή βάση και το ABox, η ίδια πληροφορία αποθηκεύεται σε πολλαπλά σημεία. Αυτό το μειονέκτημα δεν υπάρχει στη δεύτερη πιθανή προσέγγιση. Η δεύτερη προσέγγιση εστιάζει στη δημιουργία μίας αντιστοίχισης μεταξύ των δύο συστημάτων. Τα ερωτήματα μπορούν να διατυπώνονται σε οντολογική μορφή από το χρήστη και στη συνέχεια, με χρήση της προαναφερθείσας αντιστοίχησης, να μεταφράζονται σε ερωτήματα κατανοητά από το σύστημα που διαχειρίζεται τη βάση δεδομένων.

Στη συνέχεια παρουσιάζονται περισσότερες λεπτομέρειες για τη μορφή των ερωτημάτων που μπορεί να τεθούν, τα χαρακτηριστικά της αντιστοίχησης μεταξύ των σχεσιακών βάσεων δεδομένων και των οντολογιών καθώς και οι συνηθέστερες προσεγγίσεις για την απάντηση των ερωτημάτων που είναι επιθυμητό να τεθούν σε μία βάση γνώσης.

2.3.1 Συζευκτικά Ερωτήματα

Υπάρχουν διάφορα είδη ερωτημάτων που μπορεί να τεθούν σε μία βάση γνώσης. Ένα απλό είδος ερωτημάτων είναι τα ερωτήματα στιγμιοτύπων έννοιας (instance queries), τα οποία, όπως υποδεικνύει το όνομά τους, αναζητούν όλα τα άτομα που είναι στιγμιότυπα μίας πιθανώς σύνθετης έννοιας. Για παράδειγμα, ένα τέτοιο ερώτημα θα μπορούσε να αναζητά όλα τα στιγμιότυπα της έννοιας Γυναίκα. Το πρόβλημα της απάντησης τέτοιων ερωτημάτων ανάγεται στο πρόβλημα συνεπαγωγής ισχυρισμού έννοιας. Όμως, είναι εμφανές ότι τα ερωτήματα αυτά έχουν πολύ περιορισμένη εκφραστική δυνατότητα.

Από την άλλη, ένα διαφορετικό είδος ερωτημάτων που μπορεί να τεθούν είναι ερωτήματα γραμμένα σε λογική πρώτης τάξης. Η λογική πρώτης τάξης προσφέρει μεγάλη ευχέρεια στην εκφραστικότητα, όμως το πρόβλημα απάντησης τέτοιων ερωτημάτων είναι μη αποφασίσιμο (undecidable) [28]. Έτσι, είναι επιθυμητό να περιοριστεί η εκφραστικότητα του ερωτήματος ώστε το πρόβλημα να έχει καλές υπολογιστικές ιδιότητες.

Το είδος ερωτημάτων που θα μας απασχολήσουν σε αυτή τη διατριβή είναι τα συζευκτικά ερωτήματα. Τα συζευκτικά ερωτήματα είναι ερωτήματα γραμμένα σε όρους οντολογίας και μπορεί να τεθούν σε μία βάση γνώσης. Στο σώμα τους περιγράφουν τις προϋποθέσεις που πρέπει να πληροί ένα αντικείμενο για να αποτελέσει απάντηση του ερωτήματος. Όπως υποδηλώνει το όνομά τους, ενέχουν την απαίτηση οι απαντησεις τους να πληρούν όλες τις προϋποθέσεις που περιέχουν, ή αλλιώς τη σύζευξη αυτών των προϋποθέσεων. Στον παρακάτω ορισμό περιγράφεται με ακρίβεια η μορφή που μπορεί να έχει ένα συζευκτικό ερώτημα.

Ορισμός 2.7. *Ως συζευκτικό ερώτημα ορίζεται μία έκφραση λογικής πρώτης τάξης:*

$$q(x_1, ..., x_n) \leftarrow \{q_1, ..., q_m\}$$

όπου $q_i, i \in \mathbb{N}$ είναι μία ατομική έκφραση και ονομάζεται συνήθως ατομικό ερώτημα. Ένα ατομικό ερώτημα μπορεί να έχει μορφή:

$$q_i(x) = A(x)$$

ή

$$q_i(x) = R(x, y)$$

Στα συζευκτικά ερωτήματα με τα οποία θα ασχοληθούμε θα θεωρούμε ότι το A είναι ατομική έννοια και το R ατομικός ρόλος, όμως ορίζονται και ερωτήματα που περιέχουν πιο σύνθετες ατομικές εκφράσεις στο σώμα τους. Τα x, y είναι είτε μεταβλητές, είτε σταθερές και αντιστοιχίζονται (μέσω της ερμηνείας) σε αντικείμενα του κόσμου. Το $q(x_1, ..., x_n)$ ονομάζεται κεφαλή, ενώ το $\{q_1, ..., q_m\}$ ονομάζεται σώμα του ερωτήματος. Το διάνυσμα $\vec{x} = (x_1, ..., x_n)$ ονομάζεται διάνυσμα απάντησης του ερωτήματος. Οι μεταβλητές που εμφανίζονται στο σώμα του ερωτήματος, αλλά δεν ανήκουν στις μεταβλητές απάντησης θα λέγονται ελεύθερες μεταβλητές του ερωτήματος.

Παράδειγμα 2.8. Ένα συζευκτικό ερώτημα μπορεί να είναι το εξής:

$$q(x) \leftarrow \{\text{'Αντρας}(x), \text{έχειΠαιδί}(x, y), \text{'Αντρας}(y), \text{έχειΠαιδί}(y, z), \text{'Αντρας}(z)\}$$

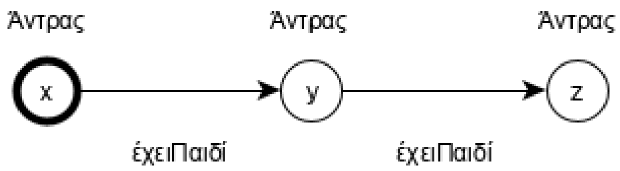
Στο ερώτημα αυτό, εάν υποτεθεί ότι οι έννοιες και ρόλοι που αναφέρονται έχουν την ερμηνεία που υποδηλώνει το όνομά τους, αναζητούμε τα άτομα του κόσμου που είναι πατέρες και έχουν έναν γιό, ο οποίος έχει επίσης κάνει γιό. Όπως και να έχει, ανεξαρτήτως ερμηνείας, το ερώτημα αναζητά κάποιον που είναι στιγμιότυπο της έννοιας 'Αντρας', συνδέεται με κάποιον που είναι, επίσης, στιγμιότυπο της έννοιας 'Αντρας', μέσω ενός ρόλου 'έχειΠαιδί' και αυτός με τη σειρά του συνδέεται μέσω ενός ρόλου 'έχειΠαιδί' με έναν τρίτο, ο οποίος είναι κι αυτός στιγμιότυπο της έννοιας 'Αντρας'.

Στα επόμενα κεφάλαια θα μελετηθούν κυρίως ερωτήματα με μονοδιάστατο διάνυσμα απάντησης, όπως αυτό του προηγούμενου παραδείγματος, δηλαδή ερωτήματα στα οποία το διάνυσμα απάντησης αποτελείται από μόνο ένα στοιχείο. Για το λόγο αυτό, είναι δυνατόν σε επόμενα κεφάλαια, σε κάποιες περιπτώσεις να γράφεται $q(x)$ όταν μας ενδιαφέρει να αποδώσουμε κάποιο προσδιοριστικό μόνο για τη μεταβλητή απάντησης του ερωτήματος, και $q(\vec{x})$ όταν θέλουμε να αναφερθούμε στο διάνυσμα όλων των μεταβλητών του ερωτήματος. Όμως, αυτή η σύμβαση δεν ισχύει για το παρόν κεφάλαιο.

Συχνά, για την καλύτερη εποπτεία του ερωτήματος, τα ερωτήματα παρουσιάζονται με τη μορφή γραφημάτων. Η αναπαράσταση αυτή θα χρησιμοποιείται συχνά σε αυτή τη διατριβή.

Ορισμός 2.8. *Ως γράφος ερωτήματος (query graph) ενός ερωτήματος q ορίζεται ένας γράφημα με ετικέτες, $G(q) = (V, E, l)$, που έχει κορυφές τις μεταβλητές του ερωτήματος, ακμές $E = \{(u, v) | R(u, v) \in q\}$ και κάθε κόμβος v έχει ετικέτα της μορφής $l(v) = \{A | A(v) \in q\}$, ενώ κάθε ακμή (u, v) έχει ετικέτα της μορφής $l(u, v) = \{R | R(u, v) \in q\}$.*

Παράδειγμα 2.9. Ο γράφος ερωτήματος του ερωτήματος του προηγούμενου παραδείγματος είναι ο ακόλουθος.



Σχήμα 2.1: Παράδειγμα γράφου ερωτήματος.

Σε επόμενη υποενότητα θα οριστεί η μορφή των απαντήσεων που ένα συζευκτικό ερώτημα επιδέχεται και θα δούμε τον τρόπο με τον οποίο μπορεί να γίνει η διαχείριση ενός τέτοιου ερωτήματος για την εύρεση των απαντήσεών του.

2.3.2 Βάσεις Σημασιολογικών Δεδομένων

Αρχικά, όμως, θα γίνει μία σύντομη περιγραφή του τρόπου με τον οποίο ένα συζευκτικό ερώτημα μπορεί να τεθεί σε μία σχεσιακή βάση δεδομένων, αφού περιγραφούν τα βασικά χαρακτηριστικά των σχεσιακών βάσεων.

Μια σχεσιακή βάση δεδομένων αποτελείται από ένα σύνολο πινάκων που διαθέτουν μοναδικό όνομα. Οι γραμμές των πινάκων ονομάζονται εγγραφές. Μπορεί να θεωρηθεί ότι κάθε εγγραφή σε έναν πίνακα υποδηλώνει μία συσχέτιση ενός αντικειμένου με κάποια άλλα αντικείμενα, μέσω της σχέσης που υπονοεί το όνομα του πίνακα αυτού [1]. Οι στήλες των πινάκων υποδηλώνουν ιδιότητες των εγγραφών.

Υπάρχουν διάφορες γλώσσες στις οποίες μπορούν να γραφούν ερωτήματα προς εκτέλεση σε μία σχεσιακή βάση δεδομένων. Η πιο ευρέως χρησιμοποιούμενη γλώσσα είναι η SQL. Μια γενική μορφή των ερωτημάτων SQL είναι η ακόλουθη:

$$\begin{array}{ll} SELECT & names \\ FROM & tables \\ WHERE & conditions \end{array} \tag{2.1}$$

Τα ερωτήματα αυτά επιδέχονται ως απαντήσεις τα στοιχεία names που ανήκουν στους πίνακες tables για τα οποία ισχύουν οι συνθήκες conditions.

Συνεπώς, προκειμένου να εκτελεστούν οντολογικά ερωτήματα σε μία βάση δεδομένων, πρέπει να είναι εφοδιασμένη με ένα σύνολο αντιστοιχίσεων ανάμεσα στα συζευκτικά ερωτήματα που μπορεί αν τεθούν σε αυτή και τα SQL ερωτήματα που τους αντιστοιχούν. Το σύνολο αυτών των αντιστοιχίσεων ονομάζεται σώμα σημασιολογικών αντιστοιχίσεων ή MBox. Μία σχεσιακή βάση εφοδιασμένη με ένα σώμα σημασιολογικών αντιστοιχίσεων ανάμεσα στα στοιχεία της και τα στοιχεία του TBox ονομάζεται βάση σημασιολογικών δεδομένων.

2.3.3 Απάντηση Ερωτημάτων

Σε αυτή την υποενότητα θα οριστούν τυπικά οι απαντήσεις που επιδέχεται ένα συζευκτικό ερώτημα. Όπως αναφέρθηκε προηγουμένως, το σώμα ενός συζευκτικού ερωτήματος περιέχει προϋποθέσεις που πρέπει να πληρούν τα άτομα του κόσμου ώστε να αποτελούν απαντήσεις του.

Πιο συγκεκριμένα, δεδομένου ενός ερωτήματος $q(\vec{x})$, όπου $\vec{x} = (x_1, ..., x_n)$, και ενός διανύσματος $\vec{a}$ μεγέθους n, αποτελούμενου από ονόματα ατόμων, θα γράφουμε ως $q(\vec{a})$ το αποτέλεσμα της αντικατάστασης κάθε εμφάνισης του x_i στο $q(\vec{x})$ με το όνομα ατόμου που βρίσκεται στην $i-$οστή θέση του διανύσματος $\vec{a}$. Υπενθυμίζεται ότι υπάρχουν ερωτήματα που δεν περιέχουν καθόλου μεταβλητές, αλλά μόνο σταθερές. Αυτά τα ερωτήματα ονομάζονται ερωτήματα αληθοτιμής (Boolean queries)[25].

Επιπλέον, έστω $J = (\Delta, I)$ μία ερμηνεία. Μία συνάρτηση a θα ονομάζεται ανάθεση εάν συσχετίζει κάθε μεταβλητή y του ερωτήματος με ένα όνομα ατόμου $a(y) \in \Delta$. Μία ανάθεση a, σε μία ερμηνεία J, θα λέμε ότι ικανοποιεί ένα συζευκτικό ερώτημα, εάν ικανοποιεί κάθε ατομική έκφραση που υπάρχει στο σώμα του ερωτήματος. Η ικανοποίηση των ατομικών εκφράσεων του ερωτήματος περιγράφεται ως εξής:

$$J \models^a A(y) \ \text{ iff } \ a(y) \in I(A)$$

$$J \models^a P(x, y) \ \text{ iff } \ (a(x), a(y)) \in I(P)$$

Επομένως, γίνεται αντιληπτό ότι για να αποτελέσει ένα διάνυσμα ονομάτων ατόμων του κόσμου απάντηση σε ένα συζευκτικό ερώτημα θα πρέπει να ικανοποιεί τους περιορισμούς που θέτει το ερώτημα. Τα ερωτήματα αληθοτιμής δεν αναμένουν διανύσματα ονομάτων ατόμων ως απάντηση, αλλά τις απαντήσεις ΝΑΙ ή ΌΧΙ εάν το σύνολο των ατομικών εκφράσεων που αποτελούν το σώμα του ερωτήματος είναι ικανοποιήσιμο ή όχι, αντίστοιχα, υπό την παρουσία της ερμηνείας J. Έτσι, παρατηρούμε ότι τα ερωτήματα αληθοτιμής δεν εξαρτώνται από κάποια ανάθεση, παρά μόνο από την υπάρχουσα ερμηνεία.

Γενικά, θα λέμε απάντηση ενός συζευκτικού ερωτήματος το διάνυσμα $\vec{a}$ εάν υπό την παρουσία κάποιας ερμηνεία και κάποιας ανάθεσης στα πλαίσια αυτής της ερμηνείας, η οποία οδηγεί στη σύνθεση του $\vec{a}$ ως αντικατάσταση του $\vec{x}$, το συζευκτικό ερώτημα ικανοποιείται.

