Εθνικό Μετσόβιο Πολυτεχνείο
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Επαναγραφή και Συνεπής Απάντηση Συζευκτικών Ερωτημάτων σε Εξελισσόμενες Βάσεις Γνώσης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

της

ΕΛΕΝΗΣ Γ. ΤΣΑΛΑΠΑΤΗ

Διπλωματούχου της Σχολής Εφαρμοσμένων Μαθηματικών &
Φυσικών Επιστημών Ε.Μ.Π. (2005)

Αθήνα, Μάρτιος 2016

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Επαναγραφή και Συνεπής Απάντηση Συζευκτικών Ερωτημάτων σε Εξελισσόμενες Βάσεις Γνώσης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

της

ΕΛΕΝΗΣ Γ. ΤΣΑΛΑΠΑΤΗ

Διπλωματούχου της Σχολής Εφαρμοσμένων Μαθηματικών & Φυσικών Επιστημών Ε.Μ.Π. (2005)

Συμβουλευτική Επιτροπή: Γεώργιος Κολέτσος
Γεώργιος Στάμου
Ευστάθιος Ζάχος

Εγκρίθηκε από την επταμελή επιτροπή την 29η Μαρτίου 2016.

. . .	. . .	. . .
Γεώργιος Κολέτσος	Γεώργιος Στάμου	Ευστάθιος Ζάχος
Καθηγητής Ε.Μ.Π.	Επ. Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.
. . .	. . .	. . .
Στέφανος Κόλλιας	Εμμανουήλ Κουμπαράκης	Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Κ.Π.Α.	Καθηγητής Ε.Μ.Π.
	. . .	
	Δημήτριος Φωτάκης	
	Επ. Καθηγητής Ε.Μ.Π.	

Αθήνα, Μάρτιος 2016

42.

Answer to the Ultimate Question of
Life, the Universe, and Everything,
Douglas Adams

Περιεχόμενα

Κατάλογος Αλγορίθμων

Κατάλογος Πινάκων

ΠΡΟΛΟΓΟΣ

Ο χώρος της Μαθηματικής Λογικής κέντρισε το ενδιαφέρον μου ήδη από τις προπτυχιακές μου σπουδές στη Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του ΕΜΠ. Στη συνέχεια, κατά τη διάρκεια των μεταπτυχιακών μου σπουδών στο Πανεπιστήμιο του Εδιμβούργου είχα την ευκαιρία να διερευνήσω την εφαρμογή της στο πεδίο της Τεχνητής Νοημοσύνης, ένας χώρος που με συγκίνησε ιδιαίτερα και στον οποίο θέλησα να εμβαθύνω. Τη δυνατότητα αυτή μου πρόσφερε ο επιβλέπων καθηγητής μου, καθηγητής του ΕΜΠ, κ. Γεώργιος Κολέτσος, τον οποίο θέλω να ευχαριστήσω ιδιαίτερα για το ενδιαφέρον του, την υποστήριξή του και την προθυμία του να βοηθήσει σε οποιοδήποτε πρόβλημα παρουσιαζόταν καθόλη τη διάρκεια της εκπόνησης της διδακτορικής διατριβής μου.

Ιδιαίτερο ρόλο στην έκβαση της διατριβής αυτής είχε, επίσης, ο επίκουρος καθηγητής του ΕΜΠ κ. Γιώργος Στάμου, ο οποίος ανήκει στην τριμελή επιτροπή και επέβλεψε στενά τη συνολική ερευνητική πορεία μου. Οι γνώσεις, οι συμβουλές, η υποστήριξη, η διάθεση και το ενδιαφέρον του ήταν για εμένα περισσότερο από πολύτιμες. Τον ευχαριστώ θερμά. Θερμές ευχαριστίες, οφείλω και στο τρίτο μέλος της τριμελούς επιτροπής, καθηγητή του ΕΜΠ, κ. Ευστάθιο Ζάχο. Θέλω, επιπλέον, να ευχαριστήσω ιδιαίτερα τον καθηγητή του ΕΜΠ κ. Στέφανο Κόλλια που με δέχθηκε στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων. Τον ευχαριστώ, όχι μόνο για την αμέριστη υποστήριξή του οποτεδήποτε αυτό χρειαζόταν, αλλά ταυτόχρονα γιατί φρόντιζε να υπάρχει πάντα στο εργαστήριο ένα υγιές, ευχάριστο και ταυτόχρονα παραγωγικό περιβάλλον εργασίας.

Τα τελευταία χρόνια της διατριβής μου είχα τη χαρά να συνεργαστώ με τον Δρ. Γιώργο Στοΐλο. Οι κατευθυντήριες γραμμές και οι εποικοδομητικές συμβουλές του ήταν καθοριστικές για την έκβαση της διατριβής μου. Το πολυτιμότερο όμως κέρδος από τη συνεργασία μας ήταν ότι έμαθα να εργάζομαι με μεθοδικότητα και επιστημονική αυστηρότητα. Τον ευχαριστώ θερμά για την επιστημονική του συμβολή, την υπομονή και επιμονή του. Ακόμη, ευχαριστώ και τα υπόλοιπα παιδιά του εργαστηρίου, πολλά από τα οποία έχω τη χαρά πλέον να θεωρώ καλούς φίλους. Η παρουσία τους στο εργαστήριο, οι συζητήσεις μας, επιστημονικού και μη περιεχομένου, έκανε την αντιμετώπιση των καθημερινών ερευνητικών δυσκολιών σημαντικά πιο εύκολη.

Η εκπόνηση μίας διδακτορικής διατριβής κρύβει πολλές στιγμές χαράς και ικανοποίησης αλλά εξίσου πολλές, ίσως και περισσότερες, δυσκολίες, προκλήσεις και απογοητεύσεις, τις οποίες είναι πραγματικά δύσκολο να υπερβεί κανείς αν δεν έχει γύρω του ανθρώπους να τον υποστηρίζουν και ενθαρρύνουν. Σε όλη αυτήν την πορεία, είχα την απίστευτη τύχη να έχω δίπλα μου τον σύντροφό μου Σάββα, χωρίς τον οποίο, η εργασία αυτή δε θα είχε πραγματοποιηθεί. Σημαντικό ρόλο, φυσικά, είχαν επίσης τα αδέρφια μου, Άγγελος, Ρένη και Παναγιώτης, καθώς και τα ανιψάκια μου, Ωκεανός και Αύγουστος, που όποτε τους κοιτάζω γίνομαι λίγο πιο ευτυχισμένη. Θέλω, να ευχαριστήσω τους φίλους μου, που, πολλές φορές παρά την απόσταση, ήταν και είναι συνεχώς με τον τρόπο τους δίπλα μου. Θέλω ιδιαίτερα να ευχαριστήσω τις φίλες μου Αλεξάνδρα, Βίκυ, Αρετή, Μαρία και Νάντια.

Τέλος, θα είμαι πάντα ευγνώμων στους γονείς μου, Γιώργο και Μαίρη, για την συνεχή, καθόλη τη διάρκεια της ζωής μου, ανιδιοτελή υποστήριξη και αγάπη τους.

Ελένη Τσαλαπάτη
Αθήνα, Μάρτιος 2016

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια το πεδίο απάντησης συζευκτικών ερωτημάτων σε μεγάλα σύνολα δεδομένων έχει καταστεί αντικείμενο συνεχούς έρευνας. Μια από τις πιο διαδεδομένες προσεγγίσεις στο πρόβλημα αυτό βασίζεται στην τεχνική της επαναγραφής ερωτημάτων. Το πρόβλημα, γενικά, ορίζεται ως εξής: Δεδομένου ενός συζευκτικού ερωτήματος και μίας οντολογίας, μία διαδικασία επαναγραφής του ερωτήματος παράγει ένα σύνολο κανόνων στο οποίο ενσωματώνει τους περιορισμούς της οντολογίας, με τέτοιο τρόπο ώστε για οποιοδήποτε σύνολο δεδομένων, η αποτίμηση του τιθέμενου ερωτήματος πάνω στην οντολογία και το σύνολο δεδομένων να επιστρέφει τις ίδιες απαντήσεις με την αποτίμηση μόνο της επαναγραφής στο ίδιο σύνολο δεδομένων.

Τα υπάρχοντα συστήματα επαναγραφής ερωτημάτων δέχονται στην είσοδο τους ένα συζευκτικό ερώτημα και μία οντολογία και υπολογίζουν μία επαναγραφή του ερωτήματος με βάση την οντολογία. Ωστόσο, τα συστήματα αυτά είναι έτσι σχεδιασμένα ώστε κάθε φορά που η οντολογία τροποποιείται -δηλαδή, επεκτείνεται ή μειώνεται κατά ένα σύνολο αξιωμάτων- να υπολογίζουν τη νέα επαναγραφή από την αρχή, χωρίς να αξιοποιούν την πληροφορία που έχει παραχθεί από τις προηγούμενες επαναγραφές. Οι οντολογίες όμως που χρησιμοποιούνται για να μοντελοποιήσουν την επιστημονική γνώση σε πραγματικά πεδία συνεχώς τροποποιούνται και συνεπώς τα υπάρχοντα συστήματα, που επαναϋπολογίζουν εξαρχής την επαναγραφή, θα καθυστερούν σημαντικά.

Στο πλαίσιο αυτό, στην παρούσα διατριβή αρχικά μελετάμε το πρόβλημα υπολογισμού μίας επαναγραφής ενός ερωτήματος με βάση μία οντολογία που έχει εξελιχθεί, αξιοποιώντας την πληροφορία που έχει παραχθεί από τον υπολογισμό μίας επαναγραφής για μία προηγούμενη έκδοση της οντολογίας. Αρχικά, το πρόβλημα μελετάται για την περίπτωση που η οντολογία *επεκτείνεται* κατά ένα σύνολο αξιωμάτων. Η προσέγγιση που ακολουθείται εστιάζει μόνο στους συμπερασμούς που πρέπει πιθανά να εφαρμοστούν εξαιτίας της προσθήκης των νέων αξιωμάτων. Στη συνέχεια, μελετάται η περίπτωση που η οντολογία *συστέλλεται* κατά ένα σύνολο αξιωμάτων. Στην αρχή, παρουσιάζουμε έναν γενικό αλγόριθμο ο οποίος, αφαιρεί με αυτόματο τρόπο τις προτάσεις που δεν παράγονται πλέον από τη νέα οντολογία και το ερώτημα και στη συνέχεια εφαρμόζει τους επιπλέον συμπερασμούς που είναι πιθανά απαραίτητοι. Επιπλέον, επιθυμώντας να ελαχιστοποιήσουμε τη συλλογιστική διαδικασία, μελετάμε αν και υπό ποιες συνθήκες είναι εφικτός ο υπολογισμός μίας νέας επαναγραφής χωρίς την εφαρμογή νέων συμπερασμών. Επίσης, βελτιστοποιούμε τους προηγούμενους αλγορίθμους εφαρμόζοντας τεχνικές που στηρίζονται σε

αναπαράσταση με τη χρήση γράφων. Για κάθε μία από τις περιπτώσεις προτείνουμε έναν νέο αλγόριθμο τον οποίο παρουσιάζουμε αναλυτικά και αποδεικνύουμε την ορθότητα του. Τέλος, αξιολογούμε πειραματικά τους προτεινόμενους αλγορίθμους και τους συγκρίνουμε με τα συστήματα Requiem και Rapid, που αποτελούν τεχνολογία αιχμής στην περιοχή της επαναγραφής με αλγόριθμους ανάλυσης. Τα αποτελέσματα της αξιολόγησης αυτής είναι ιδιαίτερα ενθαρρυντικά.

Στη συνέχεια, στο πλαίσιο της διατριβής, ασχολούμαστε με ένα από τα κυριότερα προβλήματα που εμφανίζονται κατά την συνεχή τροποποίηση των οντολογιών, δηλαδή μία πιθανή ασυνέπεια που μπορεί να εμφανιστεί στη βάση γνώσης. Συγκεκριμένα, ιδιαίτερα σε περιπτώσεις που η βάση γνώσης ανανεώνεται συνεχώς από διαφορετικούς παρόχους είναι πιθανό τα δεδομένα να είναι ασυνεπή σε σχέση με τα αξιώματα της οντολογίας. Για την επίλυση του προβλήματος αυτού προτείνονται δύο βασικές προσεγγίσεις. Η πρώτη στοχεύει στην επιδιόρθωση του συνόλου δεδομένων ώστε η βάση γνώσης να γίνει συνεπής. Η δεύτερη δεν προτείνει την τροποποίηση της βάσης γνώσης, αλλά νέους αλγόριθμους για τον υπολογισμό απαντήσεων σε περιβάλλον ασυνέπειας.

Στην παρούσα διατριβή προτείνουμε ένα πλαίσιο απάντησης ερωτημάτων που βασίζεται σε συστήματα κορεσμού δεδομένων υπό τις σημασιολογίες Τομή Διορθωμένων ABox (Intersection ABox Repair-IAR) και Τομή Διορθωμένων Κλεισμένων ABox (Intersection Closed ABox Repair- ICAR). Ένα σημαντικό πλεονέκτημα των συστημάτων αυτών είναι ότι μπορούν να διαχειριστούν με αποδοτικό τρόπο πολύ μεγάλο όγκο δεδομένων. Συγκεκριμένα, αρχικά, ακολουθώντας τη δεύτερη προσέγγιση, προτείνουμε έναν αλγόριθμο υπολογισμού των ICAR απαντήσεων. Ταυτόχρονα, αξιοποιώντας τις ιδιότητες των συστημάτων κορεσμού δεδομένων αυξάνουμε την αποδοτικότητα του προτεινόμενου αλγορίθμου. Επίσης, εισάγουμε μία νέα σημασιολογία, βασισμένη στην σημασιολογία ICAR, κατά την οποία η απάντηση ερωτημάτων ακόμα και για πιο εκφραστικές Περιγραφικές Λογικές υπολογίζεται σε πολυωνυμικό χρόνο. Προτείνουμε, επίσης, έναν αλγόριθμο υπολογισμού των απαντήσεων υπό την σημασιολογία αυτή αποδεικνύοντας την ορθότητά του. Επιπλέον, ακολουθώντας την πρώτη προσέγγιση, παρουσιάζουμε έναν αποδοτικό αλγόριθμο υπολογισμού των IAR απαντήσεων για DL-Lite$_{\mathcal{R}}$ και $\mathcal{EL}_{\bot nr}$ οντολογίες. Τέλος, παρουσιάζουμε τα πειραματικά αποτελέσματα των συστημάτων που αφορούν στον υπολογισμό απαντήσεων σε ασυνεπείς βάσεις γνώσης για βατές Περιγραφικές Λογικές. Συγκρίνοντας τα χρονικά αποτελέσματά μας με τα χρονικά αποτελέσματα των αντίστοιχων υπαρχόντων συστημάτων διαπιστώνουμε ότι τα συστήματά μας είναι ιδιαίτερα αποδοτικά.

ABSTRACT

The last years conjunctive query answering constitutes a key reasoning service for many applications which involve managing very large datasets. One of the most important reasoning techniques for query answering is query rewriting. Given a conjunctive query (CQ), a process of query rewriting produces a set of rules that captures all the information of the ontology is such a way that for every dataset, the set of answers returned from the ontology over the dataset is the same with the set of answers returned from the rewriting over the same dataset.

The existing rewriting systems accept as input a conjunctive query and an ontology and compute a rewriting of the query with respect to this ontology. However, a drawback of these techniques is that every time the initial ontology is modified –that is, new axioms are added (ontology revision) or existing ones removed (ontology contraction), they compute a new rewriting from scratch without exploiting the similarities between the different versions of the ontology. This may introduce considerable efficiency problems, as many real world applications involve frequent and relatively small modifications on the used ontologies.

In this thesis, we study the problem of computing a rewriting for a CQ over an ontology that has been modified, by reusing the information obtained by the extraction of some previous rewriting. Initially, we study the problem when the ontology is *extended* by a set of axioms. Our approach is to focus only on the new inferences introduced by the new set of axioms. Next, we study the problem when the ontology is *contracted* by as set of axioms. Initially, we provide a general algorithm which initially removes automatically the information that is no longer derivable from the new ontology and the query and then it performs some necessary new inferences. With the aim of minimizing any reasoning tasks, we investigate how a new rewriting can be produced under the constraint of not applying any inferences. Also, we provide graph-based approaches for both algorithms that optimize their performance. Finally, we evaluate experimentally the suggested algorithms and we compare their efficiency to the highly efficient systems Requiem and Rapid. The demonstrated results are very encouraging.

An issue that arises from the continuous ontology modifications, is that the knowledge base is subject to inconsistencies. Particularly, in cases where the knowledge base is being updated frequently from multiple sources it is very likely that the data will be inconsistent to the axioms of the ontology. There are two main approaches that solve this problem. The

straightforward approach is to try and resolve the inconsistencies by "cleaning" the dataset from the conflicting elements. The second approach, called consistent query answering, is to try and compute some "meaningful" answers despite the conflicting information.

In this thesis, we present a framework for efficient query answering, under the Intersection ABox Repair (IAR) and Intersection Closed ABox Repair (ICAR) semantics, that is based on highly efficient mature data saturation (triple-store) systems. This is particularly interest-ing as these systems have shown to be able to handle billions of (ontology) data. Moreover, their properties enable us to propose additional refinements and optimisations for the computation of the ICAR answers. At the same time, we suggest a new type of ICAR-like semantics which we show that can be computed in polynomial time for a very large number of highly expressive DLs, which makes them the first ever such semantics. Subsequently, we show that our framework can also be used to compute answers according to the IAR semantics for ontologies expressed in the DLs DL-Lite$_{\mathcal{R}}$ and $\mathcal{EL}_{\bot nr}$ after some data pre-processing for which task we give an optimised algorithm. Finally, we have conducted an experimental evaluation of the algorithms obtaining encouraging results as both our approaches (IAR and ICAR) are more efficient than existing IAR-answering systems.

Μέρος Ι

Θεμέλια

Κεφάλαιο 1

Εισαγωγή

Η έκρηξη της Επιστήμης των Υπολογιστών στον 20ο αιώνα, οδήγησε στην ανάγκη για εντατική μελέτη φορμαλιστικών τρόπων αναπαράστασης της ανθρώπινης γνώσης που να μπορεί να αξιοποιηθεί με αποδοτικό τρόπο από ευφυή συστήματα. Με τον όρο "ευφυή συστήματα" αναφερόμαστε στα συστήματα που έχουν την δυνατότητα να συμπεράνουν νέα γνώση από την ήδη καταχωρημένη γνώση. Ο στόχος είναι να εφαρμόζουν συλλογιστικές διαδικασίες των οποίων η ποιότητα να είναι ανάλογη της ανθρώπινης συλλογιστικής σκέψης. Για να επιτευχθεί αυτό, χρειάστηκε να οριστούν οι *γλώσσες αναπαράστασης γνώσης* (*knowledge representation languages*).

Μια πρώτη προσπάθεια προς αυτή τη κατεύθυνση ήταν ο ορισμός των Σημασιολογικών Δικτύων [108], με τα οποία η γνώση αναπαριστώνταν χρησιμοποιώντας ένα γραφικό μοντέλο. Αν και η προσέγγιση αυτή ήταν εύκολα αντιληπτή από την ανθρώπινη σκέψη, ήταν δύσκολο να γίνει επίσης κατανοητή από ένα υπολογιστικό σύστημα. Έτσι, έγινε σύντομα φανερό ότι έπρεπε να καθοριστεί μία τυπική σημασιολογία. Ενώ η χρήση της λογικής πρώτης τάξης φαινόταν να είναι μία προφανής λύση, διαπιστώθηκε ότι δεν χρειαζόντουσαν όλοι οι μηχανισμοί που παρέχει, αλλά αρκούσε ένα υποσύνολό της [24]. Μια δυνατότητα που πρόσφερε η προσέγγιση αυτή ηταν ότι ανάλογα με τα χαρακτηριστικά της γλώσσας αναπαράστασης, θα μπορούσαν να οριστούν διαφορετικά υποσύνολα της λογικής-πρώτης τάξης. Με τον τρόπο αυτό, θα μπορούσαν να εφαρμοστούν εξειδικευμένοι συλλογιστικοί μηχανισμοί, χωρίς να είναι απαραίτητη η χρήση κάποιου αυτοματοποιημένου συστήματος απόδειξης θεωρημάτων (theorem prover) της λογικής πρώτης τάξης, αποκτώντας έτσι υπολογιστικά προβλήματα των οποίων η πολυπλοκότητα εξαρτάται από την εκφραστικότητα της συγκεκριμένης γλώσσας.

Στο πλαίσιο αυτό, οι Περιγραφικές Λογικές (ΠΛ) ως *εκφραστικά* και ταυτόχρονα *εύρωστα αποφασίσιμα* (*robustly decidable*) υποσύνολα της λογικής πρώτης τάξης [23], αποτέλεσαν έναν από τους κύριους φορμαλισμούς αναπαράστασης της γνώσης. Η εκφραστική δύναμη των λογικών αυτών οδήγησε στην αξιοποίησή τους σε εφαρμογές όπως είναι η μηχανική οντολογιών (ontological engineering)· μάλιστα αποτελούν το λογικό υπόβαθρο για την γλώσσα ανάπτυξης οντολογιών OWL (Web Ontology Language) [13, 99] από το World Wide Web Consortium (W3C). Επίσης, οι ΠΛ χρησιμοποιούνται για συλλογιστικές πράξεις πανω σε σχήματα βάσεων δεδομένων και παρέχουν συλλογιστική υποστήριξη του Σημασιολογικού Ιστού [15, 123]. Ο Σημασιολογικός Ιστός αποτελεί μια νέα μορφή του Παγκόσμιου Ιστού (World Wide Web - WWW), στην οποία οι πηγές του Ιστου επισημειώνονται με όρους που ορίζονται σε μία οντολογία, καθιστώντας έτσι την πληροφορία επεξεργάσιμη από υπολογιστικά προγράμματα (πράκτορες).

Ταυτόχρονα, τα συστήματα συλλογιστικής σε Περιγραφικές Λογικές είναι ιδιαίτερα αποδοτικά σε εφαρμογές αυτόματου συμπερασμού, ακόμα σε περιπτώσεις που διαχειρίζονται μεγάλο όγκο πληροφορίας [44, 126, 55, 90]. Συνεπώς, αποτελούν τον ακρογωνιαίο λίθο σε διάφορες εφαρμογές, όπως σε πληροφοριακά συστήματα [39, 56], στη ρύθμιση παραμέτρων σε τηλεπικοινωνιακά συστήματα [14, 85], στην επεξεργασία πολυμεσικών κειμένων [12, 86, 125, 21, 105] καθώς και σε άλλους τομείς όπως στη βιολογία [124], στη γεωλογία [109], στην ιατρική [47, 110] και στη γεωγραφία [50].

1.1 Περιγραφικές Λογικές

Οι Περιγραφικες Λογικές είναι γλώσσες αναπαράστασης της γνώσης με ένα τυπικά ορισμένο συντακτικό και μία τυπικά ορισμένη σημασιολογία. Τα βασικά δομικά στοιχεία των Περιγραφικών Λογικών (ΠΛ) είναι τα *άτομα* (*individuals*), οι *έννοιες* (*concept*) που αναπαριστούν σύνολα από άτομα και οι *ρόλοι* (*roles*) που αναπαριστούν δυαδικές σχέσεις ανάμεσα στα άτομα. Κατασκευαστές, όπως είναι η τομή ($\sqcap$) ή η άρνηση ($\neg$), χρησιμοποιούνται για την κατασκευή περίπλοκων εννοιών ή ρόλων. Για παράδειγμα, μπορούμε να ορίσουμε την έννοια των προπτυχιακών φοιτητών ενός πανεπιστημίου ως το σύνολο των φοιτητών που δεν είναι μεταπτυχιακοί φοιτητές, δηλαδή με την έννοια $\mathsf{Φοιτητής} \sqcap \neg\mathsf{ΜεταπτυχιακόςΦοιτητής}$.

Μια ΠΛ *ερμηνεία* (*interpretation*), $\mathcal{I}$, ορίζεται από ένα μη-κενό σύνολο που ονομάζεται *χώρος ερμηνείας* (*domain of interpretation*) και περιέχει στοιχεία που ονομάζονται *αντικείμενα* (*objects*), και μία *συνάρτηση ερμηνείας* (*interpretation function*) που ερμηνεύει κάθε ατομική έννοια ως ένα υποσύνολο του χώρου ερμηνείας και κάθε ρόλο ως ένα υποσύνολο του καρτεσιανού γινομένου του χώρου ερμηνείας.

Οι ΠΛ δίνουν επιπλέον τη δυνατότητα της περιγραφής σχέσεων ανάμεσα σε απλές ή περίπλοκες έννοιες ή και ρόλους. Οι σχέσεις αυτές παρουσιάζονται με τη μορφή αξιωμάτων που ονομάζονται *αξιώματα ορολογίας* (*terminology axioms*). Τα αξιώματα ορολογίας μπορεί να είναι *αξιώματα υπαγωγής* (*subsumption axioms*) ή *ισοδυναμίας* (*equivalence axioms*). Ενα αξίωμα υπαγωγής έχει τη μορφή $C \sqsubseteq D$ και δηλώνει ότι η έννοια D είναι πιο γενική από την έννοια C, ενώ ένα αξίωμα ισοδυναμίας έχει τη μορφή $C \equiv D$ και δηλώνει ότι οι έννοιες C, D είναι ταυτόσημες. Λέμε ότι μία ερμηνεία ικανοποιεί ένα αξίωμα υπαγωγής $C \sqsubseteq D$ αν ερμηνεύει την έννοια D ως υπερσύνολο της έννοιας C. Το σύνολο των αξιωμάτων με βάση τα οποία ορίζονται νέες έννοιες και ρόλοι ορίζουν την *ορολογία* ή αλλιώς TBox (Terminological Box) μιας Περιγραφικής Λογικής.

Το σύνολο των ισχυρισμών ορίζουν το *σώμα ισχυρισμών* (*assertional component*) ή αλλιώς ABox (Assertion Box) μιας ΠΛ και συμβολίζεται με $\mathcal{A}$. Το σύνολο των αξιωμάτων που ορίζονται στο TBox και το σύνολο των ισχυρισμών που ορίζονται στο ABox αποτελούν μια *Βάση Γνώσης (ΒΓ)* (*Knowledge Base – KB*) $\langle \mathcal{T}, \mathcal{A} \rangle$. Μία ερμηνεία είναι *μοντέλο* (*model*) μίας ΒΓ $\langle \mathcal{T}, \mathcal{A} \rangle$, αν ικανοποιεί όλα τα αξιώματα που υπάρχουν στο $\mathcal{T}$ και όλους τους ισχυρισμούς που υπάρχουν στο $\mathcal{A}$. Μία ΒΓ είναι *ασυνεπής* (*inconsistent*) αν δεν έχει μοντέλο.

Πιο αναλυτική περιγραφή των ΠΛ παρουσιάζεται στο Κεφάλαιο 2.

1.2 Απάντηση Ερωτημάτων

Ένα από τα βασικότερα προβλήματα συλλογιστικής στις ΠΛ είναι το πρόβλημα απάντησης ερωτημάτων. Η υποβολή σύνθετων ερωτήσεων μπορεί να επιτευχθεί με τα *Συζευκτικά Ερωτήματα (ΣΕ)* (*Conjunctive Queries – CQ*) [1]. Ένα συζευκτικό ερώτημα αποτελείται από μία σύζευξη ατόμων και έχει τη μορφή:

$$\mathcal{Q}(\vec{x}) \leftarrow A_1(\vec{x}_1) \wedge \ldots \wedge A_n(\vec{x}_n)$$

όπου τα A_i είναι ατομικές έννοιες ή ατομικοί ρόλοι της ΒΓ. Οι μεταβλητές $\vec{x}$ ονομάζονται διακεκριμένες μεταβλητές. Αν όλες οι μεταβλητές σε ένα ερώτημα είναι διακεκριμένες, τότε η απάντηση στο ερώτημα αποτελείται από τη πλειάδα των ατόμων από τη ΒΓ, για τις οποίες μετά την αντικατάσταση των μεταβλητών με άτομα από την απάντηση του ερωτήματος, όλες οι έννοιες και ρόλοι του ερωτήματος είναι αληθείς για κάθε μοντέλο της ΒΓ.