Ορισμός 2.9. *Βέβαιη απάντηση ενός συζευκτικού ερωτήματος $q(\vec{x})$ είναι ένα διάνυσμα $\vec{a} = (a_1, ..., a_n)$ τέτοιο ώστε αν $J \models (A, T)$ τότε $J \models q(\vec{a})$.*

Είναι εμφανές από τον παραπάνω ορισμό ότι οι βέβαιες απαντήσεις είναι οι απαντήσεις που ικανοποιούν το συζευκτικό ερώτημα για οποιαδήποτε ερμηνεία που ικανοποιεί τη βάση γνώσης. Αυτές είναι και οι απαντήσεις που αποτελούν τη λύση στο προβλημα της απάντησης ερωτήματος (query answering problem). Οι βέβαιες απαντήσεις συχνά γράφονται και ως απαντήσεις για τις οποίες ισχύει $(A, T) \models q(\vec{a})$, δηλαδή χωρίς να γίνει αναφορά στην ερμηνεία που μπορεί να έχει αποδοθεί.

Συχνά, δεδομένης μίας βάσης γνώσης, χρησιμοποιείται η έκφραση $ans(q)$ για να περιγράψει τη συνάρτηση που λαμβάνει ως είσοδο ένα συζευκτικό ερώτημα και επιστρέφει ως έξοδο το σύνολο των διανυσμάτων απάντησης που αποτελούν βέβαιες απαντήσεις του ερωτήματος αυτού. Στην περίπτωση των ερωτημάτων με μοναδιαίο διάνυσμα απάντησης, το $ans(q)$ θα επιστρέφει ένα σύνολο από ονόματα ατόμων (individual names) που αποτελούν βέβαιες απαντήσεις του ερωτήματος.

Στο προηγούμενο τμήμα της ενότητας περιγράφηκαν οι απαντήσεις που αναζητώνται και δόθηκε μία τυπική περιγραφή του προβλήματος της εύρεσης των απαντήσεων των ερωτημάτων. Στην υπόλοιπη ενότητα θα συζητηθούν οι συνηθέστερες προσεγγίσεις που υιοθετούνται για την επίλυση αυτού του προβλήματος.

Μία αφελής προσέγγιση, όπου θα έδινε βέβαιες απαντήσεις βασιζόμενη μόνο στα στιγμιότυπα του ABox δεν θα παρείχε πλήρη λύση, καθώς με αυτόν τον τρόπο αγνοείται η υπονοούμενη γνώση που μπορεί να εξαχθεί με τις διάφορες μεθόδους συλλογιστικής, οι οποίες συνδυάζουν τη γνώση που είναι αποθηκευμένη στο ABox με τους κανόνες του TBox για την εκμαίευση νέας γνώσης. Το φαινόμενο αυτό φαίνεται ξεκάθαρα στο ακόλουθο παράδειγμα.

Παράδειγμα 2.10. Έστω ABox:

$$A = \{\text{Πατέρας}(\text{Κωνσταντίνος}), \text{Μητέρα}(\text{Μαρία})\}$$

και TBox:

$$T = \{\text{Πατέρας} \sqsubseteq \text{Γονιός}, \text{Μητέρα} \sqsubseteq \text{Γονιός}\}$$

Ας υποθέσουμε ότι ψάχνουμε τα στιγμιότυπα της έννοιας Γονιός. Τότε, το ερώτημα που θέτουμε μπορεί να είναι το

$$q: \quad q(x) \leftarrow \{\text{Γονιός}(x)\}$$

Εάν το ερώτημα απαντηθεί μόνο με βάση τη γνώση που είναι αποθηκευμένη στο A τότε θα επιστρέψει το κενό σύνολο ως απάντηση, ενώ στην πραγματικότητα οι βέβαιες απαντήσεις είναι οι $\{\text{Κωνσταντίνος}, \text{Μαρία}\}$.

Έτσι, το πρόβλημα απάντησης ερωτημάτων μετατρέπεται σε σύνθετο πρόβλημα συλλογιστικής που απαιτεί την εύρεση των απαντήσεων που προκύπτουν άμεσα από το ABox, αλλά και αυτών που υπονούνται από τα στοιχεία της βάσης γνώσης. Οι βασικότερες προσεγγίσεις που προκύπτουν για την επίλυση αυτού του προβλήματος είναι δύο, η υλοποίηση του ABox και η επαναγραφή του ερωτήματος.

Η υλοποίηση του ABox (ABox materialisation) περιλαμβάνει την εξαγωγή και την αποθήκευση όλης της γνώσης που υπονοείται, αλλά δεν είναι ρητά καταγεγγραμμένη στο ABox. Προφανώς, ύστερα από αυτή τη διαδικασία το ερώτημα θα μπορεί να απαντηθεί ορθά ελέγχοντας μόνο το ABox, όμως συχνά ο αριθμός των αξιωμάτων που να πρέπει να αποθηκευτούν για να υλοποιηθεί η υπονοούμενη γνώση είναι απαγορευτικός.

Για τους λόγους που αναφέρθηκαν προηγουμένως, συχνά προτιμάται η δεύτερη προσέγγιση, κατά την οποία η αποθηκευμένη γνώση δε μεταβάλλεται, ενώ όλες οι απαιτούμενες μεταβολές γίνονται στο ερώτημα που τίθεται στη βάση γνώσης. Στην προσέγγιση αυτή, το TBox χρησιμοποιείται μόνο κατά την επαναγραφή του ερωτήματος, ενώ μετά από αυτή τη διαδικασία καθίσταται και πάλι άχρηστο. Η προσέγγιση αυτή θα μελετηθεί με λεπτομέρεια στην επόμενη υποενότητα.

2.3.4 Επαναγραφή Ερωτημάτων

Όπως αναφέρθηκε, η απλη εφαρμογή ενός ερωτήματος στο ABox δεν φέρνει πάντα όλες τις βέβαιες απαντήσεις. Αυτό συμβαίνει επειδή η βάση γνώσης περιέχει υπονούμενη γνώση, η οποία δεν είναι ρητά καταγεγγραμένη. Κατά την μέθοδο της επαναγραφής ερωτημάτων, η γνώση του TBox αξιοποιείται σε βαθμό που το TBox καθίσταται άχρηστο κατά την απάντηση των ερωτημάτων. Η ιδέα της επαναγραφής ερωτημάτων περιγράφηκε στο [10] και θα φανεί καλύτερα στο επόμενο παράδειγμα.

Παράδειγμα 2.11. Έστω ABox:

$$A = \{\text{Πατέρας}(\text{Κωνσταντίνος}), \text{Μητέρα}(\text{Μαρία})\}$$

και TBox:

$$T = \{\text{Πατέρας} \sqsubseteq \text{Γονιός}, \text{Μητέρα} \sqsubseteq \text{Γονιός}\}$$

Έστω ότι και πάλι ψάχνουμε τα στιγμιότυπα της έννοιας Γονιός, θέτοντας το ερώτημα

$$q(x) \leftarrow \{\text{Γονιός}(x)\}$$

Εξετάζοντας το TBox γίνεται αντιληπτό ότι εάν κάποιος είναι στιγμιότυπο της έννοιας Πατέρας, τότε είναι επιθυμητό να ανήκει στο σύνολο απαντήσεων. Έτσι, το αρχικό ερώτημα ξαναγράφεται, αντικαθιστώντας την έννοια Γονιός, με την έννοια Πατέρας και κατασκευάζεται το ερώτημα:

$$q_1(x) \leftarrow \{\text{Πατέρας}(x)\}$$

Με την ίδια διαδικασία κατασκευάζεται το ερώτημα:

$$q_2(x) \leftarrow \{\text{Μητέρα}(x)\}$$

Τελικά, το σύνολο των ερωτημάτων q, q_1, q_2 θα τεθεί στο ABox επιστρέφοντας τα αποτελέσματα που αναζητούμε, χωρίς να χρειάζεται να εξεταστεί ξανά το TBox.

Στο παραπάνω παράδειγμα φάνηκε ότι αντί για το αρχικό ερώτημα, στη βάση γνώσης τέθηκε μία ένωση συζευκτικών ερωτημάτων (Union of Conjunctive Queries - UCQ). Επομένως, το πρόβλημα φαίνεται να μπορεί να λυθεί εάν κάποιος, κάνοντας επαναγραφή του ερωτήματος, μπορεί να υπολογίσει ένα νέο τέτοιων σύνολο ερωτημάτων που μπορούν να τεθούν στη βάση γνώσης, καθιστώντας μη απαραίτητο το TBox. Αξίζει να σημειωθεί ότι η επαναγραφή που περιγράφηκε δεν είναι πάντα δυνατή.

Ορισμός 2.10. *FOL-Επαναγραψιμότητα (FOL − rewritability/FOL − reducibility): Δεδομένου ενός TBox T και ενός ερωτήματος q, μπορεί να υπολογιστεί ένα ερώτημα q^T τέτοιο ώστε, για κάθε ABox A, ισχύει:*

$$ans(q, (A, T)) = ans(q^T, (A, \emptyset))$$

Επομένως, ένα σημαντικό πλεονέκτημα των περιγραφικών λογικών DL-Lite είναι ότι έχουν την ιδιότητα της επαναγραψιμότητας. Μάλιστα, στο [10] δείχθηκε ότι ένα συζευκτικό ερώτημα γραμμένο σε DL-Lite μπορεί να επαναγραφεί σε ένα σύνολο από συζευκτικά ερωτήματα (UCQ) και προτάθηκε ο αλγόριθμος PerfectRef για τον υπολογισμό της επαναγραφής αυτής. Ο αλγόριθμος PerfectRef ουσιαστικά εξετάζει τις σχέσεις υπαγωγής ρόλων και εννοιών που υπάρχουν στο TBox για να αντικαταστήσει έναν ρόλο (ή έννοια) από κάποιον άλλο ρόλο (ή έννοια) που υπάγεται στον πρώτο. Ο αλγόριθμος παίρνει όλους τους πιθανούς συνδυασμούς αντικαταστάσεων για να κατασκευάσει νέα υπο-ερωτήματα, τα οποία τελικά θα συντελέσουν στην κατασκευή ενός UCQ. Παράλληλα, στο [9] αποδείχθηκε ότι η οικογένεια λογικών DL-Lite διαθέτει μεγιστοτική εκφραστικότητα τέτοια ώστε να επιτρέπει τα ερωτήματα που τίθενται σε αυτή να είναι επαναγράψιμα. Τέλος, αποδείχθηκε ότι η πολυπλοκότητα της απάντησης ερωτημάτων έχει εκθετική πολυπλοκότητα χειρότερης περίπτωσης, λόγω του πιθανού εκθετικού μεγέθους του UCQ [12].

Παράδειγμα 2.12. Ας υποθέσουμε ότι έχουμε το ABox:

$$A = \{\text{Γονιός}(\text{Κωνσταντίνος}), \text{Άντρας}(\text{Κωνσταντίνος})\}$$

και TBox:

$$T = \{\text{Γονιός} \sqsubseteq \exists\text{έχειΠαιδί}, \text{Άντρας} \sqsubseteq \text{Άνθρωπος}\}$$

Αυτή τη φορά ψάχνουμε τα στιγμιότυπα της έννοιας 'Ανθρωπος' και του ρόλου 'έχειΠαιδί', θέτοντας το συζευκτικό ερώτημα:

$$q(x) \leftarrow \{\text{έχειΠαιδί}(x,y), \text{Άνθρωπος}(x)\}$$

Ο αλγόριθμος PerfectRef εξετάζοντας το TBox θα εκμεταλλευόταν το πρώτο αξίωμα του και θα κατασκεύαζε το ερώτημα

$$q_1: \ q(x) \leftarrow \{\text{Γονιός}(x), \text{Άνθρωπος}(x)\}$$

Στη συνέχεια, εξετάζοντας το δεύτερο αξίωμα του TBox θα κατασκεύαζε το ερώτημα

$$q_2: \ q(x) \leftarrow \{\text{Γονιός}(x), \text{Άντρας}(x)\}$$

Συνεχίζοντας τη διαδικασία, τελικά θα καταλήγαμε με τα ερωτήματα:

$$\begin{aligned} q: \ q(x) &\leftarrow \{\text{έχειΠαιδί}(x,y), \text{Άνθρωπος}(x)\} \\ q_1: \ q(x) &\leftarrow \{\text{Γονιός}(x), \text{Άνθρωπος}(x)\} \\ q_2: \ q(x) &\leftarrow \{\text{Γονιός}(x), \text{Άντρας}(x)\} \\ q_3: \ q(x) &\leftarrow \{\text{έχειΠαιδί}(x,y), \text{Άντρας}(x)\} \end{aligned}$$

Από το προηγούμενο παράδειγμα φαίνεται ότι το μέγεθος του UCQ μπορεί να είναι εκθετικό σε σχέση με το μέγεθος του αρχικού ερωτήματος. Επιπλέον, μια σημαντική παρατήρηση είναι ότι κάποια υπο-ερωτήματα μπορεί να είναι υποσύνολα άλλων ερωτημάτων ή να παράγονται πολλές φορές από τον αλγόριθμο. Ο αναγνώστης παραπέμπεται στην τελευταία στήλη του πίνακα 2.1 που δείχνει την πολυπλοκότητα της απάντησης ερωτημάτων σε διάφορες διαλέκτους της οικογένειας DL-Lite. Βάσει των ατελειών που αναφέρθηκαν, πολλές ερευνητικές εργασίες έχουν επικεντρωθεί στην βελτίωση του αλγορίθμου επαναγραφής ερωτημάτων, ώστε να παρουσιάζει καλύτερη απόδοση.

Στο [13] προτείνεται το σύστημα Rapid που παρουσιάζει μία σειρά από βελτιστοποιήσεις με στόχο να μειωθεί ο όγκος των επαναγραφών που παράγονται, χωρίς να επηρεάζεται η πληρότητα της απάντησης. Περισσότερες λεπτομέρειες για τον αλγόριθμο θα αναφερθούν στη συνέχεια. Το σύστημα QuOnto [2] περιέχει τον αλγόριθμο PerfectRef που αναφέρθηκε προηγουμένως, με κάποιες βελτιστοποιήσεις. Το σύστημα OntOP [30] είναι ένα πλήρες σύστημα ανάκτησης δεδομένων βασισμένης σε οντολογίες (OBDA) και εστιάζει σε βελτστοποιήσεις που σχετίζονται με τη μορφή των αξιωμάτων του TBox. Επιπλέον, υπάρχουν συστήματα όπως το Quest [32], που υποστηρίζουν ερωτήματα σε γλώσσα SPARQL. Μία επιπλέον επέκταση του PerfectRef που βασίζεται στη γλώσσα Datalog είναι υλοποιημένη στο σύστημα Nyaya [18]. Άλλες προσεγγίσεις, εστιάζουν στην επαναγραφή του ερωτήματος όχι σε UCQ, αλλά σε εκφραστικότερες μορφές. Στον αλγόριθμο του συστήματος Presto χρησιμοποιείται επαναγραφή του ερωτήματος σε μη αναδρομική Datalog [33]. Προφανώς, υπάρχουν και προσεγγίσεις που δουλεύουν και για πιο εκφραστικές γλώσσες ερωτημάτων, όπως οι [35],[29]. Στο σύστημα Requiem [29] χρησιμοποιείται επαναγραφή του ερωτήματος σε αναδρομική Datalog.

Το σύστημα Rapid

Το Rapid είναι ένα σύστημα επαναγραφής ερωτημάτων που περιέχει διάφορες βελτιστοποιήσεις με στόχο την παραγωγή ενός συνόλου επαναγραφών χωρίς περιττά στοιχεία, όπως επαναγραφές που αποτελούν υπερσύνολα άλλων, ελαχιστοποιώντας τον αριθμό τω τελικών ελέγχων για περιττές επαναγραφές. Αν και έχει επεκταθεί για να υποστηρίζει και πιο σύνθετες περιγραφικές λογικές, εδώ θα μας απασχολήσει η επαναγραφή ερωτημάτων γραμμένων σε $DL - Lite_R$.

Το Rapid βασίζει τη λειτουργία του σε ανάλυση λογικής πρώτης τάξης First Order Resolution. Ξεκινά με τη μετάφραση των όρων του TBox, οι οποίοι είναι γραμμένοι σε $DL - Lite_R$, σε προτάσεις λογικής πρώτης τάξης, ώστε να μπορεί να χρησιμοποιηθεί Resolution για την παραγωγή των νέων ερωτημάτων.

Κάθε αξίωμα υπαγωγής του TBox μεταφράζεται σε μία ή δύο προτάσεις, όπως φαίνεται στον πίνακα 2.2. Όπως φαίνεται στον πίνακα, ένας όρος του TBox μπορεί να μετατραπεί σε δύο προτάσεις όταν έχει την ιδιομορφία της $DL - Lite_R$ που αναφέρθηκε σε προηγούμενη ενότητα, δηλαδή όταν επιτρέπεται η ρητή εγγραφή όρων υπαγωγής που έχουν στο δεξί μέλος εκφράσεις της μορφής $\exists R.B$. Γενικά φαίνεται ότι, αν και το αρχικό ερώτημα δεν περιέχει συναρτησιακούς όρους, τέτοιοι όροι μπορεί να εισαχθούν κατά τη διαδικασία επαναγραφής του ερωτήματος. Σε κάθε περίπτωση, το Rapid εγγυάται ότι τελικά θα παραχθούν επαναγραφές ελεύθερες από συναρτησιακούς όρους.

Αξίωμα	Πρόταση
$A \sqsubseteq B$	$B(x) \leftarrow A(x)$
$P \sqsubseteq S$	$S(x,y) \leftarrow P(x,y)$
$P^- \sqsubseteq S^-$	$S(x,y) \leftarrow P(x,y)$
$\exists P \sqsubseteq A$	$A(x) \leftarrow P(x,y)$
$A \sqsubseteq \exists P$	$P(x, f_P^A(x)) \leftarrow A(x)$
$A \sqsubseteq \exists P.B$	$P(x, f_{P.B}^A(x)) \leftarrow A(x)$ $B(f_{P.B}^A(x)) \leftarrow A(x)$
$P \sqsubseteq S^-$	$S(x,y) \leftarrow P(y,x)$
$P^- \sqsubseteq S$	$S(x,y) \leftarrow P(y,x)$
$\exists P^- \sqsubseteq A$	$A(x) \leftarrow P(y,x)$
$A \sqsubseteq \exists P^-$	$P(f_P^A(x), x) \leftarrow A(x)$
$A \sqsubseteq \exists P^-.B$	$P(f_{P.B}^A(x), x) \leftarrow A(x)$ $B(f_{P.B}^A(x)) \leftarrow A(x)$

Table 2.2: Κανόνες μετάφρασης αξιωμάτων $DL - Lite_R$

Παράδειγμα 2.13. Στο παράδειγμα 2.12, το TBox:

$$T = \{\text{Γονιός} \sqsubseteq \exists \text{έχειΠαιδί}, \text{Άντρας} \sqsubseteq \text{Άνθρωπος}\}$$

θα μεταφραζόταν στις προτάσεις:

$$\text{έχειΠαιδί}(x', f^{\text{Γονιός}}_{\text{έχειΠαιδί}}(x')) \leftarrow \text{Γονιός}(x') \quad (1)$$

$$\text{Άνθρωπος}(x') \leftarrow \text{Άντρας}(x') \quad (2)$$

Το Rapid στηρίζεται σε δύο βασικές διαδικασίες για την παραγωγή των επαναγραφών. Η πρώτη είναι ο κανόνας Unfolding, κατά την εφαρμογή του οποίου κάποιο ατομικό ερώτημα του ερωτήματος αντικαθίσταται με κάποιο άλλο που περιέχει υποέννοια ή υπορόλο, όπως ορίζουν τα αξιώματα υπαγωγής της γλώσσας.

$$\frac{q \quad C}{q'\sigma}$$

όπου στον παραπάνω κανόνα, q είναι το ερώτημα στο οποίο εφαρμόζεται ο κανόνας Unfolding, C είναι κάποια πρόταση της μορφής που υποδηλώνει η δεύτερη στήλη του πίνακα 2.2, το q' είναι το παραγόμενο ερώτημα και το σ είναι μία αντικατάσταση μεταβλητών από μεταβλητές του C σε μεταβλητές του ερωτήματος. Κατά την εφαρμογή αυτού του κανόνα είναι εμφανές ότι το ερώτημα που θα προκύψει ως αποτέλεσμα θα έχει ίδιο μέγεθος με το προηγούμενο ερώτημα, εκτός εάν το ατομικό ερώτημα που εισάχθηκε στο νέο ερώτημα ως αντικαταστάτης υπήρχε ήδη στο προηγούμενο.

Παράδειγμα 2.14. Στο παράδειγμα 2.12, με βάση την πρόταση (2) που παράχθηκε στο παράδειγμα 2.13 το ερώτημα q_2:

$$q(x) \leftarrow \{\text{Γονιός}(x), \text{Άντρας}(x)\}$$

κατασκευάζεται από το q_1:

$$q(x) \leftarrow \{\text{Γονιός}(x), \text{Άνθρωπος}(x)\}$$

σύμφωνα με την παρακάτω εφαρμογή του κανόνα unfolding:

$$\frac{q(x) \leftarrow \{\text{Γονιός}(x), \text{Άνθρωπος}(x)\} \quad \text{Άνθρωπος}(x') \leftarrow \text{Άντρας}(x')}{q(x) \leftarrow \{\text{Γονιός}(x), \text{Άντρας}(x)\}}$$

όπου $\sigma = \{x' \mapsto x\}$.

Η δεύτερη βασική διαδικασία στην οποία στηρίζεται το Rapid είναι ο κανόνας Shrinking. Ο κανόνας αυτός στοχεύει στην ενοποίηση διαφόρων συναρτησιακών συμβόλων με σκοπό την απαλοιφή τους, η οποία τελικά οδηγεί σε μικρότερα ερωτήματα.

$$\frac{q \quad C_1[C_2]}{q'\sigma}$$

όπου στον παραπάνω κανόνα, q είναι το ερώτημα στο οποίο εφαρμόζεται ο κανόνας Shrinking, C_1, C_2 είναι προτάσεις της μορφής που υποδηλώνει η δεύτερη στήλη του πίνακα 2.2, το q' είναι το παραγόμενο ερώτημα και το σ είναι μία αντικατάσταση μεταβλητών από μεταβλητές των προτάσεων C_1, C_2 σε μεταβλητές του ερωτήματος. Η πρόταση C_2 είναι προαιρετική και αντιστοιχεί στις περιπτώσεις του πίνακα που η δεύτερη στήλη περιέχει δύο προτάσεις. Τελικά, το Rapid με μια σειρά βελτιστοποιήσεων εξασφαλίζει την παραγωγή ενός συνόλου επαναγραφής που δεν περιέχει περιττές επαναγραφές.

Κεφάλαιο 3

Σχετικές Εργασίες

Το πρόβλημα που μελετάται στα πλαίσια της παρούσας διατριβής είναι το πρόβλημα της κατασκευής ελαχιστοτικών ABox υπό την παρουσία συζευκτικών ερωτημάτων. Το πρόβλημα δεν είναι ισοδύναμο με την κλασσική ελαχιστοποίηση ενός ABox, καθώς στην περίπτωσή μας δεν μας ενδιαφέρει εάν το τελικό ABox είναι ισοδύναμο με το αρχικό. Το ζητούμενο είναι, το τελικό ABox να δίνει ισοδύναμες απαντήσεις ως προς ένα σύνολο συζευκτικών ερωτημάτων που μπορεί να τεθούν σε αυτό. Αυτό θα γίνει καλύτερα κατανοητό στο επόμενο κεφάλαιο, στο οποίο παρουσιάζεται αναλυτικά το πρόβλημα προς επίλυση.

3.1 Ελαχιστοποίηση ABox

Το πρόβλημα της ελαχιστοποίησης ABox εστιάζει στην παραγωγή ενός νέου ABox, το οποίο δεν περιέχει περιττές πληροφορίες και διατηρεί την γνώση του αρχικού, δηλαδή είναι ισοδύναμο με το αρχικό. Αρχικά, ένα ABox ορίζεται ως ισοδύναμο ενός άλλου όταν έχει αποθηκευμένη την ίδια πληροφορία με το αρχικό. Προφανώς, η πληροφορία αυτή δεν είναι απαραίτητο να είναι ρητά καταγεγγραμμένη, αλλά μπορεί να υπονοείται από τη γνώση.

Ορισμός 3.1. *Έστω $K = (A, T)$ μία βάση γνώσης. Το ABox A' χαρακτηριζεται ως ισοδύναμο του ABox A εάν:*

$$(A, T) \models \varphi \Leftrightarrow (A', T) \models \varphi$$

όπου φ είναι ένας ισχυρισμός.

Όπως αναφέρθηκε, το πρόβλημα της ελαχιστοποίησης ABox εστιάζει στη δημιουργία ενός νέου ABox, το οποίο είναι ισοδύναμο με το αρχικό και επιπλέον, δεν περιέχει περιττούς ισχυρισμούς. Γενικά, ένας ισχυρισμός χαρακτηριζεται ως περιττός εάν η αφαίρεσή του δεν περιορίζει τη γνώση. Τότε, είναι εμφανές ότι ένας ισχυρισμός είναι περιττός όταν δεν χρειάζεται να είναι ρητά καταγεγγραμμενος στο ABox, δηλαδή όταν η γνώση που προσφέρει μπορεί ήδη να εξαχθεί από την υπόλοιπη.

Ορισμός 3.2. *Ένας ισχυρισμός ψ του ABox χαρακτηρίζεται ως περιττός (redundant) εάν ισχύει $(A - \{\psi\}, T) \models \psi$.*

Αξίζει να διακριθεί η περίπτωση του ελαχιστοτικού ABox (minimal ABox), η οποία αναφέρθηκε παραπάνω, από την περίπτωση του ελάχιστου ABox (minimum ABox), το οποίο είναι ένα minimal ABox με την ελάχιστη δυνατή πληθικότητα. Επιπλέον, αξίζει να σημειωθεί ότι το ελαχιστοτικό ABox δεν είναι μοναδικό. Αυτό γίνεται εμφανές στο ακόλουθο παράδειγμα.

Παράδειγμα 3.1. Εάν υποθέσουμε ότι η βάση γνώσης αποτελείται από τα:

$$TBox\ T = \{\text{έχειΠατέρα} \sqsubseteq \text{έχειΠαιδί}^-, \text{έχειΠαιδί} \sqsubseteq \text{έχειΠατέρα}^-\}$$

$$ABox\ A = \{\text{έχειΠατέρα}(a, b), \text{έχειΠαιδί}(b, a)\}$$

Τότε υπάρχουν δύο δυνατά ελαχιστοτικά $ABox$, τα:

$$A = \{\text{έχειΠαιδί}(b, a)\}$$

$$A = \{\text{έχειΠατέρα}(a, b)\}$$

Το πρόβλημα της εύρεσης ενός ελαχιστοτικού ABox έχει ήδη μελετηθεί στη βιβλιογραφία. Στο [26] το πρόβλημα της ελαχιστοποίησης ABox μελετάται στα πλαίσια επίλυσης ενός λίγο διαφοροποιημένου προβλήματος. Συγκεκριμένα, η εργασία ασχολείται με την εξαγωγή στοιχείων μίας τοπικής οντολογίας (local/private ontology) σε μία μοιραζόμενη (shared/domain ontology), όπου η πρώτη συνήθως είναι επέκταση της δεύτερης. Μία περίπτωση χρήσης ενός τέτοιου συστήματος μπορεί να είναι η εξαγωγή ευαίσθητων ιατρικών δεδομένων από μία τοπική βάση γνώσης που διαθέτει μεγάλη περιγραφική δυνατότητα για την αποθήκευση μεγάλης λεπτομέρειας, σε μία μοιραζόμενη οντολογία, η οποία διαθέτει πιο περιορισμένη περιγραφική δυνατότητα που επιτρέπει την απόκρυψη ιδιωτικών πληροφοριών και την αποθήκευση γενικών στατιστικών σε αυτή.

Στο πλαίσιο αυτής της εξαγωγής, παρουσιάζεται και ένας αλγόριθμος ελαχιστοποίησης των πληροφοριών που εξάγονται. Ο αλγόριθμος ασχολείται με αξιώματα εκφρασμένα σε OWL (DL) για να ορίσει κάποιους κανόνες που χαρακτηρίζουν τους περιττούς ισχυρισμούς. Η OWL (DL) διαφέρει από την $DL - Lite_R$ σε κάποια στοιχεία της, όπως στο γεγονός ότι επιτρέπει ισχυρισμούς ισότητας ή ανισότητας ατόμων (individuals) λόγω του ότι δεν υιοθετεί την υπόθεση μοναδικού ονόματος, σε αντίθεση με την οικογένεια $DL - Lite$ που συνήθως υιοθετεί την υπόθεση αυτή. Επομένως, σε επόμενα κεφάλαια, από τους κανόνες που προτείνονται, θα χρησιμοποιηθούν αυτοί που έχουν νόημα στην $DL - Lite_R$.