Για παράδειγμα, έστω η ΒΓ που αποτελείται από το TBox:

$$\mathsf{ΜεταπτυχιακόςΦοιτητής} \sqsubseteq \mathsf{Φοιτητής} \tag{1.1}$$
$$\mathsf{παρακολουθείΜεταπτυχιακόΜάθημα} \sqsubseteq \mathsf{παρακολουθείΜάθημα} \tag{1.2}$$

που δηλώνει ότι καθε μεταπτυχιακός φοιτητής είναι φοιτητής και ότι αν κάποιος παρακολουθεί ένα μεταπτυχιακό μάθημα τότε παρακολουθεί αυτό το μάθημα. Έστω επίσης το ABox:

$$\mathsf{ΜεταπτυχιακόςΦοιτητής}(\mathsf{Γιάννης}), \tag{1.3}$$
$$\mathsf{παρακολουθείΜεταπτυχιακόΜάθημα}(\mathsf{Γιάννης}, \mathsf{Λογική}) \tag{1.4}$$

και έστω το ερώτημα:

$$\mathcal{Q}(x, y) \leftarrow \mathsf{Φοιτητής}(x) \wedge \mathsf{παρακολουθείΜάθημα}(x, y)$$

Τότε το ζεύγος $(\mathsf{Γιάννης}, \mathsf{Λογική})$ είναι μία απάντηση, αφού ο ισχυρισμός $\mathsf{Φοιτητής}(\mathsf{Γιάννης})$ προκύπτει από τα (1.3), (1.1) και το $\mathsf{παρακολουθείΜάθημα}(\mathsf{Γιάννης}, \mathsf{Λογική})$ από τα (1.4), (1.2).

Αν το ερώτημα περιέχει επίσης μεταβλητές που δεν είναι διακεκριμένες, τότε οι μεταβλητές αυτές αντιμετωπίζονται σαν να είναι υπαρξιακά ποσοτικοποιημένες, δηλαδή αρκεί η ύπαρξη ενός κατάλληλου στοιχείου στο μοντέλο, το οποίο δεν χρειάζεται να αντιστοιχεί σε κάποιο συγκεκριμένο άτομο του ABox. Για παράδειγμα, έστω μία ΒΓ με TBox:

$$\mathsf{ΠροπτυχιακόςΦοιτητής} \sqcup \mathsf{ΜεταπτυχιακόςΦοιτητής} \sqsubseteq \exists \mathsf{παρακολουθεί}.\mathsf{Μάθημα}$$

και έστω ότι το ABox περιέχει μόνο τον ισχυρισμό $\mathsf{ΜεταπτυχιακόςΦοιτητής}(\mathsf{Γιάννης})$. Τότε, στο ερώτημα

$$\mathcal{Q}(x) \leftarrow \mathsf{παρακολουθεί}(x, y)$$

το άτομο Γιάννης αποτελεί μία ορθή απάντηση. Αυτό ισχύει επειδή κάθε άτομο της έννοιας ΜεταπτυχιακόςΦοιτητής, και συνεπώς και το άτομο Γιάννης, παρακολουθεί κάποιο μάθημα που είναι άτομο της έννοιας Μάθημα.

Είναι σημαντικό να σημειωθεί ότι θεωρούμε ότι αν κάποιος ισχυρισμός δεν βρίσκεται μέσα στο ABox αυτό δεν σημαίνει ότι ο ισχυρισμός δεν ισχύει, η υπόθεση αυτή ονομάζεται *υπόθεση ανοικτού κόσμου* (*open world assumption -OWA*) και αποτελεί μία από τις βασικές διαφορές με τα συστήματα Βάσεων Δεδομένων όπου ισχύει η *υπόθεση κλειστού κόσμου* (*closed world assumption*). Με την υπόθεση κλειστού κόσμου, αν κάτι δεν είναι ρητά δηλωμένο, θεωρείται ότι δεν ισχύει.

Επίσης, στις ΒΔ ισχύει η *υπόθεση μοναδικών ονομάτων* (*Unique Name Assumption-UNA*), με την οποία διαφορετικά ονόματα ατόμων αντιστοιχίζονται σε διακριτά στοιχεία ενός μοντέλου. Καθώς στα περισσότερα συστήματα ΠΛ δεν υιοθετείται η UNA, διαφορετικά ονόματα ατόμων μπορεί να αντιστοιχηθούν στο ίδιο στοιχείο ενός μοντέλου. Από την OWA και την έλλειψη της UNA, προκύπτει ότι συνήθως μια ΒΓ της ΠΛ έχει περισσότερα από ένα μοντέλα, σε αντίθεση με μία ΒΔ που πάντα έχει ένα μοναδικό μοντέλο. Αυτό έχει καθιστά το πρόβλημα απάντησης ερωτημάτων σε μία ΒΓ υπολογιστικά αρκετά πιο δύσκολο από το πρόβλημα απάντησης ερωτημάτων σε ΒΔ, εφόσον στη ΒΔ αρκεί να ελεγχθεί ένα μοντέλο ενώ στη ΒΓ πρέπει να ελεγχθούν όλα τα μοντέλα.

1.2.1 Απάντηση Ερωτημάτων μέσω Επαναγραφής Ερωτημάτων

Έχει αποδειχθεί ότι το πρόβλημα απάντησης συζευκτικών ερωτημάτων για γλώσσες υψηλής εκφραστικότητας είναι στην χειρότερη περίπτωση coNP-complete [80] *ως προς τα δεδομένα* (*data complexity*)–σε σχέση δηλαδή με το μέγεθος του ABox– είναι, δηλαδή, απαγορευτική για πρακτικές εφαρμογές. Για τον λόγο αυτό αναπτύχθηκαν γλώσσες Περιγραφικών Λογικών που είναι *βατές* (*tractable*), όπως οι οικογένειες γλωσσών DL-Lite [29], Description Logic Programs (DLP) [54] και $\mathcal{EL}$ [6], οι οποίες αποτελούν το υπόβαθρο των προφίλ OWL 2 QL, OWL 2 RL και OWL 2 EL της OWL 2 [73].

Μία από τις πιο διαδεδομένες τεχνικές απάντησης ερωτημάτων σε αυτές τις γλώσσες είναι η *επαναγραφή ερωτήματος* (*query rewriting*). Δεδομένου ενός ερωτήματος $\mathcal{Q}$ και μίας οντολογίας $\mathcal{T}$, μία επαναγραφή $\mathcal{R}$ του $\mathcal{Q}$ με βάση το $\mathcal{T}$ είναι μία ένωση συζευκτικών ερωτημάτων ή ένα πρόγραμμα datalog το οποίο περιλαμβάνει όλη τη σχετική πληροφορία από το $\mathcal{T}$ για την απάντηση του $\mathcal{Q}$ πάνω σε ένα οποιοδήποτε σύνολο δεδομένων $\mathcal{A}$.

Για παράδειγμα, η επαναγραφή $\mathcal{R}$ του ερωτήματος:

$$\mathcal{Q}(x,y) \leftarrow \textsf{Φοιτητής}(x) \wedge \textsf{παρακολουθεί}(x,y)$$

με βάση το TBox:

$$\mathcal{T} = \{\textsf{ΜεταπτυχιακόςΦοιτητής} \sqsubseteq \textsf{Φοιτητής}, \textsf{παρακολουθείΜετΜάθημα} \sqsubseteq \textsf{παρακολουθεί}\}$$

είναι το σύνολο των ερωτημάτων:

$$\begin{aligned}
\mathcal{Q}(x,y) &\leftarrow \textsf{Φοιτητής}(x) \wedge \textsf{παρακολουθεί}(x,y)\\
\mathcal{Q}(x,y) &\leftarrow \textsf{ΜεταπτυχιακόςΦοιτητής}(x) \wedge \textsf{παρακολουθεί}(x,y)\\
\mathcal{Q}(x,y) &\leftarrow \textsf{Φοιτητής}(x) \wedge \textsf{παρακολουθείΜετΜάθημα}(x,y)\\
\mathcal{Q}(x,y) &\leftarrow \textsf{ΜεταπτυχιακόςΦοιτητής}(x) \wedge \textsf{παρακολουθείΜετΜάθημα}(x,y)
\end{aligned}$$

Το σύνολο των ερωτημάτων αυτών προκύπτει από την διαδοχική αντικατάσταση του ατόμου $\textsf{Φοιτητής}(x)$ με το άτομο $\textsf{ΜεταπτυχιακόςΦοιτητής}(x)$ και του ατόμου $\textsf{παρακολουθεί}(x,y)$ με το άτομο $\textsf{παρακολουθείΜετΜάθημα}(x,y)$.

Με τον τρόπο αυτό, το $\mathcal{R}$ μπορεί να μεταφραστεί σε ένα SQL ερώτημα το οποίο να εκτελεστεί σε μία επαγωγική βάση δεδομένων. Έτσι, το πρόβλημα απάντησης ερωτήματος σε μία ΒΓ ανάγεται στο πρόβλημα απάντησης ενός ερωτήματος σε μία ΒΔ. Συνεπώς, δεδομένου ότι η πολυπλοκότητα για την εκτέλεση ερωτημάτων πρώτης-τάξης σε μία ΒΔ ανήκει στην κλάση πολυπλοκότητας AC_0 [96, 135] και ενός datalog προγράμματος είναι PTIME-πλήρης [103] ως προς τα δεδομένα, συμπεραίνουμε ότι το πρόβλημα απάντησης ερωτήματος για βατές ΠΛ είναι αρκετά χαμηλής πολυπλοκότητας.

1.2.2 Απάντηση Ερωτημάτων σε Ασυνεπείς ΒΓ

Με τη κλασσική σημασιολογία της λογικής πρώτης τάξης το πρόβλημα απάντησης ερωτημάτων σε μία ασυνεπή ΒΓ δεν έχει νόημα, αφού από την ασυνέπεια εξάγεται οποιαδήποτε πρόταση. Ωστόσο, στη πραγματικότητα η ασυνέπεια σε μία ΒΓ είναι ένα πολύ συχνό φαινόμενο, ειδικά σε περιπτώσεις που οι ΒΓ είναι διαχειρίσιμες από πολλούς διαφορετικούς φορείς ή είναι αποτέλεσμα ενοποίησης μικρότερων ΒΓ. Για να επιλυθεί το πρόβλημα αυτό, υπάρχουν δύο βασικές προσεγγίσεις. Με την πρώτη, και ίσως πιο προφανή, προσέγγιση εντοπίζονται οι ισχυρισμοί που προκαλούν την ασυνέπεια και απομακρύνονται. Ωστόσο, η λύση αυτή είναι αρκετά χρονοβόρα, αφού συνήθως ο όγκος των δεδομένων είναι αρκετά μεγάλος. Επιπρόσθετα, δεν είναι πάντα εφικτή, γιατί μπορεί να μην υπάρχει πάντα πρόσβαση στα δεδομένα. Με τη δεύτερη προσέγγιση, η οποία ονομάζεται *συνεπής απάντηση ερωτημάτων* (*consistent query answering*), εξάγονται απαντήσεις με νόημα, παρά την ασυνέπεια της ΒΓ.

Η συνεπής απάντηση ερωτημάτων άρχισε να μελετάται στο πεδίο των ΠΛ [75, 115, 20, 18] αρκετά πρόσφατα. Ένα θεμελιώδες εργαλείο για την ανάκτηση συνεπούς πληροφορίας από μία ασυνεπή ΒΓ είναι η έννοια της *διόρθωσης* (*repair*) [16, 3, 33, 27] που προέρχεται από τον χώρο των ΒΔ. Μία διόρθωση μίας βάσης δεδομένων που είναι ασυνεπής ως προς τους περιορισμούς ακεραιότητας (integrity constraints) είναι μία βάση δεδομένων που υπολογίζεται εφαρμόζοντας ένα "ελάχιστο" σύνολο αλλαγών στην αρχική βάση δεδομένων ώστε να εξασφαλιστεί η συνέπεια. Υπάρχουν πολλές ερμηνείες της έννοιας του "ελαχίστου" οι οποίες οδηγούν σε διαφορετικές σημασιολογίες. Υπο τις περισσότερες ερμηνείες της έννοιας του "ελαχίστου" υπάρχουν πολλές διορθώσεις μίας βάσης δεδομένων και επομένως το πρόβλημα της συνεπούς απάντησης ερωτημάτων ανάγεται στο πρόβλημα εύρεσης όλων των πλειάδων που αποτελούν απάντηση στο ερώτημα σε όλες τις διορθώσεις.

Αντίστοιχα, στις ΒΓ έχει οριστεί η έννοια της *ABox διόρθωσης* (*ABox Repair -AR*) [75] και της *σημασιολογίας της ABox διόρθωσης* (*AR semantics*). Για μία ΒΓ που αποτελείται από ένα TBox $\mathcal{T}$ και ένα ABox $\mathcal{A}$, η ABox διόρθωση είναι ένα μέγιστο υποσύνολο του $\mathcal{A}$ που είναι συνεπές με το $\mathcal{T}$. Δυστυχώς όμως, έχει δειχθεί ότι ακόμα για οντολογίες που ανήκουν στην οικογένεια της DL-Lite, η συνεπής απάντηση ερωτημάτων υπό την AR σημασιολογία είναι coNP-complete σε σχέση με τα δεδομένα και συνεπώς πρακτικά μη-υπολογίσιμη [75]. Για να επιλυθεί το πρόβλημα αυτό, οι Lembo et al. [75] εισήγαγαν την σημασιολογία της *Τομής των ABox Διορθώσεων* (*Intersection ABox Repair-IAR*), σύμφωνα με την οποία αντί να λαμβάνονται υπόψη όλες οι δυνατές διορθώσεις του ABox, λαμβάνεται υπόψη μόνο η τομή των διορθώσεων αυτών για την απάντηση του τιθέμενου ερωτήματος. Προφανώς, αφού κάθε διόρθωση είναι συνεπής ως προς την οντολογία, και η τομή τους θα είναι συνεπής ως προς την οντολογία. Για την DL-Lite, η πολυπλοκότητα του προβλήματος της συνεπούς απάντησης ερωτημάτων υπό την IAR σημασιολογία ανήκει στην κλάση AC^0 σε σχέση με τα δεδομένα, και συνεπώς επιλύεται εύκολα.

Ένα αρνητικό χαρακτηριστικό που έχει η IAR σημασιολογία είναι ότι δεν λαμβάνει υπόψη τους ισχυρισμούς που μπορεί να συμπεραχθούν από την ΒΓ αλλά μόνο τους ισχυρισμούς που εμφανίζονται σε αυτή. Για να ξεπεραστεί το πρόβλημα αυτό οι οι Lembo et al. [75] εισήγαγαν επίσης την σημασιολογία της *Τομής Διορθώσεων του Κλειστού ABox* (*Intersection Closed ABox Repair (ICAR)*), σύμφωνα με την οποία λαμβάνονται υπόψη και οι ισχυρισμοί που συμπεραίνονται από τη ΒΓ. Επομένως, η διόρθωση στην ICAR σημασιολογία αποτελείται από τους ισχυρισμούς που ανήκουν στη τομή των διορθώσεων του συνόλου των ισχυρισμών που συμπεραίνονται από τη ΒΓ. Όπως και στην περίπτωση της IAR σημασιολογίας, η πολυπλοκότητα του προβλήματος της συνεπούς απάντησης ερωτημάτων υπό την ICAR σημασιολογία ανήκει στην κλάση AC^0 σε σχέση με τα δεδομένα.

1.3 Παρουσίαση Προβλήματος

Στις περισσότερες πρακτικές εφαρμογές οι ΒΓ δεν είναι σταθερές αλλά τροποποιούνται συχνά, δηλαδή, επεκτείνονται ή συστέλλονται κατά ένα σύνολο αξιωμάτων. Χαρακτηριστικά παραδείγματα εμφανίζονται σε ιατρικές εφαρμογές που διαχείριζονται μεγάλες οντολογίες, όπως είναι ο NCI Thesaurus (NCI) [1], το Foundational Model of Anatomy (FMA) [2] και η SNOMED [3]. Στην NCI Thesaurus, για παράδειγμα, που έχει πρόσβαση ένα μεγάλο πλήθος διαφορετικών φορέων, προστίθενται περίπου 900 αξιώματα τον μήνα, ενώ ταυτόχρονα υπήρχαν περιπτώσεις που μεταξύ δύο διαδοχικών εκδόσεων της οντολογίας αφαιρέθηκαν 10.000 αξιώματα [49].

Αν υποθέσουμε ότι κάθε νέα ΒΓ που προκύπτει από τις τροποποιήσεις είναι συνεπής, τότε για την απάντηση ενός ερωτήματος θεωρητικά αρκεί να υπολογιστεί μία νέα επαναγραφή κάθε φορά που αλλάζει η οντολογία, η οποία θα εκτελεστεί στη βάση δεδομένων. Τα τελευταία χρόνια έχει αναπτυχθεί πλήθος αλγορίθμων και συστημάτων που αντιμετωπίζουν το πρόβλημα υπολογισμού επαναγραφής ενός ερωτήματος. Μερικά από τα πιο γνωστά εργαλεία που βασίζονται στους αλγορίθμους αυτούς είναι το Presto [116], το Requiem [103], το Mastro [28], το Quest [113], Clipper [41], το Blackout [104], το PURE [72], το Ontop [114], το kyrie [89], το IQAROS [137] και το Rapid [133].

Παρά τα θετικά αποτελέσματα των συστημάτων αυτών, το πρόβλημα υπολογισμού μίας επαναγραφής, ειδικά σε περιπτώσεις που η οντολογία είναι αρκετά υψηλής εκφραστικότητας ή πολύ μεγάλη, παραμένει αρκετά δύσκολο αφού ακόμα και με τα πιο εξελιγμένα συστήματα μπορεί ο υπολογισμός της επαναγραφής να απαιτεί έως και κάποιες ώρες. Συνεπώς, σε εφαρμογές [37, 22, 111] που η οντολογία τροποποιείται σχετικά συχνά δεν είναι ιδιαίτερα πρακτικό να αξιοποιήσει κανείς αυτά τα συστήματα αφού θα υπολογίζουν κάθε φορά μία νέα επαναγραφή από την αρχή με βάση τη νέα έκδοση της οντολογίας.

Αυτή ακριβώς η δυναμική φύση των οντολογιών επιτάσσει την ανάγκη για περαιτέρω έρευνα στο πεδίο των οντολογιών που τροποποιούνται, τόσο σε θεωρητικό όσο και σε πρακτικό επίπεδο [94, 57, 106, 45, 32, 138, 64, 48, 36]. Στην Τεχνητή Νοημοσύνη, η διαδικασία ένταξης νέας πληροφορίας σε μία βάση γνώσης λέγεται "επέκταση" ή "αναθεώρηση" ενώ η αφαίρεση της ονομάζεται "συστολή". Στην διατριβή αυτή και οι δύο έννοιες περικλείονται στην έννοια "εξέλιξη".

Για παράδειγμα, έστω το ερώτημα

$$\mathcal{Q} = Q(x, y) \leftarrow \text{Φοιτητής}(x) \wedge \text{παρακολουθεί}(x, y)$$

και το TBox $\mathcal{T}$:

$$\mathcal{T} = \{\text{ΜεταπτυχιακόςΦοιτητής} \sqsubseteq \text{Φοιτητής}, \text{παρακολουθείΜετΜάθημα} \sqsubseteq \text{παρακολουθεί}\}$$

[1] https://wiki.nci.nih.gov/display/VKC/NCI+Thesaurus+Terminology

[2] http://sig.biostr.washington.edu/projects/fm/AboutFM.html

[3] http://bioportal.bioontology.org/ontologies/1353

τότε, σύμφωνα με την Ενότητα 1.2 η επαναγραφή του $\mathcal{Q}$ μ.β.τ. $\mathcal{T}$ είναι το σύνολο $\mathcal{R}$. Αν υποθέσουμε τώρα ότι το $\mathcal{T}$ αναθεωρείται προσθέτοντας το αξίωμα

$$c = \mathsf{ΥποψήφιοςΔιδάκτορας} \sqsubseteq \mathsf{ΜεταπτυχιακόςΦοιτητής}$$

τότε μία νέα επανεγραφή του $\mathcal{Q}$ μ.β.τ. $\mathcal{T}' = \mathcal{T} \cup \{c\}$ θα είναι το σύνολο:

$$\mathcal{R}' = \mathcal{R} \cup \{\mathcal{Q}(x,y) \leftarrow \mathsf{ΥποψήφιοςΔιδάκτορας}(x) \wedge \mathsf{παρακολουθεί}(x,y),$$
$$\mathcal{Q}(x,y) \leftarrow \mathsf{ΥποψήφιοςΔιδάκτορας}(x) \wedge \mathsf{παρακολουθείΜετΜάθημα}(x,y)\}$$

όπως είναι φανερό, για να υπολογιστεί η νέα επαναγραφή δεν είναι απαραίτητο να την υπολογίσει κάποιος από την αρχή, αρκεί μόνο να εστιάσει στον υπολογισμο των καινούριων ερωτημάτων που προκύπτουν από την εισαγωγή του νέου αξιώματος. Αντίστοιχα, αν θεωρήσουμε ότι αργότερα κάποιος αποφασίσει να αφαιρέσει το αξίωμα c, τότε πάλι δεν χρειάζεται να υπολογιστεί από την αρχή η νέα επαναγραφή, αρκεί μόνο να αφαιρεθεί το ερώτημα (ή τα ερωτήματα) για τα οποία είναι υπεύθυνο το c. Δηλαδή, στο παράδειγμα αυτό, αρκεί να αφαιρεθούν τα ερωτήματα που μόλις προστέθηκαν.

Ένα ακόμα όμως πρόβλημα που προκύπτει από την συχνή τροποποίηση των οντολογιών είναι η ασυνέπεια που μπορεί να εμφανιστεί στη ΒΓ. Παρά την σημαντική θεωρητική μελέτη που έχει εμφανιστεί στη βιβλιογραφία για την συνεπή απάντηση ερωτημάτων [75, 115, 20, 18], αυτή τη στιγμή το πρόβλημα σχεδιασμού αποδοτικών συστημάτων για μεγάλο όγκο δεδομένων παραμένει ανοιχτό. Ειδικά, για την συνεπή απάντηση ερωτημάτων υπό την ICAR σημασιολογία, υπάρχει στη βιβλιογραφία μία πρώτη προσπάθεια από τους Masotti et al. [84] η οποία όμως αξιολογήθηκε σε λίγα δεδομένα και η απόδοση που εμφάνισε δεν ήταν ιδιαίτερα καλή. Επίσης, αν η ασυνέπεια εμφανιστεί σε ΒΓ που αποτελούνται από βατές οντολογίες που όμως δεν ανήκουν στην οικογένεια της DL-Lite, για παράδειγμα σε $\mathcal{EL}$ οντολογίες, τότε το πρόβλημα της συνεπούς απάντησης ερωτημάτων υπό την ICAR σημασιολογία γίνεται δυσεπίλυτο.

Στην παρούσα διατριβή στην αρχή θεωρούμε ότι η ΒΓ που έχει προκύψει μετά την τροποποίηση της οντολογίας είναι συνεπής. Μελετάμε το πρόβλημα της επαναγραφής ερωτημάτων πάνω σε οντολογίες που έχουν είτε αναθεωρηθεί είτε συσταλεί. Πιο συγκεκριμένα, δεδομένου ενός αρχικού ερωτήματος $\mathcal{Q}$, μίας οντολογίας $\mathcal{T}$, η οποία μπορεί να επεκταθεί με ένα σύνολο αξιωμάτων $\mathcal{T}^+$ ή να συσταλεί κατά ένα σύνολο αξιωμάτων $\mathcal{T}^-$ και μιας ήδη υπολογισμένης επαναγραφής $\mathcal{R}$ για τα $\mathcal{Q}$, $\mathcal{T}$, ο στόχος είναι να υπολογιστεί αποδοτικά μία επαναγραφή του $\mathcal{Q}$ μ.β.τ. $\mathcal{T}'$, όπου $\mathcal{T}' = \mathcal{T} \cup \mathcal{T}^+$ ή $\mathcal{T}' = \mathcal{T} \setminus \mathcal{T}^-$. Αυτό μπορεί να επιτευχθεί αξιοποιώντας την πληροφορία που υπάρχει ήδη στο $\mathcal{R}$, χωρίς να επαναληφθούν οι υπολογισμοί που πραγματοποιήθηκαν για την παραγωγή της αρχικής επαναγραφής.

Σε κάθε μία περίπτωση, προσεγγίζουμε το πρόβλημα θεωρητικά και στη συνέχεια σχεδιάζουμε ένα αποδοτικό αλγόριθμο. Στην περίπτωση που η οντολογία επεκτείνεται, τότε ο αλγόριθμος εστιάζει στην υλοποιήση των συλλογισμών που οφείλονται μόνο στην εισαγωγή των αξιωμάτων νέων $\mathcal{O}^+$. Με τον τρόπο αυτό, ο αλγόριθμος μας εκτελεί *μόνο* την πρόσθετη εργασία που είναι απαραίτητη για τον υπολογισμό μιας επαναγραφής για την επεκτεταμένη οντολογία $\mathcal{O} \cup \mathcal{O}^+$. Στην περίπτωση που η οντολογία συστέλλεται κατά ένα σύνολο αξιωμάτων, τότε αρχικά προτείνουμε έναν γενικό αλγόριθμο υπολογισμού μίας νέας επαναγραφής, ο οποίος είναι πιθανό να περιλαμβάνει και την εφαρμογή νέων συμπερασμών. Στη συνέχεια, μελετάμε τις συνθήκες υπό τις οποίες είναι εφικτός ο υπολογισμός μίας νέας επαναγραφής χωρίς την εφαρμογή συμπερασμών. Επίσης, στο ίδιο πλαίσιο, αξιοποιώντας τις ιδιότητες των γράφων, βελτιστοποιούμε τους προηγούμενους αλγορίθμους. Αξίζει να σημειώσουμε ότι απ' όσο γνωρίζουμε αυτή είναι η πρώτη θεωρητική και πρακτική μελέτη του προβλήματος αυτού, τόσο στον χώρο των οντολογιών όσο και στον χώρο της τροποποίησης των οντολογιών.

Στη συνέχεια της διατριβής, θεωρούμε ότι η ΒΓ που έχει προκύψει μετά την τροποποίηση της οντολογίας είναι ασυνεπής. Μελετάμε το πρόβλημα υπολογισμού συνεπών απαντήσεων από τη ΒΓ υπό την IAR και την ICAR σημασιολογία. Αρχικά, ασχολούμαστε με την ICAR σημασιολογία και προτείνουμε ένα πλαίσιο εύρεσης απαντήσεων που βασίζεται στα ιδιαίτερα αποδοτικά συστήματα κορεσμού δεδομένων. Ένα σημαντικό προτέρημα της προσέγγισης αυτής είναι ότι τα συστήματα αυτά μπορούν να διαχειριστούν πολύ μεγάλους όγκους δεδομένων. Επίσης, αξιοποιώντας τις ιδιότητές τους βελτιστοποιούμε περαιτέρω τον αρχικό αλγόριθμο. Οι αλγόριθμοι που προτείνουμε είναι ορθοί και πλήρεις για DL-Lite οντολογίες. Ωστόσο, όπως έχει ήδη αναφερθεί, για οντολογίες διαφορετικής εκφραστικότητας, όπως είναι η $\mathcal{EL}$, το πρόβλημα συνεπούς απάντησης ερωτημάτων είναι δυσεπίλυτο, συνεπώς ο αλγόριθμός μας μπορεί να υπολογίσει μόνο προσεγγιστικά τις ICAR απαντήσεις. Για αυτόν τον λόγο, εισάγουμε μία νέα ICAR-τύπου σημασιολογία και αποδεικνύουμε ότι μπορεί το πρόβλημα να επιλυθεί σε πολυωνυμικό χρόνο για έναν μεγάλο αριθμό εκφραστικών ΠΛ. Τέλος, παρουσιάζουμε ένα ιδιαίτερα αποδοτικό αλγόριθμο για τον υπολογισμό των IAR διορθώσεων για ένα υποσύνολο των βατών ΠΛ.

1.4 Δομή της Διατριβής

Η διατριβή αυτή οργανώνεται ως εξής:

- Στο Κεφάλαιο 2 παρουσιάζουμε το μαθηματικό υπόβαθρο που είναι απαραίτητο για την κατανόηση του υπόλοιπου της διατριβής. Πιο συγκεκριμένα, παρουσιάζουμε μια σύντομη εισαγωγή στη λογική πρώτης τάξης και πιο συγκεκριμένα στους υπαρξιακούς κανόνες, τις προτάσεις και στον λογισμό της ανάλυσης. Ακολουθεί μια σύντομη εισαγωγή στο συντακτικό και τη σημασιολογία των ΠΛ καθώς και στις υπηρεσίες συλλογιστικής που παρέχουν. Τέλος, εισάγουμε το θεωρητικό υπόβαθρο που αφορά στα συζευκτικά ερωτήματα, για τα οποία ορίζουμε το πρόβλημα της απάντησης ερωτημάτων μέσω επαναγραφής για συνεπείς και ασυνεπείς ΒΓ.

- Στο Κεφάλαιο 3 μελετάμε το πρόβλημα υπολογισμού επαναγραφής για κάποιο ερώτημα με βάση μία οντολογία συνεπή με τη βάση γνώσης που έχει είτε επεκταθεί (με ένα σύνολο αξιωμάτων που διατηρεί την συνέπεια της ΒΓ) είτε μειωθεί κατά ένα σύνολο αξιωμάτων, δεδομένης μιας επαναγραφής με βάση την αρχική οντολογία. Παρουσιάζουμε μια σειρά από παραδείγματα που μας καθοδηγούν διαισθητικά στους αλγορίθμους. Επίσης παρουσιάζουμε ένα σύνολο από αναλυτικούς αλγόριθμους με τις αποδείξεις τους.