Μία σημαντική παρατήρηση είναι ότι ο αλγόριθμος που προτίνεται στο [26] δεν εγγυάται την πλήρη εξάλειψη των περιττών ισχυρισμών. Σε επόμενο κεφάλαιο προτείνεται αλγόριθμος που εγγυάται την πλήρη εξάλειψη των περιττών ισχυρισμών στην περίπτωση της $DL - Lite$. Αρχικά, στην εργασία αυτή οι ισχυρισμοί του ABox περιορίζονται στις μορφές:

$$C(a),\ R(a, b),\ a \approx b,\ a \not\approx b$$

όπου C, R ονοματική έννοια και ονοματικός ρόλος και a, b ονόματα ατόμων. Παράλληλα, ορίζεται η έννοια του ελαχιστοτικού συνόλου ισχυρισμών του ABox που αφορούν ένα άτομο, καθώς και η έννοια της ελάχιστης εξαγωγής ισχυρισμών ως εξής:

Ορισμός 3.3. *Το σύνολο A_i όλων των αξιωμάτων που αφορούν το άτομο i είναι ελαχιστοτικό εάν δεν υπάρχει αξίωμα $\psi \in A_i$ τέτοιο ώστε $\{(A - \psi, T) \models \psi\}$.*

Ορισμός 3.4. *Ως ελάχιστη προβολή/εξαγωγή P_{min} ορίζεται ένα σύνολο ισχυρισμών $P_{min} = \bigcup_{i=1}^{|IN|} A_i$, όπου το A_i είναι ένα ελαχιστοτικό σύνολο αξιωμάτων για το άτομο i.*

Στη συνέχεια, προτείνεται ένας αλγόριθμος που βλέπει τη γνώση σαν μαύρο κουτί (black box algorithm) και απλώς χρησιμοποιεί τις λειτουργίες συλλογιστικής (reasoner) που παρέχονται για να αφαιρέσει τα περιττά αξιώματα. Συγκεκριμένα, ο αλγόριθμος ρωτά τον reasoner για κάθε αξίωμα εάν είναι περιττό και αν ναι, το αφαιρεί από τη γνώση. Όμως, η πολυπλοκότητα ενός τέτοιου αλγορίθμου τον καθιστά μη εφικτό.

Για το λόγο αυτό, γίνεται μελέτη των ιδιοτήτων των ισχυρισμών που μπορεί να οδηγήσουν σε περιττές πληροφορίες στη γνώση και ορίζονται κάποιοι κανόνες, η χρήση των οποίων μπορεί να βοηθήσει στον γρηγορότερο χαρακτηρισμό των περιττών ισχυρισμών. Συγκεκριμένα:

Θεώρημα 3.1.

1. *Ένας ισχυρισμός έννοιας $C(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός έννοιας $D(x)$ στο A και το $(A, T) \models D \sqsubseteq C$.*

2. *Ένας ισχυρισμός ρόλου $R(x, y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $S(x, y)$ στο A και το $(A, T) \models S \sqsubseteq R$.*

3. *Ένας ισχυρισμός έννοιας $C(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $R(x, y)$ στο A και το $(A, T) \models \exists R \sqsubseteq C$.*

4. *Ένας ισχυρισμός ρόλου $R(x, y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $R^{-}(y, x)$ στο A.*

5. *Ένας ισχυρισμός ρόλου $R(x, y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $R(y, x)$ στο A και ο R έχει οριστεί ως συμμετρικός ρόλος.*

6. *Ένας ισχυρισμός ρόλου $R(x, y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $S(x, z)$ στο A και $(A, T) \models S \sqsubseteq R$ και $(A, T) \models y \approx z$.*

7. *Ένας ισχυρισμός ρόλου $R(x, x)$ είναι περιττός στο A εάν ο ρόλος είναι ανακλαστικός και υπάρχει $C(x)$ στο A ώστε $(A, T) \models C \sqsubseteq \exists R.Self$.*

8. *Ένας ισχυρισμός ισότητας ατόμων $a \approx b$ είναι περιττός στο A εάν υπάρχουν ισχυρισμοί ισότητας ατόμων $a \approx c, c \approx b$ στο A.*

9. *Ένας ισχυρισμός ισότητας ατόμων $a \approx b$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ισότητας ατόμων $b \approx a$ στο A.*

10. *Έστω a, c άτομα τέτοια ώστε $a \approx c$. Κάθε ισχυρισμός για τον a είναι περιττός στο A εάν υπάρχει ένα άλλο αντίστοιχο αξίωμα για τον c στο A.*

11. *Ένας ισχυρισμός ρόλου $R(x, y)$ είναι περιττός στο A εάν υπάρχουν ισχυρισμοί ρόλου $R(x, z), R(z, y)$ στο A και ο R είναι μεταβατικός ρόλος.*

12. *Ένας ισχυρισμός έννοιας $C(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $R(y, x)$ στο A και το $(A, T) \models \top \sqsubseteq \forall R.C$.*

13. *Ένας ισχυρισμός έννοιας $C(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $R(y, x)$ στο A και υπάρχει περιορισμός $b \sqsubseteq \forall R.C$.*

14. *Ένας ισχυρισμός ισότητας ατόμων $a \approx b$ είναι περιττός στο A εάν υπάρχει functional ισχυρισμός ρόλου $R(a, x), R(c, y)$ και ισχύει $x \approx y$.*

Υπάρχουν άλλα είδη περιττών ισχυρισμών, όπως ισχυρισμοί που προκύπτουν από αλυσίδες ρόλων, αλλά στο [26] επιλύεται το πρόβλημα μόνο για τους παραπάνω. Αποδεικνύεται ότι οι περιττοί ισχυρισμοί που προκύπτουν από τους κανόνες 11 έως 14 δεν μπορούν να εξαλειφθούν με έναν επαυξητικό (incremental) αλγόριθμο, που εκτελεί έλεγχο περιττότητας για κάθε ισχυρισμό. Συνεπώς, προτείνεται επαυξητικός αλγόριθμος που αφαιρεί τους περιττούς ισχυρισμούς τύπου 1 έως 10, ενώ για η αφαίρεση των υπόλοιπων γίνεται μία μετεπεξεργασία στο αποτέλεσμα του επαυξητικού αλγορίθμου.

3.2 Απαγωγή ABox - ABox Abduction

Ένας σχετικός με το πρόβλημά μας τομέας είναι το πρόβλημα της απαγωγής ABox. Ο τομέας αυτός σχετίζεται με την επίλυση προβλημάτων όπως την εύρεση όλων των ελαχιστοτικών συνόλων ισχυρισμών ώστε να ισχύει μία συγκεκριμένη πρόταση. Το πρόβλημα αυτό έχει άμεση σχέση με το πρόβλημά μας και μπορεί να βοηθησει στην επίλυσή του. Συγκεκριμένα, το πρόβλημά μας μπορεί να αναχθεί σε ένα πρόβλημα απαγωγής ABox, εάν θεωρήσουμε ότι οι προτάσεις που επιθυμούμε να ισχύουν είναι ολές οι ρητά και μη αποθηκευμένες στη γνώση πληροφορίες για κάθε άτομο. Όμως, η περίπτωση αυτή είναι μία τετριμμένη περίπτωση για το πρόβλημα της απαγωγής ABox. Οι μέθοδοι που έχουν προταθεί για την επίλυση προβλημάτων απαγωγής στοχεύουν στην επίλυση μη τετριμμένων περιπτώσεων και απαιτούν τη χρήση εξειδικευμένων εργαλείων συλλογιστικής, η περιπλοκότητα λειτουργίας των οποίων τα καθιστά μη αποδοτική την επίδοσή τους για το δικό μας σενάριο χρήσης.

Το πρόλημα της απαγωγής ABox ορίστηκε αρχικά στο [23], στο οποίο για την επίλυση του προβλήματος χρησιμοποιούνται δύο κλασσικά εργαλεία συλλογιστικής.Αρχικά, το πρόβλημα ορίζεται ως εξής:

Ορισμός 3.5. *Έστω L_1, L_2 δύο περιγραφικές λογικές, $K = (A, T)$ μία βάση γνώσης στην L_1 και Φ ένα σύνολο ισχυρισμών ABox στην L_2, το οποίο χαρακτηρίζεται ως* ***απαγωγικό ερώτημα (abductive query)****. Το (K, Φ) ονομάζεται πρόβλημα απαγωγής εάν $K \not\models \Phi$ και $K \cup \Phi \not\models \bot$.*

Ορισμός 3.6. *Έστω L_s μία περιγραφική λογική και A ένα σύνολο ισχυρισμών ABox γραμμένων στην L_s. Το A είναι απάντηση ενός απαγωγικού προβλήματος (K, Φ) ανν $K \cup A \models \Phi$. Επιπλέον, το A καλείται:*

1. *συνεπές, ανν $K \cup A \not\models \bot$.*

2. *ελάχιστο, ανν δεν υπάρχει λύση B του (K, Φ) που να είναι ελαχιστοτική σε σχέση με το A.*

Μία σημαντική παρατήρηση είναι ότι επειδή ο χώρος των απαγωγικών λύσεων είναι άπειρος, μπορούν να εισαχθούν επιπλέον περιορισμοί ώστε να μειωθούν οι επιτρεπτές λύσεις που μπορεί να βρεθούν. Για το λόγο αυτό ορίζονται οι έννοιες της συνέπειας και της ελαχιστότητας για τις απαγωγικές απαντήσεις.

Αρχικά, στο [23] περιγράφονται οι κατάλληλοι μετασχηματισμοί των προτάσεων που είναι εκφρασμένες σε περιγραφική λογική ALC, ώστε να μπορούν να χρησιμοποιηθούν τα εργαλεία συλλογιστικής λογικής πρώτης τάξης. Συγκεκριμένα, τα αξιώματα ΠΛ γράφονται σε Κανονική Μορφή Άρνησης (Negation Normal Form), δηλαδή σε μορφή τέτοια ώστε οι αρνήσεις ¬ να υπάρχουν μόνο μπροστά από ονοματικές έννοιες. Στη συνέχεια, οι προτάσεις είναι απαραίτητο να γραφούν σε συζευκτική κανονική μορφή (conjunctive normal form) καθώς και να έχουν υποστεί τη διαδικασία του Skolemization, δηλαδή της απαλοιφής των υπαρξιακών ποσοδεικτών και την εισαγωγής του όρου που υποδήλωνε ο ποσοδείκτης. Ο όρος αυτός ονομάζεται σταθερά Skolem.

Οι μετασχηματισμοί αυτοί είναι απαραίτητη για να ακολουθήσει η εφαρμογή εργαλείων συλλογιστικής. Συγκεκριμένα, στο [23] προτείνεται δύο μέθοδοι επίλυσης, καθένας από τους οποίους χρησιμοποιεί κάποιο βασικό εργαλείο συλλογιστικής. Τα δύο αυτά εργαλεία είναι οι μέθοδοι ταμπλώ (semantic tableaux) και ο αλγόριθμος ανάλυσης (resolution). Και στις δύο αυτές περιπτώσεις, εάν το ζητούμενο είναι η εύρεση απόδειξης για το $K \models \Phi$, η μέθοδος επίλυσης αρχικά δίνει την πρόταση $\neg\Phi$ σε οποιοδήποτε από τα δύο εργαλεία που αναφέρθηκαν και προσπαθεί να αποδείξει ότι $K \cup \neg\Phi \vdash \bot$. Έτσι, εάν κάποιος από αυτούς τους αλγορίθμους καταλήξει, για παράδειγμα, στην πρόταση A, μπορούμε να υποθέσουμε ότι το $\neg A$ είναι η απόδειξη που ψάχνω.

Εκτός από τη χρήση παραδοσιακών εργαλείων συλλογιστικής για την επίληση του προβλήματος της απαγωγής $ABox$, όπως είδαμε παραπάνω. έχουν προταθεί και εναλλακτικές προσεγγίσεις για το πρόβλημα αυτό. Αρχικά, όπως αναφέρθηκε, στην προηγούμενη προσέγγιση το πρόβλημα προς επίλυση έχει πιθανώς άπειρες απαγωγικές λύσεις. Επιπλέον, για να εγγυηθεί την ελαχιστότητα των απαγωγικών λύσεων, ο αλγόριθμος που προτάθηκε στην προηγούμενη προσέγγιση πρέπει να περιμένει την παραγωγή όλων των λύσεων, με αποτέλεσμα να μην είναι εγγυημένη η εύρεση μίας ελαχιστοτικής απαγωγικής λύσης σε πεπερασμένο χρόνο.

Έτσι, στο [14] προτείνεται ένας νέος ορισμός απαγωγικού προβλήματος, ο οποίος παρουσιάζει πεπερασμένο αριθμό πιθανών λύσεων. Συγκεκριμένα, οι λύσεις περιορίζονται σε ονοματικές έννοιες και ρόλους. Στη συνέχεια, για να επιλυθεί το πρόβλημα αυτό, χρησιμοποιούνται έτοιμα συστήματα απαγωγικής συλλογιστικής γλωσσών λογικού προγραμματισμού, όπως Prolog engines. Για να χρησιμοποιηθούν τα εργαλεία αυτά, η γνώση πρέπει να μετατραπεί σε λογικό πρόγραμμα.

Πιο πρόσφατα, το προβλημα της απαγωγής ABox έχει χρησιμοποιηθεί για να βρεθεί απάντηση στο γιατί ένα διάνυσμα δεν αποτελεί απάντηση ενός συζευκτικού ερωτήματος [11]. Έτσι, εισάγεται ένα νέο είδος προβλήματος, το πρόβλημα απαγωγής ερωτήματος (query abduction problem). Συγκεκριμένα, το ερώτημα είναι ένα δίτιμο συζευκτικό ερώτημα και στην προσέγγιση αυτή είναι επιτρεπτό να συμμετέχουν νέα ονόματα ατόμων στις απαντήσεις.

Μία άλλη προσέγγιση για την επίλυση του προβλήματος της απαγωγής ABox εντοπίζεται στο [15]. Συγκεκριμένα, προτίνεται μία μέθοδος υπολογισμού των ελαχιστοτικών απαγωγικών απαντήσεων, στην περίπτωση που η πρόταση το TBox είναι δυνατόν να γραφεί με χρήση λογικής πρώτης τάξης.

Μία άλλη προσέγγιση για την επίλυση του προβλήματος της απαγωγής ABox εντοπίζεται στο [15]. Συγκεκριμένα, προτίνεται μία μέθοδος υπολογισμού των απαγωγικών απαντήσεων, στην περίπτωση που η παρατήρηση που επιθυμούμε να αποδείξουμε είναι έ΄να δίτιμο συζευκτικό ερώτημα . Επιπλέον, για να περιοριστούν οι πιθανώς άπειρες λύσεις του προβλήματος αυτού, εισάγεται η απαίτηση το TBox να μπορεί να γραφεί με χρήση λογικής πρώτης τάξης (first order rewritable).

3.3 Module extraction

Ένας άλλος τομέας που είναι σχετικός με το πρόβλημα της ελαχιστοποίησης ABox περιλαμβάνει αντίστοιχες προσεγγίσεις για modular οντολογίες [34]. Συγκεκριμένα, ο τομάς module extraction σχετίζεται στενά με το πρόβλημά μας. Η γενική ιδέα είναι να εξαχθεί μία ελαχιστοτική υπο-οντολογία που αφορά σε ένα συγκεκριμένο σύνολο ονομάτων (ατόμων, εννοιών και ρόλων) και διατηρεί τη γνώση της αρχικής οντολογίας σε σχέση με το συγκεκριμένο σύνολο ονομάτων. Το ελαχιστοτικό αυτό υποσύνολο της οντολογίας συχνά ονομάζεται module. Έτσι, φαίνεται ότι το πρόβλημα του module extraction λύνει ένα γενικότερο πρόβλημα από το πρόβλημα της ελαχιστοποίησης ABox.

Στο [19] δείχθηκε ότι το πρόβλημα module extraction είναι μη αποφασίσιμο (undecidable) για την οικογένεια γλωσσών OWL DL. Συνεπώς, οι προσεγγίσεις για την επίλυση του προβλήματος αυτού συνήθως περιλαμβάνουν προσεγγιστικές λύσεις ή εισάγουν επιπλέον περιορισμούς. Επειδή το ζητούμενο είναι η εξαγωγή υπό-οντολογιών τα αποτελέσματα που παράγονται συνήθως περιλαμβάνουν και έννοιες της αρχικής οντολογίας που περιγράφουν έννοιες της νέας, όμως δεν αποτελούν στοιχεία της.

Στο [19] εισάγονται κάποιες έννοιες όπως conservative extension, safety και module και τις συσχετλισεις αυτών των χαρακτηριστικών με σκοπό να οριστεί τυπικά το πρόβλημα modular reuse ως πρόβλημα συλλογιστικής. Επειδή τα προβλήματα που ορίζονται αποδεικνύεται ότι είναι μη επιλύσιμα, ορίζονται κάποιες κλάσεις, οι οποίες ονομαζονται κάσεις ασφάλειας (safety classes). Οι κάσσεις αυτές περιέχουν συνθήκες που είναι ικανές για να εξασφαλίσουν την επιλυσιμότητα των προβλημάτων που προτείνονται. Τέλος, προτείνονται διάφοροι αλγόριθμοι που εξαρτώνται από τα χαρακτηριστικά κάθε κλάσης ασφάλειας.

Στο [24] μελετάται η σημασιολογική έννοια modularity στις πειγραφικές λογικές και προβλήματα συλλογιστικής που σχετίζονται με την έννοια αυτή. Παρέχονται δύο ορισμοί των module καθώς και αλγόριθμοι για $\mathcal{ALC}$ και $\mathcal{EL}$ που ελέγχουν αν ένα τμήμα αξιωμάτων είναι module. Επιπλέον, προτείνονται αλγόριθμοι για module extraction και μελετάται η πολυπλοκότητα των προβλημάτων που επιλύουν.

Επιπλέον, ένα παρόμοιο πρόβλημα που σχετίζεται με modular ονρολογίες παρουσιάζεται στο [27]. Εδώ, με αφορμή το γεγονός ότι συχνά η γνώση που αποθηκεύεται σχετικά με ένα γνωστικό αντικείμενο, όπως ιατρικά δεδομένα, μπορεί να είναι τεράστια σε όγκο, ενώ τα συχνότερα σενάρια χρήσης μπορεί να σχετίζονται με ένα συγκεκριμένο θέμα εντός αυτού του γνωστικού αντικειμένου, προτείνεται η έννοια του traversal view, δηλαδή μίας όψης (view) όπου ο χρήστης μπορεί να ορίσει κάποιες συγκεκριμένες έννοιες ενδιαφέροντος και ίσως κάποιο βάθος έως το οποίο μη ενδιαφέρουσες έννοιες, οι οποίες όμως σχετίζονται με τις ενδιαφέρουσες, μπορούν να συμπεριλαμβάνονται στην όψη. Τέλος, περιγράφονται τα εργαλεία που μπορεί να χρησιμοποιηθούν για να επιτευχθεί η κατασκευή τέτοιων όψεων.

3.4 Απλούστευση ABox - ABox Reduction

Μία κατηγορία προβλημάτων που εμπεριέχουν την έννοια της ελαχιστοποίησης ABox σχετίζονται με την εύρεση ενός υποσυνόλου του αρχικού ABoξ, το οποίο συχνά ονομάζεται απλουστευμένο/περιληπτικό ABox (summary ABox). Το νέο αυτό ABox προορίζεται συνήθως για ελέγχους από τους οποίους θα εξαχθούν συμπεράσματα για το αρχικό. Συχνά δημιουργούνται τέτοια ABox για να επιτευχθούν έλεγχοι συνέπειας, ένα πρόβλημα που είναι δυσεπίλυτο για εκφραστικές περιγραφικές λογικές. Μία σημαντική παρατήρηση είναι ότι το νέο ABox συχνά δεν είναι ισοδύναμο με το αρχικό, απλώς ίσως δίνει ισοδύναμα συμπεράσματα ως προς μία κατηγορία ελέγχων που μπορεί να απαιτεί η κάθε εφαρμογή.

Το πρόβλημα αυτό έχει μελετηθεί για SHIN ABox στο [16]. Συγκεκριμένα, η εργασία αυτή ασχολείται με το θέμα των ελέγχων συνέπειας σε μεγάλα ABox. Για να επιτευχθούν αποδοτικά αυτοί οι έλεγχοι παράγεται ένα νέο ABox, το οποίο ονομάζεται proxy ABox και αποδεικνύεται ότι αυτό είναι συνεπές (δηλαδή δεν περιέχει αντιφατικούς ισχυρισμούς) αν και μόνο εάν το αρχικό είναι συνεπές. Για παράδειγμα, εάν ο ρόλος R ή ο αντίστροφος R^- δεν συμμετέχει σε κανένα περιορισμό καθολικότητας $\forall P.C$ και σε κανέναν περιορισμό πληθικότητας $\geq nP.C$, τότε οι ισχυρισμοί ρόλου $R(a, b)$ μπορούν να αφαιρεθούν γιατί δε μπορούν να δημιουργήσουν αντιφάσεις στο SHIN ABox.

Με το πρόβλημα των ελέγχων συνέπειας σε μεγάλα ABox ασχολείται και το [17]. Σε αυτή την προσέγγιση, το μέγεθος του νέου ABox, το οποίο ονομάζεται περιληπτικό ABox (summary ABox), μειώνεται δραματικά με ακόμα περισσότερες επεμβάσεις σε αυτό. Συγκεκριμένα, στο νέο ABox είναι δυνατόν να έχουν ομαδοποιηθεί ονόματα ατόμων που είναι στιγμιότυπα της ίδιας έννοιας. Στην περίπτωση αυτή, οι ασυνέπειες που μπορεί να εντοπίζονται στο summary ABox μπορεί να οφείλονται είτε σε ασυνέπειες του αρχικού, είτε στη διαδικασία ενοποίησης που περιγράφηκε. Δηλαδή, εάν το περιληπτικό ABox είναι συνεπές τότε και το αρχικό είναι συνεπές. Η απόδειξη για τον ισχυρισμό αυτό βασίζεται στο ότι εάν το περιληπτικό ABox, στο οποίο έχει γίνει ενοποίηση ονομάτων ατόμων, επιδέχεται μοντέλο, τότε αυτό θα είναι μοντέλο και του αρχικού, εάν διατηρηθεί η ίδια αντιστοιχία ονομάτων ατόμων με αντικείμενα του κόσμου. Έτσι, σε περίπτωση που εντοπιστεί ασυνέπεια στο περιληπτικό ABox, αρκεί να πραγματοποιηθούν έλεγχοι συνέπειας στο τμήμα του αρχικού που αντιστοιχεί σε αυτό στο οποίο εντοπίστηκε ασυνέπεια στο νέο.

Αν και στοιχεία των προσεγγίσεων που αναφέρθηκαν μπορούν να χρησιμοποιηθούν στην ελαχιστοποίηση ABox, οι προσεγγίσεις αυτές αποκλίνουν αρκετά από το πρόβλημα επειδή εστιάζουν σε διαφορετικούς κανόνες ελαχιστοποίησης, οι οποίοι επιτρέπουν την δημιουργία ενός νέου ABox το οποίο δεν είναι απαραίτητα ισοδύναμο με το αρχικό.

Κεφάλαιο 4

Ελαχιστοποίηση ABox υπό την παρουσία συζευκτικών ερωτημάτων

Σε αυτό το κεφάλαιο θα παρουσιαστεί το πρόβλημα προς επίλυση καθώς και οι προσεγγίσεις που ακολουθήθηκαν για τη λύση του. Σε αντίθεση με άλλους τομείς της Αναπαράτασης Γνώσης και Συλλογιστικής, το πρόβλημα της εύρεσης ελαχιστοτικών ισοδύναμων υποσυνόλων, υπό την παρουσία ενός συνόλου ερωτημάτων στερείται μελέτης έως τώρα. Παρ΄ όλα αυτά, η εύρεση ελαχιστοτικών ισοδύναμων υποσυνόλων ως προς ένα σύνολο ερωτημάτων μπορεί να οδηγήσει σε σημαντική μείωση του μεγέθους των ABox που παράγονται, σε σχέση με το μέγεθος του αρχικού. Στο πλαίσιο αυτής της διατριβής θα μελετηθεί το πρόβλημα της εύρεσης όλων των ισοδύναμων ελαχιστοτικών υποσυνόλων ενός ABox υπό την παρουσία συζευκτικών ερωτημάτων, καθώς και το πρόβλημα της εύρεσης ενός ή όλων των ισοδύναμων ελαχιστοτικών υποσυνόλων ενός ABox χωρίς την παρουσία συνόλου ερωτημάτων. Το έναυσμα για την δημιουργία ενός τέτοιου συνόλου αποτέλεσαν οι προσπάθειες για τη συσχέτιση του τομέα αναπαράστασης γνώσης και συλλογιστικής με τη μηχανική μάθηση. Συγκεκριμένα, μία πιθανή χρήση ενός τέτοιου συνόλου είναι η παραγωγή δεδομένων εκμάθησης για ένα Νευρωνικό Δίκτυο (Neural Network).

Στην πρώτη ενότητα εισάγεται το πρόβλημα την εύρεσης όλων των ελαχιστοτικά ισοδύναμων υποσυνόλων ενός ABox. Το πρόβλημα της εύρεσης ενός ελαχιστοτικού, ισοδύναμου ABox έχει μελετηθεί στη βιβλιογραφία, στο πλαίσιο της δημιουργία μία ελαχιστοτικής προβολής της βάσης γνώσης, και έχει προταθεί μία μέθοδος επίλυσής του προβλήματος όταν το ABox είναι γραμμένο σε OWL(DL), η οποία όμως λόγω της εκφραστικότητας της αναπαράστασης αυτής, δεν εγγυάται την ελαχιστότητα του ABox που παράγεται από την προτεινόμενη μέθοδο, αλλά πετυχαίνει σημαντική μείωση του μεγέθους του αρχικού ABox. Στην πρώτη ενότητα παρουσιάζεται ένας αλγόριθμος ελαχιστοποίησης του ABox καθώς και ένας αλγόριθμος παραγωγής όλων των ελαχιστοτικά ισοδύναμων υποσυνόλων του αρχικού ABox, και αποδεικνύεται η ορθότητά τους.

Στη δεύτερη ενότητα εισάγεται η έννοια της ισοδυναμίας ABox υπό την παρουσία ενός συνόλου συζευκτικών ερωτημάτων και αναλύονται οι βασικές δυσκολίες του προβλήματος. Τέλος, παρουσιάζεται ένας αλγόριθμος εύρεσης όλων των ελαχιστοτικά ισοδύναμων υποσυνόλων ενός ABox υπό την παρουσία ενός αυθαίρετου συνόλου συζευκτικών ερωτημάτων Q.

4.1 Ελαχιστοποίηση ABox

Αν και γενικά το πρόβλημα που εξετάζει αυτή η διατριβή είναι η εύρεση όλων των ελαχιστοτικών υποσυνόλων ενός DL-Lite$_{\mathrm{R}}$ ABox που είναι ισοδύναμα με το αρχικό υπό την παρουσία ενός συνόλου ερωτημάτων, για αρχή θα μελετηθεί το ίδιο πρόβλημα όταν το σύνολο των ερωτημάτων είναι αυθαίρετο. Είναι εμφανές ότι το πρόβλημα εύρεσης ενός ελαχιστοτικού ισοδύναμου ABox όσον αφορά σε κάθε πιθανό ερώτημα σχετικό με τη βάση γνώσης, είναι το ίδιο με το πρόβλημα της ελαχιστοποίησης ABox διατηρώντας όλη την πληροφορία του.

Το πρόβλημα της ελαχιστοποίησης ABox περιλαμβάνει την εύρεση ισοδύναμων ελαχιστοτικών υποσυνόλων ενός ABox, δηλαδή ABox τα οποία διατηρούν τη γνώση που είναι αποθηκευμένη στο αρχικό ABox και δεν περιέχουν περιττές πληροφορίες. Το πρόβλημα αυτό σχετίζεται με την ύπαρξη περιττών ισχυρισμών στο αρχικό ABox, δηλαδή ισχυρισμών που δεν προσφέρουν νέα δεδομένα στη γνώση. Για ακριβέστερη περιγραφή των περιττών ισχυρισμών ο αναγνώριστης παραπέμπεται στον ορισμό 3.2.

Το πρόβλημα εύρεσης ενός ελαχιστοτικού ισοδύναμου υποσυνόλου ενός ABox έχει μελετηθεί στο [26]. Στην εργασία αυτή παρουσιάζονται τα είδη των περιττών ισχυρισμών που μπορεί να υπάρχουν σε ένα σύνολο αξιωμάτων OWL (DL) και προτείνεται ένας αλγόριθμος εξαγωγής ενός ισοδύναμου υποσυνόλου της γνώσης με λιγότερους περιττούς ισχυρισμούς. Βέβαια, η προσέγγιση αυτή δεν εγγυάται την ελαχιστότητα του εξαγόμενου υποσυνόλου. Περισσότερες πληροφορίες για την προσέγγιση αυτή παρουσιάστηκαν λεπτομερώς στην προηγούμενη ενότητα.

Εδώ το πρόβλημα της εύρεσης ελαχιστοτικών ισοδύναμων ABox επεκτείνεται, καθώς το πρόβλημά μας περιλαμβάνει την εύρεση όλων των ελαχιστοτικών ABox που είναι ισοδύναμα με ένα αρχικό ABox. Αρχικά, υπενθυμίζεται ότι η $DL\!-\!Lite_R$ ισοδυναμεί με την περιγραφική λογική $DL - Lite^{H}_{core}$ και μπορεί να περιέχει τα ακόλουθα στοιχεία:

$$R ::= P_k \mid P_k^-$$

$$B ::= A_k \mid \exists R \mid \bot$$

$$C ::= B \mid \neg C$$

όπου P_k όνομα ρόλου, A_k όνομα έννοιας, B, C σύνθετες έννοιες και R σύνθετος ρόλος. Οι σύνθετες έννοιες της μορφής B θα ονομάζονται βασικές. Το $DL - Lite^{H}_{core}$ ABox μπορεί να περιέχει αξιώματα της μορφής:

$$A_k(a), \quad \neg A_k(a), \quad P_k(a_i, a_j), \quad \neg P_k(a_i, a_j)$$

Τέλος, σε ένα *core TBox* τα αξιώματα υπαγωγής εννοιών επιτρέπεται να έχουν τη μορφή:

$$B_1 \sqsubseteq B_2, \quad B_1 \sqsubseteq \neg B_2$$

όπου οι έννοιες B_1, B_2 είναι βασικές. Επιπλεόν, τα αξιώματα υπαγωγής ρόλων, εφόσον επιτρέπονται στη γλώσσα, μπορούν να έχουν τη μορφή:

$$R_1 \sqsubseteq R_2$$

όπου οι ρόλοι R_1, R_2 είναι σύνθετοι.