- Στο Κεφάλαιο 4 παρουσιάζουμε μια αναλυτική αξιολόγηση των προτεινόμενων αλγορίθμων που παρουσιάζονται στο Κεφάλαιο 3, αντιπαραβάλλοντάς τους με τα αντίστοιχα συστήματα στα οποία βασίστηκε η υλοποίησή τους. Η αξιολόγηση έχει πραγματοποιηθεί με βάση πειραματικές αλλά και ρεαλιστικές οντολογίες.

- Στο Κεφάλαιο 5 μελετάμε το πρόβλημα της συνεπούς απάντησης ερωτημάτων υπό την ICAR σημασιολογία. Αρχικά, παρουσιάζουμε έναν αλγόριθμο αποδοτικού υπολογισμού των ICAR απαντήσεων για $DL - Lite_{\mathcal{R}}$ βάσεις γνώσης, τον οποίο στη συνέχεια βελτιστοποιούμε. Επίσης, εισάγουμε μία νέα σημασιολογία που προσεγγίζει άνωθεν την ICAR σημασιολογία για πιο εκφραστικές ΠΛ και προτείνουμε έναν αλγόριθμο υπολογισμού απαντήσεων με βάση τη νέα σημασιολογία. Τέλος, προτείνουμε έναν αλγόριθμο υπολογισμού της IAR διόρθωσης. Όπως και στο Κεφάλαιο 3 παρουσιάζουμε ένα σύνολο παραδειγμάτων που εξηγούν διαισθητικά τα προβλήματα και στη συνέχεια παρουσιάζουμε τους αντίστοιχους αλγορίθμους και τις αποδείξεις τους.

- Στο Κεφάλαιο 6 παρουσιάζουμε μια πειραματική αξιολόγηση των αλγορίθμων που παρουσιάζονται στο Κεφάλαιο 5, και τους συγκρίνουμε την απόδοσή τους με τα αντίστοιχα διαθέσιμα συστήματα.

- Στο Κεφάλαιο 7, παρουσιάζουμε τη βιβλιογραφία που είναι σχετική με τα προβλήματα που πραγματευόμαστε στη διατριβή αυτή και, τέλος, με το Κεφάλαιο 8 κλείνει η διατριβή συζητώντας την συνεισφορά μας και παρουσιάζοντας θέματα για μελλοντική έρευνα.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Στο κεφάλαιο αυτό παρουσιάζεται περιληπτικά το απαραίτητο θεωρητικό υπόβαθρο για την κατανόηση της διατριβής αυτής. Πιο συγκεκριμένα, στην Ενότητα 2.1 παρουσιάζουμε τους βασικούς ορισμούς από τη λογική πρώτης τάξης που απαιτούνται για την ορθή κατανόηση των προβλημάτων που μελετώνται στα επόμενα κεφάλαια. Στην ενότητα αυτή επίσης παρουσιάζεται ο κανόνας συμπερασμού ανάλυσης ο οποίος χρησιμοποιείται για την εξαγωγή συμπερασμάτων από δοσμένες υποθέσεις. Στη συνέχεια, στην Ενότητα 2.2 γίνεται μια σύντομη εισαγωγή στο συντακτικό και τη σημασιολογία των Περιγραφικών Λογικών καθώς και στις υπηρεσίες συλλογιστικής που παρέχουν. Επίσης, παρουσιάζονται οι βατές Περιγραφικές Λογικές με τις οποίες θα ασχοληθούμε εκτενώς στη διατριβή αυτή. Στην Ενότητα 2.3 γίνεται μία σύντομη εισαγωγή στους υπαρξιακούς κανόνες που μπορούν να χρησιμοποιηθούν για την αναπαράσταση οντολογιών. Στη συνέχεια, στην Ενότητα 2.4, παρουσιάζεται η σύνταξη και η σημασιολογία των Συζευκτικών Ερωτημάτων. Στην ίδια ενότητα, περιγράφεται το πρόβλημα απάντησης ερωτημάτων μέσω της διαδικασίας της επαναγραφής για συνεπείς βάσεις γνώσης. Επίσης, παρουσιάζονται οι σημασιολογίες, με τις οποίες θα ασχοληθούμε στη διατριβή αυτή, που υποστηρίζουν την απάντηση ερωτημάτων σε ασυνεπείς βάσεις γνώσης. Για κάθε περίπτωση, παρουσιάζονται οι πολυπλοκότητες του αντίστοιχου προβλήματος για τις εκφραστικότητες που μας ενδιαφέρουν. Τέλος, καθώς οι κατευθυνόμενοι γράφοι αποτελούν εργαλείο επίλυσης κάποιων απο των προβλημάτων που αντιμετωπίζουμε στα επόμενα κεφάλαια, στην ενότητα 2.5 παρουσιάζουμε μια μικρή εισαγωγή στις βασικά χαρακτηριστικά τους και τις ιδιότητές τους.

2.1 Λογική Πρώτης Τάξης

Στην ενότητα αυτή γίνεται μία σύντομη εισαγωγή στους ορισμούς από τη λογική πρώτης τάξης. Εκτενέστερη ανάλυση υπάρχει στα [43, 87]. Η γλώσσα *Λογικής Πρώτης Τάξης-ΛΠΤ* (*First Order Logic*) δηλώνεται ως $\Sigma(\mathcal{P}, \mathcal{F}, \mathcal{C}, \mathcal{V})$, όπου $\mathcal{P}$ σύνολο από *κατηγορήματα* (*predicates*), $\mathcal{F}$ σύνολο από *συναρτησιακά σύμβολα* (*function symbols*), $\mathcal{C}$ από *σταθερές* (*constants*), και $\mathcal{V}$ σύνολο από *μεταβλητές* (*variables*). Τα σύνολα $\mathcal{P}, \mathcal{F}, \mathcal{C}$ είναι πεπερασμένα ή μετρήσιμα ενώ το σύνολο $\mathcal{V}$ είναι άπειρο.

Το σύνολο των *όρων* (*terms*), $O(\Sigma)$, ορίζεται ως εξής: Μια σταθερά είναι όρος. Μια μεταβλητή είναι όρος. Αν f είναι μια n-αδική συνάρτηση και $t_1, \ldots, t_n$ είναι n στο πλήθος όροι, τότε το $f(t_1, \ldots, t_n)$ είναι όρος. Οι όροι κατασκευάζονται με τους τρεις αυτούς κανόνες και μόνον αυτούς. Όροι της μορφής $f(t_1, \ldots, t_n)$ ονομάζονται *συναρτησιακοί όροι* (*functional terms*). Για τη συνέχεια της διατριβής θεωρούμε μόνο συναρτησιακούς όρους που έχουν βαθμό 1. Ένας όρος λέγεται *βασικός* (*ground term*) εάν δεν περιέχει μεταβλητές.

Κάθε κατηγόρημα σχετίζεται με ένα θετικό ακέραιο n που ονομάζεται *βαθμός του κατηγορήματος* (*arity*). Ένας *καλά σχηματισμένος τύπος* (*well formed formula*) ή απλά *τύπος* (*formula*) κατασκευάζεται με βάση τους κανόνες: Αν P είναι ένα n-δικό κατηγόρημα και $t_1, \ldots, t_n$ είναι n στο πλήθος όροι, τότε το $P(t_1, \ldots, t_n)$ είναι μία *ατομική φόρμουλα* (*atomic formula*) ή αλλιώς *άτομο* (*atom*). Αν το ϕ είναι τύπος, τότε και ο $\neg\phi$ είναι τύπος. Αν ϕ, ψ είναι τύποι, τότε και οι $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, $\phi \leftrightarrow \psi$ είναι τύποι. Αν ϕ είναι τύπος και x μεταβλητή, τότε και οι $\exists x.\phi$, $\forall x.\phi$ είναι τύποι. Οι τύποι κατασκευάζονται με τους τέσσερις προηγούμενους κανόνες και μόνον αυτούς. Στους τύπους $\exists x.\phi$, $\forall x.\phi$, το ϕ ονομάζεται *εμβέλεια* (*scope*) των ποσοδεικτών $\exists$, $\forall$. Μια *εμφάνιση* (*occurrence*) κάποιας μεταβλητής αμέσως μετά τον ποσοδείκτη και μέσα στην εμβέλειά του, ονομάζεται *δεσμευμένη* (*bound*), ενώ η εμφάνισή της έξω από την εμβέλεια οποιουδήποτε ποσοδείκτη, ονομάζεται *ελεύθερη* (*free*). Επίσης, ένα *λεκτικό* (*literal*) L ειναι ένα άτομο (*θετικό λεκτικό στοιχείο* (*positive literal*)) ή η άρνηση ενός ατόμου (*αρνητικό λεκτικό στοιχείο* (*negative literal*)).

Μια *αντικατάσταση* (*substitution*) είναι μια συνάρτηση με πεδίο ορισμού το $\mathcal{V}$ και πεδίο τιμών το $\mathcal{T}(\Sigma)$ η οποία μπορεί να αναπαρασταθεί ως το σύνολο των αντιστοιχίσεων $\{x_1/t_1, \ldots, x_n/t_n\}$, όπου $x_i \in \mathcal{V}$ και $t_i \in \mathcal{T}(\Sigma)$. Η εφαρμογή μιας αντικατάστασης σ σε έναν όρο t, γράφεται $t\sigma$. Επίσης, $c\sigma = c$ εάν $c \in \mathcal{C}$, και $x\sigma = \sigma$ εάν $x \in \mathcal{V}$. Η εφαρμογή μιας αντικατάστασης σ σε ένα άτομο $P(t_1, \ldots, t_n)$ ορίζεται ως $P(t_1\sigma, \ldots, t_n\sigma)$. Μια αντικατάσταση σ είναι πιο γενική (more general) από μια αντικατάσταση δ εάν υπάρχει μια αντικατάσταση λ τ.ω. $\delta = \sigma\lambda$. Επιπλέον, μια αντικατάσταση σ ονομάζεται *ενοποιητής* (*unifier*) των όρων s και t εάν ισχύει $s\sigma = t\sigma$ ενώ εάν η σ είναι πιο γενική από κάθε άλλο ενοποιητή (για τα άτομα αυτά) τότε ονομάζεται *πιο γενικός ενοποιητής* (*most general unifier-mgu*).

Μια *πρόταση* (*clause*) είναι μια διάζευξη λεκτικών $L_1 \vee \ldots \vee L_n$. Μια πρόταση θα λέγεται *πρόταση Horn* (*Horn clause*) εάν το πολύ ένα από τα λεκτικά είναι θετικό. Για παράδειγμα, η πρόταση $\neg R(x,y) \vee \neg A(y) \vee B(x)$ είναι μια πρόταση Horn, όπου τα $\neg R(x,y)$ και $\neg A(y)$ είναι αρνητικά λεκτικά και το $B(x)$ είναι θετικό λεκτικό. Η πρόταση αυτή γράφεται ισοδύναμα και ως: $R(x,y) \wedge A(y) \rightarrow B(x)$. Σε μια πρόταση Horn θεωρούμε πως οι ελεύθερες μεταβλητές των λεκτικών προέρχονται από καθολικούς ποσοδείκτες που εμφανίζονται στην αρχή του τύπου.

Για να παραχθούν συμπεράσματα από μία βάση γνώσης εκφρασμένη σε ΛΠΤ χρειάζεται να εφαρμοστεί ένα σύνολο συμπερασμών. Ένας *κανόνας συμπερασμού* (*inference rule*) είναι μία $n+1$ σχέση προτάσεων. Συμβολίζουμε τα στοιχεία μίας τέτοιας σχέσης ως

$$\frac{c_1, \ldots, c_n}{c}$$

ή, για συντομία, ως $\langle c_1, \ldots, c_n; c\rangle$ και τα ονομάζουμε *συμπερασμούς* (*inferences*). Οι προτάσεις $c_1, \ldots, c_n$ ονομάζονται *υποθέσεις* (*premises*) και το c *συμπέρασμα* (*conclusion*) του συμπερασμού. Στην συνέχεια, θα παρουσίασουμε τον κανόνα συμπερασμού ανάλυσης.

2.1.1 Κανόνας Συμπερασμού Ανάλυσης

Ο *λογισμός της ανάλυσης* (*resolution calculus*) [112], είναι μια μέθοδος που εισήχθη από τον Robinson το 1965 [112] και χρησιμοποιείται ευρέως στην απόδειξη θεωρημάτων (theorem proving).

Ο λογισμός της ανάλυσης στηρίζεται σε δύο *κανόνες συμπερασμού* (*inference rules*). Τον *κανόνα της δυαδικής ανάλυσης* (*binary resolution*) και τον *κανόνα της παραγοντοποίησης* (*factoring*)

$$\frac{C \vee A \qquad D \vee \neg B}{(C \vee D)\sigma}\ \text{BR} \qquad\qquad \frac{C \vee A \vee B}{(C \vee A)\sigma}\ \text{PF}$$

όπου C και D είναι προτάσεις και σ είναι ο πιο γενικός ενοποιητής των ατόμων A και B. Οι κανόνες αυτοί δοθέντων κάποιων υποθέσεων παρέχουν τη δυνατότητα εξαγωγής νέων προτάσεων (συμπερασμάτων) που ικανοποιούν τις υποθέσεις αυτές.

Μία επέκταση του λογισμού της ανάλυσης είναι ο *λογισμός της υπερανάλυσης* (*hyper-resolution calculus*) που βασίζεται σε συμπερασμούς που περιέχουν περισσότερες από δύο υποθέσεις επιστρέφοντας ένα μοναδικό συμπέρασμα. Πρακτικά, ο λογισμός της υπερανάλυσης είναι ισοδύναμος με τον λογισμό της ανάλυσης, με τη διαφορά ότι μία σειρά από κανόνες συμπερασμού πραγματοποιείται σε έναν συμπερασμό. Δηλαδή, για μία πρόταση c της μορφής $C \vee \neg B_1 \vee \ldots \vee \neg B_n$ και προτάσεις c_i της μορφής $D_i \vee B'_i$, ο κανόνας συμπερασμού της υπερανάλυσης είναι μία $n + 2$ σχέση μεταξύ προτάσεων και έχει την παρακάτω μορφή:

$$\frac{C \vee \neg B_1 \vee \ldots \vee \neg B_n \qquad D_1 \vee B'_1\ \ldots\ D_n \vee B'_n}{(C \vee D_1 \vee \ldots D_n)\sigma}$$

με σ τέτοιο ώστε $B_i\sigma = B'_i\sigma$. Λέμε ότι έναν συμπερασμός $\langle c_1, \ldots c_n; c\rangle$ είναι *ορθός* (*sound*) αν $\{c_1, \ldots, c_n\} \models c$. Ένα *σύστημα συμπερασμού* (*inference system*) Γ είναι μια συλλογή από κανόνες συμπερασμού για ένα σύνολο από προτάσεις N. *Απόδειξη* (*proof*) ή *παραγωγή* (*derivation*) μίας πρότασης c από ένα σύνολο προτάσεων μ.β.τ. Γ είναι μια σειρά από προτάσεις $c_1, \ldots, c_m$ τ.ω. $c = c_m$ και κάθε c_i είτε ανήκει στο N είτε είναι συμπέρασμα ενός συμπερασμού από το Γ με υποθέσεις στο σύνολο $N \cup \{c_1, \ldots c_{i-1}\}$. Για ένα σύνολο προτάσεων N και μία πρόταση c συμβολίζουμε με $N \vdash_\Gamma c$ αν υπάρχει μία απόδειξη για το c από το N χρησιμοποιώντας κανόνες από το σύστημα συμπερασμού Γ [9].

Λέμε ότι ένα σύνολο N είναι *κορεσμένο* (*saturated*) ως προς το Γ αν κάθε συμπερασμός υπό το Γ με υποθέσεις από το N ανήκει στο N (μόντουλο τη μετονομασία μεταβλητών).

2.2 Περιγραφικές Λογικές

Οι Περιγραφικές Λογικές (ΠΛ) αποτελούν μία οικογένεια γλωσσών αναπαράστασης της γνώσης που χρησιμοποιούνται ευρέως για τη μοντελοποιήση των οντολογιών. Πρώτη φορά χρησιμοποιήθηκαν στα μέσα του 1980 σαν γλώσσα μοντελοποίησης. Σήμερα, αποτελούν τη βάση της γλώσσας οντολογίας OWL Web όπως έχει οριστεί από το World Wide Web Consortium (W3C).

Οι ΠΛ είναι ένα αποφασίσιμο υποσύνολο της λογικής πρώτης τάξης. Αποτελούν ένα μέσο για την μοντελοποίηση των σχέσεων μεταξύ των οντοτήτων του πεδίου ενδιαφέροντος. Το αλφάβητο των ΠΛ ορίζεται από τρία ξένα μεταξύ τους σύνολα. Το σύνολο των *ατομικών εννοιών* (*atomic concepts*) **C**, ή στη ΛΠΤ των μοναδιαίων κατηγορημάτων, που περιέχει ονόματα που αναφέρονται σε τύπους, κατηγορίες ή κλάσεις οντοτήτων που χαρακτηρίζονται από κάποια κοινή ιδιότητα. Το σύνολο *ατομικών ρόλων* (*atomic roles*) ή αλλιώς *σχέσεων* (*relations*) **R**, ή στη ΛΠΤ των δυαδικών κατηγορημάτων, που περιέχει τις δυαδικές σχέσεις μεταξύ των οντοτήτων, και το σύνολο *ατόμων ή στιγμιοτύπων* (*individuals*) **I**, ή στη ΛΠΤ των σταθερών, που περιέχει όλες τις οντότητες του πεδίου ενδιαφέροντος. Συνήθως οι ατομικές έννοιες αναπαριστώνται με τα γράμματα A, B, οι ρόλοι με τα γράμματα R, S και τα άτομα με τα γράμματα a, b. Οι πρωτογενείς έννοιες σε συνδυασμό με τους κατασκευαστές εννοιών των ΠΛ μπορούν να δημιουργήσουν *περιγραφές εννοιών* (*concept descriptions*) ή αλλιώς *περίπλοκες έννοιες* (*complex concepts*) και συνήθως αναπαριστόνται με τα γράμματα C, D.

Οι Schmidt-Schauss και Smolka [122] όρισαν ένα σύστημα ονοματοποίησης των ΠΛ και ονόμασαν την βασική γλώσσα $\mathcal{AL}$ (Attributive concept Language).Οι περιγραφές εννοιών στη γλώσσα $\mathcal{AL}$ ορίζονται επαγωγικά από την ακόλουθη αφηρημένη σύνταξη:

$$C, D \rightarrow A \mid \top \mid \bot \mid \neg A \mid C \sqcap D \mid \forall R.C \mid \exists R.\bot$$

Όπου R μπορεί να είναι ο *αντίστροφος ρόλος* (*inverse roles*) [60] ενός ατομικού ρόλου P και σε αυτήν την περίπτωση συμβολίζεται ως P^{-}. Η ύπαρξη των αντίστροφων ρόλων σε μία γλώσσα συμβολίζεται με το γράμμα $\mathcal{I}$.

Προσθέτοντας επιπλέον κατασκευαστές εννοιών στην απλή περιγραφική λογική $\mathcal{AL}$ δημιουργούνται περισσότερο εκφραστικές γλώσσες. Ο κατασκευαστής ένωσης (union) (ο οποίος συμβολίζεται με το γράμμα $\mathcal{U}$) επιτρέπει την ένωση δυο εννοιών C, D και συμβολίζεται ως $C \sqcup D$, οπότε σε αυτήν την περίπτωση έχουμε την γλώσσα $\mathcal{ALU}$. Με την επέκταση του περιορισμένου υπαρξιακού περιορισμού στον πλήρη υπαρξιακό περιορισμό (ο οποίος συμβολίζεται με το γράμμα $\mathcal{E}$), έχουμε την γλώσσα $\mathcal{ALE}$. Η επέκταση της $\mathcal{AL}$ που να επιτρέπει την άρνηση και των περίπλοκων εννοιών, με τη σύνταξη $\neg C$, και ονομάζεται $\mathcal{ALC}$, δηλαδή, η ΠΛ που επιτρέπει αυτόν τον κατασκευαστή συμβολίζεται με το γράμμα $\mathcal{C}$. Από τους κανόνες De Morgan εξάγεται εύκολα ότι οι γλώσσες $\mathcal{ALC}$ και $\mathcal{ALUEC}$ ταυτίζονται. Επίσης, οι γλώσσες αυτές μπορούν να επεκταθούν με τον *προσοντούχο περιοριστή πληθυκότητας* (*qualified number restriction ή qualified cardinality restriction*) που συμβολίζεται με το γραμμα $\mathcal{Q}$. Η σύνταξη της μορφής το-πολύ είναι $\leq nR.C$, όπου n ένας φυσικός αριθμος, R ένας ρόλος και C είναι μια οποιαδήποτε έννοια της ΠΛ. Παρόμοια ορίζεται και η μορφή το-λιγότερο. Έτσι, αν υποθέσουμε ότι επεκτείνουμε την $\mathcal{AL}$ με τον προσοντούχο περιοριστή πληθυκότητας τότε έχουμε την γλώσσα $\mathcal{ALQ}$.

Οι ΠΛ προσφέρουν επιπλέον τη δυνατότητα περιγραφής σχέσεων μεταξύ εννοιών, μεταξύ ρόλων καθώς και υποθέσεις όσον αφορά τα άτομα του πεδίου ενδιαφέροντος, αυτό επιτυγχάνεται μέσω του ορισμού των αξιωμάτων. Παρά το γεγονός ότι στη λογική πρώτης τάξης δεν υπάρχει διάκριση μεταξύ των διαφορετικών τύπων αξιωμάτων, στις ΠΛ τα διαχωρίζουμε σε τρεις κατηγορίες: στα *αξιώματα ορολογίας* (*terminological axioms*), *αξιώματα σχέσεων* (*relational axioms*) και στα *αξιώματα ισχυρισμών* (*assertional axioms*).

Τα αξιώματα ορολογίας χρησιμοποιούνται για να εκφράσουν τις σχέσεις μεταξύ εννοιών. Ένα αξίωμα ορολογίας έχει τη μορφή:

$$C \sqsubseteq D \quad \text{ή} \quad C \equiv D$$

Αξιώματα του πρώτου τύπου ονομάζονται *αξιώματα υπαγωγής* (*subsumption axioms ή inclusion axioms*) και δηλώνουν ότι η έννοια C είναι ύπο-έννοια της έννοιας D. Ένα παράδειγμα ενός τέτοιου αξιώματος είναι το εξής: $\mathsf{ΜεταπτυχιακόςΦοιτητής} \sqsubseteq \mathsf{Φοιτητής}$. Αξιώματα του δεύτερου τύπου ονομάζονται *αξιώματα ισοδυναμίας* (*equivalence axioms*) και δηλώνουν ότι οι έννοιες C, D είναι ισοδύναμες. Για παράδειγμα, το αξίωμα $\mathsf{Γυναίκα} \equiv \mathsf{Άνθρωπος} \sqcap \mathsf{Θηλυκό}$ είναι ένα αξίωμα ισοδυναμίας. Όπως φαίνεται από τον ορισμό των αξιωμάτων υπαγωγής, οι έννοιες στο αριστερό αλλά και στο δεξί μέρος των αξιωμάτων ορολογίας μπορούν να είναι οποιαδήποτε περίπλοκη έννοια. Στην περίπτωση αυτή, τα αξιώματα υπαγωγής ονομάζονται *υπαγωγές γενικών εννοιών* (*general concept inclusions – GCIs*) ή *γενικευμένα αξιώματα* (*general axioms*). Ένα σύνολο από αξιώματα υπαγωγής ή ισοδυναμίας αποτελούν το *σώμα ορολογίας* (*TBox – Terminological Box*) ή απλώς μια *ορολογία* (*terminology*) η οποία συμβολίζεται με το γράμμα $\mathcal{T}$.

Τα αξιώματα σχέσεων μπορεί να είναι *αξιώματα υπαγωγής ρόλων* (*role inclusion axioms -RIAs*) [60]. Η ύπαρξή τους συμβολίζεται με το γράμμα $\mathcal{H}$. Τα αξιώματα υπαγωγής ρόλων δηλώνονται συντακτικά με τον ίδιο τρόπο που δηλώνονται και τα αξιώματα υπαγωγής εννοιών. Δηλαδή, αν R και S δύο ρόλοι, τότε ένα αξίωμα υπαγωγής ρόλων έχει τη μορφή:

$$R \sqsubseteq S$$

και δηλώνει ότι ο ρόλος R είναι υπο-ρόλος του S. Το σύνολο των αξιωμάτων υπαγωγής ρόλων ονομάζεται *σώμα ρόλων* (*RBox-Relational Box*). Για συντομία, στη διατριβή αυτή όταν αναφερόμαστε στο TBox θα εννοούμε και το RBox και για την ένωσή τους θα χρησιμοποιούμε τον όρο *οντολογία* (*ontology*).

Ένα από τα αξιώματα ρόλων που προσφέρει μεγάλη εκφραστική δυνατότητα είναι τα *αξιώματα μεταβατικών ρόλων* (*transitive role axioms*) [119]. Ένας ρόλος λέγεται μεταβατικός αν όποτε ισχύει $R(a, b)$ και $R(b, c)$ τότε ισχύει και $R(a, c)$. Η επέκταση της $\mathcal{ALC}$ που περιέχει και μεταβατικούς ρόλους συμβολίζεται με το γράμμα $\mathcal{S}$.

Το σύνολο των GCIs και RIAs μπορούν να διαχωριστούν σε δύο κατηγορίες: στα *θετικά αξιώματα υπαγωγής* (*positive inclusion*) και στα *αρνητικά αξιώματα υπαγωγής* (*negative inclusion*). Τα αρνητικά αξιώματα υπαγωγής είναι της μορφής $C \sqsubseteq \neg D$ ή $C \sqcap D \sqsubseteq \bot$, ανάλογα με την εκφραστικότητα της γλώσσας.

Τα αξιώματα ισχυρισμών δίνουν τη δυνατότητα καθορισμού *σχέσεων στιγμιοτύπου* (*instance relations*) ανάμεσα σε ένα άτομο (ζευγάρι ατόμων) και μια έννοια (ρόλο), τα οποία ονομάζονται *ισχυρισμοί* (*assertions*). Υπάρχουν δυο είδη ισχυρισμών, οι *ισχυρισμοί εννοιών* (*concept assertions*), που έχουν τη σύνταξη $a : C$ ή $C(a)$ και οι *ισχυρισμοί ρόλων* (*role assertion*) που έχουν τη σύνταξη $(a, b) : R$ ή $R(a, b)$, όπου a, b άτομα. Ένα παράδειγμα αξιώματος ισχυρισμού είναι το Γυναίκα(Ελένη). Το σύνολο των ισχυρισμών αποτελεί το *σώμα ισχυρισμών* (*Assertional Box-ABox*).

Μια επιπλέον εκφραστική δυνατότητα που απολαμβάνουν κάποιες ΠΛ είναι η δημιουργία εννοιών με ρητή απαρίθμηση των μελών τους. Ο κατασκευαστής αυτός ονομάζεται *ονοματική έννοια* (*nominal concept*)[121]. Με αυτόν τον τρόπο, αν για παράδειγμα η ΠΛ επιτρέπει τη χρήση ένωσης μπορούμε να περιγράψουμε την έννοια των ημερών της εβδομάδας ως, $\text{ΜέρεςΕβδομάδος} \equiv \{\text{Δευτέρα}\} \sqcup \{\text{Τρίτη}\} \ldots \sqcup \{\text{Κυριακή}\}$. Η ύπαρξη του κατασκευαστή αυτού σε μια ΠΛ δηλώνεται με το γράμμα $\mathcal{T}$.

Δεδομένης μίας ΠΛ γλωσσας $\mathcal{L}$, μία *Βάση Γνώσης* (*Knowledge Base*)-$\mathcal{L}$ ή $\mathcal{L}$-ΒΓ είναι ένα ζεύγος $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, όπου $\mathcal{T}$ το σώμα των σχέσεων υπαγωγής, δηλαδή το TBox, και όπου $\mathcal{A}$ το σώμα των ισχυρισμών, δηλαδή το ABox. Στην εργασία αυτή θεωρούμε ότι όλοι οι ισχυρισμοί κατασκευάζονται πάνω σε ατομικά κατηγορήματα και επομένως η γλώσσα $\mathcal{L}$ προσδιορίζει αποκλειστικά την μορφή των αξιωμάτων του σώματος της ορολογίας.

2.2.1 Σημασιολογία

Μια ΠΛ *ερμηνεία* (*interpretation*) I ορίζεται από ένα ζεύγος $(\Delta^I, \cdot^I)$, όπου Δ^I είναι ένα μη-κενό σύνολο που ονομάζεται *χώρος ερμηνείας* (*domain of interpretation*) και περιέχει στοιχεία που ονομάζονται *αντικείμενα* (*objects*), και $\cdot^I$ είναι μια *συνάρτηση ερμηνείας* (*interpretation function*) που ερμηνεύει κάθε ατομική έννοια ως ένα υποσύνολο του χώρου ερμηνείας και κάθε ρόλο ως υποσύνολο του καρτεσιανού γινομένου του χώρου ερμηνείας. Συγκεκριμένα, ερμηνεύει κάθε έννοια ή ρόλο με τον τρόπο που εμφανίζεται στον Πίνακα 2.1. Η συνάρτηση ερμηνείας μπορεί να επεκταθεί για να δώσει ερμηνεία και σε αξιώματα ισοδυναμίας, υπαγωγής ή ισχυρισμούς. Στον Πίνακα 2.2 παρουσιάζεται η ερμηνεία βασικών αξιωμάτων.