Από τα παραπάνω φαίνεται ότι τα αξιώματα υπαγωγής εννοιών του TBox επιτρέπεται να περιέχουν άρνηση μόνο στο δεξί μέλος της σχέσης. Επιπλέον, επισημαίνεται ότι συχνά στην $DL-Lite_R$ επιτρέπεται η ρητή καταγραφή αξιωμάτων της μορφής $\exists R.C$ στο δεξί μέλος των αξιωμάτων υπαγωγής του TBox, καθώς αυτά τα αξιώματα μπορούν να υπαινιχθούν από τα υπόλοιπα στοιχεία της περιγραφικής λογικής [12], όμως στο πλαίσιο αυτής της διατριβής δεν θεωρείται ότι τέτοια αξιώματα είναι ρητά καταγεγγραμμένα στη βάση γνώσης.

Δεδομένων των επιτρεπτών συμβόλων της DL-Lite$_{\mathrm{R}}$, και εμπνευσμένοι από το θεώρημα 3.1, μπορούμε με τη σειρά μας να προσδιορίσουμε τύπους περιττών ισχυρισμών που μπορεί να υπάρχουν σε ένα DL-Lite$_{\mathrm{R}}$ ABox.

Θεώρημα 4.1.

1. *Ένας ισχυρισμός έννοιας $A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός έννοιας $A_2(x)$ στο A και το $(A,T) \models A_2 \sqsubseteq A_1$.*

2. *Ένας ισχυρισμός έννοιας $\neg A_2(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός έννοιας $\neg A_1(x)$ στο A και το $(A,T) \models A_2 \sqsubseteq A_1$.*

3. *Ένας ισχυρισμός έννοιας $\neg A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός έννοιας $A_2(x)$ στο A και το $(A,T) \models A_2 \sqsubseteq \neg A_1$.*

4. *Ένας ισχυρισμός έννοιας $A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_1(x,y)$ στο A και το $(A,T) \models \exists P_1 \sqsubseteq A_1$.*

5. *Ένας ισχυρισμός έννοιας $\neg A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_1(x,y)$ στο A και το $(A,T) \models \exists P_1 \sqsubseteq \neg A_1$.*

6. *Ένας ισχυρισμός έννοιας $A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_1(y,x)$ στο A και το $(A,T) \models \exists P_1^- \sqsubseteq A_1$.*

7. *Ένας ισχυρισμός έννοιας $\neg A_1(x)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_1(y,x)$ στο A και το $(A,T) \models \exists P_1^- \sqsubseteq \neg A_1$.*

8. *Ένας ισχυρισμός ρόλου $P_1(x,y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_2(x,y)$ στο A και το $(A,T) \models P_2 \sqsubseteq P_1$.*

9. *Ένας ισχυρισμός ρόλου $P_1(x,y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_2(y,x)$ στο A και το $(A,T) \models P_2^- \sqsubseteq P_1$.*

10. *Ένας ισχυρισμός ρόλου $P_1(x,y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_2(y,x)$ στο A και το $(A,T) \models P_2 \sqsubseteq P_1^-$.*

11. *Ένας ισχυρισμός ρόλου $P_1(x,y)$ είναι περιττός στο A εάν υπάρχει ισχυρισμός ρόλου $P_2(x,y)$ στο A και το $(A,T) \models P_2^- \sqsubseteq P_1^-$.*

Απόδειξη.

1. Έστω ότι υπάρχουν οι ισχυρισμοί έννοιας $A_1(x)$ και $A_2(x)$ στο A και το $(A,T) \models A_2 \sqsubseteq A_1$. Τότε ισχύει $(A-\{A_1(x)\},T) \models A_1(x)$, επομένως ο ισχυρισμός $A_1(x)$ είναι περιττός σύμφωνα με τον ορισμό 3.2.

2. Η απόδειξη των υπόλοιπων σημείων είναι όμοια με την απόδειξη του προηγούμενου σημείου.

□

Βάσει των παραπάνω κανόνων αξίζει να παρατηρηθεί ότι ένας ισχυρισμός ρόλου μπορεί να χαρακτηριστεί ως περιττός λόγω άλλων ισχυρισμών ρόλου μεταξύ των ίδιων ατόμων. Αντίθετα, ένας ισχυρισμός έννοιας μπορεί να χαρακτηριστεί ως περιττός λόγω της παρουσίας άλλων ισχυρισμών εννοιών ή/και ρόλων που εμπλέκουν τα ίδια άτομα.

Για την εύρεση ενός ελαχιστοτικού ισοδύναμου ABox σε $DL - Lite_R$ μπορεί να χρησιμοποιηθεί ένας επαυξητικός αλγόριθμος (incremental algorithm) που ακολουθεί την ίδια λογική με αυτόν που προτείνεται στο [26].

Algorithm 1 One minimal equivalent Abox w.r.t. any query

```
procedure OneMin(A,T)                                ▷ A is the Abox, T is the TBox
    M ← ∅
    for each assertion A_k ∈ A do
        if A_k is not redundant w.r.t. theorem 4.1 then M ← M ∪ {A_k}
        end if
    end for
    return M
end procedure
```

Θεώρημα 4.2. *Ο αλγόριθμος 1 βρίσκει ένα ελαχιστοτικό ισοδύναμο υποσύνολο του αρχικού ABox.*

Απόδειξη. Αρκεί να δείξουμε ότι το σύνολο περιττών ισχυρισμών που παρουσιάζεται στο θεώρημα 4.1 είναι πλήρες, δηλαδή περιγράφει όλες τις περιπτώσεις περιττών ισχυρισμών σε ένα $DL{-}Lite_R$ $ABox$. Αρχικά, αν εξετάσουμε τη δομή του $DL{-}Lite_R$ $TBox$, παρατηρούμε ότι αυτό μπορεί να περιέχει αξιώματα υπαγωγής εννοιών και ρόλων. Από εξαντλητική καταγραφή όλων των ειδών αξιωμάτων υπαγωγής του TBox παρατηρούμε ότι κάποια από αυτά μπορεί να χρησιμοποιηθούν για το χαρακτηρισμό περιττών ισχυρισμών, ενώ άλλα μόνο για ελέγχους συνέπειας, όπως τα αξιώματα της μορφής $A_k \sqsubseteq \neg\exists R$ ή $\exists R \sqsubseteq \bot$. □

Το πρόβλημα που εξετάζεται στην παρούσα διατριβή είναι η εύρεση όλων των ελαχιστοτικών υποσυνόλων ενός ABox που είναι ισοδύναμα με το αρχικό, υπό την παρουσία συζευκτικών ερωτημάτων. Στην παρούσα ενότητα εξετάζεται η περίπτωση που επιθυμούμε τα παραγόμενα ABox να δίνουν ισοδύναμες απαντήσεις όταν τεθούν σε αυτά αυθαίρετα ερωτήματα. Αρχικά, παρατηρείται ότι οι ακόλουθες ιδιότητες ισχύουν για τους ισχυρισμούς $DL - Lite_R$.

Λήμμα 4.1. *Έστω A ένα $DL - Lite_R$ $ABox$ και $R(a,b) \in A$ ένας ισχυρισμός ρόλου. Κάθε ελαχιστοτικό σύνολο ισχυρισμών S το οποίο είναι ισοδύναμο με το $\{R(a,b)\}$, είναι μονοσύνολο και περιέχει είτε $P(a,b)$ είτε $P(b,a)$, για κάποιο ρόλο P ή P^- ισοδύναμο με τον R, αντίστοιχα.*

Απόδειξη. Όσον αφορά στους ρόλους, το $DL-Lite_R$ *core TBox* επιτρέπει μόνο αξιώματα υπαγωγής ρόλων ανάμεσα σε ατομικούς ρόλους P ή ανάστροφους ρόλους P^-. Επομένως, σχετικά με τους ρόλους, ισοδυναμίες μπορούν να προκύψουν σύμφωνα με τις ακόλουθες μορφές: $P_1 \equiv P_2$, $P_1 \equiv P_2^-$ (και $P_1^- \equiv P_2^-$), όπου P_1, P_2 είναι ατομικοί ρόλοι.

- Ας υποθέσουμε ότι το S δεν περιέχει ισχυρισμό ρόλου $P(a,b)$ ή $P(b,a)$, για κάποιο ρόλο P ή P^- ισοδύναμο με τον R, αντίστοιχα. Με άλλα λογια, μπορεί να περιέχει ισχυρισμούς εννοιών με ή χωρίς άρνηση, ισχυρισμούς ρόλων μεταξύ των a, b για ρόλους που δεν είναι ισοδύναμοι με τον R, αρνήσεις ισχυρισμών ρόλων μεταξύ των a, b, ή/και ισχυρισμούς ρόλων μεταξύ άλλων ατόμων. Επειδή τα $S, R(a,b)$ είναι ισοδύναμα, το $R(a,b)$ θα πρέπει να υπονοεί (υπάγεται σε) κάθε στοιχείο του S. Αντίστοιχα, ο συνδυασμός (σύζευξη) των στοιχείων του S θα πρέπει να υπάγεται στο $R(a,b)$.
  Επειδή ο $R(a,b)$ είναι ισχυρισμός ρόλου, δε μπορεί να υπονοείται από ισχυρισμούς εννοιών. Επίσης, λόγω της δομής του $TBox$ (έλλειψη μεταβατικών ρόλων, κ.λ.π) δε μπορεί να υπονοείται από ισχυρισμούς ρόλων που εμπλέκουν άλλα ονόματα ατόμων. Επιπλέον, λόγω της δομής του $TBox$, η οποία εάν επιτρέπει άρνηση ρόλων, αυτό συμβαίνει μόνο στο δεξί μέλος των αξιωμάτων υπαγωγής, δε μπορεί να υπονοείται από αρνήσεις ισχυρισμών ρόλων. Παρόλα αυτά, μπορεί να υπαινιχθεί από ισχυρισμούς ρόλων $Q(a,b)$, για ρόλους Q τπου δεν είναι ισοδύναμοι με τον R. Τότε, θα πρέπει να ισχύει $Q \sqsubseteq R$. Επειδή το K υπονοεί όλα τα στοιχεία του S, θα πρέπει επίσης να ισχύει $R \sqsubseteq Q$. Συνεπώς, τα R και Q προκύπτουν ισοδύναμα, το οποίο αντιτίθεται στην αρχική υπόθεση.

- Ας υποθέσουμε ότι το S περιέχει ισχυρισμούς ρόλων $P(a,b)$ ή $P(b,a)$, για κάποιους ρόλους P ή P^- ισοδύναμους με τον R, αλλά περιέχει και επιπλέον στοιχεία. Τότε, το S δεν είναι ελαχιστοτικό, επειδή εάν τα επιπλέον στοιχεία αφαιρούνταν, θα προέκυπτε ένα μικρότερο, ισυδύναμο με το $R(a,b)$, σύνολο.

□

Λήμμα 4.2. *Έστω A ένα $DL-Lite_R$ ABox και ένας ισχυρισμός ρόλου με άρνηση $\neg R(a,b) \in A$. Κάθε ελαχιστοτικό σύνολο ισχυρισμών S το οποίο είναι ισοδύναμο με το $\{\neg R(a,b)\}$, είναι μονοσύνολο και περιέχει είτε $\neg P(a,b)$ είτε $\neg P(b,a)$, για κάποιο ρόλο P ή P^- ισοδύναμο με τον R, αντίστοιχα.*

Απόδειξη. Υπενθυμίζεται ότι τα $DL-Lite_R$ *TBox* περιέχουν αξιώματα υπαγωγής ρόλων μόνο μεταξύ ατομικών ρόλων P ή ρόλων της μορφής P^-. Εφόσον η άρνηση δεν εμφανίζεται στα αξιώματα υπαγωγής ρόλων, αξιώματα ισοδυναμίας μπορεί να προκύψουν μόνο ως εξής: $P_1 \equiv P_2$, $P_1 \equiv P_2^-$ (και $P_1^- \equiv P_2^-$). Έτσι, η απόδειξη προκύπτει άμεσα από το λήμμα 4.1 □

Λήμμα 4.3. *Έστω A ένα $DL-Lite_R$ ABox και ισχυρισμός έννοιας $A_k(a) \in A$, όπου A_k ατομική έννοια. Κάθε ελαχιστοτικό σύνολο ισχυρισμών S το οποίο είναι ισοδύναμο με το $\{A_k(a)\}$, είναι μονοσύνολο και περιέχει κάποιον ισχυρισμό έννοιας $B(a)$, για κάποια ατομική έννοια B ισοδύναμη με την έννοια A_k.*

Απόδειξη. Αυτή η περίπτωση είναι λίγο διαφορετική από την περίπτωση των ισχυρισμών ρόλων επειδή, παρόλο που το TBox επιτρέπει υπαγωγές εννοιών μεταξύ απλών ή σύνθετων εννοιών, σύμφωνα με το θεώρημα 4.1 μία έννοια μπορεί να χαρακτηριστεί ως περιττή λόγω της παρουσίας άλλων ισχυρισμών εννοιών ή/και ρόλων στο ABox. Επίσης αξίζει να σημειωθεί ότι οι ισοδυναμίες εννοιών που περιλαμβάνουν ονόματα εννοιών περιορίζονται στις μορφές: $A_1 \equiv A_2$, $A_1 \equiv \exists R$ (και $A_1 \equiv \bot$, η οποία μπορεί να χρησιμοποιηθεί για ελέγχους συνέπειας του ABox), όπου A_1, A_2 είναι ατομικές έννοιες και R είναι σύνθετος ρόλος. Αυτές οι ισοδυναμίες εννοιών δεν μπορούν να περιέχουν άρνηση, λόγω της δομής του $TBox$ που επιτρέπει την εμφάνιση αρνήσεων μόνο στο δεξί μέλος των αξιωμάτων υπαγωγής εννοιών.