Υπάρχουν περιπτώσεις που δεν υπάρχει καμία ερμηνεία που να ερμηνεύει μία έννοια ως μη κενή, όπως για παράδειγμα η έννοια $C \sqcap \neg C$. Οι έννοιες αυτές ονομάζονται *μη-ικανοποιήσιμες* (*unsatisfiable*).

Μια ερμηνεία I ικανοποιεί μία οντολογία $\mathcal{T}$ αν ικανοποιεί όλα τα αξιώματα υπαγωγής και ισοδυναμίας που υπάρχουν στο $\mathcal{T}$. Τότε λέμε ότι η I είναι *μοντέλο* (*model*) της $\mathcal{T}$. Μια ερμηνεία I ικανοποιεί ένα σώμα ισχυρισμών $\mathcal{A}$ αν ικανοποιεί όλους τους ισχυρισμούς που βρίσκονται στο $\mathcal{A}$. Τότε λέμε ότι η I είναι μοντέλο του $\mathcal{A}$. Τέλος, μια ερμηνεία I ικανοποιεί

Πίνακας 2.1: *Συντακτικό και Σημασιολογία βασικών Κατασκευαστών*

	Συντακτικό	Σημασιολογία
όνομα ατόμου	α	α^I
Ρόλοι:		
ατομικός ρόλος	R	R^I
αντίστροφος ρόλος	R^-	$\{\langle x,y\rangle \| \langle y,x\rangle \in R^I\}$
καθολικός ρόλος	U	$\Delta^I \times \Delta^I$
Έννοιες:		
ατομική έννοια	A	A^I
τομή	$C \sqcap D$	$C^I \cap D^I$
ένωση	$C \sqcup D$	$C^I \cup D^I$
συμπληρωματική	$\neg C$	$\Delta^I \setminus C^I$
καθολική έννοια	$\top$	Δ^I
κενή έννοια	$\bot$	$\emptyset$
υπαρξιακός περιορισμός	$\exists R.C$	$\{x \mid \exists y \in \Delta^I$ τ.ω. $\langle x,y\rangle \in R^I$ και $y \in C^I\}$
καθολικός περιορισμός	$\forall R.C$	$\{x \mid \forall y \in \Delta^I$ τ.ω. $\langle x,y\rangle \in R^I \to y \in C^I\}$
ονοματική έννοια	$\{a\}$	$\{a^I\}$

Πίνακας 2.2: *Συντακτικό και Σημασιολογία βασικών Αξιωμάτων*

	Συντακτικό	Σημασιολογία
ABox:		
ισχυρισμός έννοιας	$C(\alpha)$	$\alpha^I \in C^I$
ισχυρισμός ρόλου	$R(\alpha,\beta)$	$\langle \alpha^I, \beta^I\rangle \in R^I$
TBox:		
υπαγωγή εννοιών	$C \sqsubseteq D$	$C^I \subseteq D^I$
ισοδυναμία εννοιών	$C \equiv D$	$C^I = D^I$
RBox:		
υπαγωγή ρόλων	$R \sqsubseteq S$	$R^I \subseteq S^I$
ισοδυναμία ρόλων	$R \equiv S$	$R^I = S^I$

ένα σώμα ισχυρισμών $\mathcal{A}$ με βάση το σώμα ορολογίας $\mathcal{T}$ αν η I είναι μοντέλο του $\mathcal{A} \cup \mathcal{T}$, οπότε σε αυτήν την περίπτωση λέμε ότι είναι μοντέλο της βάσης γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ και το συμβολίζουμε με $Mod(\mathcal{K})$.

2.2.2 Υπηρεσίες Εξαγωγής Συμπερασμάτων

Οι ΠΛ παρέχουν υπηρεσίες πάνω στα σώματα ορολογίας και στα σώματα ισχυρισμών. Έστω $\mathcal{T}$ ένα TBox και $\mathcal{A}$ ένα ABox.

- *Ικανοποιησιμότητα* (*satisfiability*): Η έννοια C είναι ικανοποιήσιμη (satisfiable) μβτ $\mathcal{T}$ αν υπάρχει μοντέλο I του $\mathcal{T}$ τέτοιο ώστε $C^I \neq \emptyset$.

- *Υπαγωγή* (*subsumption*): Η έννοια C υπάγεται στην έννοια D (subsumed by) μβτ $\mathcal{T}$ αν $C^I \subseteq D^I$ για κάθε μοντέλο I του $\mathcal{T}$. Σε αυτήν την περίπτωση γράφουμε $\mathcal{T} \models C \sqsubseteq D$.

- *Ισοδυναμία* (*equivalence*): Η έννοια C είναι ισοδύναμη με την έννοια D μβτ $\mathcal{T}$ αν $C^I = D^I$ για κάθε μοντέλο I του $\mathcal{T}$. Σε αυτήν την περίπτωση γράφουμε $\mathcal{T} \models C \equiv D$.

- *Ξένες Έννοιες* (*disjointness*): Η έννοια C είναι ξένη με την έννοια D (disjoint with) μβτ $\mathcal{T}$ αν $C^I \cap D^I = \emptyset$ για κάθε μοντέλο I του $\mathcal{T}$.

- *Συνέπεια* (*consistency*): Το $\mathcal{A}$ είναι συνεπές (consistent) μβτ $\mathcal{T}$ αν υπάρχει μοντέλο του $\mathcal{T}$ το οποίο είναι και μοντέλο του $\mathcal{A}$.

- *Συνεπαγωγή* (*entailment*): Το $\mathcal{A}$ συνεπάγεται (entails) έναν ισχυρισμό ϕ μβτ $\mathcal{T}$, αν κάθε μοντέλο του $\mathcal{A}$ και του $\mathcal{T}$ ικανοποιεί τον ισχυρισμό. Σε αυτήν την περίπτωση γράφουμε $\mathcal{T}, \mathcal{A} \models \phi$

Όλες οι υπηρεσίες εξαγωγής συμπερασμάτων που αφορούν στο σώμα ορολογίας μπορούν να αναχθούν στο πρόβλημα της μη-ικανοποιησιμότητας [40]. Επίσης, το πρόβλημα της συνεπαγωγής μπορεί να αναχθεί στο πρόβλημα της συνέπειας ενός ABox, δηλαδή, ισχύει ότι $\mathcal{A} \models \phi$ μβτ TBox $\mathcal{T}$ ανν το $\mathcal{A} \cup \{\neg\phi\}$ είναι ασυνεπές (inconsistent) μβτ $\mathcal{T}$.

2.2.3 Βατές Περιγραφικές Λογικές

Οι συλλογιστικοί αλγόριθμοι που αφορούν στη γλώσσα $\mathcal{ALC}$, η οποία φαίνεται σχετικά απλή, είναι στην χειρότερη περίπτωση εκθετικού χρόνου. Για τον λόγο αυτό, τα τελευταία χρόνια έχει αναπτυχθεί ένα σύνολο γλωσσών το οποίο έχει καλές υπολογιστικές ιδιότητες ενώ ταυτόχρονα διατηρεί μια σχετική εκφραστικότητα. Οι τρεις βασικές οικογένειες γλωσσών που ανήκουν στο σύνολο αυτό είναι οι $\mathcal{EL}$ [7], DLP [54] και DL-Lite[30], οι οποίες αντιστοιχούν στις γλώσσες OWL EL, OWL RL και OWL QL της Web Ontology

Language. Στη συνέχεια, θα παρουσιάσουμε περιεκτικά τις γλώσσες και τις επεκτάσεις τους στις οποίες θα αναφερθούμε στη διατριβή αυτή.

Η οικογένεια των $\mathcal{EL}$ χαρακτηρίζεται από την απεριόριστη χρήση υπαρξιακών ποσοδεικτών και τομής εννοιών. Επίσης, επιτρέπει την έννοια $\top$ αλλά δεν επιτρέπει τον σύνδεσμο της ένωσης, τον κατασκευαστή της άρνησης ή καθολικούς ποσοδείκτες. Η επέκταση της $\mathcal{EL}$ στην $\mathcal{EL}_{\bot}$ ορίζεται ως εξής:

$$R \rightarrow P$$
$$C \rightarrow A \mid \bot \mid C_1 \sqcap C_2 \mid \exists R.C$$

όπου P είναι ένας ατομικός ρόλος. Ένα $\mathcal{EL}_{\bot}$ TBox αποτελείται από αξιώματα της μορφής $C_1 \sqsubseteq C_2$, $R_1 \sqsubseteq R_2$. Στην εκφραστικότητα αυτή, οι περισσότερες υπηρεσίες εξαγωγής συμπερασμάτων πραγματοποιούνται στην χειρότερη περίπτωση σε πολυωνυμικό χρόνο. Επίσης, μία γλώσσα που θα μας απασχολήσει είναι η $\mathcal{EL}_{\bot nr}$ [115]. Μία ατομική έννοια A ονομάζεται *αναδρομική* (*recursive*) σε ένα TBox $\mathcal{T}$, αν το $\mathcal{T}$ συνεπάγεται ένα αξίωμα της μορφής $C \sqsubseteq A$, όπου το C περιέχει τουλάχιστον μία εμφάνιση της A εμφωλευμένη σε μία περίπλοκη έννοια ποσοτικοποιημένη με ένα υπαρξιακό ποσοδείκτη και δεν υπάρχει έννοια C' τέτοια ώστε $\mathcal{T} \models C' \sqsubseteq A$ και $\mathcal{T} \models C' \sqsubseteq C$. Διαφορετικά, η έννοια A ονομάζεται *μη-αναδρομική* (*non-recursive*). Η γλώσσα $\mathcal{EL}_{\bot nr}$ είναι το υποσύνολο της γλώσσας $\mathcal{EL}_{\bot}$ τέτοιο ώστε κάθε έννοια C_i, που συμμετέχει σε ένα αξίωμα της μορφής $C_1 \sqcap \ldots \sqcap C_n \sqsubseteq \bot$ που υπάγεται από το $\mathcal{T}$, είναι μη-αναδρομική.

Αν επεκταθεί η $\mathcal{EL}$ με αξιώματα υπαγωγής ρόλων, αρνητικών αξιωμάτων υπαγωγής, τη δυνατότητα ορισμού αντίστροφων ρόλων και ορισμού ονοματικών εννοιών, τότε, σύμφωνα με τα προηγούμενα, έχουμε τη γλώσσα $\mathcal{ELHIO}^{\neg}$. Τυπικά, η γλώσσα $\mathcal{ELHIO}^{\neg}$ ορίζεται ως εξής:

$$R \rightarrow P \mid P^{-} \qquad E \rightarrow R \mid \neg R$$
$$B \rightarrow A \mid \{a\} \mid \top \mid B_1 \sqcap B_2 \mid \exists R.B \qquad C \rightarrow B \mid \neg B$$

Ένα $\mathcal{ELHIO}^{\neg}$ TBox αποτελείται από ένα πεπερασμένο σύνολο αξιωμάτων της μορφής $B \sqsubseteq C$ και της μορφής $R \sqsubseteq E$.

Μια DLP γλώσσα της οποίας αξιοποιούμε τις υπολογιστικές δυνατότητες είναι η $\mathcal{RL}$. Το συντακτικό της ορίζεται ως εξής:

$$C \rightarrow A \mid C_1 \sqcap C_2 \mid C_1 \sqcup C_2 \mid \exists R.C$$
$$D \rightarrow A \mid D_1 \sqcap D_2 \mid \forall R.D$$

και τα αξιώματα που μπορούν να κατασκευαστούν στην $\mathcal{RL}$ είναι τα εξής: $C \sqsubseteq D$, $C_1 \equiv C_2$, $\top \sqsubseteq \forall P^{-}.D$, $P \sqsubseteq Q$, $P \equiv Q^{-}$,$P \equiv Q$.

Μια DL-Lite γλώσσα στην οποία θα αναφερθούμε αρκετές φορές, είναι η DL-Lite$_{\mathcal{R}}$ η οποία αποτελεί ένα υποσύνολο της $\mathcal{ELHIO}^{\neg}$ και ορίζεται ως εξής:

$$R \rightarrow P \mid P^{-} \qquad E \rightarrow R \mid \neg R$$
$$B \rightarrow A \mid \exists R \qquad C \rightarrow B \mid \neg B \mid \exists R.B$$

ενώ τα αξιώματα που μπορούν να σχηματιστούν στην DL-Lite$_{\mathcal{R}}$ είναι της μορφής $B \sqsubseteq C$, $R \sqsubseteq E$.

Μία ακόμα γλώσσα που, υπό ορισμένες συνθήκες, έχει καλές υπολογιστικές ιδιότητες είναι η Horn $- \mathcal{SHIQ}$. Η γλώσσα αυτή αποτελεί ένα υποσύνολο της γλώσσας $\mathcal{SHIQ}$ καθώς το συντακτικό της είναι περιορισμένο με τέτοιον τρόπο ώστε να μην επιτρέπεται η ένωση. Παρά το γεγονός ότι η συνήθεις συλλογιστικές υπηρεσίες είναι εκθετικής πολυπλοκότητας [62], εξαιτίας της απουσίας του κατασκευαστή της ένωσης είναι πολυωνυμικής πολυπλοκότητας ως προς τα δεδομένα. [74]. Στην Horn$-\mathcal{SHIQ}$ μπορούμε να έχουμε μόνο τα εξής αξιώματα:

$$\begin{array}{ccc} A \sqcap B \sqsubseteq C & A \sqsubseteq \forall R.B & A \sqsubseteq \geq mS.B \\ \exists R.A \sqsubseteq B & A \sqsubseteq \exists R.B & A \sqsubseteq \leq 1S.B \end{array}$$

όπου τα A, B, C είναι ονόματα εννοιών, το R μπορεί να είναι ένας ατομικός ρόλος P ή ο ανάστροφός του P^{-}, το S ένας απλός ρόλος, και $m \geq 1$.

2.3 Υπαρξιακοί Κανόνες

Ένας *υπαρξιακός κανόνας* (*existential rule*) (ή απλά κανόνας) [11, 26], που συχνά ονομάζεται και αξίωμα, είναι μια φόρμουλα της μορφής

$$\exists \vec{y}.\psi(\vec{x},\vec{y}) \leftarrow \phi(\vec{x},\vec{z}) \tag{2.1}$$

όπου τα $\vec{x}$, $\vec{z}$ είναι καθολικά ποσοτικοποιημένα (universally quantified), τα $\phi(\vec{x},\vec{z})$ και $\psi(\vec{x},\vec{y})$ είναι συζεύξεις από άτομα και τα $\vec{x}$, $\vec{y}$ και $\vec{z}$ είναι ανά δύο διαφορετικά. Στη διατριβή αυτή θα μας απασχολήσουν μόνο άτομα βαθμού το πολύ δύο. Η φόρμουλα ϕ ονομάζεται *σώμα* (*body*), η φόρμουλα ψ ονομάζεται *κεφαλή* (*head*) και οι καθολικοί ποσοδείκτες συχνά παραλείπονται.

Όταν το διάνυσμα $\vec{y}$ είναι κενό τότε ο κανόνας ονομάζεται *κανόνας Datalog* (*Datalog rule*). Ένα *πρόγραμμα Datalog* (*Datalog program*) είναι ένα πεπερασμένο σύνολο από κανόνες Datalog. Αξίζει να σημειωθεί ότι εξ' ορισμού οι κανόνες Datalog είναι ασφαλείς (safe)−−δηλαδή, όλες οι μεταβλητές που εμφανίζονται στο διάνυσμα $\vec{x}$ της κεφαλής του εμφανίζονται και στο σώμα του. Πολλές δημοφιλείς γλώσσες οντολογιών Horn, όπως είναι και η οικογένεια γλωσσών της DL-Lite [31] , η $\mathcal{ELHI}$ [103], καθώς και η Datalog$^{\pm}$ [26] αποτελούν αποφασίσιμα υποσύνολα των υπαρξιακών κανόνων.

Πρέπει να σημειωθεί, ότι, θεωρητικά, οι γλώσσες DL-Lite και Datalog$^{\pm}$ έχουν μεγαλύτερη εκφραστικότητα. Για παράδειγμα, περιέχουν επίσης αρνητικούς περιορισμούς της μορφής $\perp \leftarrow C(x) \wedge D(x)$ καθώς και αξιώματα ισοδυναμίας της μορφής $y \approx z \leftarrow R(x,y) \wedge R(x,z)$. Από τη στιγμή όμως που η γλώσσα ερωτήματος δεν επιτρέπει της ισοδυναμία ατόμων και υποθέτουμε ότι η οντολογία είναι συνεπής ως προς τα δεδομένα, η εκφραστικότητα αυτή δεν σχετίζεται με τους αλγορίθμους απάντησης ερωτημάτων για συνεπείς βάσεις γνώσης που μελετώνται αργότερα και συνεπώς δεν λαμβάνεται υπόψη.

Κάθε υπαρξιακός κανόνας μπορεί να μετασχηματιστεί σε ένα ισοδύναμο σύνολο προτάσεων με τη χρήση τεχνικών κανονικοποιήσης και των κανόνων skolem [8] στους κανόνες που περιέχουν υπαρξιακούς ποσοδείκτες στην κεφαλή. Με τη τεχνικής της κανονικοποίησης, κάθε περίπλοκη έννοια αντικαθίσταται με συστηματικό τρόπο από μία ατομική. Για παράδειγμα, ένας κανόνας $\exists y.R(x, y) \wedge C(x) \wedge D(x) \leftarrow A(x)$ μπορεί να μετασχηματιστεί σε δύο κανόνες $\exists y.R(x, y) \wedge A'(x) \leftarrow A(x), C(x) \wedge D(x) \leftarrow A'(x)$, όπου A' ένα νέο κατηγόρημα. Με την εφαρμογή του skolemization ο κανόνας $\exists y.P(x, y) \wedge A'(y) \leftarrow A(x)$ μετασχηματίζεται στο ζεύγος προτάσεων $P(x, f(x)) \leftarrow A(x)$, $A'(f(x)) \leftarrow A(x)$.

Ονομάζουμε τις προτάσεις που δεν περιέχουν συναρτησιακούς όρους ως *συναρτησιακά-ελέυθερες* (*function-free*).

Συχνά δεν κάνουμε καμία διάκριση μεταξύ ενός υπαρξιακού κανόνα και της αντίστοιχης αναπαράστασής του σε μία ή δύο προτάσεις.

2.4 Συζευκτικά Ερωτήματα

Ο τύπος των ερωτημάτων που θα μας απασχολήσουν στη διατριβή αυτή είναι τα συζευκτικά ερωτήματα. Ένα *σύνολο συζευκτικών ερωτημάτων-ΣΣΕ* (*Union of Conjunctive Queries – UCQ*) είναι ένα σύνολο κανόνων Datalog οι οποίοι περιέχουν το κατηγόρημα Q στην κεφαλή αλλά όχι στο σώμα τους. Ένα ΣΣΕ λέγεται *συζευκτικό ερώτημα-ΣΕ* (*conjunctive query-CQ*) αν περιέχει μόνο έναν κανόνα. Για λόγους απλοποίησης, στη συνέχεια της διατριβής μας κάνουμε συχνά κατάχρηση του συμβολισμού και χρησιμοποιούμε το $\mathcal{Q}$ για να αναφερθούμε στον μοναδικό κανόνα ενός ΣΕ.

Ορισμός 1. Μια πλειάδα $\vec{a}$ ατόμων είναι *απάντηση* (*certain answer*) σε ένα ΣΕ $\mathcal{Q}$ με βάση μία οντολογία $\mathcal{T}$ και ένα στιγμιότυπο $\mathcal{I}$ αν το μέγεθος της $\vec{a}$ ταυτίζεται με τον βαθμό του $\mathcal{Q}$ και $\mathcal{Q} \cup \mathcal{T} \cup \mathcal{I} \models Q(\vec{a})$. Συμβολίζουμε με $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{I})$ τις απαντήσεις στο $\mathcal{Q}$ με βάση το $\mathcal{T} \cup \mathcal{I}$.

Με βάση τα παραπάνω, μπορούμε να ορίσουμε επιπλέον την εξής υπηρεσία συλλογιστικής:

- *Απάντηση ερωτήματος* (*query answering*): Για μια ΒΓ $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ και ένα ερώτημα $\mathcal{Q}$, να υπολογιστεί το σύνολο $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{I})$.

2.4.1 Απάντηση Ερωτημάτων μέσω Επαναγραφής για Συνεπείς Βάσεις Γνώσης

Μία από τις πιο κοινές προσεγγίσεις για την απάντηση ερωτημάτων (query answering) είναι μέσω της διαδικασίας της επαναγραφής των ερωτημάτων (query rewriting) [30, 103]. Διαισθητικά, μία *επαναγραφή* ενός ερωτήματος $\mathcal{Q}$ με βάση μία οντολογία $\mathcal{T}$ είναι ένα Datalog πρόγραμμα που περιέχει όλη την πληροφορία του $\mathcal{T}$ σχετική με το ερώτημα $\mathcal{Q}$ πάνω σε ένα οποιοδήποτε στιγμιότυπο $\mathcal{I}$.

Ορισμός 2. Μια *επαναγραφή Datalog* (*Datalog rewriting*) για ένα ΣΕ $\mathcal{Q}$ με βάση μία οντολογία $\mathcal{T}$ είναι ένα πρόγραμμα Datalog $\mathcal{R}$ το οποίο μπορεί να χωριστεί σε δύο ανεξάρτητα σύνολα $\mathcal{R}_D \cup \mathcal{R}_Q$, με $\mathcal{R}_D$ ένα σύνολο από κανόνες Datalog που δεν περιέχουν το κατηγόρημα Q και $\mathcal{R}_Q$ ένα ΣΣΕ με κατηγόρημα ερωτήματος Q και για κάθε στιγμιότυπο συνεπές ώς προς την $\mathcal{T}$ και χρησιμοποιώντας κατηγορήματα μόνο από την $\mathcal{T}$ ισχύει:

$$\mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}_Q, \mathcal{R}_D \cup \mathcal{I})$$

Η επαναγραφή $\mathcal{R}$ είναι μια *επαναγραφή ΣΣΕ* (*UCQ rewriting*) εάν $\mathcal{R}_D = \emptyset$.

Με βάση την έννοια της επαναγραφής μπορούμε να διαχωρίσουμε τις ΠΛ με βάση την επαναγραψιμότητά τους.

Ορισμός 3. Μία ΠΛ εκφραστικότητας $\mathcal{L}$ ονομάζεται *datalog επαναγράψιμη* (*datalog rewritable*) αν για κάθε $\mathcal{L}$-TBox $\mathcal{T}$ υπάρχει ένα datalog-πρόγραμμα $\mathcal{R}_D$ τέτοιο ώστε για κάθε ερώτημα $\mathcal{Q}$ που δεν περιέχει υπαρξιακές μεταβλητές το $\mathcal{R} = \langle \{\mathcal{Q}\}, \mathcal{R}_{\mathcal{D}} \rangle$ είναι μία datalog-επαναγραφή του $\mathcal{Q}$ μ.β. το $\mathcal{T}$.

Μια επαναγραφή μπορεί να υπολογιστεί εφαρμόζοντας μετασχηματισμούς που διατηρούν την ισοδυναμία (equivalence preserving transformations) για τα δεδομένα εισόδου $\mathcal{T}$ και $\mathcal{Q}$. Διάφορες τεχνικές και αλγόριθμοι έχουν προταθεί μέχρι στιγμής στην βιβλιογραφία για διάφορες γλώσσες οντολογιών όπως είναι η DL-Lite, η $\mathcal{ELHI}$, οι υπαρξιακοί κανόνες και άλλες [30, 2, 116, 35, 51, 95, 132, 137] και πολλοί από αυτούς διαφέρουν σημαντικά μεταξύ τους.

Ένα από τα πρώτα συστήματα επαναγραφής που αναπτύχθηκαν στο πλαίσιο των ΠΛ είναι το Requiem το οποίο βασίζεται στη μέθοδο της ανάλυσης. Αποτελεί ένα από τα βασικά συστήματα επαναγραφής, πάνω στο οποίο έχουν βασιστεί πολλά επόμενα συστήματα επαναγραφής, όπως είναι το Blackout [104], το Kyrie [89] και το Rapid [133]. Στη συνέχεια, θα παρουσιάσουμε περιεκτικά τον αλγοριθμό του.

Αρχικά, δέχεται στην είσοδο μία οντολογία $\mathcal{T}$ και ένα ερώτημα $\mathcal{Q}$. Στη συνέχεια, μετασχηματίζει τα αξιώματα της οντολογίας σε προτάσεις Horn. Στο επόμενο βήμα, εφαρμόζει όλους τους δυνατούς συμπερασμούς στο σύνολο $\mathcal{Q} \cup \mathcal{T}$, με βάση το σύστημα συμπερασμού του (που συμβολίζουμε με Γ^{Rq}), δηλαδή, υπολογίζει τον κορεσμό του συνόλου $\mathcal{Q} \cup \mathcal{T}$. Τέλος, επιστρέφει τις συναρτησιακά ελεύθερες προτάσεις.

Για παράδειγμα, έστω η οντολογία $\mathcal{T} = \{A \sqsubseteq \exists R.C,\, R \sqsubseteq S\}$ και το ερώτημα $\mathcal{Q} = Q(x) \leftarrow S(x,y) \land C(y)$ το Requiem θα υπολογίσει μία επαναγραφή ως εξής:

Αρχικά, θα μετασχηματίσει την $\mathcal{T}$ στο παρακάτω σύνολο από προτάσεις:

$$R(x, f(x)) \leftarrow A(x) \tag{2.2}$$

$$C(f(x)) \leftarrow A(x) \tag{2.3}$$

$$S(x, y) \leftarrow R(x, y) \tag{2.4}$$

όπου f είναι μια συνάρτηση skolem. Στη συνέχεια, σύμφωνα με το Γ^{Rq} οι προτάσεις $\mathcal{Q}$, (2.3) θα αναλυθούν, ενοποιώντας τα άτομα $C(y)$ και $C(f(x))$:

$$\frac{Q(x) \leftarrow S(x,y) \wedge C(y) \quad C(f(x)) \leftarrow A(x)}{Q(x) \leftarrow S(x,f(x)) \wedge A(x)} \tag{2.5}$$

ενώ στο επόμενο βήμα θα αναλύσει τις προτάσεις (2.2), (2.4), ενοποιώντας τα άτομα $R(x,y)$ και $R(x,f(x))$:

$$\frac{S(x,y) \leftarrow R(x,y) \quad R(x,f(x)) \leftarrow A(x)}{S(x,f(x)) \leftarrow A(x)} \tag{2.6}$$

Μετά, θα αναλύσει τα συμπεράσματα των συμπερασμών (2.5),(2.6):

$$\frac{Q(x) \leftarrow S(x,f(x)) \wedge A(x) \quad S(x,f(x)) \leftarrow A(x)}{Q(x) \leftarrow A(x)} \tag{2.7}$$

Με τον τρόπο αυτό, υπολογίζει τον κορεσμο $\mathcal{S}$ του συνόλου $\mathcal{Q} \cup \mathcal{T}$:

$$\mathcal{S} = \mathcal{T} \cup \{Q(x) \leftarrow S(x,f(x)) \wedge A(x), S(x,f(x)) \leftarrow A(x), Q(x) \leftarrow A(x)\} \tag{2.8}$$

Τέλος, θα επιστρέψει την επαναγραφή $\mathcal{R}$ του $\mathcal{Q}$ μ.β.τ. $\mathcal{T}$ αφαιρώντας τις προτάσεις που περιέχουν συναρτησιακούς όρους, δηλαδή, $\mathcal{R} = \{\mathcal{Q}, Q(x) \leftarrow A(x), S(x,y) \leftarrow R(x,y)\}$.

Κριτήριο Περιττότητας

Πολλοί αλγόριθμοι επαναγραφής για να βελτιστοποιήσουν τόσο την απόδοσή τους εφαρμόζουν όσο και τον χρόνο απάντησης του τελικού ερωτήματος, εφαρμόζουν τεχνικές αφαίρεσης περιττών προτάσεων. Διαισθητικά, μία πρόταση είναι περιττή είτε αν είναι ταυτολογία, για παράδειγμα, είναι της μορφής $A(x) \leftarrow A(x)$, όπου δεν προσφέρει κάποια πληροφορία, είτε αν υπάγεται από μία υπάρχουσα πρόταση. Για παράδειγμα, έστω τα ερωτήματα $c_1 = Q(x) \leftarrow A(x)$, $c_2 = Q(x) \leftarrow A(x) \wedge B(x)$, είναι προφανές ότι όταν το ερώτημα c_1 αποτιμηθεί, θα επιστραφούν περισσότερες απαντήσεις από αυτές που θα επιστραφούν από την αποτίμηση του c_2. Συνεπώς, το ερώτημα c_2 είναι περιττό, καθώς και όλες οι προτάσεις που υπολογίζονται με βάση αυτό.