- Ας υποθέσουμε ότι το S δεν περιέχει κάποιον ισχυρισμό έννοιας $B(a)$ όπου B ατομική έννοια ισοδύναμη με την A_k. Έτσι, μπορεί να περιέχει ισχυρισμούς έννοιας $C(a)$, για έννοιες C που δεν είναι ισοδύναμες με την A_k, αρνήσεις ισχυρισμών έννοιας $\neg C(a)$, ή/και ισχυρισμούς ρόλων που μπορεί να περιέχουν αρνήσεις ή όχι. Επειδή τα $S, \{A_k(a)\}$ είναι ισοδύναμα, το $A_k(a)$ θα πρέπει να υπονοεί κάθε στοιχείο του S. Επίσης, ο συνδυασμός (σύζευξη) των στοιχείων του S θα πρέπει να υπονοεί τον $A_k(a)$. Επειδή το $A_k(a)$ είναι ισχυρισμός έννοιας, λόγω της δομής του TBox δε μπορεί να υποννοείται από ισχυρισμούς εννοιών που αφορούν σε άλλα ονόματα ατόμων, ούτε από αρνήσεις ισχυρισμών εννοιών ή ρόλων που εμπλέκουν τον a. Παρόλα αυτά μπορεί να υπονοείται από ισχυρισμούς εννοιών $C(a)$, για έννοιες C που δεν είναι ισοδύναμες με την A_k. Σε αυτή την περίπτωση, πρέπει να ισχύει $C \sqsubseteq A_k$. Επίσης, επειδή το A_k υπάγεται σε κάθε στοιχείο του S, θα πρέπει να ισχύει και $A_k \sqsubseteq C$, δηλαδή $C \equiv A_k$, το οποίο, όμως, αντιτίθεται στην αρχική υπόθεση. Επιπλέον, το $A_k(a)$ μπορεί να υπονοείται από την ύπαρξη ισχυρισμών ρόλων $R(a, b)$ (ή $R(b, a)$) για ρόλους R τέτοιους ώστε $T \models \exists R \sqsubseteq A$ (ή $T \models \exists R^- \sqsubseteq A$). Όμως, όπως εξηγήθηκε στο λήμμα 4.1, ένα αξίωμα υπαγωγής ρόλου δε μπορεί να υπονοείται από κάποιο αξίωμα υπαγωγής έννοιας, το οποίο αντιτίθεται στην αρχική υπόθεση ότι το $A_k(a)$ θα πρέπει να υπονοεί όλα τα στοιχεία του S.
- Ας υποθέσουμε ότι το S περιέχει έναν ισχυρισμό έννοιας $B(a)$, για κάποια ατομική έννοια B ισοδύναμη με την A, αλλά περιέχει και επιπλέο στοιχεία. Τότε, το S δεν είναι ισοδύναμο γιατί εάν αφαιρεθούν τα επιπλέον στοιχεία θα προκύψει ένα μικρότερο ισοδύναμο με το $A_k(a)$ σύνολο, το οποίο είναι υποσύνολο του S.

□

Λήμμα 4.4. *Έστω A ένα $DL - Lite_R$ ABox και ισχυρισμός έννοιας $\neg A_k(a) \in A$, όπου A_k ατομική έννοια. Κάθε ελαχιστοτικό σύνολο ισχυρισμών S το οποίο είναι ισοδύναμο με το $\{\neg A_k(a)\}$, είναι μονοσύνολο και περιέχει κάποιον ισχυρισμό έννοιας $\neg B(a)$, για κάποια ατομική έννοια B ισοδύναμη με την έννοια A_k.*

Απόδειξη. Όπως αναφέρθηκε στο λήμμα 4.3 ισοδυναμίες ατομικών εννοιών δεν μπορούν να περιέχουν αρνήσεις, λόγω της δομής του TBox, το οποίο επιτρέπει αρνήσεις μόνο στη δεξιά πλευρά των αξιωμάτων υπαγωγής εννοιών. Συνεπώς, οι ισοδυναμίες εννοιών που μπορεί να περιέχουν έννοιες με άρνηση μπορεί να εμφανιστούν μόνο στις μορφές: $\neg A_1 \equiv \neg A_2$ και $\neg A_1 \equiv \neg \exists R_1$, όπου A_1, A_2 ατομικές έννοιες και R_1 ατομικός ή σύνθετος ρόλος. Έτσι, η απόδειξη προκύπτει άμεσα από το 4.3.

□

Βάσει των ιδιοτήτων που αποδείχθηκαν προηγουμένως στα λήμματα 4.1, 4.2, 4.3, 4.4 μπορεί να προταθεί ένας αλγόριθμος που επιλύει το πρόβλημα που αναφέρθηκε και να αποδειχθεί η ορθότητά του. Αρχικά, ο αλγόριθμος αυτός βασίζεται στην ύπαρξη ενός ελαχιστοτικού ισοδύναμου ABox, το οποίο μπορεί να παραχθεί με τον αλγόριθμο 1 που περιγράφηκε προηγουμένως, και το χρησιμοποιεί για να παράξει όλα τα υπόλοιπα ισοδύναμα ελαχιστοτικά υποσύνολα ενός αρχικού ABox.

Algorithm 2 All minimal equivalent Aboxes w.r.t. every query

```
procedure AllMin(A,M,T)                          ▷ A is the ABox, T is the TBox
                                                 ▷ M is a minimal equivalent Abox
    S ← ∅
    for each assertion A1 ∈ A equivalent to the first assertion of M do
        for each assertion A2 ∈ A equivalent to the second assertion of M do
            for ... do
                for each assertion An ∈ A equivalent to the last assertion of M do
                    S ← S ∪ {A1, A2, ..., An}
                end for
            end for
        end for
    end for
    return S
end procedure
```

Θεώρημα 4.3. *Ο αλγόριθμος 2 παράγει όλα τα ελαχιστοτικά υποσύνολα του αρχικού ABox που είναι ισοδύναμα με το αρχικό και μόνο αυτά.*

Απόδειξη. Ο αλγόριθμος εξ' ορισμού παράγει ABox που είναι ισοδύναμα με το αρχικό, αφού κάθε ισχυρισμός κάθε παραγόμενου ABox είναι ισοδύναμος με κάποιον ξεχωριστό ισχυρισμό του αρχικού. Ας υποθέσουμε ότι ο αλγόριθμος παράγει ένα ABox που δεν είναι ελαχιστοτικό. Τότε αυτό περιέχει κάποιο στοιχείο K που μπορεί να αφαιρεθεί επειδή υπονοείται από άλλα στοιχεία του. Έστω L το ισοδύναμο στοιχείο του K που ανήκει στο δοσμένο ελαχιστοτικό ABox. Τότε, το L μπορεί επίσης να υπαινιχθεί από τα στοιχεία του δοσμένου ελαχιστοτικού ABox, που είναι ισοδύναμα με αυτά του παραγόμενου. Αυτό αντιτίθεται στην αρχική υπόθεση ότι το αρχικό ABox είναι ελαχιστοτικό. Ας υποθέσουμε τώρα ότι υπάρχει ένα ελαχιστοτικό ισοδύναμο υποσύνολο του αρχικού ABox, B, το οποίο δεν παράγεται από τον αλγόριθμο 2. Επειδή το B είναι ισοδύναμο με το M και είναι και τα δύο ελαχιστοτικά, και με χρήση των 4.1, 4.2, 4.3, 4.4 προκύπτει ότι κάθε στοιχείο του B είναι ισοδύναμο με ακριβώς ένα στοιχείο του M και αντίστροφα. Εξ' ορισμού, ο αλγόριθμος 2 παράγει το B.

□

4.2 Ελαχιστοποίηση ABox υπό την παρουσία ενός συνόλου συζευκτικών ερωτημάτων

Το πρόβλημα της εύρεσης ελαχιστοτικών ισοδύναμων ABox αλλάζει όταν επιθυμούμε τα παραγόμενα ABox να είναι ισοδύναμα υπό την παρουσία ενός συγκεκριμένου συνόλου συζευκτικών ερωτημάτων. Αυτό φαίνεται καλύτερα στο παρακάτω παράδειγμα.

Παράδειγμα 4.1. Για παράδειγμα, έστω ότι έχουμε το ABox:

$$A = \{\text{Γονιός(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ναυσικά)},\ \text{Άνθρωπος(Μαρία)}\}$$

και το TBox:

$$T = \{\exists\text{έχειΠαιδί} \sqsubseteq \text{Γονιός}\}$$

και μας ενδιαφέρει το ερώτημα:

$$q(x) \leftarrow \{\text{Γονιός}(x)\}$$

Τότε, τα ελαχιστοτικά ισοδύναμα ABox ως προς το q είναι τα ακόλουθα:

$$A_1 = \{\text{Γονιός(Μαρία)}\}$$

$$A_2 = \{\text{έχειΠαιδί(Μαρία,Ναυσικά)}\}$$

Μέσω του παραπάνω παραδείγματος γίνεται εμφανές ότι τα παραγόμενα ABox δεν είναι απαραίτητα ισοδύναμα με το αρχικό, ούτε μεταξύ τους. Παραπάνω, το A_2 περιέχει περισσότερη πληροφορία από το A_1. Είναι εμφανές ότι αυτό που μας ενδιαφέρει είναι, τα παραγόμενα ABox να δίνουν ίδιο σύνολο απαντήσεων με το αρχικό ABox, όταν σε αυτά τεθούν τα ερωτήματα που μας ενδιαφέρουν.

Η κεντρική ιδέα της παρούσας προσέγγισης για την επίλυση αυτού του προβλήματος βασίζεται στο γεγονός ότι όταν ψάχνουμε όλα τα ελαχιστοτικά υποσύνολα ενός ABox ως προς ένα σύνολο από ερωτήματα, ουσιαστικά ψάχνουμε όλους τους διαφορετικούς τρόπους με τους οποίους μπορούν να απαντηθούν αυτά τα ερωτήματα. Αυτό σχετίζεται άμεσα με το πρόβλημα της επαναγραφής ερωτημάτων, που παρουσιάστηκε σε προηγούμενο κεφάλαιο, κατά την οποία παράγονται όλα τα εναλλακτικά συζευκτικά ερωτήματα που μπορούν να δώσουν ισοδύναμη απάντηση με ένα αρχικό ερώτημα. Έτσι, η κεντρική ιδέα της παρούσας προσέγγισης περιλαμβάνει τη χρήση των επαναγραφών των δεδομένων ερωτημάτων για να παραχθούν τα ισοδύναμα, ως προς τα ερωτήματα αυτά, υποσύνολα ενός ABox. Τα παραπάνω μπορούν να γίνουν πιο κατανοητά με ένα παράδειγμα.

Παράδειγμα 4.2. Για παράδειγμα, έστω ότι έχουμε το ABox:

$$A = \{\text{Γυναίκα(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ναυσικά)}, \text{έχειΠαιδί(Μαρία,Ελένη)},$$

$$\text{Γυναίκα(Ναυσικά)},\ \text{Άνθρωπος(Ελένη)}, \text{Ψηλός(Ελένη)}\}$$

και το TBox:

$$T = \{\text{Γυναίκα} \sqsubseteq \text{Άνθρωπος}\}$$

και μας ενδιαφέρει το ερώτημα:

$$q(x) \leftarrow \{\text{Άνθρωπος}(x),\ \text{έχειΠαιδί}(x, y),\ \text{Άνθρωπος}(y)\}$$

Οι επαναγραφές του ερωτήματος q είναι:

$$q_1 : q(x) \leftarrow \{\text{Γυναίκα}(x),\ \text{έχειΠαιδί}(x, y),\ \text{Άνθρωπος}(y)\}$$

$$q_2 : q(x) \leftarrow \{\text{Άνθρωπος}(x),\ \text{έχειΠαιδί}(x, y),\ \text{Γυναίκα}(y)\}$$

$$q_3 : q(x) \leftarrow \{\text{Γυναίκα}(x),\ \text{έχειΠαιδί}(x, y),\ \text{Γυναίκα}(y)\}$$

Το q_1 μπορεί να απαντηθεί από το $\{\text{Γυναίκα(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ελένη)}, \text{Άνθρωπος(Ελένη)}\}$ και το q_3 μπορεί να απαντηθεί από το $\{\text{Γυναίκα(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ναυσικά)}, \text{Γυναίκα(Ναυσικά)}\}$, ενώ το q_2 δε μπορεί να απαντηθεί χωρίς τη χρήση του TBox. Έτσι, μπορούν να προκύψουν τα ακόλουθα ελαχιστοτικά ισοδύναμα ως προς το q ABox:

$$A_1 = \{\text{Γυναίκα(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ελένη)}, \text{Άνθρωπος(Ελένη)}\}$$

$$A_2 = \{\text{Γυναίκα(Μαρία)},\ \text{έχειΠαιδί(Μαρία,Ναυσικά)}, \text{Γυναίκα(Ναυσικά)}\}$$

Έτσι, βάσει της ιδέας που παρουσιάζεται στο παραπάνω παράδειγμα, προκύπτει η ανάγκη για το σαφή προσδιορισμό των ABox που προκύπτουν από τη διαδικασία απάντησης ενός ερωτήματος. Ουσιαστικά, το ABox κατασκευάζεται παίρνοντας όλα τα στοιχεία (ατομικούς τύπους) του συζευκτικού ερωτήματος, όταν σε αυτό έχει γίνει αντικατάσταση των ελεύθερων και δεσμευμένων μεταβλητών του με σταθερές (εδώ ονόματα ατόμων), σύμφωνα με ένα διάνυσμα απάντησης.

Ορισμός 4.1. *Έστω $q(\vec{x})$ ένα συζευκτικό ερώτημα και $\vec{a}$ ένα διάνυσμα απάντησης του ερωτήματος q. Ως υποστηρικτικό (supporting) ABox ορίζεται το ABox που προκύπτει από τη διαδικασία απάντησης ενός συζευκτικού ερωτήματος ως εξής:*

$$A_{supportive} = \{A_i : A_i \text{ ατομικός τύπος του } q(\vec{a})\}$$

Επομένως, οι λύσεις που αναζητούμε μπορούν να βρεθούν ανάμεσα στα υποστηρικτικά ABox του συζευκτικού ερωτήματος. Η ιδέα αυτή, αν και αρχικά φαίνεται απλή, κρύβει κάποιες επιπλοκές. Αρχικά, μία σημαντική παρατήρηση είναι ότι τα ABox που κατασκευάζονται με τον τρόπο που περιγράφηκε παραπάνω δεν είναι κατ΄ ανάγκη ελαχιστοτικά. Αυτό μπορεί να συμβαίνει είτε επειδή οι περιορισμοί που απαιτεί το ερώτημα περιέχουν περιττές πληροφορίες, είτε επειδή ο τρόπος με τον οποίο οι μεταβλητές του ερωτήματος αντιστοιχίστηκαν σε άτομα δημιουργεί περιττούς ισχυρισμούς. Στα παρακάτω παραδείγματα γίνονται κατανοητές οι περιπτώσεις αυτές.

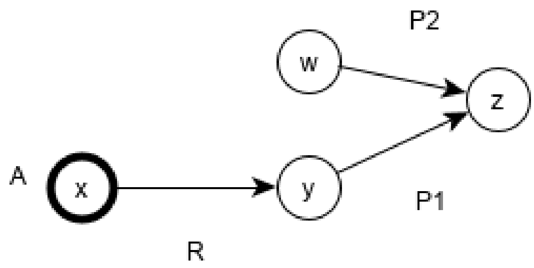
Παράδειγμα 4.3. Για παράδειγμα, έστω ότι το ABox περιέχει τα:

$$\{A(a), R(a, b), P1(b, c), P2(b, c)\}$$

και πιθανώς επιπλέον στοιχεία, και το TBox περιέχει:

$$\{P1 \sqsubseteq P2\}$$

Μας ενδιαφέρει το ερώτημα του σχήματος 4.1. Το υποστηρικτικό ABox που παράγεται από άμεση απάντηση του ερωτήματος του σχήματος είναι το $\{A(a), R(a, b), P1(b, c), P2(b, c)\}$, το οποίο δεν είναι ελαχιστοτικό επειδή περιέχει τον περιττό ισχυρισμό ρόλου $P2(b, c)$.