Τυπικά, μία πρόταση είναι συντακτική ταυτολογία αν περιέχει δύο λεκτικά του τύπου A και $\neg A$. Δεδομένων δύο προτάσεων c_1, c_2 λέμε ότι το c_1 *υπάγει* (*subsumes*) το c_2 αν υπάρχει αντικατάσταση σ τέτοια ώστε κάθε λεκτικό που εμφανίζεται στο $c_1\sigma$ να εμφανίζεται και στο c_2. Επίσης, μία πρόταση c είναι *περιττή* (*redundant*) σε ένα σύνολο προτάσεων $\mathcal{S}$ αν το c είναι ταυτολογία ή υπάγεται από κάποια άλλη πρόταση που ανήκει στο $\mathcal{S}$. Είναι προφανές ότι ισχύει η παρακάτω πρόταση:

Λήμμα 4. *Έστω c_1, c_2 προτάσεις και έστω ότι η c_1 υπάγει τη c_2. Τότε, για κάθε πρόταση c που παράγεται από την c_2 και ένα σύνολο προτάσεων $\mathcal{T}$, υπάρχει πρόταση c' που παράγεται από την c_1 και το $\mathcal{T}$ τέτοια ώστε η c' να υπάγει τη c.*

Σε ένα σύστημα επαναγραφής οι έλεγχοι υπαγωγής μπορεί να εφαρμοστούν με εμπρόσθιο ή οπισθοδρομικό τρόπο. Με τον *εμπρόσθιο έλεγχο υπαγωγής* (*forward subsumption checking*) κάθε νέα πρόταση c αφαιρείται αν είναι περιττή σε σχέση με το σύνολο των ήδη υπάρχουσων προτάσεων $\mathcal{S}$, ενώ με τον *οπισθοδρομικό έλεγχο υπαγωγής* (*backward subsumption checking*) αφαιρούνται οι προτάσεις του $\mathcal{S}$ που υπάγονται από το c.

Ένα *κριτήριο περιττότητας* (*redundancy elimination*) $\mathcal{D}$ είναι μία αντιστοίχιση που συσχετίζει με κάθε σύνολο $\mathcal{S}$ ένα μέγιστο υποσύνολο του $\mathcal{S}$ που δεν περιέχει περιττές προτάσεις. Αξίζει να σημειωθεί, ότι το υποσύνολο αυτό δεν είναι μοναδικό, για παράδειγμα, μπορεί μία πρόταση να υπάγει την άλλη, και αντιστρόφως, οπότε οποιαδήποτε από τις δύο μπορεί να αφαιρεθεί.

Έστω το σύστημα συμπερασμού Γ και το κριτήριο περιττότητας $\mathcal{D}$, λέμε ότι ένα σύνολο προτάσεων $\mathcal{S}$ είναι κορεσμένο και ελάχιστο ως προς τις σχέσεις υπαγωγής, και το συμβολίζουμε με $\mathcal{S}_{\Gamma,\mathcal{D}}$, αν όλοι οι συμπερασμοί του Γ με μη-περιττές υποθέσεις από το $\mathcal{S}$ παράγουν συμπεράσματα που είναι περιττά στο $\mathcal{S}$.

Ζητήματα Πολυπλοκότητας

Ένα ενδιαφέρον ζήτημα που προκύπτει κατά την κατασκευή ενός αλγορίθμου είναι η συμπεριφορά του σε διαφορετικές εισόδους. Εφόσον τερματίζει, μας ενδιαφέρει ο χρόνος που απαιτείται για την επίλυση του προβλήματος. Ο χρόνος αυτός μπορεί να σχετίζεται με γραμμικό, πολυωνυμικό, εκθετικό ή με πιο περίπλοκες συναρτήσεις με τα δεδομένα εισόδου. Σε προβλήματα όπως είναι η απάντηση ερωτημάτων η πολυπλοκότητα υπολογίζεται με βάση τον όγκο των δεδομένων γιατί συνήθως τόσο το μέγεθος του ερωτήματος όσο και το μέγεθος της οντολογίας είναι σημαντικά μικρότερα από το μέγεθος του ABox. Η δομή όμως της επαναγραφής καθορίζει την πολυπλοκότητα του προβλήματος της απάντησης ερωτημάτων.

Η δομή επαναγραφής εξαρτάται άμεσα από την γλώσσα της οντολογίας. Σύμφωνα με τους Pérez-Urbina et al. [103], αν το $\mathcal{T}$ είναι εκφραστικότητας $\mathcal{ELHIO}$, τότε η επαναγραφή ενός ερωτήματος $\mathcal{Q}$ μ.β.τ. $\mathcal{T}$ είναι στη χειρότερη περίπτωση μια επαναγραφή Datalog, εάν είναι εκφραστικότητας γραμμικής Datalog$^{\pm}$[26], τότε στη χειρότερη περίπτωση θα είναι μια επαναγραφή μη-αναδρομικής Datalog, εάν είναι εκφραστικότητας DL-Lite^{+} [5], τότε η επαναγραφή αποτελείται στη χειρότερη περίπτωση από ένα ΣΣΕ και από μια επαναγραφή γραμμικής Datalog, ενώ τέλος εάν η $\mathcal{T}$ είναι εκφραστικότητας DL-Lite$_{\mathcal{R}}$, τότε η επαναγραφή είναι ένα ΣΣΕ.

Το πρόβλημα της εκτέλεσης ενός προγράμματος Datalog εμφανίζει πολυπλοκότητα PTIME ως προς τα δεδομένα [103], ενός γραμμικού datalog προγράμματος NLogSpace ως προς τα δεδομένα [103], ενός ΣΣΕ ανήκει στην κλάσση πολυπλοκότητας AC^0 [96, 134, 135] που περιέχεται στην LOGSPACE, ως προς τα δεδομένα. Παρατηρούμε δηλαδή, ότι η απάντηση ερωτημάτων για απλές ΠΛ είναι χαμηλής πολυπλοκότητας.

2.4.2 $\mathcal{RL}$ Συστήματα Απάντησης Ερωτημάτων

Όπως έχει ήδη σημειωθεί, η τεχνική της επαναγραφής αποτελεί μία από τις πιο διαδεδομένες μεθόδους υπολογισμού απάντησης ερωτημάτων. Ωστόσο, ένα αρνητικό χαρακτηριστικό της προσέγγισης αυτής είναι ότι υπολογίζονται επαναγραφές πολύ μεγάλου μεγέθους των οποίων η εκτέλεση απαιτεί αρκετό χρόνο. Για παράδειγμα, το μέγιστο μέγεθος μίας επαναγραφής ενός ερωτήματος $|Q|$ με βάση μία $\mathcal{ELHIO}$ οντολογία $|\mathcal{T}|$ είναι $O(2^{|\mathcal{T}|\cdot|Q|})$. Μια διαφορετική προσέγγιση στο πρόβλημα της απάντησης ερωτημάτων, για οντολογίες χαμηλής εκφραστικότητας, είναι μέσω της "υλοποίησης" (materialization) της βάσης γνώσης.

Τα περισσότερα συστήματα απάντησης ερωτημάτων που υποστηρίζουν τα $\mathcal{RL}$-αξιώματα, υπολογίζουν το κορεσμένο ABox λαμβάνοντας υπόψη μόνο το $\mathcal{RL}$-τμήμα του TBox $\mathcal{T}$ εισόδου, το οποίο και συμβολίζουμε με $\mathcal{T}|_{\mathsf{rl}}$.

Ορισμός 5. Ένα $\mathcal{RL}$ *ABox-σύστημα κορεσμού* ans είναι μία διαδικασία που παίρνει στη είσοδο ένα TBox $\mathcal{T}$, ένα ABox $\mathcal{A}$, ένα ΣΕ $\mathcal{Q}$ και επιστρέφει το σύνολο $\mathsf{ans}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{cert}(\mathcal{Q}, \mathcal{A}_s)$. Το σύνολο $\mathcal{A}_s \supseteq \mathcal{A}$, ονομάζεται *κορεσμένο* (*saturation*), υπολογίζεται από το σύνολο $\mathcal{T}|_{\mathsf{rl}} \cup \mathcal{A}$ και είναι τέτοιο ώστε $\mathsf{cert}(\mathcal{Q}, \mathcal{A}_s) = \mathsf{cert}(\mathcal{Q}, \mathcal{T}|_{\mathsf{rl}} \cup \mathcal{A})$.

Τα περισσότερα γνωστά συστήματα απάντησης ερωτημάτων, όπως είναι το GraphDB (πρώην OWLim), το Semantic Graph της Oracle, το RDFox, είναι $\mathcal{RL}$ ABox-συστήματα κορεσμού και είναι ιδιαίτερα αποδοτικά. Είναι προφανές ότι αυτή η προσέγγιση εύρεσης απαντήσεων είναι πλήρης μόνο για γλώσσες χαμηλής εκφραστικότητας όπως είναι η $\mathcal{RL}$.

2.4.3 Απάντηση Ερωτημάτων μέσω Επαναγραφής για Ασυνεπείς Βάσεις Γνώσης

Ένα αρκετά συχνό φαινόμενο, ειδικά σε περιπτώσεις που η βάση γνώσης ανανεώνεται από διαφορετικούς φορείς, είναι ασυνέπεια που μπορεί να εμφανιστεί σε μία βάση γνώσης. Για παράδειγμα, σε μια βαση γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, εκφρασμένη στη γλώσσα DL-Lite$_{\mathcal{R}}$, η ασυνέπεια μπορεί να εμφανιστεί για τους εξής λόγους:

- $\mathcal{T} \models A \sqsubseteq \neg A$ και υπάρχει $A(a) \in \mathcal{A}$,
- $\mathcal{T} \models P \sqsubseteq \neg P^-$ ή $\mathcal{T} \models \exists P \sqsubseteq \neg \exists P^-$ και υπάρχει $P(a, a) \in \mathcal{A}$,
- Υπάρχει αρνητικό αξίωμα υπαγωγής που συνεπάγεται από το $\mathcal{T}$ και το $\mathcal{A}$ είναι ασυνεπές με βάση το αξίωμα αυτό. Για παράδειγμα, αν $\{A(d), P(d, c)\} \subseteq \mathcal{A}$ και $\mathcal{T} \models A \sqsubseteq \neg \exists P$.

Στις ερευνητικές περιοχές της Λογικής, της Τεχνητής Νοημοσύνης και των Βάσεων Δεδομένων, έχουν προταθεί διάφοροι τρόποι αντιμετώπισης της ασυνέπειας σε μία βάση γνώσης [16]. Στην εργασία αυτή, ακολουθούνται δύο βασικές προσεγγίσεις που προέρχονται από τον χώρο των ΒΔ. Με την πρώτη προσέγγιση η βάση διορθώνεται εφαρμόζοντας τις ελάχιστες δυνατές αλλαγές ώστε να εξασφαλιστεί η συνέπεια. Με τη δεύτερη προσέγγιση εξάγονται συνεπείς απαντήσεις παρά την ασυνέπεια της βάσης, η προσέγγιση αυτή καλείται εύρεσης συνεπών απαντήσεων σε βάσεις δεδομένων (consistent query answering in databases) [33]. Βασικό εργαλείο και των δύο προσεγγίσεων είναι η έννοια της *διόρθωσης της βάσης δεδομένων* (*database repair*). Είναι φανερό ότι η διόρθωση μίας βάσης δεδομένων δεν είναι μοναδική, επομένως κάτι είναι ορθό στην ΒΔ μόνο αν είναι ορθό σε όλες τις διορθώσεις της. Αξιοποιώντας την έννοια της διόρθωσης, το πρόβλημα εύρεσης απαντήσεων σε ασυνεπείς βάσεις γνώσεις ανάγεται στο πρόβλημα εύρεσης απαντήσεων σε κάθε διόρθωση.

Επεκτείνοντας την προσέγγιση αυτή, οι Lembo et al. [78] εισήγαγαν την έννοια του *Διορθωμένου ABox* (*ABox Repair (AR)*):

Ορισμός 6. Έστω $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ μία βάση γνώσης. Μία *ABox διόρθωση* (*ABox repair (AR)*) του $\mathcal{K}$ είναι ένα σύνολο $\mathcal{A}'$ ισχυρισμών τέτοιο ώστε:

- $\mathcal{A}' \subseteq \mathcal{A}$
- $Mod(\langle \mathcal{T}, \mathcal{A}' \rangle) \neq \emptyset$
- δεν υπάρχει $\mathcal{A}''$ τ.ω. $\mathcal{A}' \subsetneq \mathcal{A}'' \subseteq \mathcal{A}$ και $Mod(\langle \mathcal{T}, \mathcal{A}'' \rangle) \neq \emptyset$

Συμβολίζουμε με $\mathcal{A}_R(\mathcal{K})$ το σύνολο των ABox διορθώσεων του $\mathcal{K}$.

Διαισθητικά, μία ABox διόρθωση είναι ένα μέγιστο υποσύνολο του ABox συνεπές με το TBox.

Παράδειγμα 7. Έστω η βάση γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$, όπου $\mathcal{T}, \mathcal{A}$:

$$\mathcal{T} \;=\; \{A \sqsubseteq \neg A', B \sqsubseteq A\}, \qquad \mathcal{A} \;=\; \{A(a), A'(a), B(b)\}$$

Τότε, υπάρχουν οι εξής ABox διορθώσεις:

$$r_1 \;=\; \{A(a), B(b)\}, \qquad r_2 \;=\; \{A'(a), B(b)\}$$

συνεπώς, $\mathcal{A}_R(\mathcal{K}) = \{r_1, r_2\}$. $\diamondsuit$

Με βάση τον ορισμό της ABox διόρθωσης, ορίζεται το μοντέλο της ABox διόρθωσης.

Ορισμός 8 ([120]). Έστω η DL βάση γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. Μία ερμηνεία $\mathcal{I}$ είναι ένα *μοντέλο μίας ABox διόρθωσης* (*ABox repair model*), ή απλά ένα *AR-μοντέλο* (*AR-model*), του $\mathcal{K}$, αν υπάρχει $\mathcal{A}' \in \mathcal{A}_R(\mathcal{K})$ τέτοιο ώστε $I \models \langle \mathcal{T}, \mathcal{A}' \rangle$. Το σύνολο των AR-μοντέλων συμβολίζεται ως $\mathcal{A}_R - Mod(K)$.

Η παρακάτω έννοια της συνεπούς συνεπαγωγής (consistent entailment) αποτελεί γενίκευση του του ορισμού της συνεπαγωγής για τη σημασιολογία της ABox διόρθωσης.

Ορισμός 9 ([120]). Έστω η DL βάση γνώσης $\mathcal{K}$ και έστω ϕ μία πρόταση της ΛΠΤ. Λέμε ότι η ϕ *AR-συνεπάγεται* (*AR-entailed*) από το $\mathcal{K}$ και συμβολίζεται με $\mathcal{K} \models_{AR} \phi$, αν $I \models \phi$ για κάθε $I \in \mathcal{A}_R - Mod(\mathcal{K})$

Όπως αναφέρουμε και στην υποενότητα "Ζητήματα Πολυπλοκότητας" που ακολουθεί, η AR-συνεπαγωγή μίας ΣΣΕ είναι δυσεπίλυτο (intractable) πρόβλημα ακόμα και για γλώσσες πολύ χαμηλής εκφραστικότητας. Για τον λόγο αυτό, οι Lembo et al. εισήγαγαν μία προσέγγιση της AR-σημασιολογίας, την IAR-σημασιολογία.

Ορισμός 10 ([120]). Έστω η DL βάση γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. Η *Τομή των ABox Διορθώσεων* (*Intersection ABox Repair-IAR*) για το $\mathcal{K}$ είναι το σύνολο ισχυρισμών $\mathcal{A}'$ τέτοιο ώστε $\mathcal{A}' = \bigcap_{\mathcal{A}_I \in \mathcal{A}_R(\mathcal{K})} \mathcal{A}_I$. Συμβολίζουμε το σύνολο που περιέχει την τομή των ABox διορθώσεων για το $\mathcal{K}$ με $\mathcal{A}_I(\mathcal{K})$.

Το *IAR-μοντέλο* (*IAR-model*) και η *IAR-συνεπαγωγή* (*IAR-entailement*) ορίζονται αντίστοιχα όπως στους Ορισμούς 8, 9.

Παράδειγμα 11. Έστω η ΒΓ $\mathcal{K}$ του Παραδείγματος 7. Τότε, $\mathcal{A}_I(\mathcal{K}) = r_1 \cap r_2 = \{B(b)\}$. Έστω τώρα μία δεύτερη ΒΓ $\mathcal{K}' = \langle \mathcal{T}', \mathcal{A}' \rangle$, όπου $\mathcal{T}' = \mathcal{T}$ και $\mathcal{A}' = \mathcal{A} \cup \{A(b)\}$. Τότε, $\mathcal{A}_I(\mathcal{K}') = \mathcal{A}_I(\mathcal{K}) \cup \{A(b)\}$. $\diamondsuit$

Στο προηγούμενο παράδειγμα, υποδεικνύεται ότι ένα μειονέκτημα της IAR σημασιολογίας είναι ότι εξαρτάται από τα αξιώματα που εμφανίζονται στην οντολογία. Συγκεκριμένα, είναι φανερό ότι ο επιπλέον ισχυρισμός $A(b)$ συνεπάγεται από το $\mathcal{T}$ και το $\{B(b)\}$, που είναι ένα υποσύνολο του $\mathcal{A}$ συνεπές ως προς το $\mathcal{T}$. Επομένως, το αναμενόμενο θα ήταν το $A(b)$ να ανήκει και στο $\mathcal{A}_I(\mathcal{K})$, αλλά όπως φαίνεται αυτό δεν ισχύει. Η "δυσλειτουργία" αυτή της IAR σημασιολογίας οδήγησε στην ανάγκη εισαγωγής της Τομή Διορθώσεων του Κλειστού ABox (Intersection Closed ABox Repair -ICAR) [120]. Με την ICAR διόρθωση υπολογίζονται πρώτα όλοι οι ισχυρισμοί που μπορούν να συμπεραχθούν από τη ΒΓ και στη συνέχεια αφαιρούνται οι ισχυρισμοί που προκαλούν ασυνέπεια.

Ορισμός 12. Έστω η DL βάση γνώσης $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ και έστω $clc(\mathcal{T}, \mathcal{A}) = \{\alpha \mid$ υπάρχει συνεπές υποσύνολο $S \subseteq \mathcal{A}$ τέτοιο ώστε $\mathcal{T} \cup S \models \alpha\}$. Η *Τομή Διορθώσεων του Κλειστού ABox* (*Intersection Closed ABox Repair (ICAR)*) είναι ένα σύνολο $\mathcal{A}'$ ισχυρισμών τέτοιο ώστε:

- $\mathcal{A}' \subseteq clc(\mathcal{T}, \mathcal{A})$
- $Mod(\langle \mathcal{T}, \mathcal{A}' \rangle) \neq \emptyset$
- δεν υπάρχει $\mathcal{A}''$ τ.ω. $\mathcal{A}' \subsetneq \mathcal{A}'' \subseteq clc(\mathcal{T}, \mathcal{A})$ και $Mod(\langle \mathcal{T}, \mathcal{A}'' \rangle) \neq \emptyset$

Συμβολίζουμε με $\mathcal{A}_C(\mathcal{K})$ την τομή διορθώσεων του κλειστού ABox του $\mathcal{K}$.

Παράδειγμα 13. ΄Έστω οι ΒΓ $\mathcal{K}, \mathcal{K}'$ των Παραδειγμάτων 7,11.
Τότε, $clc(\mathcal{T}, \mathcal{A}) = clc(\mathcal{T}, \mathcal{A}') = \mathcal{A} \cup \{A(b)\}$ και $\mathcal{A}_C(\mathcal{K}) = \mathcal{A}_C(\mathcal{K}') = \{A(b), B(b)\}$. $\diamondsuit$

Το παρακάτω λήμμα είναι ισχύει προφανώς.

Λήμμα 14. *Έστω η DL βάση γνώσης $\mathcal{K}$. Ισχύει ότι $\mathcal{A}_I(\mathcal{K}) \subseteq \mathcal{A}_C(\mathcal{K})$.*

Σε αντιστοιχία με τον Ορισμό 1, ορίζουμε τις IAR και ICAR-απαντήσεις του $\mathcal{Q}$ με βάση τη ΒΓ $\mathcal{K}$.

Ορισμός 15. Έστω $\mathcal{Q}$ ένα ΣΕ και έστω $\mathcal{A}_I$ το IAR του $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. Η πλειάδα $\vec{a}$ είναι μία *IAR-απάντηση* (*IAR-answer*) του $\mathcal{Q}$ με βάση το $\mathcal{K}$ αν $\vec{a} \in \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}_I)$. Συμβολίζουμε με $\mathsf{cert}^{\mathsf{ia}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A})$ όλες τις IAR-απαντήσεις του $\mathcal{Q}$ με βάση το $\mathcal{K}$.

Ορισμός 16. Έστω $\mathcal{Q}$ ένα ΣΕ και έστω $\mathcal{A}_C$ το ICAR του $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. Η πλειάδα $\vec{a}$ είναι μία *ICAR-απάντηση* (*ICAR-answer*) του $\mathcal{Q}$ με βάση το $\mathcal{K}$ αν $\vec{a} \in \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}_C)$. Συμβολίζουμε με $\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A})$ όλες τις ICAR-απαντήσεις του $\mathcal{Q}$ με βάση το $\mathcal{K}$.

Στη διατριβή αυτή θα μας απασχολήσει η μεθοδολογία που έχουν προτείνει οι Lembo et al. [76], [77] για τον υπολογισμό των IAR-/ICAR-απαντήσεων σε DL-Lite οντολογίες. Έστω $\mathcal{R}$ μία ΣΣΕ επαναγραφή ενός ερωτήματος $\mathcal{Q}$ με βάση το TBox $\mathcal{T}$. Αφού κατα τον υπολογισμό του $\mathcal{R}$ αφαιρούνται όλα τα αρνητικά αξιώματα υπαγωγής, κάθε ερώτημα στο $\mathcal{R}$ πρέπει να επεκταθεί με ένα σύνολο από αρνητικά άτομα τα οποία αποτρέπουν τον υπολογισμό μη-έγκυρων απαντήσεων. Συμβολίζουμε τη διαδικασία αυτή με $\mathsf{ref}(\mathcal{R}, \mathcal{T})$. Στη συνέχεια, παρουσιάζουμε περιληπτικά τον αλγόριθμο που προτείνουν οι Lembo et al. [77] για τον υπολογισμο του ref περιορισμένη για DL-Lite$_\mathcal{R}$ TBoxes.

Έστω $\mathcal{T}^+$ το σύνολο των θετικών υπαγωγών του $\mathcal{T}$. Ο αλγόριθμος αρχικά χρησιμοποιεί τη συνάρτηση ϕ που συνδέει τα αρνητικά αξιώματα επαγωγής με ενώσεις από Boolean ΣΕ (ΕΒΣΕ).

$$\begin{aligned} \phi(B_1 \sqsubseteq \neg B_2) = &\quad \exists x.\gamma(B_1, x) \wedge \gamma(B_2, x) \\ \phi(R_1 \sqsubseteq \neg R_2) = &\quad \exists x_1, x_2.\eta(R_1, x_1, x_2) \wedge \eta(R_2, x_1, x_2) \end{aligned}$$

Στον παραπάνω ορισμό τα x, x_1, x_2 είναι μεταβλητές. Στη συνέχεια, το y_{new} θα συμβολίζει μία νέα μεταβλητή, δηλαδή μία μεταβλητή που δεν εμφανίζεται στην ΕΒΣΕ.

$$\gamma(B, x) = \begin{cases} A(x) & \text{if } B = A, \\ \exists y_{new}.P(x, y_{new}) & \text{if } B = \exists P, \\ \exists y_{new}.P(y_{new}, x) & \text{if } B = \exists P^- \end{cases}$$

$$\eta(R, x_1, x_2) = \begin{cases} P(x_1, x_2) & \text{if } R = P, \\ P(x_2, x_1) & \text{if } R = P^-. \end{cases}$$

$$\mathsf{UnsatQuery}(\mathcal{T}) = \quad \bigvee\nolimits_{\alpha \in \mathcal{T}} \mathsf{rew}(\phi(\alpha), \mathcal{T}^+)$$

όπου η συνάρτηση $\mathsf{rew}(x, y)$ υπολογίζει την ΣΣΕ-επαναγραφή του x μ.β.τ. y. Εξαιτίας της δομής της DL-Lite$_\mathcal{R}$ η ΣΣΕ-επαναγραφή είναι μοναδική (μόντουλο την μετονομασία των μεταβλητών). Η συνάρτηση UnsatQuery υπολογίζει την επαναγραφή κάθε ερωτήματος που θα φτιαχτεί με τη συνάρτηση ϕ από τα αρνητικά αξιώματα υπαγωγής του $\mathcal{T}$ με βάση το $\mathcal{T}^+$. Η ένωση των επαναγραφών αυτών κωδικοποιεί όλα τα ξένα αξιώματα που συμπεραίνονται από το $\mathcal{T}$. Για ένα άτομο $A(\vec{x})$ ορίζεται το σύνολο των ξένων αξιωμάτων στα οποία συμμετέχει ως εξής:

$$\mathsf{S}(A(\vec{x}), \mathcal{T}) = \{q\sigma \mid q \in \mathsf{UnsatQuery}(\mathcal{T}) \text{ και υπάρχει } \sigma \text{ τ.ω. } A\sigma \in q\}$$

Τώρα μπορεί να οριστεί η συνάρτηση ref. Έστω ένα ερώτημα $\mathcal{Q} = \exists z_1, \ldots, z_k. \bigwedge_{i=1}^{n} A_i(t_i^1) \wedge \bigwedge_{i=1}^{m} P_i(t_i^2, t_i^3)$. Τότε το ref θα είναι:

$$\begin{aligned}\mathsf{ref}(\mathcal{Q}, \mathcal{T}) = \ & \exists z_1, \ldots, z_k. \\ & \bigwedge_{i=1}^{n} A_i(t_i^1) \wedge \neg(\textstyle\bigvee_{q \in \mathsf{S}(A_i(t_i^1), \mathcal{T})} q) \\ & \bigwedge_{i=1}^{m} P_i(t_i^2, t_i^3) \wedge \neg(\textstyle\bigvee_{q \in \mathsf{S}(P_i(t_i^2, t_i^3), \mathcal{T})} q)\end{aligned}$$

Επίσης, ορίζεται η συνάρτηση $\mathsf{ref}^{\mathsf{ic}}$:

$$\begin{aligned}\mathsf{ref}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T}) = \ & \exists z_1, \ldots, z_k. \\ & \bigwedge_{i=1}^{n} \mathsf{rew}(A_i(t_i^1), \mathcal{T}|_{\mathsf{rl}}) \wedge \neg(\textstyle\bigvee_{q \in \mathsf{S}(A_i(t_i^1), \mathcal{T})} q) \\ & \bigwedge_{i=1}^{m} \mathsf{rew}(P_i(t_i^2, t_i^3), \mathcal{T}|_{\mathsf{rl}}) \wedge \neg(\textstyle\bigvee_{q \in \mathsf{S}(P_i(t_i^2, t_i^3), \mathcal{T})} q)\end{aligned}$$

όπου εδώ για τον υπολογισμό της επαναγραφής rew δεν χρησιμοποιούνται αξιώματα του $\mathcal{T}$ που έχουν υπαρξιακό ποσοδείκτη στα δεξιά.

Έστω $\mathcal{R}$ μία ΣΣΕ-επαναγραφή του $\mathcal{Q}$ μβτ $\mathcal{T}^+$. Τότε, μία *IAR-επαναγραφή* (*IAR-rewriting*) του $\mathcal{Q}$ μβτ $\mathcal{T}$ ορίζεται ως: $\mathcal{R}_i = \bigvee_{q \in \mathcal{R}} \mathsf{ref}(q, \mathcal{T})$ και μία *ICAR-επαναγραφή* (*ICAR-rewriting*) του $\mathcal{Q}$ μβτ $\mathcal{T}$ ορίζεται ως: $\mathcal{R}_c = \bigvee_{q \in \mathcal{R}} \mathsf{ref}^{\mathsf{ic}}(q, \mathcal{T})$.

Παράδειγμα 17. Έστω η ΒΓ $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ του Παραδείγματος 7, έστω $\mathcal{K}'' = \langle \mathcal{T}'', \mathcal{A} \rangle$, με $\mathcal{T}'' = \mathcal{T} \cup \{B \sqsubseteq \neg B'\}$, και έστω το ΣΕ $Q(x) \leftarrow A(x)$. Τότε, σύμφωνα με τα προηγούμενα, $\mathsf{ref}(\mathcal{Q}, \mathcal{T}'') = \{Q(x) \leftarrow A(x) \wedge \neg A'(x), Q(x) \leftarrow B(x) \wedge \neg A'(x) \wedge \neg B'(x)\}$. Επίσης, $\mathsf{ref}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T}'') = \mathsf{ref}(\mathcal{Q}, \mathcal{T}) \cup \{Q(x) \leftarrow B(x) \wedge \neg A'(x)\}$. ♢

Ισχύει το παρακάτω λήμμα:

Λήμμα 18 ([76])**.** *Έστω ένα DL-Lite*$_\mathcal{R}$ *TBox* $\mathcal{T}$*, ένα ΣΕ* $\mathcal{Q}$*, και μία επαναγραφή UCQ* $\mathcal{R}$ *για τα* $\mathcal{T}, \mathcal{Q}$*. Τότε για κάθε* $\mathcal{A}$*, ισχύει ότι* $\mathsf{cert}^{\mathsf{ia}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{cert}(\mathsf{ref}(\mathcal{R}, \mathcal{T}), \mathcal{A})$.

Ζητήματα Πολυπλοκότητας

Για την AR-σημασιολογία, το πρόβλημα της συνεπαγωγής ενός ερωτήματος από μία ΒΓ στην DL-Lite$_{core}$, που είναι η απλούστερη γλώσσα της οικογένειας της DL-Lite (υποσύνολο της DL-Lite$_\mathcal{R}$ που δεν περιέχει αξιώματα μεταξύ ρόλων), είναι coNP-complete ως προς τα δεδομένα [75]. Επίσης, στην $\mathcal{EL}_\perp$ είναι coNP-hard ως προς τα δεδομένα. Στην $\mathcal{ALC}$ και στην $\mathcal{SHIQ}$ είναι Π_2^p ως προς τα δεδομένα [115].

Για την IAR-σημασιολογία, στην DL-Lite$_\mathcal{R}$ έχει πολυπλοκότητα AC^0 [76]. Ωστόσο, για τις $\mathcal{EL}_\perp$, $\mathcal{ALC}$, $\mathcal{SHIQ}$ έχει την ίδια πολυπλοκότητα με την AR-σημασιολογία [115].

Για την ICAR-σημασιολογία, στην DL-Lite$_\mathcal{R}$ έχει επίσης πολυπλοκότητα AC^0. Στην $\mathcal{EL}_\perp$ έχει $\Delta_2^p[O(\log n)]$ και στις $\mathcal{ALC}$, $\mathcal{SHIQ}$ έχει $\Delta_3^p[O(\log n)]$ ως προς τα δεδομένα. Όπου, $\Delta_2^p[O(\log n)]$ (αντ. $\Delta_3^p[O(\log n)]$) είναι η κλάση των προβλημάτων που λύνονται σε πολυωνυμικό χρόνο με $O(\log n)$ κλήσεις σε ένα NP-μαντείο (αντ. Σ_2^p-μαντείο) [115].

Λόγω της δυσεπιλυσιμότητας του προβλήματος της συνεπαγωγής ενός ερωτήματος από μία ΒΓ για τις σημασιολογίες IAR και ICAR σε γλώσσες λίγο πιο εκφραστικές από την DL-Lite$_\mathcal{R}$, όπως είναι η $\mathcal{EL}_\perp$, ο Rosati [115] εισήγαγε τη γλώσσα $\mathcal{EL}_{\perp nr}$, της οποίας το συντακτικό περιγράφεται στην Ενότητα 2.2.3. Στην $\mathcal{EL}_{\perp nr}$, για την IAR-σημασιολογία το πρόβλημα είναι PTIME-complete ενω για την ICAR-σημασιολογία παραμένει coNP-hard ως προς τα δεδομένα.

2.5 Κατευθυνόμενοι Γράφοι

Ένας κατευθυνόμενος γράφος είναι ένα ζεύγος $\mathcal{G} = \langle V, E \rangle$, με V ένα μη-κενό *σύνολο κόμβων* και $E \subseteq V \times V$ ένα, ενδεχομένως, κενό σύνολο από διατεταγμένα ζεύγη κόμβων, τις *κατευθηνόμενες ακμές ή βέλη* που ορίζονται πάνω στο V και λέγεται *σύνολο ακμών*. Στη συνέχεια της διατριβής αυτής κάνουμε συχνά κατάχρηση του συμβολισμού και χρησιμοποιούμε τον γράφο $\mathcal{G}$ για να αναφερθούμε στο σύνολο σχέσεων του γράφου. Στην περίπτωση αυτή το V είναι ακριβώς το σύνολο των παραμέτρων της E. Συνεπώς, προσθέτοντας ένα ζευγάρι $\langle a, b \rangle$ στον $\mathcal{G}$ προκύπτει ο νέος γράφος $\mathcal{G} = \langle V \cup \{a, b\}, E \cup \{\langle a, b \rangle\} \rangle$. Αν ισχύει $\langle a, b \rangle \in V$, τότε λέμε ότι ο κόμβος b είναι *παιδί* (*child*) του κόμβου a.

Μονοπάτι (*path*) από έναν κόμβο σε έναν άλλο ενός γράφου ονομάζεται μια ακολουθία κόμβων, όπου κάθε κόμβος της ακολουθίας συνδέεται με τον επόμενό του μέσω ακμής. Αν υπάρχει ένα μονοπάτι μεταξύ δύο κόμβων a και b, τότε οι κόμβοι αυτοί λέγονται *προσπελάσιμοι* (*reachable*) ο ένας από τον άλλον. Λέμε ότι ένας γράφος είναι *μεταβατικά μειωμένος* (*transitively reduced*), αν έχει τις λιγότερες δυνατές ακμές τέτοιες ώστε η προσπελασιμότητα μεταξύ οποιωνδήποτε δύο κόμβων παραμένει ίδια. Δηλαδή, αν υπάρχει μονοπάτι από τον κόμβο a στον κόμβο b στο $\mathcal{G}$ θα πρέπει να υπάρχει μονοπάτι από το a στο b και στον μεταβατικά μειωμένο γράφο του $\mathcal{G}$.

Μέρος II

Επαναγραφή Ερωτημάτων σε Συνεπείς Εξελισσόμενες Βάσεις Γνώσης

Κεφάλαιο 3

Επαναγραφή Ερωτημάτων σε Συνεπείς Εξελισσόμενες Βάσεις Γνώσης

Στο κεφάλαιο αυτό μελετάμε το πρόβλημα του υπολογισμού μιας επαναγραφής ενός συζευκτικού ερωτήματος για μία οντολογία οι οποία έχει τροποποιηθεί, δηλαδή έχουν προστεθεί ή έχουν αφαιρεθεί αξιώματα υπαγωγής από αυτήν. Θεωρούμε ότι έχει ήδη υπολογιστεί μία επαναγραφή για το ερώτημα αυτό και την αρχική οντολογία και δείχνουμε πώς μπορούμε να υπολογίσουμε μια νέα επαναγραφή για το ίδιο ερώτημα αλλά με τροποποιημένη την αρχική οντολογία. Ο στόχος είναι να αξιοποιήσουμε την αρχική επαναγραφή ώστε να μην επαναληφθούν υπολογισμοί που έχουν ήδη πραγματοποιηθεί.

Έστω ένα ερώτημα $\mathcal{Q}$, μία οντολογία $\mathcal{T}$ και ένα σύνολο αξιωμάτων $\mathcal{T}'$. Αρχικά, στην Ενότητα 3.2 μελετάμε το πρόβλημα υπολογισμού μίας επαναγραφής ενός ερωτήματος $\mathcal{Q}$ με βάση την οντολογία $\mathcal{T} \cup \mathcal{T}^{+}$, δεδομένης μίας επαναγραφής $\mathcal{R}$ του $\mathcal{Q}$ με βάση την $\mathcal{T}$. Καταρχήν, δείχνουμε ότι δεν αρκεί η αρχική επαναγραφή για τον υπολογισμό της νέας επαναγραφής, αλλά απαιτείται το σύνολο $\mathcal{S}$ ($\mathcal{S} \supseteq \mathcal{R}$) που υπολογίστηκε κατά τον υπολογισμό του $\mathcal{R}$. Στη συνέχεια, δείχνουμε ότι για τον υπολογισμό μίας νέας επαναγραφής χρειάζεται η εφαρμογή νέων συμπερασμών με τη χρήση κάποιου συστήματος επαναγραφής. Στο πλαίσιο αυτό, προτείνουμε έναν νέο αλγόριθμο και αποδεικνύουμε την ορθότητά του.

Στη συνέχεια, στην Ενότητα 3.3 μελετάμε το πρόβλημα υπολογισμού μίας επαναγραφής ενός ερωτήματος $\mathcal{Q}$ με βάση την οντολογία $\mathcal{T} \cup \mathcal{T}^{-}$, δεδομένης μίας επαναγραφής $\mathcal{R}$ του $\mathcal{Q}$ με βάση την $\mathcal{T}$. Στην αρχή, παρουσιάζουμε έναν γενικό αλγόριθμο ο οποίος, όπως στη περίπτωση των οντολογιών που επεκτείνονται, απαιτεί το σύνολο $\mathcal{S}$ που υπολογίστηκε κατά τον υπολογισμό του $\mathcal{R}$ και εφαρμόζει κάποιους επιπλέον συμπερασμούς. Στη συνέχεια, στην Ενότητα 3.4 μελετάμε αν και υπό ποιες συνθήκες είναι εφικτός ο υπολογισμός μίας νέας επαναγραφής χωρίς την εφαρμογή νέων συμπερασμών. Τέλος, στην Ενότητα 3.5 βελτιστοποιούμε τους προηγούμενους αλγορίθμους με τη χρήση γράφων. Για κάθε περίπτωση προτείνουμε έναν νέο αλγόριθμο τον οποίο παρουσιάζουμε αναλυτικά και αποδεικνύουμε την ορθότητα του.

Καθώς τα περισσότερα συστήματα επαναγραφής διαχειρίζονται προτάσεις και όχι γενικευμένα αξιώματα, στο κεφάλαιο υποθέτουμε ότι οι οντολογίες είναι σε προτασιακή μορφή. Είναι σημαντικό, τέλος, να σημειωθεί ότι υποθέτουμε ότι η ΒΓ είναι συνεπής και συνεπώς αυτό μας επιτρέπει να παραβλέψουμε τα αρνητικά αξιώματα υπαγωγής που μπορεί να περιέχει η οντολογία.

3.1 Αλγόριθμοι Επαναγραφής βασισμένοι στη Μέθοδο της Ανάλυσης

Τα τελευταία χρόνια έχει προταθεί στη βιβλιογραφία ένα πλήθος αλγορίθμων επαναγραφής (για παράδειγμα, [116, 103, 28, 113, 41, 104, 137]), πολλοί από τους οποίους χρησιμοποιώντας το σύστημα συμπερασμού Γ που έχουν ορίσει, υπολογίζουν μία επαναγραφή βαθμιαία, με την έννοια ότι ξεκινούν από το σύνολο $\mathcal{Q} \cup \mathcal{T}$ και σε κάθε βήμα εφαρμόζεται ένας νέος συμπερασμός μέχρι τελικά να υπολογιστεί το σύνολο $(\mathcal{Q} \cup \mathcal{T})_\Gamma$. Στην εργασία αυτή λέμε ότι οι αλγόριθμοι αυτοί είναι βασισμένοι στη μέθοδο της ανάλυσης.

Ορισμός 19. Ένας *αλγόριθμος επαναγραφής* (*rewriting algorithm*) για μία γλώσσα οντολογίας είναι μία υπολογίσιμη συνάρτηση τέτοια ώστε, για κάθε $\mathcal{T} \in \mathcal{L}$ και κάθε ΣΕ $\mathcal{Q}$ υπολογίζει σε ένα πεπερασμένο αριθμό βημάτων μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$. Ο αλγόριθμος είναι *βασισμένος στην μέθοδο της ανάλυσης* (*resolution based*) αν βασίζεται σε ένα σύστημα συμπερασμού Γ και ένα κριτήριο περιττότητας $\mathcal{D}$ για να υπολογίσει μια σειρά $\mathcal{R}_1, \ldots, \mathcal{R}_n$ από σύνολα προτάσεων τέτοια ώστε:

- $\mathcal{R}_1 = \mathcal{Q} \cup \mathcal{T}$,
- για κάθε $i \in \{1, \ldots, n-1\}$, $\mathcal{R}_{i+1} = \mathcal{D}(\mathcal{R}_i \cup \{c'\})$, όπου το c' είναι τ.ώ. $\langle c_1, \ldots, c_m, c' \rangle$ με $c_1, \ldots, c_m \in \mathcal{R}_i$, είναι ένας συμπερασμός του Γ,
- για κάθε $c_1, \ldots, c_m \in \mathcal{R}_m$ και κάθε συμπερασμό $\langle c_1, \ldots, c_m; c' \rangle$ του Γ, το c' είναι ισοδύναμο ως προς την αλλαγή μεταβλητών με κάποιο $c'' \in \mathcal{R}_n$ και
- το μέγιστο συναρτησιακά ελεύθερο υποσύνολο του $\mathcal{R}_n$ είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$.

Στο κεφάλαιο αυτό, υποθέτουμε ότι η αρχική επαναγραφή $\mathcal{R}$ έχει υπολογιστεί από κάποιον αλγόριθμο βασισμένο στη μέθοδο της ανάλυσης που χρησιμοποιεί ένα ορθό και πλήρες σύστημα συμπερασμού Γ και ένα κριτήριο περιττότητας $\mathcal{D}$. Με την παραδοχή αυτή περιλαμβάνεται ένας μεγάλος αριθμός αλγορίθμων επαναγραφής ερωτημάτων, αφού οι περισσότεροι είτε είναι βασισμένοι ή μπορούν να αναχθούν στη μέθοδο (υπέρ-)ανάλυσης. Για παράδειγμα, ο αλγόριθμος του συστήματος Requiem είναι πρακτικά η εφαρμογή του αλγορίθμου επαναγραφής βασισμένου στην ανάλυση ενώ ταυτόχρονα αποτελεί πυρήνα των συστημάτων Blackout και Kyrie. Το Mastro βασίζεται στον αλγόριθμο CGLLR, το οποίο επίσης μπορεί να αναχθεί στην μέθοδο της ανάλυσης. Τέλος τα συστήματα Rapid, Nyaya και Clipper, που υπολογίζουν αποδοτικά επαναγραφές για πιο εκφραστικές ΠΛ, βασίζονται στη μέθοδο της (υπερ-)ανάλυσης.

3.2 Επαναγραφή Ερωτημάτων πάνω σε Οντολογίες που Επεκτείνονται

Στην ενότητα αυτή θα παρουσιάσουμε έναν αλγόριθμο υπολογισμού επαναγραφής ενός ερωτήματος $\mathcal{Q}$ με βάση μία οντολογία $\mathcal{T}$ η οποία επεκτείνεται με ένα σύνολο αξιωμάτων $\mathcal{T}^+$. Θεωρούμε ότι έχει ήδη υπολογιστεί ένα σύνολο $\mathcal{R}$ για τον υπολογισμό μίας επαναγραφής του $\mathcal{Q}$ με βάση την $\mathcal{T}$. Ο στόχος μας είναι να αξιοποιήσουμε στο μέγιστο δυνατό την πληροφορία του $\mathcal{R}$ και να υπολογίσουμε με τον τρόπο αυτό μόνο τα νέα στοιχεία που εισάγονται εξαιτίας της προσθήκης του συνόλου $\mathcal{T}^+$.

Στο παρακάτω παράδειγμα παρουσιάζονται οι βασικές ιδέες του αλγορίθμου που θα παραθέσουμε στη συνέχεια.

Παράδειγμα 20. Έστω μια οντολογία $\mathcal{T}$ που αποτελείται από τις εξής προτάσεις:

$$\begin{array}{rcrcl} c_1 & = & P(x,y) & \leftarrow & S(x,y) \\ c_2 & = & S(x,f(x)) & \leftarrow & A(x) \end{array}$$

και το ερώτημα $\mathcal{Q} = Q(x) \leftarrow R(x,y) \wedge P(x,y)$. Έστω επίσης το σύστημα επαναγραφής βασισμένο στη μέθοδο της ανάλυσης, Requiem, το οποίο υπολογίζει τον κορεσμό του συνόλου εισόδου, $\mathcal{Q} \cup \mathcal{T}$, με τη μέθοδο της ανάλυσης με ελεύθερη επιλογή (συμβολίζουμε με Γ^{Rq} το σύστημα συμπερασμού που χρησιμοποιείται από το Requiem). Αρχικά, σύμφωνα με το Γ^{Rq} οι προτάσεις c_1, c_2 θα αναλυθούν (ενοποιώντας τα άτομα $S(x,y)$ και $S(x,f(x))$) παράγοντας την πρόταση c_3:

$$\frac{c_1 \quad c_2}{c_3} \quad \text{όπου} \quad c_3 = P(x,f(x)) \leftarrow A(x) \tag{3.1}$$

και στην συνέχεια θα υλοποιηθεί ο επόμενος συμπερασμός:

$$\frac{\mathcal{Q} \quad c_3}{\mathcal{Q}'} \quad \text{όπου} \quad \mathcal{Q}' = Q(x) \leftarrow R(x,f(x)) \wedge A(x) \tag{3.2}$$

Σύμφωνα με το Γ^{Rq}, δεν μπορούν να παραχθούν άλλες προτάσεις, συνεπώς ο κορεσμός του συνόλου $\mathcal{T} \cup \mathcal{Q}$ από το Γ^{Rq} θα αποτελείται από το σύνολο των προτάσεων $\mathcal{R}_n = \{c_3, \mathcal{Q}, \mathcal{Q}'\} \cup \mathcal{T}$. Στη συνέχεια, το Requiem θα επιστρέψει την επαναγραφή $\mathcal{R} = \{\mathcal{Q}, c_1\}$ του $\mathcal{Q}$ με βάση την $\mathcal{T}$ εξάγοντας το συναρτησιακά-ελεύθερο υποσύνολο του $\mathcal{R}_n$.

Ας υποθέσουμε τώρα ότι η οντολογία $\mathcal{T}$ αναθεωρείται από κάποιον εξειδικευμένο στο πεδίο ο οποίος προσθέτει την πρόταση $c_4 = R(x,y) \leftarrow P(x,y)$, αποκτώντας έτσι την οντολογία $\mathcal{T}' = \mathcal{T} \cup \{c_4\}$. Μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}'$ αποτελείται από το σύνολο $\mathcal{R}' = \{\mathcal{Q}, \mathcal{Q}_1, c_1, c_4\}$, όπου $\mathcal{Q}_1 = Q(x) \leftarrow A(x)$. Οποιοδήποτε υπάρχον συστήμα επαναγραφής θα υπολόγιζε το $\mathcal{R}'$ από την αρχή. Για παράδειγμα, το Requiem με τη χρήση του Γ^{Rq} στην αρχή θα υλοποίησει τους συμπερασμούς (3.1), (3.2), όπως και προηγουμένως, και στη συνέχεια θα εκτελέσει τους συμπερασμούς:

$$\frac{c_4 \quad c_3}{c_5} \quad \text{όπου} \quad c_5 = R(x, f(x)) \leftarrow A(x),$$
$$\frac{\mathcal{Q} \quad c_5}{\mathcal{Q}''} \quad \text{όπου} \quad \mathcal{Q}'' = Q(x) \leftarrow A(x) \wedge P(x, f(x)),$$
$$\frac{\mathcal{Q}' \quad c_5}{Q(x) \leftarrow A(x)}, \text{και} \qquad \frac{\mathcal{Q}'' \quad c_3}{Q(x) \leftarrow A(x).}$$

Όπως είναι φανερό, το $\mathcal{R}'$ μπορεί να υπολογιστεί είτε από την αρχή, επαναλαμβάνοντας έτσι τους συμπερασμούς (3.1) και (3.2), είτε "συνεχίζοντας" από το $\mathcal{R}_n$ εφαρμόζοντας λιγότερους συμπερασμούς. $\diamondsuit$

Μία σημαντική παρατήρηση που προκύπτει από το Παράδειγμα 20 είναι ότι δεν αρκεί η αρχική επαναγραφή με την νέα πρόταση για να υπολογιστεί η νέα επαναγραφή. Συγκεκριμένα, μόνο από την αρχική επαναγραφή και τη νέα οντολογία δεν μπορούν να συμπεραχθούν οι ενδιάμεσες προτάσεις c_3, c_5 και $\mathcal{Q}''$ από το $\mathcal{R} \cup \{c_4\}$, ώστε τελικά να συμπεραχθεί και η πρόταση $Q(x) \leftarrow A(x)$ που ανήκει στη νέα επαναγραφή $\mathcal{R}'$. Προφανώς, αν επίσης χρησιμοποιηθεί για τον υπολογισμό τη νέας επαναγραφής η οντολογία $\mathcal{T}$, τότε θα υπολογιστεί μεν το $\mathcal{R}'$, αλλά θα είναι αδύνατο να διακριθούν οι απαραίτητοι από τους μη-απαραίτητους συμπερασμούς.

Συνεπώς, για να μην επαναληφθούν οι συμπερασμοί που έχουν υπολογιστεί για την αρχική επαναγραφή υποθέτουμε ότι δίνεται στην είσοδο ο κορεσμός του συνόλου $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$. Με τον τρόπο αυτό εφαρμόζονται μόνο οι συμπερασμοί που "ενεργοποιούνται" από την προσθήκη των νέων προτάσεων. Αρχικά, θα είναι συμπερασμοί που θα έχουν στις υποθέσεις τους κάποια από τις νέες προτάσεις. Στη συνέχεια όμως τα συμπεράσματα των συμπερασμών αυτών θα μπορέσουν να αναλυθούν περαιτέρω με προτάσεις που ανήκουν στο σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}} \cup \mathcal{T}^+$ ενεργοποιώντας έτσι νέους συμπερασμούς, τα συμπεράσματα των οποίων θα ενεργοποιήσουν επίσης νέους συμπερασμούς κ.ο.κ. Η διαδικασία θα τερματίσει όταν πλέον δεν θα παράγονται νέα συμπεράσματα.

Τέλος, είναι σημαντικό να παρατηρήσουμε ότι η νέα επαναγραφή μπορεί να υπολογιστεί αποδοτικά, μόνο αν αξιοποιήσουμε συστήματα επαναγραφής τα οποία βασίζονται στη μέθοδο της ανάλυσης. Με τον τρόπο αυτό, μπορούμε να θεωρήσουμε ότι το σύνολο $(\mathcal{Q} \cup \mathcal{T})_\Gamma$ έχει υπολογιστεί σε ένα ενδιάμεσο στάδιο και να συνεχίσουμε με την εφαρμογή των νέων συμπερασμών που προκύπτουν από την εισαγωγή του $\mathcal{T}^+$ ώστε να υπολογίσουμε μία τελική επαναγραφή.

3.2.1 Αλγόριθμος

Ο Αλγόριθμος 1 ($\mathsf{Add}_{\Gamma,\mathcal{D}}$) υπολογίζει την επαναγραφή ενός ερωτήματος $\mathcal{Q}$ με βάση μια οντολογία $\mathcal{T}$ που επεκτείνεται από ένα σύνολο προτάσεων $\mathcal{T}^+$. Ο $\mathsf{Add}_{\Gamma,\mathcal{D}}$ δέχεται στην είσοδο το σύνολο προτάσεων $\mathcal{S} = (\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ και ένα σύνολο προτάσεων $\mathcal{T}^+$. Επίσης,

Algorithm 1 $\mathsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{S}, \mathcal{T}^+)$

Input: Τα σύνολα προτάσεων $\mathcal{S}, \mathcal{T}^+$

Output: Μια επαναγραφή για τη νέα οντολογία

```
1: Αρχικοποίησε το U := T⁺
2: while U \ S ≠ ∅ do
3:     if υπάρχει μη-κενό σύνολο προτάσεων U′ ⊆ U και ένα σύνολο προτάσεων
       S′ ⊆ S τ.ω.
       1. ⟨c1, ..., cn; c⟩, με {c1, ..., cn} = U′ ∪ S′, είναι ένας συμπερασμός του Γ και
       2. δεν υπάρχει c′ ∈ U ∪ S που να υπάγει το c then
4:         U := U ∪ {c}
5:     else
6:         S := D(S ∪ U)
7:     end if
8: end while
9: return {c ∈ S | c είναι συναρτησιακά ελεύθερο}
```

χρησιμοποιεί εσωτερικά το ορθό σύστημα συμπερασμού Γ για τον υπολογισμό των νέων συμπερασμών.

Στη γραμμή 1, αρχικοποιεί ένα νέο σύνολο $\mathcal{U}$ το οποίο περιέχει τις προτάσεις το νέο σύνολο αξιωμάτων $\mathcal{T}^+$. Στη συνέχεια, επιλέγει ένα κατάλληλο υποσύνολο του $\mathcal{U}$ και ένα (πιθανά μηδενικό) υποσύνολο του $\mathcal{S}$ για να εφαρμόσει έναν συμπερασμό παράγοντας μία νέα πρόταση με το σύστημα συμπερασμού Γ (γραμμή 3), η οποία, αν δεν υπάγεται από κάποια άλλη υπάρχουσα πρόταση, προστίθεται στο $\mathcal{U}$ (γραμμή 4). Η διαδικασία αυτή επαναλαμβάνεται μέχρι να σταματήσουν να παράγονται καινούριες προτάσεις. Τότε, όλες οι προτάσεις του $\mathcal{U}$ προστίθενται στο $\mathcal{S}$ και από το τελικό σύνολο απομακρύνονται οι περιττές προτάσεις (γραμμή 6). Στο σημείο αυτό, θα τερματίσει ο while-βρόγχος, δίοτι κάθε πρόταση c που θα ανήκει στο $\mathcal{U}$ αλλά όχι στο $\mathcal{S}$ θα υπάγεται από κάποια πρόταση του $\mathcal{S}$. Συνεπώς, από τις ιδιότητες της υπαγωγής κάθε πρόταση που θα παράγεται από το c θα υπάγεται από το κάποια πρόταση του $\mathcal{S}$ και άρα δε θα προστίθενται στο $\mathcal{U}$ για περαιτέρω επεξεργασία. Όταν η διαδικασία αυτή τερματίσει, τότε ο αλγόριθμος επιστρέφει το συναρτησιακά-ελεύθερο υποσύνολο του συνόλου $\mathcal{S}$, το οποίο αποτελεί μία επαναγραφή ως προς τη νέα οντολογία. Ωστόσο, για να χρησιμοποιηθεί το αποτέλεσμα του αλγορίθμου και για μελλονικές επεκτάσεις της οντολογίας, θα πρέπει ο αλγόριθμος να επιστρέψει το $\mathcal{S}$.

Ο έλεγχος υπαγωγής που εφαρμόζεται στη συνθήκη 2 της γραμμής 3 εξασφαλίζει ότι ο while-βρόγχος δεν θα είναι ατέρμονος και ταυτόχρονα βελτιστοποιεί την απόδοση του αλγορίθμου διότι περιορίζει την παραγωγή περιττών προτάσεων. Αν για παράδειγμα, δεν υπήρχε ο έλεγχος αυτός και στη γραμμή 4 προστίθεντο μία πρόταση c για την οποία υπήρχε $c' \in \mathcal{U}$ που να υπάγει την c, τότε κάθε άλλη πρόταση που θα παραγόταν στην συνέχεια από την c θα υπαγόταν από την c' ή από κάποια άλλη πρόταση που θα ήταν παράγωγο της c'.

Αν θεωρήσουμε ότι στο κορεσμένο σύνολο εισόδου $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ έχουν εφαρμοστεί όλοι οι έλεγχοι υπαγωγής, τότε στη γραμμή 6 στην ουσία εφαρμόζεται μόνο οπισθοδρομικός έλεγχος, διότι ο εμπρόσθιος έχει ήδη εφαρμοστεί στη συνθήκη 2 της γραμμής 3. Η απόδοση του αλγορίθμου μπορεί να βελτιωθεί περαιτέρω αν κατα οπισθοδρομικό έλεγχο υπαγωγής αποφύγουμε τους ελέγχους μεταξύ των προτάσεων της αρχικής επαναγραφής και εστιάσουμε μόνο στους υπόλοιπους.

Θα πρέπει να σημειωθεί ότι ο Αλγόριθμος 1 είναι αρκετά γενικός και αφαιρετικός. Για να εφαρμοστεί σε ένα συγκεκριμένο αλγόριθμο επαναγραφής θα πρέπει να προσαρμοστεί στο σύστημα συμπερασμού του. Για παράδειγμα, εφόσον το Requiem χρησιμοποιεί μόνο τη μέθοδο δυαδικής ανάλυσης (Γ^{Rq}), στη γραμμή 3 μπορεί απλά να επιλέγει δύο προτάσεις $c_1 \in \mathcal{U}$ και $c_2 \in \mathcal{S}$ τέτοιες ώστε το $\langle c_1, c_2; c\rangle$ να είναι ένας συμπερασμός του Γ^{Rq}. Αντίστοιχα, το σύστημα συμπερασμού που χρησιμοποιεί το Rapid για την DL-Lite$_R$ έχει μία βασική υπόθεση και το πολύ δύο πλάγιες υποθέσεις, δηλαδή οι συμπερασμοί που εφαρμόζει είναι της μορφής $\langle c, c_1, c_2; c'\rangle$. Συνεπώς, η γραμμή 3 μπορεί να προσαρμοστεί έτσι ώστε να επιλέγει τουλάχιστον μία πρόταση από το $\mathcal{U}$ και τις υπόλοιπες από το $\mathcal{S}$.