Σχήμα 4.1: Ερώτημα με περιττούς περιορισμούς

Παράδειγμα 4.4. Σε αυτό το παράδειγμα παρουσιάζεται ένα ερώτημα που δεν περιέχει περιττές απαιτήσεις, αλλά κατά την απάντησή του μπορεί να προκύψει μη ελαχιστοτικό υποστηρικτικό ABox. Έστω ότι το ABox περιέχει τα

$$\{A(a), R(a,b), P(b,c), W(c,d), R1(a,b), P1(b,c)\}$$

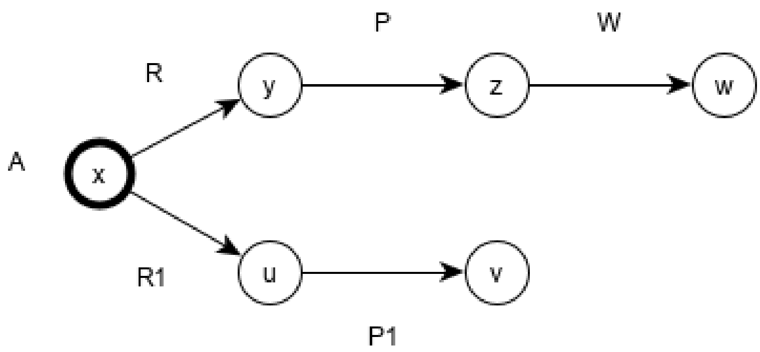
και το TBox τα:

$$\{P1 \sqsubseteq P,\ R1 \sqsubseteq\ R\}$$

και μας ενδιαφέρει το ερώτημα του σχήματος 4.2. Τότε, το σύνολο

$$\{A(a), R(a,b), P(b,c), W(c,d), R1(a,b), P1(b,c)\}$$

μπορεί να προκύψει ως απάντηση του ερωτήματος, το οποίο δεν είναι ελαχιστοτικό γιατί περιέχει τους περιττούς ισχυρισμούς $R(a,b), P(b,c)$.



Σχήμα 4.2: Ερώτημα μπορεί να δημιουργήσει μη ελαχιστοτικό ABox

Για την εύρεση όλων των ελαχιστοτικών υποσυνόλων ενός ABox που είναι ισοδύναμα με το αρχικό υπό την παρουσία ενός συνόλου συζευκτικών ερωτημάτων q προτείνεται ο αλγόριθμος 3. Ο αλγόριθμος αυτός βασίζεται στην αξιοποίηση των επαναγραφών των δεδομένων ερωτημάτων για την εύρεση των υποστηρικτικών ABox, ενώ παράλληλα εκτελεί τους κατάλληλους ελέγχους που εξασφαλίζουν την ελαχιστότητα των λύσεων.

Algorithm 3 All minimal equivalent ABoxes w.r.t. a set of queries Q

procedure PROC1(Q,A,T) ▷ $Q = \{q_1, ..., q_n\}$ contains all queries of interest
▷ A is the ABox, T the TBox
    $S = \emptyset$
    $Ans = \bigcup_{i=1}^{n} ans(q_i)$
    for $a \in Ans$ **do**
        $Q_a = \{q_i \in Q : a \in ans(q_i)\}$
        $q_a = join(Q_a)$
    end for
    $S = \{\bigcup_{a \in Ans} S_a$: S_a is a supporting ABox of some $q \in rewr(q_a)$ leading to the answer $a\}$
    return S
end procedure

procedure PROC2(S) ▷ S is a set of ABoxes
    for $A \in S$ **do**
        for $B \in S$ **do**
            if $A \subset B$ **then**
                $S \leftarrow S - B$
                break
            end if
        end for
    end for
    return S
end procedure

$V \leftarrow Proc1(Q, A)$
$S \leftarrow Proc2(V)$

Θεώρημα 4.4. *Ο αλγόριθμος 3 παράγει όλα τα ελαχιστοτικά ισοδύναμα υποσύνολα ενός ABox ως προς ένα σύνολο συζευκτικών ερωτημάτων.*

Απόδειξη. Η πληρότητα του αλγορίθμου 3 ανάγεται στην πληρότητα του συστήματος επαναγραφής ερωτημάτων. Με άλλα λόγια, ας υποθέσουμε ότι υπάρχει ένα ελαχιστοτικό υποσύνολο του ABox, M, ισοδύναμο με το αρχικό ως προς το δεδομένο σύνολο ερωτημάτων Q, που όμως δεν έχει παραχθεί από τον αλγόριθμο. Επειδή το M θα πρέπει να δίνει ισοδύναμες απαντήσεις ως προς το Q, το M θα πρέπει να περιέχει κάποια υποστηρικτικά σύνολα ισχυρισμών, όχι απαραίτητα ξένα μεταξύ τους, που θα αντιστοιχούν στις απαντήσεις του αρχικού ABox ως προς τα ερωτήματα του Q ή κάποιες επανεγγραφές τους. Εξόρισμού η διαδικασία $Proc1$ παράγει αυτό το ABox. Έπειτα, το M δεν μπορεί να αφαιρεθεί κατά την $Proc2$, γιατί αλλιώς δεν θα ήταν ελαχιστοτικό.

Ας θεωρήσουμε τώρα ότι ο αλγόριθμος παράγει κάποιο ABox B που δεν είναι ελαχιστοτικά ισοδύναμο ως προς το Q. Τότε, κάποιο υποσύνολό του, M, θα είναι ελαχιστοτικά ισοδύναμο με το αρχικό ως προς το Q. Συνεπώς, όπως αποδείχθηκε προηγουμένως, το M θα παραχθεί από την $Proc1$, και εφόσον $M \subset B$, το B θα αφαιρεθεί από την $Proc2$.

□

Κεφάλαιο 5

Μελλοντική Εργασία

Όπως αναφέρθηκε προηγουμένως, το έναυσμα για την εργασία αυτήν ήταν η σύνδεση της Αναπαράστασης Γνώσης και της Μηχανικής Μάθησης. Επομένως, ένα επόμενο βήμα θα μπορούσε να είναι η χρήση των ισοδύναμων υποσυνόλων ως δεδομένα εκμάθησης για την εκπαίδευση ενός συστήματος που μπορεί να απαντάει συζευκτικά ερωτήματα.

Μία διαφορετική κατεύθυνση είναι η βελτιστοποίηση των προτεινόμενων αλγορίθμων. Συγκεκριμένα, ο αλγόριθμος 3 χρησιμοποιεί τα συστήματα απάντησης και επαναγραφής ερωτημάτων σας μαύρα κουτιά, αφού χρησιμοποιεί τις εξόδους τους αυτούσιες. Το γεγονός αυτό, οπως είναι λογικό, πιθανώς αυξάνει κατά πολύ την πολυπλοκότητα του αλγορίθμου που προτείνεται. Αντίθετα, ο αλγόριθμος 1 ελαχιστοποιεί το δεδομένο ABox αξιοποιώντας τη δομή του TBox και κάνοντας χαρακτηρισμό των περιττών ισχυρισμών. Έτσι, μία ενδιαφέρουσα κατεύθυνση είναι η προσπάθεια αντικατάστασης του Black Box αλγορίθμου 3 με κάποιον που θα αναλύει τη δομή των επιθυμητών ερωτημάτων και θα προσπαθεί να αποφεύγει τη δημιουργία περιττών λύσεων, οι οποίες είναι βέβαιο ότι θα απαλειφθούν στη συνέχεια.

Βιβλιογραφία

[1] P. Abraham Silberschatz, H.F. Korth, and S. Sudarshan. *Database System Concepts*. McGraw-Hill Education, 2010. ISBN: 9780073523323. URL: https://books.google.co.in/books?id=re4YQAAACAAJ.

[2] Andrea Acciarri et al. "QuOnto: Querying Ontologies." In: vol. 4. Jan. 2005, pp. 1670–1671.

[3] P.B. Andrews. *An introduction to mathematical logic and type theory: to truth through proof*. Computer science and applied mathematics. Academic Press, 1986. URL: https://books.google.gr/books?id=8qUoAQAAMAAJ.

[4] A. Artale et al. "The DL-Lite Family and Relations". In: *Journal of Artificial Intelligence Research* 36 (Oct. 2009), pp. 1–69. ISSN: 1076-9757. DOI: 10.1613/jair.2820. URL: http://dx.doi.org/10.1613/jair.2820.

[5] Franz Baader. *The Description Logic Handbook: Theory, Implementation, and Applications*. Jan. 2003.

[6] Samantha Bail, Bijan Parsia, and Uli Sattler. "Extracting Finite Sets of Entailments from OWL Ontologies." In: vol. 745. Jan. 2011.

[7] J. Barwise. "An Introduction to First-Order Logic". In: *Handbook of Mathematical Logic*. Ed. by J. Barwise. Amsterdam: North-Holland, 1977, pp. 5–46.

[8] Ronal Brachman and Hector Levesque. *Knowledge Representation and Reasoning*. May 2004. ISBN: 1558609326.

[9] Diego Calvanese et al. "Data Complexity of Query Answering in Description Logics." In: vol. 195. Jan. 2005. DOI: 10.1016/j.artint.2012.10.003.

[10] Diego Calvanese et al. "DL-Lite: Tractable Description Logics for Ontologies." In: vol. 2. Jan. 2005, pp. 602–607.

[11] Diego Calvanese et al. "Reasoning about Explanations for Negative Query Answers in DL-Lite". In: *J. of Artificial Intelligence Research* 48 (2013), pp. 635–669. DOI: 10.1613/jair.3870.

[12] Diego Calvanese et al. "Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family". In: *Journal of Automated Reasoning* 39 (Oct. 2007), pp. 385–429. DOI: 10.1007/s10817-007-9078-x.

[13] Alexandros Chortaras, Despoina Trivela, and Giorgos Stamou. "Optimized Query Rewriting for OWL 2 QL". In: vol. 6803. July 2011, pp. 192–206. DOI: 10.1007/978-3-642-22438-6_16.

[14] J.F. Du et al. "Towards practical ABox abduction in large OWL DL ontologies". In: vol. 2. Jan. 2011.

[15] Jianfeng Du, Kewen Wang, and Yi-Dong Shen. "A Tractable Approach to Abox Abduction over Description Logic Ontologies". In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI'14. Québec City, Québec, Canada: AAAI Press, 2014, pp. 1034–1040. URL: http://dl.acm.org/citation.cfm?id=2893873.2894034.

[16] Achille Fokoue, Aaron Kershenbaum, and Li Ma. "SHIN ABox Reduction". In: (Jan. 2006).

[17] Achille Fokoue et al. "The Summary Abox: Cutting Ontologies Down to Size". In: Nov. 2006, pp. 343–356. DOI: 10.1007/11926078_25.

[18] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. "Ontological Queries: Rewriting and Optimization (Extended Version)". In: (Dec. 2011).

[19] Bernardo Grau et al. "Modular Reuse of Ontologies: Theory and Practice". In: *J. Artif. Intell. Res. (JAIR)* 31 (Jan. 2008), pp. 273–318. DOI: 10.1613/jair.2375.

[20] Frank Harmelen, Vladimir Lifschitz, and Bruce Porter. "The Handbook of Knowledge Representation". In: *Elsevier Science San Diego, USA* (Jan. 2007), p. 1034.

[21] D. Hilbert and W. Ackermann. *Principles of Mathematical Logic*. Chelsea scientific books. Chelsea Publishing Company, 1950. URL: https://books.google.gr/books?id=XqAXAAAAIAAJ.

[22] Aditya Kalyanpur et al. "Finding All Justifications of OWL DL Entailments". In: Jan. 2007, pp. 267–280. DOI: 10.1007/978-3-540-76298-0_20.

[23] Szymon Klarman, Ulle Endriss, and Stefan Schlobach. "ABox Abduction in the Description Logic ALC". In: *J. Autom. Reasoning* 46 (2011), pp. 43–80.

[24] Boris Konev et al. "Semantic Modularity and Module Extraction in Description Logics". In: Jan. 2008, pp. 55–59. DOI: 10.3233/978-1-58603-891-5-55.

[25] Roman Kontchakov and Michael Zakharyaschev. "An Introduction to Description Logics and Query Rewriting". In: *Reasoning Web. Reasoning on the Web in the Big Data Era: 10th International Summer School 2014, Athens, Greece, September 8-13, 2014. Proceedings*. Ed. by Manolis Koubarakis et al. Cham: Springer International Publishing, 2014, pp. 195–244. ISBN: 978-3-319-10587-1. DOI: 10.1007/978-3-319-10587-1_5. URL: https://doi.org/10.1007/978-3-319-10587-1_5.

[26] Julius Köpke, Johann Eder, and Michaela Schicho. "Efficient Projection of Ontologies". In: vol. 4. Sept. 2013. DOI: 10.1007/978-3-642-41030-7_48.

[27] Natasha Noy and Mark Musen. "Specifying Ontology Views by Traversal". In: Nov. 2004, pp. 713–725. DOI: 10.1007/978-3-540-30475-3_49.

[28] Magdalena Ortiz and Mantas Šimkus. "Reasoning and Query Answering in Description Logics". In: (Sept. 2012). DOI: 10.1007/978-3-642-33158-9_1.

[29] Hector Perez-Urbina, Ian Horrocks, and Boris Motik. "Practical Aspects of Query Rewriting for OWL 2." In: vol. 529. Jan. 2009.

[30] M. Rodríguez-Muro, R. Kontchakov, and Michael Zakharyaschev. "Query rewriting and optimisation with database dependencies in Ontop". In: *CEUR Workshop Proceedings* 1014 (Jan. 2013), pp. 917–929.

[31] Mariano Rodriguez-Muro and Diego Calvanese. "Dependencies: Making Ontology Based Data Access Work." In: vol. 749. Jan. 2011.

[32] Mariano Rodríguez-Muro and Diego Calvanese. "Quest, a System for Ontology Based Data Access". In: *CEUR Workshop Proceedings* 849 (Jan. 2012).

[33] Riccardo Rosati. "Prexto: Query Rewriting under Extensional Constraints in DL-Lite". In: *The Semantic Web: Research and Applications*. Ed. by Elena Simperl et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 360–374. ISBN: 978-3-642-30284-8.

[34] Heiner Stuckenschmidt, Christine Parent, and Stefano Spaccapietra. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*. Vol. 5445. Jan. 2009. DOI: 10.1007/978-3-642-01907-4.

[35] Despoina Trivela et al. "Optimising resolution-based rewriting algorithms for OWL ontologies". In: *Journal of Web Semantics* 33 (2015). Ontology-based Data Access, pp. 30–49. ISSN: 1570-8268. DOI: https://doi.org/10.1016/j.websem.2015.02.001. URL: http://www.sciencedirect.com/science/article/pii/S1570826815000141.

[36] Moshe Vardi. "The Complexity of Relational Query Languages (Extended Abstract)". In: Jan. 1982, pp. 137–146. DOI: 10.1145/800070.802186.