Παράδειγμα 21. Έστω μια οντολογία $\mathcal{T}$ που αποτελείται από τις εξής προτάσεις:

$$\begin{aligned} c_1 &= R(x, f(x)) \leftarrow C(x) \\ c_2 &= A(x) \leftarrow B(x) \end{aligned}$$

και το ερώτημα $\mathcal{Q} = Q(x) \leftarrow R(x,y) \wedge A(y)$. Μία επαναγραφή για τα $\mathcal{Q}$, $\mathcal{T}$ είναι η εξής: $\mathcal{R} = \{\mathcal{Q}, \mathcal{Q}_3\}$, όπου $\mathcal{Q}_3 = Q(x) \leftarrow R(x,y) \wedge B(y)$. Έστω επίσης το αξίωμα $c_3 = B(f(x)) \leftarrow C(x)$ και $\mathcal{T}^+ = \{c_3\}$.

Αν το Γ είναι το Γ^{Rq}, τότε $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}} = \{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3\}$, όπου $\mathcal{Q}_1 = Q(x) \leftarrow C(x) \wedge A(f(x))$, $\mathcal{Q}_2 = Q(x) \leftarrow C(x) \wedge B(f(x))$. Συνεπώς, ο $\mathsf{Add}_{\Gamma,\mathcal{D}}$ θα επιλέξει τις δύο προτάσεις $\mathcal{Q}_2, c_3$ και θα εφαρμόσει τον συμπερασμό:

$$\frac{\mathcal{Q}_2 \quad c_3}{\mathcal{Q}_4} \quad \text{όπου} \quad \mathcal{Q}_4 = Q(x) \leftarrow C(x)$$

Αν το Γ είναι το σύστημα συμπερασμού του Rapid, τότε $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}} = \{\mathcal{Q}, \mathcal{Q}_3\}$. Συνεπώς, ο $\mathsf{Add}_{\Gamma,\mathcal{D}}$ θα χρησιμοποιήσει για κύρια υπόθεση το $\mathcal{Q}_3$, για πλάγιες υποθέσεις τα c_1, c_3 και θα εφαρμόσει τον συμπερασμό:

$$\frac{\mathcal{Q}_3 \quad c_1, c_3}{\mathcal{Q}_4}$$

Αν χρησιμοποιηθεί οποιοδήποτε σύστημα από τα δύο, θα υπολογιστεί τελικά η νέα επαναγραφή $\mathcal{R}' = \{\mathcal{Q}, \mathcal{Q}_3, \mathcal{Q}_4\}$. $\diamondsuit$

Η δομή του αλγορίθμου μοιάζει πολύ με τους συνήθεις αλγορίθμους επαναγραφής που βασίζονται στη μέθοδο της ανάλυσης και μπορούν να βρεθούν στη βιβλιογραφία [9]. Ανάλογα με το σύστημα συμπερασμού που χρησιμοποιείται μπορεί να τρέξει σε εκθετικό χρόνο ως προς την είσοδο. Στη συνέχεια αποδεικνύουμε την ορθότητα του αλγορίθμου 1.

3.2.2 Απόδειξη Ορθότητας

Θεώρημα 22. *Έστω $\mathcal{T} \in \mathcal{L}$ μία οντολογία, $\mathcal{Q}$ ένα ΣΕ. Έστω $\mathcal{S}$ το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ όπως υπολογίζεται από έναν αλγόριθμο επαναγραφής βασισμένο στη μέθοδο ανάλυσης για το $\mathcal{L}$ και έστω $\mathcal{T}^+$ ένα πεπερασμένο σύνολο προτάσεων. Αν ο αλγόριθμος επαναγραφής τερματίζει για οποιοδήποτε σύνολο εισόδου, τότε ο Αλγόριθμος 1 επίσης τερματίζει. Έστω $\mathcal{R}'$ το σύνολο των προτάσεων που παράγεται από τον αλγόριθμο, τότε το $\mathcal{R}'$ είναι μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+$.*

Απόδειξη: Αρχικά θα δείξουμε ότι ο αλγόριθμος τερματίζει. Αρκεί να παρατηρήσουμε ότι το σύστημα Γ, στο οποίο βασίζεται ο αλγόριθμος 1, είναι ορθός και παράγει πεπερασμένο αριθμό προτάσεων όταν εφαρμόζεται σε μία οντολογία εκφρασμένη στη $\mathcal{L}$. Επίσης, λόγω της συνθήκης τερματισμού στη γραμμή 2 σε συνδυασμό με τη δεύτερη συνθήκη στη γραμμή 3, αποτρέπονται οι ατέρμονοι βρόγχοι. Για παράδειγμα, αν υπάρχει ένας κύκλος συμπερασμών, π.χ. μια ακολουθία συμπερασμών της μορφής: $\langle c_1, \ldots ; c_n \rangle$, $\langle c_n, \ldots ; c'_{n'} \rangle, \ldots, \langle c''_n, \ldots ; c'''_{n'''} \rangle$, με $c_n = c'''_{n'''}\sigma$ τότε ο τελευταίος συμπερασμός δε θα εφαρμοστεί αφού το c_n υπάγει το $c'''_{n'''}$ και συνεπώς το $c'''_{n'''}$ δε θα προστεθεί στο $\mathcal{U}$.

Για να δείξουμε ότι το $\mathcal{R}'$ είναι μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+$, αρκεί να δείξουμε ότι για κάθε στιγμιότυπο $\mathcal{I}$ ισχύει ότι $\mathsf{cert}(\mathcal{R}'_Q, \mathcal{R}'_D \cup \mathcal{I}) = \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+ \cup \mathcal{I})$, όπου $\mathcal{R}'_Q$, $\mathcal{R}'_D$ είναι τέτοια ώστε $\mathcal{R}' = \mathcal{R}'_Q \cup \mathcal{R}'_D$. Είναι προφανές ότι $\mathsf{cert}(\mathcal{R}'_Q, \mathcal{R}'_D \cup \mathcal{I}) \subseteq \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+ \cup \mathcal{I})$, αφού ο αλγόριθμος βασίζεται σε ένα ορθό σύστημα συμπερασμού και συνεπώς παράγει μόνο ορθές προτάσεις.

Τώρα θα δείξουμε ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+ \cup \mathcal{I}) \subseteq \mathsf{cert}(\mathcal{R}'_Q, \mathcal{R}'_D \cup \mathcal{I})$. Έστω $\mathcal{R}'' = \mathcal{R}''_Q \cup \mathcal{R}''_D$ μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+$ όπως προκύπτει από έναν αλγόρθμο επαναγραφής ερωτημάτων βασισμένο στην μέθοδο ανάλυσης και οποίος βασίζεται στο σύστημα συμπερασμού Γ. Εφόσον $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+ \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}''_Q, \mathcal{R}''_D \cup \mathcal{I})$, αρκεί να δείξουμε ότι για κάθε πρόταση c τ.ω. $\mathcal{R}'' \vdash_\Gamma c$ ισχύει ότι $\mathcal{R}' \vdash_\Gamma c$. Για να το δείξουμε αυτό, θα εφαρμόσουμε επαγωγή στα βήματα που ακολουθεί το Γ με είσοδο το $\mathcal{Q}, \mathcal{T} \cup \mathcal{T}^+$.

Έστω $\mathcal{R}''_i$ το σύνολο των προτάσεων που παράγεται από το Γ στο βήμα i. Έστω επίσης, ότι τα $\mathcal{U}_0$, $\mathcal{S}_0$ είναι τα σύνολα $\mathcal{U}$, $\mathcal{S}$ όπως έχουν σχηματιστεί στη γραμμή 1, δηλαδή, $\mathcal{U}_0 = \mathcal{T}^+$, $\mathcal{S}_0 = (\mathcal{Q} \cup \mathcal{T})_\Gamma$. Επίσης, έστω ότι ο αλγόριθμος τερματίζει μετά από n επαναλήψεις του βρόγχου while και ότι τα σύνολα $\mathcal{S}_n, \mathcal{U}_n$ είναι τα σύνολα $\mathcal{U}$, $\mathcal{S}$ όπως σχηματίζονται στη γραμμή 7, συνεπώς $\mathcal{S}_n = \mathcal{D}(\mathcal{S}_0 \cup \mathcal{U}_n)$. Είναι προφανές ότι $\mathcal{U}_0 \subseteq \ldots \subseteq \mathcal{U}_n$ και ότι για κάθε $c \in \mathcal{S}_0 \cup \mathcal{U}_n$ υπάρχει $c' \in \mathcal{S}_n$ τέτοιο ώστε το c' να υπάγει το c. Αρκεί να δείξουμε ότι για κάθε $i \geq 0$ και $c \in \mathcal{R}''_i$ υπάρχει κάποιο $c' \in \mathcal{S}_n$ τέτοιο ώστε το c' να υπάγει το c.

- *Βασικό Βήμα (i=0):* Από τον ορισμό των συστημάτων επαναγραφής που είναι βασισμένοι στη μέθοδο της ανάλυσης ισχύει ότι $\mathcal{R}''_0 = \mathcal{Q} \cup \mathcal{T} \cup \mathcal{T}^+$. Αν $c \in \mathcal{Q} \cup \mathcal{T} \cup \mathcal{T}^+$, τότε ο ισχυρισμός προκύπτει απευθείας απο το γεγονός ότι $\mathcal{S}_0 = (\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ και $\mathcal{T}^+ \subseteq \mathcal{U}_n$.

- *Επαγωγική Υπόθεση:* Έστω ότι για κάθε $c \in \mathcal{R}''_i$, στο βήμα i, υπάρχει $c' \in \mathcal{S}_n$ τέτοιο ώστε το c' να υπάγει το c. Τότε στο βήμα $i+1$ κάποια νέα πρόταση c_{new} θα παραχθεί από ένα συμπερασμό του οποίου οι υποθέσεις θα ανήκουν στο $\mathcal{R}_i''$. Στη συνέχεια, θα εξετάσουμε τις περιπτώσεις με βάση τα σύνολα στα οποία μπορεί να ανήκουν οι υποθέσεις αυτές.

1. Έστω ότι οι υποθέσεις ανήκουν μόνο στην $\mathcal{T}$. Εφόσον $\mathcal{S}_0 = (\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$, θα υπάρχει μία πρόταση c' στο $\mathcal{S}_0$, συνεπώς επίσης στο $\mathcal{S}_n$, που να υπάγει την c_{new}.
2. Έστω ότι οι υποθέσεις ανήκουν μόνο στο $\mathcal{T}^+$ και έστω ότι το $\langle c_1, \ldots, c_r; c_{new} \rangle$ είναι ο συμπερασμός. Προκύπτει από τις γραμμές 2-8 ότι εφαρμόζονται όλοι οι δυνατοί συμπερασμοί ανάμεσα στις προτάσεις του $\mathcal{U}_0$, ($\mathcal{U}_0 = \mathcal{T}^+$) και, εκτός αν τα συμπεράσματα είναι περιττά ως προς το $\mathcal{U} \cup \mathcal{S}$, όπως το $\mathcal{U}$ δημιουργείται τη στιγμή του συμπερασμού, προστίθενται στο $\mathcal{U}$. Συνεπώς, το c_{new} θα προστεθεί στο $\mathcal{U}$ και άρα θα ανήκει στο $\mathcal{S}_n$, εκτός αν υπάρχει κάποιο c' στο $\mathcal{U} \cup \mathcal{S}$, και άρα και στο $\mathcal{S}_n$, που το υπάγει.
3. Αν οι υποθέσεις ανήκουν και στα δύο σύνολα, $\mathcal{T}, \mathcal{T}^+$, ο ισχυρισμός προκύπτει, όπως προηγουμένως, από τις γραμμές 2-8, αφού $\mathcal{T}^+ \subseteq \mathcal{U}_0$ και $\mathcal{T} \subseteq \mathcal{S}_0$ και μεταξύ των προτάσεων των $\mathcal{U}_0, \mathcal{S}_0$ εφαρμόζονται όλοι οι δυνατοί συμπερασμοί.
4. Έστω ότι οι υποθέσεις ανήκουν στο $\mathcal{R}_i''$, δηλαδή, υπάρχει ένα σύνολο προτάσεων $\{c_1, \ldots, c_r\} \subseteq \mathcal{R}_i''$ τέτοιο ώστε $\langle c_1, \ldots, c_r; c_{new} \rangle$. Αν $\{c_1, \ldots, c_r\} \subseteq \mathcal{S}_0$ (όπου το $\mathcal{S}_0$ σύνολο προτάσεων εισόδου) τότε προφανώς θα υπάρχει μία πρόταση $c' \in \mathcal{S}_0$ που θα υπάγει το c_{new}. Αν υπάρχουν προτάσεις $c_u, \ldots, c_v \notin \mathcal{S}_0$, με $1 \leq u, v \leq r$, τότε προκύπτει από την επαγωγική υπόθεση ότι υπάρχει ένα σύνολο προτάσεων $c'_u, \ldots, c'_v \in \mathcal{S}_n$ που υπάγει τα $c_u, \ldots, c_v$. Αν $c'_u, \ldots, c'_v \in \mathcal{S}_0$ τότε προκύπτει εύκολα ο ισχυρισμός. Αν $c'_u, \ldots, c'_v \notin \mathcal{S}_0$ τότε θα υπάρχει μία επανάληψη k του βρόγχου 2-8 τέτοια ώστε $c'_u, \ldots, c'_v \in \mathcal{U}_k$ (όπου το $\mathcal{U}_k$ είναι το σύνολο $\mathcal{U}$ όπως σχηματίζεται στην k επανάληψη) και από τις ιδιότητα της υπαγωγής προκύπτει ότι υπάρχει ένα συμπερασμός $\langle c'_1, \ldots, c'_r; c' \rangle$ τέτοιος ώστε το c' να υπάγει το c_{new}. Προκύπτει από τη δομή του αλγορίθμου ότι υπάρχει ένα $m > k$ τέτοιο ώστε $c' \in \mathcal{U}_m$ και το c' να υπάγει το c_{new}.

■

3.3 Επαναγραφή Ερωτημάτων πάνω σε Οντολογίες που Συστέλλονται με τη χρήση Συμπερασμών

Στην ενότητα αυτή μελετάμε το πρόβλημα υπολογισμού μιας επαναγραφής ενός ΣΕ $\mathcal{Q}$ με βάση μια οντολογία $\mathcal{T}$ μειωμένη κατά ένα σύνολο αξιωμάτων $\mathcal{T}^-$, δοθείσας μίας επαναγραφής του $\mathcal{Q}$ με βάση μία οντολογία $\mathcal{T}$. Όπως και στην προηγούμενη ενότητα, μελετάμε πως μπορεί να γίνει αυτό εφικτό αξιοποιώντας τη μέγιστη δυνατή πληροφορία από τον υπολογισμό του $\mathcal{R}$ αποφεύγοντας να επαναλάβουμε την εφαρμογή συμπερασμών. Στην αρχή θα μελετήσουμε το πρόβλημα σε ένα θεωρητικό επίπεδο με τη βοήθεια χαρακτηριστικών παραδειγμάτων που τονίζουν τα σημαντικά σημεία του προβλήματος. Στη συνέχεια, παρουσιάζουμε αναλυτικά τον αλγόριθμο που έχουμε κατασκευάσει.

Παράδειγμα 23. Έστω μία οντολογία $\mathcal{T}$ που αποτελείται από τις εξής προτάσεις:

$$\begin{array}{rcrcl} c_1 & = & P(x,f(x)) & \leftarrow & A(x), \\ c_2 & = & B(f(x)) & \leftarrow & A(x), \\ c_3 & = & R(x,g(x)) & \leftarrow & B(x), \\ c_4 & = & B(x) & \leftarrow & P(x,y) \wedge B(y) \end{array}$$

και έστω το ερώτημα $\mathcal{Q} = Q(x) \leftarrow P(x,y) \wedge R(y,z)$. Τότε, μία επαναγραφή $\mathcal{R}$ του $\mathcal{Q}$ με βάση την $\mathcal{T}$ μπορεί να υπολογιστεί από ένα οποιοδήποτε σύγχρονο σύστημα επαναγραφής ερωτημάτων το οποίο να υποστηρίζει τη γλώσσα DL $\mathcal{EL}$. Για παράδειγμα, ένας τρόπος με τον οποίο μπορεί το Requiem να υπολογίσει την επαναγραφή αυτή είναι εφαρμόζοντας τους εξής συμπερασμούς:

$$\begin{array}{ll} \dfrac{\mathcal{Q} \quad c_3}{\mathcal{Q}_1} & \text{όπου} \quad \mathcal{Q}_1 = Q(x) \leftarrow P(x,y) \wedge B(y), \\ \dfrac{\mathcal{Q}_1 \quad c_1}{c'} & \text{όπου} \quad c' = Q(x) \leftarrow A(x) \wedge B(f(x)), \\ \dfrac{c' \quad c_2}{\mathcal{Q}_2} & \text{όπου} \quad \mathcal{Q}_2 = Q(x) \leftarrow A(x), \\ \dfrac{\mathcal{Q} \quad c_1}{c''} & \text{όπου} \quad c'' = Q(x) \leftarrow A(x) \wedge R(f(x),z), \\ \dfrac{c_4 \quad c_1}{c'''} & \text{όπου} \quad c''' = B(x) \leftarrow A(x) \wedge B(f(x)), \\ \dfrac{c''' \quad c_2}{c_5} & \text{όπου} \quad c_5 = B(x) \leftarrow A(x) \end{array}$$

και στη συνέχεια, αφαιρώντας τις προτάσεις που περιέχουν συναρτησιακούς όρους θα επιστρέψει την επαναγραφή $\mathcal{R} = \{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2, c_4, c_5\}$.

Έστω τώρα ότι αφαιρείται το αξίωμα c_2 από την $\mathcal{T}$ αποκτώντας με τον τρόπο αυτό την οντολογία $\mathcal{T}' = \mathcal{T} \setminus \{c_2\}$. Μπορεί να υπολογιστεί μια επαναγραφή $\mathcal{R}'$ για το $\mathcal{Q}$ με βάση την $\mathcal{T}'$ χρησιμοποιώντας το ίδιο σύστημα επαναγραφής. Η επαναγραφή σε αυτή την περίπτωση θα αποτελείται από τα ερωτήματα $\mathcal{Q}$, $\mathcal{Q}_1$ και την πρόταση c_4. $\diamondsuit$

Είναι προφανές ότι μπορεί να υπολογιστεί μια νέα επαναγραφή χρησιμοποιώντας κάποιο από τα γνωστά συστήματα επαναγραφής με είσοδο τα $\mathcal{Q}, \mathcal{T}'$, ωστόσο βλέπουμε ότι με αυτόν τον τρόπο θα επαναληφθεί ο υπολογισμός του ερωτήματος $\mathcal{Q}_1$ (συγκεκριμένα, το Requiem θα υπολογίσει ξανα τις προτάσεις c', c'' και c'''). Επίσης, παρατηρούμε ότι μπορεί να επιτευχθεί ο υπολογισμός της νέας επαναγραφής κατευθείαν από το $\mathcal{R}$ απλά αφαιρώντας το ερώτημα $\mathcal{Q}_2$ και την πρόταση c_5, αφού για την εξαγωγή των $\mathcal{Q}_2$, c_5 απαιτείται η πρόταση c_2. Οι παρατηρήσεις αυτές μας οδηγούν στο συμπέρασμα ότι αν κάθε στοιχείο της αρχικής επαναγραφής σημειωθεί με τα ελάχιστα υποσύνολα της οντολογίας από τα οποία μπορεί να παραχθεί, τότε θα μπορεί εύκολα να υπολογιστεί μία επαναγραφή ως προς τη νέα οντολογία.

Ορισμός 24. Έστω $\mathcal{Q}$ ένα ΣΕ, $\mathcal{T}$ μία οντολογία, $\mathcal{S}$ το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ και έστω κάποια πρόταση $c \in \mathcal{S}$. Λέμε ότι το σύνολο $\mathcal{T}' \subseteq \mathcal{T}$ είναι *ελάχιστο υποσύνολο* (*minimal subset*) για το c με βάση το $\mathcal{T}$ αν ισχύουν οι παρακάτω συνθήκες:

1. $c \in (\mathcal{Q} \cup \mathcal{T}')_{\Gamma,\mathcal{D}}$.

2. Για κάθε $\mathcal{T}'' \subset \mathcal{T}'$ δεν ισχύει η συνθήκη 1.

Διαισθητικά, το $\mathcal{T}'$ είναι ελάχιστο για κάποιο c αν το c εμφανίζεται σε κάποια επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}'$, αλλά αν αφαιρέσουμε μία οποιαδήποτε πρόταση από το $\mathcal{T}'$ τότε το c δεν θα εμφανίζεται σε καμία επαναγραφή του $\mathcal{Q}$ με βάση την τροποποιημένη οντολογία.

Στο τρέχον παράδειγμα (Παράδειγμα 23) έχουμε τα εξής ελάχιστα υποσύνολα για κάθε στοιχείο της επαναγραφής $\mathcal{R}$:

$$\begin{array}{rll}
\emptyset & \text{είναι ελάχιστο για το} & \mathcal{Q} \\
\{c_3\} & \text{είναι ελάχιστο για το} & \mathcal{Q}_1 \\
\{c_1, c_3\} & \text{είναι ελάχιστο για το} & c' \\
\{c_1, c_2, c_3\} & \text{είναι ελάχιστο για το} & \mathcal{Q}_2 \\
\{c_1\} & \text{είναι ελάχιστο για το} & c'' \\
\{c_1, c_4\} & \text{είναι ελάχιστο για το} & c''' \\
\{c_1, c_2, c_4\} & \text{είναι ελάχιστο για το} & c_5
\end{array}$$

και το $\{c_i\}$ είναι ελάχιστο για το c_i με βάση την $\mathcal{Q}$, για κάθε $1 \leq i \leq 4$. Αφού το ελάχιστο υποσύνολο για το $\mathcal{Q}_2$ περιέχει την πρόταση c_2, η οποία αφαιρέθηκε από την οντολογία $\mathcal{T}$ αποκτώντας έτσι την οντολογία $\mathcal{T}'$, συνεπάγεται ότι το $\mathcal{Q}_2$ δεν μπορεί να ανήκει στη νέα επαναγραφή για τα $\mathcal{Q}, \mathcal{T}'$. Αντίστοιχα, το ίδιο ισχύει και για την c_5.

Είναι σημαντικό να σημειώσουμε ότι για μία πρόταση c μίας επαναγραφής μπορεί να υπάρχουν περισσότερα από ένα ελάχιστα υποσύνολα. Για παράδειγμα, το ερώτημα $\mathcal{Q}_1 = Q(x) \leftarrow A(x) \wedge C(x)$ που περιέχεται στην επαναγραφή ενός ΣΕ $\mathcal{Q} = Q(x) \leftarrow A(x) \wedge B(x) \wedge C(x)$ με βάση την οντολογία $\mathcal{T} = \{B(x) \leftarrow A(x), B(x) \leftarrow C(x)\}$, έχει τα εξής δύο ελάχιστα υποσύνολα: $\{B(x) \leftarrow A(x)\}$, $\{B(x) \leftarrow C(x)\}$. Συνεπώς κάθε πρόταση μίας επαναγραφής θα πρέπει αν σημειωθεί με το *σύνολο* των ελάχιστων υποσυνόλων της.

Ορισμός 25. Έστω $\mathcal{Q}$ ένα ΣΕ, έστω $\mathcal{T}$ μία οντολογία και έστω $\mathcal{S}$ το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$. Η *επισημείωση* (*labelling*) του $\mathcal{S}$ είναι ένα ζεύγος $\mathcal{S}_{\mathcal{L}} = \langle \mathcal{S}, \rho \rangle$, όπου ρ μία αντιστοίχιση από το $\mathcal{S}$ στο $2^{\mathcal{T}}$ τέτοια ώστε για κάθε $c \in \mathcal{S}$, το $\rho(c)$ θα περιέχει όλα τα $\mathcal{T}' \subseteq \mathcal{T}$ τα οποία είναι ελάχιστα για το c.

Αξίζει να σημειωθεί ότι για να υπολογιστεί η επισημείωση του $\mathcal{S}_{\mathcal{L}}$ ενός κορεσμένου συνόλου προτάσεων θα πρέπει να τροποποιηθούν οι εσωτερικές διαδικασίες του αλγορίθμου επαναγραφής. Αυτό μπορεί να επιτευχθεί αρχικοποιώντας ένα κενό σύνολο $\emptyset$ ως το ελάχιστο σύνολο για το $\mathcal{Q}$ και το μοναδιαίο σύνολο $\{c\}$ για κάθε πρόταση $c \in \mathcal{T}$ και στη συνέχεια αποθηκεύοντας τα αξιώματα που χρησιμοποιήθηκαν για να παραχθεί κάθε πρόταση της τελικής επαναγραφής. Για παράδειγμα, αν το c' παράχθηκε αναλύοντας το c_1 με το c_2, τότε $\rho(c') = \rho(c') \cup \{\mathcal{T}_1 \cup \mathcal{T}_2 \mid \mathcal{T}_1 \in \rho(c_1) \text{ and } \mathcal{T}_2 \in \rho(c_2)\}$.

Ένα σημαντικό, τεχνικής φύσης, ερώτημα που προκύπτει είναι αν είναι εφικτός ο υπολογισμός μίας νέας επαναγραφής για ένα ΣΕ ερώτημα με βάση μία μειωμένη οντολογία $\mathcal{T}'$ δεδομένης μίας οποιασδήποτε επαναγραφής με βάση την αρχική οντολογία $\mathcal{T} \supseteq \mathcal{T}'$. Όπως φαίνεται στο παρακάτω παράδειγμα αυτό δεν είναι πάντα εφικτό.

Παράδειγμα 26. Έστω η οντολογία $\mathcal{T}$:

$$\begin{aligned} c_1 &= R(x, f(x)) \leftarrow A(x), \\ c_2 &= R(x, g(x)) \leftarrow B(x) \end{aligned}$$

και το ΣΕ $\mathcal{Q} = Q(x) \leftarrow R(x,y) \wedge A(x)$. Το σύνολο $\mathcal{R} = \{Q, \mathcal{Q}_1, \mathcal{Q}_2\}$, όπου $\mathcal{Q}_1 = Q(x) \leftarrow A(x)$, $\mathcal{Q}_2 = Q(x) \leftarrow A(x) \wedge B(x)$ είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}_1$. Ωστόσο, το $\mathcal{Q}_1$ υπάγει τα $\mathcal{Q}$, $\mathcal{Q}_2$, οπότε το $\mathcal{R}' = \{\mathcal{Q}_1\}$, είναι επίσης μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$.

Ας υποθέσουμε τώρα ότι η πρόταση c_1 αφαιρείται από την $\mathcal{T}$ παίρνοντας έτσι την οντολογία $\mathcal{T}'$. Μία επαναγραφή $\mathcal{R}''$ του $\mathcal{Q}$ με βάση την $\mathcal{T}'$ είναι η $\mathcal{R}'' = \{\mathcal{Q}, \mathcal{Q}_2\}$. Ωστόσο, ο απευθείας υπολογισμός της $\mathcal{R}''$ από την επαναγραφή $\mathcal{R}'$ δεν είναι εφικτός, αφού η $\mathcal{R}'$ δεν περιέχει τα ερωτήματα $\mathcal{Q}$, $\mathcal{Q}_2$. Επίσης, ακόμα και αν εφαρμοστούν νέοι συμπερασμοί, πάλι δεν εφικτός ο υπολογισμός της $\mathcal{R}''$. Αντίθετα, η $\mathcal{R}''$ μπορεί να εξαχθεί από την $\mathcal{R}$ που περιέχει την $\mathcal{Q}$ αφαιρώντας απλά την $\mathcal{Q}_1$. $\diamondsuit$

Το προηγούμενο παράδειγμα υποδεικνύει ότι αν από την επαναγραφή $\mathcal{R}$ ως προς την αρχική οντολογία $\mathcal{T}$ έχουν αφαιρεθεί όλες οι περιττές προτάσεις, τότε υπάρχει η πιθανότητα να μην είναι εφικτή η απευθείας παραγωγή από το $\mathcal{R}$ μίας επαναγραφής ως προς την μειωμένη οντολογία $\mathcal{T}'$. Δύο είναι οι λόγοι που μπορεί να οδηγήσουν σε μια τέτοια κατάσταση. Ο πρώτος λόγος είναι ότι οι προτάσεις που αρχικά υπήγαν τις περιττές προτάσεις να μην μπορούν πλέον να παραχθούν από τα $\mathcal{Q}, \mathcal{T}'$, και ο δεύτερος είναι ότι οι προτάσεις που ήταν αρχικά περιττές να μην έχουν πλέον αυτήν την ιδιότητα. Συνεπώς συμπεραίνουμε ότι απαιτούνται και κάποιες από τις περιττές προτάσεις. Στο Παράδειγμα 26, αν υπάρχει η πληροφορία ότι το $\mathcal{Q}_1$ (το οποίο υπολογίστηκε αρχικά για το $\mathcal{R}$) υπάγει το $\mathcal{Q}$, τότε μπορεί να υπολογιστεί το $\mathcal{R}''$ αντιγράφοντας το $\mathcal{Q}$, διότι το ελάχιστο σύνολό του

Algorithm 2 $\mathsf{Delete}_{\Gamma,\mathcal{D}}(\mathcal{S}_{\mathcal{L}}, \mathcal{T}^{-}, \lambda_{\mathcal{S}})$

Input: $\mathcal{S}_{\mathcal{L}} = \langle \mathcal{S}, \rho \rangle$ ένα σημειωμένο σύνολο προτάσεων, $\mathcal{T}^{-}$ σύνολο προτάσεων, $\lambda_{\mathcal{S}}$ αντιστοίχιση

Output: $\mathcal{R}'$ μία επαναγραφή για τη μειωμένη οντολογία

```
R' := ∅ και ρ' μία κενή αντιστοίχιση.
for all c ∈ S τ.ω. υπάρχει T_i ∈ ρ(c) με T_i ∩ T⁻ = ∅ do
    R' := R' ∪ {c}
    ρ'(c) := {T_j ∈ ρ(c) | T_j ∩ T⁻ = ∅}
end for
T⁺ := {c' | 1. c' ∈ λ_R(c) με c ∈ S \ R'
            2. υπάρχει T_i ∈ ρ(c') και T_i ∩ T⁻ = ∅
            3. δεν υπάρχει c'' ∈ R' με c' ∈ λ_R(c'')}
R' := Add_{Γ,D}(R', T⁺)
return R'
```

είναι το $\{\mathcal{Q}\}$, διαγράφοντας το $\mathcal{Q}_1$, διότι το ελάχιστο σύνολό του είναι το $\{c_1\}$ και, τέλος, συμπεραίνοντας το $\mathcal{Q}_2$ από τα $\mathcal{Q}$ και c_2.

Σύμφωνα με τα προηγούμενα, για να εξαχθεί μία νέα επαναγραφή από μία προηγούμενη επαναγραφή $\mathcal{R}$, η οποία είναι μειωμένη ως προς τη σχέση υπαγωγής, θα πρέπει να έχουν αποθηκευθεί οι σχέσεις υπαγωγής που υπολογίστηκαν κατά την παραγωγή της $\mathcal{R}$.

Ορισμός 27. Έστω $\mathcal{R}_1, \ldots, \mathcal{R}_n$ μία ακολουθία από σύνολα προτάσεων που παράγεται από έναν αλγόριθμο επαναγραφής. Τότε για κάθε $c \in \mathcal{S}$, όπου $\mathcal{S} = \mathcal{R}_n$, συμβολίζουμε με $\lambda_{\mathcal{S}}(c)$ το ελάχιστο υποσύνολο $D = \bigcup_{1 \leq i < n} \mathcal{R}_i$ που περιέχει όλα τα $c_i \in D$ που υπάγονται από το c.

Επίσης, σύμφωνα με το Παράδειγμα 26, από τη στιγμή που θα εντοπιστούν οι περιττές προτάσεις που δεν υπάγονται πλέον από κάποια άλλη πρόταση, αφού η πρόταση αυτή δεν μπορεί να παραχθεί από τα $\mathcal{Q}, \mathcal{T}'$, θα πρέπει να εφαρμοστούν νέοι συμπερασμοί. Στο σημείο αυτό πρέπει να σημειωθεί ότι για να εφαρμοστούν μόνο οι απαραίτητοι νέοι συμπερασμοί θα πρέπει να χρησιμοποιηθεί ο αλγόριθμος $\mathsf{Add}_{\Gamma,\mathcal{D}}$ που παρουσιάζεται στην Ενότητα 3.2. Συνεπώς, αντί για την αρχική επαναγραφή, θα πρέπει να αξιοποιήσουμε το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$.

3.3.1 Αλγόριθμος

Ο Αλγόριθμος 2 υπολογίζει μία επαναγραφή για ένα ΣΕ $\mathcal{Q}$ και μία μειωμένη οντολογία $\mathcal{T} \setminus \mathcal{T}^{-}$ αξιοποιώντας την πληροφορία που έχει παραχθεί κατα τον υπολογισμό της επαναγραφής του $\mathcal{Q}$ με βάση την $\mathcal{T}$. Συγκεκριμένα, δοθείσας της επισημείωσης $\mathcal{S}_{\mathcal{L}}$ του συνόλου $\mathcal{S} = (\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$, μίας αντιστοίχισης $\lambda_{\mathcal{S}}$, όπως ορίζεται στον Ορισμό 27, και του συνόλου των αξιωμάτων $\mathcal{T}^{-}$ που έχει αφαιρεθεί από το $\mathcal{T}$, ο αλγόριθμος αρχικά ελεγχει ποιές από τις προτάσεις του $\mathcal{S}$ μπορούν ακόμα να συμπεραχθούν από το $\mathcal{T} \setminus \mathcal{T}^-$, τις προσθέτει στο σύνολο $\mathcal{R}'$ (γραμμή 3) και ανανεώνει τις σημειώσεις τους (γραμμή 4). Στη συνέχεια, συλλέγει στο σύνολο $\mathcal{T}^+$ τις προτάσεις για τις οποίες ισχύουν τα ακόλουθα: i) συμπεραίνονται από την μειωμένη οντολογία και ii) ήταν περιττές στο $\mathcal{S}$ εξαιτίας μίας πρότασης που τις υπήγε και η οποία δεν συμπεραίνεται πλέον από το $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$. Προφανώς, οι προτάσεις αυτές πρέπει να αναλυθούν περαιτέρω αφου μπορεί να συμπεράνουν νέες προτάσεις. Αυτό επιτυγχάνεται με τη χρήση του αλγορίθμου $\mathsf{Add}_{\Gamma,\mathcal{D}}$. Συνεπώς, η πολυπλοκότητα του Αλγορίθμου $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ βασίζεται στην πολυπλοκότητα του $\mathsf{Add}_{\Gamma,\mathcal{D}}$ η οποία εξαρτάται από την πολυπλοκότητα του αλγορίθμου επαναγραφής που χρησιμοποιείται.

Παράδειγμα 28. Έστω η οντολογία $\mathcal{T}$, το ΣΕ $\mathcal{Q}$ του Παραδείγματος 26, $\mathcal{T}^- = \{c_1\}$ και έστω $\mathcal{S} = \{\mathcal{Q}_1, c_1, c_2\}$. Τότε, σύμφωνα με τον αλγόριθμο $\mathsf{Delete}_{\Gamma,\mathcal{D}}$, το $\mathcal{R}'$ στη γραμμή 5 θα είναι το σύνολο $\{c_2\}$, διότι $\rho(c_1) = \rho(\mathcal{Q}_1) = \{c_1\}$. Επίσης, ισχύει ότι $\mathcal{Q} \in \lambda_{\mathcal{R}}(\mathcal{Q}_1)$, $\rho(\mathcal{Q}) = \{\mathcal{Q}\}$ και το c_2 δεν υπάγει το $\mathcal{Q}$, άρα $\mathcal{T}^+ = \{\mathcal{Q}\}$. Αν εφαρμόσουμε τα $\mathcal{R}', \mathcal{T}^+$ στην $\mathsf{Add}_{\Gamma,\mathcal{D}}$ θα παραχθεί η τελική επαναγραφή $\mathcal{R}''$.

Ο αλγόριθμος $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ μπορεί να χρησιμοποιηθεί ξανά για μελλοντικές συστολές της οντολογίας αν ο αλγόριθμος $\mathsf{Add}_{\Gamma,\mathcal{D}}$ τροποποιηθεί με τέτοιον τρόπο ώστε να αποθηκεύει σε κάθε συμπερασμό τα αξιώματα που χρησιμοποιήθηκαν για την παραγωγή μίας πρότασης. Επίσης, όπως περιγράφεται στην Ενότητα 3.2, ο αλγόριθμος $\mathsf{Add}_{\Gamma,\mathcal{D}}$ μπορεί να τροποποιηθεί ώστε να υπολογίζει το σύνολο $\mathcal{S}$. Με τις τροποποιήσεις αυτές, το αποτέλεσμα του αλγορίθμου $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ θα είναι το σημειωμένο σύνολο $\mathcal{S}'_{\mathcal{L}} = \langle (\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-)_{\Gamma,\mathcal{D}}, \rho' \rangle$, όπου το ρ' είναι η νέα αντιστοίχιση.

3.3.2 Απόδειξη Ορθότητας

Θεώρημα 29. *Έστω η οντολογία $\mathcal{T} \in \mathcal{L}$ και το ΣΕ $\mathcal{Q}$. Έστω $\mathcal{S}$ το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ όπως υπολογίζεται από έναν αλγόριθμο επαναγραφής για την $\mathcal{L}$ που βασίζεται στη μέθοδο της ανάλυσης και έστω $\mathcal{S}_{\mathcal{L}} = \langle \mathcal{S}, \rho \rangle$ η επισημείωση του $\mathcal{S}$. Έστω επίσης το $\lambda_{\mathcal{S}}$ όπως ορίζεται στον Ορισμό 27 για κάθε $c \in \mathcal{S}$ και έστω το $\mathcal{T}^-$ υποσύνολο της $\mathcal{T}$. Αν για οποιαδήποτε είσοδο ο αλγόριθμος επαναγραφής τερματίζει τότε ο αλγόριθμος $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ όταν εφαρμόζεται στα $\mathcal{S}_{\mathcal{L}}$, $\mathcal{T}^-$ και $\lambda_{\mathcal{S}}$ τερματίζει και το αποτέλεσμα $\mathcal{R}'$ είναι μία επαναγραφή για το $\mathcal{Q}$, $\mathcal{T} \setminus \mathcal{T}^-$.*

Απόδειξη: Είναι προφανές ότι οι γραμμές 2–5 εκτελούνται μόνο μία φορά και το σύνολο $\mathcal{T}^+$ κατασκευάζεται σε πεπερασμένο αριθμό βημάτων. Επίσης, από το Θεώρημα 22 προκύπτει ότι η γραμμή 7 τερματίζει, συνεπώς ο αλγόριθμος τερματίζει $\mathsf{Delete}_{\Gamma,\mathcal{D}}$.

Στη συνέχεια θα δείξουμε ότι το σύνολο $\mathcal{R}'$ που υπολογίζεται από τον αλγόριθμο είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T} \setminus \mathcal{T}^-$. Για να το δείξουμε αυτό αρκεί να δείξουμε ότι για όλους τους ισχυρισμούς $\mathcal{I}$ ισχύει ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$, όπου τα $\mathcal{R}'_D$, $\mathcal{R}'_Q$ είναι ο διαμερισμός του $\mathcal{R}'$, δηλαδή, $\mathcal{R}' = \mathcal{R}'_Q \cup \mathcal{R}'_D$.

Αρχικά, δείχνουμε ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) \supseteq \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$. Όπως φαίνεται από τη δομή του αλγορίθμου, μία πρόταση c ανήκει στο $\mathcal{R}'$ αν ισχύει μία από τις ακόλουθες περιπτώσεις: i) $c \in \mathcal{S}$ και δεν υπάρχει $\mathcal{T}_i \in \rho(c)$ τέτοιο ώστε $\mathcal{T}_i \cap \mathcal{T}^- = \emptyset$ (συμβολίζουμε με $\mathcal{R}_t$ το σύνολο των προτάσεων για τις οποίες ικανοποιείται αυτή η πρόταση, είναι προφανές ότι $\mathcal{R}_t \subseteq \mathcal{S}$), ii) το c ανήκει στο σύνολο προτάσεων $\mathcal{T}^+$, δηλαδή, υπάρχει πρόταση $c' \in \mathcal{S} \setminus \mathcal{R}_t$ που υπάγει το c, το οποίο συμπεραίνεται από το $\mathcal{R}_t$, και δεν υπάρχει πρόταση στο $\mathcal{R}_t$ που να υπάγει το c, iii) το c περιέχεται στο αποτέλεσμα του αλγορίθμου $\mathsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{R}_t, \mathcal{T}^+)$.

Αν ισχύει η πρώτη περίπτωση, τότε εφόσον ισχύει ότι $\mathcal{T}_i \cap \mathcal{T}^- = \emptyset$, τότε ισχύει ότι $\mathcal{T}_i \subseteq \mathcal{T} \setminus \mathcal{T}^-$ και συνεπώς $\mathcal{T} \setminus \mathcal{T}^- \models \mathcal{T}_i$. Επίσης, από τον ορισμό του ρ ισχύει ότι $\mathcal{T}_i \cup \mathcal{Q} \models c$. Από τις δύο συνθήκες αυτές και την ιδιότητα της μονοτονίας προκύπτει ότι $\mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{Q} \models c$. Αυτό ισχύει για όλα τα μέρη του $\mathcal{R}_t$ και συνεπώς ισχύει ότι $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{R}_t$. Αν ισχύει η δεύτερη περίπτωση, τότε το c υπάγεται απο μία πρόταση του $\mathcal{S}$, άρα ισχύει ότι $\mathcal{S} \models c$. Επίσης, υπάρχει $\mathcal{T}_i \in \rho(c)$ τέτοιο ώστε $\mathcal{T}_i \subseteq \mathcal{T} \setminus \mathcal{T}^-$, άρα σύμφωνα με το i) ισχύει και πάλι ότι $\mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{Q} \models c$ και συνεπώς ότι $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{T}^+$. Τέλος αν ισχύει η περίπτωση iii), τότε από τις i) και ii) έχουμε ότι $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{R}_t \cup \mathcal{T}^+$ και από το Θεώρημα 22 ισχύει ότι αν $c \in \mathsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{R}_t, \mathcal{T}^+)$ (συμβολίζουμε το σύνολο αυτό με $\mathcal{R}'$) τότε $\mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{Q} \models c$. Αυτό ισχύει για όλα τα στοιχεία του $\mathcal{R}'$ άρα $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{R}'$ από το οποίο προκύπτει ότι, για κάθε $\mathcal{I}$, οι απαντήσεις του $\mathcal{R}'$ πάνω στο $\mathcal{I}$ είναι επίσης απαντήσεις του $\mathcal{Q}$ πάνω στο $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{I}$.

Τώρα, θα δείξουμε ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) \subseteq \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$. Έστω $\mathcal{R}''$ μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T} \setminus \mathcal{T}^-$. Έστω επίσης $\mathcal{R}'' = \mathcal{R}''_Q \cup \mathcal{R}''_D$. Εφόσον, $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}''_Q, \mathcal{R}''_D \cup \mathcal{I})$, αρκεί να δείξουμε ότι για κάθε πρόταση c τέτοια ώστε $c \in \mathcal{R}''$ υπάρχει πρόταση $c' \in \mathcal{R}'$ τέτοια ώστε η c' να υπάγει τη c.

Αν $c \in \mathcal{R}''$ τότε $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^- \models c$, και άρα λόγω της ιδιότητας της μονοτονίας και του γεγονότος ότι ισχύει $\mathcal{S} = (\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ είτε υπάρχει μία πρόταση $c' \in \mathcal{S}$ που είναι ισοδύναμη με το c, είτε υπάρχουν δύο προτάσεις $c_0, c'' \in \mathcal{R}''$ τέτοιες ώστε το c'' να υπάγει το c_0 και υπάρχει μία ακολουθία συμπερασμών $\langle c_0, \ldots ; c_1 \rangle, \ldots, \langle c_{n-1}, \ldots ; c_n \rangle$ με $c_n = c\sigma$ για κάποιο σ, και $c_1, \ldots, c_n \in \mathcal{R}''$. Στην πρώτη περίπτωση, προκύπτει από τη δομή του αλγορίθμου ότι $c' \in \mathcal{R}_t$ συνεπώς $c' \in \mathcal{R}'$ και άρα ο ισχυρισμός έχει αποδειχθεί. Στη δεύτερη περίπτωση υπάρχουν δύο ενδεχόμενα για το c'': Αν υπάρχει $\mathcal{T}_i \in \rho(c'')$ τέτοιο ώστε $\mathcal{T}_i \cap \mathcal{T}^- = \emptyset$, τότε $c'' \in \mathcal{R}_t$, άρα $c'' \in \mathcal{R}'$ και συνεπώς από τις ιδιότητες της υπαγωγής, υπάρχει μία πρόταση c''' που συμπεραίνεται από το c'', και υπάγει το c. Διαφορετικά, για κάθε $\mathcal{T}_i \in \rho(c'')$, $\mathcal{T}_i \cap O^- \neq \emptyset$ και άρα σύμφωνα με τον αλγόριθμο $c'' \in \mathcal{S} \setminus \mathcal{R}_t$. Αν υπάρχει πρόταση $c''' \in \mathcal{R}_t$ που υπάγει το c_0 τότε ο ισχυρισμός έχει αποδειχθεί. Διαφορετικά, είναι προφανές από τη γραμμή 6 του αλγορίθμου ότι $c_0 \in \mathcal{T}^+$. Συνεπώς, προκύπτει από το Θεώρημα 22 ότι υπάρχει πρόταση $c''' \in \mathcal{R}'$ τέτοια ώστε το c''' να υπάγει το c. ∎

3.3.3 Συζήτηση

Είναι σημαντικό να σημειωθεί υπάρχει σχέση μεταξύ του μεγέθους το αρχικού συνόλου $\mathcal{S}_{\mathcal{L}}$ $((\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}})$ και του αριθμού των νέων συμπερασμών που θα εφαρμοστούν στο βήμα 7 του αλγορίθμου. Από τη μία πλευρά, αν το $\mathcal{S}_{\mathcal{L}}$ είναι αρκετά μεγάλο, τότε μπορεί να μην χρειαστεί να εφαρμοστούν νέοι συμπερασμοί. Στο Παράδειγμα 26, αν το δοθέν σύνολο είναι το $\{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2\}$, τότε μπορεί να υπολογιστεί μία επαναγραφή για την $\mathcal{T}'$ μόνο από τα βήματα 2–5. Στην περίπτωση αυτή όμως, το $\mathcal{S}_{\mathcal{L}}$ μπορεί να είναι πολύ μεγάλο.

Από την άλλη πλευρά, αν ο αλγόριθμος επαναγραφής έχει εφαρμόσει εμπρόσθιο και οπισθοδρομικό έλεγχο υπαγωγής, τότε το αρχικό σύνολο θα είναι μεν μικρό, αλλά θα πρέπει να εφαρμοστούν πολλοί νέοι συμπερασμοί, ή ακόμα και να επαναληφθεί ο υπολογισμός κάποιων. Για παράδειγμα, αν υποθέσουμε ότι ένας αλγόριθμος επαναγραφής παράγει σειριακά τις προτάσεις $c_1, c_2, \ldots, c_n$ και υποθέσουμε ότι το c_n υπάγει το c_1, τότε λόγω οπισθοδρομικού ελέγχου υπαγωγής, το c_1 θα προστεθεί στο $\lambda(c_n)$ και όλες οι προτάσεις $c_2, \ldots, c_{n-1}$ μπορούν να αφαιρεθούν για να μειωθεί το μέγεθος της επαναγραφής. Αν υποθέσουμε τώρα, ότι αφαιρείται κάποια πρόταση και το c_n δεν μπορεί να παραχθεί πια, τότε ο Αλγόριθμος 2 θα πρέπει να προσθέσει το c_1 στο σύνολο $\mathcal{T}^+$ και στη συνέχεια να υπολογίσει ξανά τις προτάσεις $c_2, \ldots, c_{n-1}$ στο βήμα 7.

Ανάμεσα στις δύο αυτές περιπτώσεις, η βέλτιστη λύση είναι να εφαρμοστούν μερικώς οι έλεγχοι υπαγωγής, δηλαδή να εφαρμοστεί μόνο εμπρόσθιος έλεγχος υπαγωγής. Για παράδειγμα, στην προηγούμενη περίπτωση, καμία από τις προτάσεις $c_2, \ldots, c_{n-1}$ δεν θα είχε αφαιρεθεί, αφού παρήχθησαν πριν το c_n και άρα ο Αλγόριθμος 2 δε θα εφαρμόσει αυτούς τους συμπερασμούς. Ωστόσο, εξαιτίας του εμπρόσθιου ελέγχου, κάποιες περιττές προτάσεις θα απουσιάζουν από το αρχικό σύνολο $\mathcal{S}$.

3.4 Επαναγραφή Ερωτημάτων πάνω σε Οντολογίες που Συστέλλονται χωρίς τη χρήση Συμπερασμών

Στην ενότητα αυτή, παρουσιάζουμε την συνθήκη που απαιτείται ώστε να υπολογιστεί η νέα επαναγραφή χωρίς να εφαρμοστούν τα βήματα 6–7 του αλγορίθμου $\mathsf{Delete}_{\Gamma,\mathcal{D}}$, δηλαδή χωρίς να εφαρμοστούν νέοι συμπερασμοί. Η συνθήκη που παρουσιάζεται είναι κάπως καλύτερη από το να μην εφαρμοστούν καθόλου έλεγχοι υπαγωγής κατά τον υπολογισμό της αρχικής επαναγραφής. Εφόσον δεν πρόκειται να εφαρμοστούν νέοι συμπερασμοί, δηλαδή να κληθεί ο αλγόριθμος $\mathsf{Add}_{\Gamma,\mathcal{D}}$, το σύνολο εισόδου μπορεί να είναι η αρχική επαναγραφή και όχι το κορεσμένο σύνολο $\mathcal{S}$.

Διαισθητικά, μία επαναγραφή για μία μειωμένη οντολογία $\mathcal{T}' \subseteq \mathcal{T}$ μπορεί να υπολογιστεί από την επαναγραφή $\mathcal{R}$ για την $\mathcal{T}$ χωρίς την εφαρμογή συμπερασμών, αν το $\mathcal{R}$ περιέχει μία επαναγραφή για κάθε υποσύνολο της $\mathcal{T}$.

Ορισμός 30. Η επαναγραφή $\mathcal{R}$ ενός ΣΕ $\mathcal{Q}$ με βάση την οντολογία $\mathcal{T}$ είναι *κλειστή ως προς τα υποσύνολα* (*subset closed*) αν για κάθε $\mathcal{T}' \subseteq \mathcal{T}$ υπάρχει $\mathcal{R}' \subseteq \mathcal{R}$ τέτοιο ώστε το σύνολο $\mathcal{R}'$ είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}'$.

Ένας προφανής τρόπος για να υπολογιστεί μία επαναγραφή κλειστή ως προς τα υποσύνολα είναι να μην εφαρμοστούν κατά τον υπολογισμό της έλεγχοι υπαγωγής. Ωστόσο, σύμφωνα με τον Ορισμό 30, αν εφαρμοστούν κάποιοι συγκεκριμένοι έλεγχοι υπαγωγής,

Algorithm 3 Delete($\mathcal{R}_\mathcal{L}, \mathcal{T}^-$)

Input: $\mathcal{R}_\mathcal{L} = \langle \mathcal{R}, \rho \rangle$ η σημειωμένη αρχική επαναγραφή, $\mathcal{T}^-$ σύνολο προτάσεων
Output: $\mathcal{R}'$ μία επαναγραφή για τη μειωμένη οντολογία

$\mathcal{R}' := \emptyset$ και ρ' μία κενή αντιστοίχιση.
for all $c \in \mathcal{R}$ τ.ω. $\exists \mathcal{T}_i \in \rho(c)$ με $\mathcal{T}_i \cap \mathcal{T}^- = \emptyset$ **do**
  $\mathcal{R}' := \mathcal{R}' \cup \{c\}$
  $\rho'(c) := \{\mathcal{T}_j \in \rho(c) \mid \mathcal{T}_j \cap \mathcal{T}^- = \emptyset\}$
end for
return $\mathcal{R}'$

τότε και πάλι μπορεί να υπολογιστεί μία επαναγραφή κλειστή ως προς τα υποσύνολα. Συγκεκριμένα, μία πρόταση $c \in \mathcal{R}$ μπορεί να αφαιρεθεί μόνο αν υπάρχει πρόταση $c' \in \mathcal{R}$ που να υπάγει τη c και δεν υπάρχει επαναγραφή $\mathcal{R}' \subseteq \mathcal{R}$ του $\mathcal{Q}$ με βάση την $\mathcal{T}' \subseteq \mathcal{T}$ τέτοια ώστε $c \in \mathcal{R}'$ και $c' \notin \mathcal{R}'$.

Παράδειγμα 31. Έστω η οντολογία $\mathcal{T}$ και το ΣΕ $\mathcal{Q}$ από το Παράδειγμα 26. Για $\mathcal{T}' = \emptyset$ δεν υπάρχει υποσύνολο του $\mathcal{R}'$ το οποίο να είναι επαναγραφή για τα $\mathcal{Q}, \mathcal{T}'$. Αντίθετα, για το υποσύνολο $\{\mathcal{Q}\}$ της επαναγραφής του $\mathcal{R}$ είναι μία επαναγραφή των $\mathcal{Q}, \mathcal{T}'$. Συνεπώς, το $\mathcal{R}$ είναι κλειστό ως προς τα υποσύνολα ενώ το $\mathcal{R}'$ όχι. Έστω ότι η πρόταση c_1 αφαιρείται από την $\mathcal{T}$. Είναι προφανές ότι μπορούμε να υπολογίσουμε μία νέα επαναγραφή $\mathcal{R}''$ του $\mathcal{Q}$ με βάση τη νέα οντολογία, απλά αφαιρώντας την πρόταση $\mathcal{Q}_1$ από το $\mathcal{R}$. $\diamondsuit$

Σύμφωνα με το προηγούμενο παράδειγμα, μία νέα επαναγραφή μπορεί να υπολογιστεί από μία αρχική επαναγραφή κλειστή ως προς τα υποσύνολα, αφαιρώντας απλά τις προτάσεις που δεν μπορούν πλέον να παραχθούν από το ερώτημα και την μειωμένη οντολογία. Συνεπώς, για τον υπολογισμό της νέας επαναγραφής αρκεί να έχει σημειωθεί κάθε πρόταση της αρχικής κλειστής ως προς τα υποσύνολα επαναγραφής με τον τρόπο που προτείνεται στην προηγούμενη ενότητα. Οι ορισμοί του ελαχίστου υποσυνόλου και σημειώσης (Ορισμοί 24 και 25, αντίστοιχα) μπορούν να προσαρμοστούν στην περίπτωση των κλειστών ως προς τα υποσύνολα επαναγραφών αν αντικατασταθεί το "$(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$" που αναφέρεται στους ορισμούς με το "μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$". Τότε, ο Αλγόριθμος 2 μετασχηματίζεται στον Αλγόριθμο 3, ο οποίος υπολογίζει τη νέα επαναγραφή με βάση την αρχική κλειστή ως προς τα υποσύνολα επαναγραφή. Στη συνέχεια, δείχνουμε την ορθότητα του αλγορίθμου.

Θεώρημα 32. *Έστω $\mathcal{T}$ μία οντολογία, $\mathcal{Q}$ μία ΣΕ και $\mathcal{R}_\mathcal{L} = \langle \mathcal{R}, \rho \rangle$ μία σημειωμένη επαναγραφή $\mathcal{Q}$ με βάση την $\mathcal{T}$ που είναι κλειστή ως προς τα υποσύνολα. Έστω επίσης $\mathcal{T}^-$ ένα υποσύνολο του $\mathcal{T}$. Έστω $\mathcal{R}'_\mathcal{L} = \langle \mathcal{R}', \rho' \rangle$ το ζεύγος που παράγεται από τον αλγόριθμο* Delete, *τότε, το $\mathcal{R}'$ είναι μία κλειστή ως προς τα υποσύνολα επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T} \setminus \mathcal{T}^-$.*

Απόδειξη: Αρχικά, θα δείξουμε ότι το $\mathcal{R}'$, όπως υπολογίζεται από τον αλγόριθμο, είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T} \setminus \mathcal{T}^-$. Αρκεί να δείξουμε ότι για όλα τα στιγμιότυπα $\mathcal{I}$ ισχύει ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$, όπου τα $\mathcal{R}'_Q$, $\mathcal{R}'_D$ είναι τέτοια ώστε $\mathcal{R}' = \mathcal{R}'_Q \cup \mathcal{R}'_D$.

Αρχικά, θα δείξουμε ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) \supseteq \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$. Από την κατασκευή του αλγορίθμου προκύπτει ότι μία πρόταση c ανήκει στο $\mathcal{R}'$ μόνο αν υπάρχει $\mathcal{T}' \in \rho(c)$ τ.ω. $\mathcal{T}' \subseteq \mathcal{T} \setminus \mathcal{T}^-$. Συνεπώς, $\mathcal{T} \setminus \mathcal{T}^- \models \mathcal{T}'$. Επίσης, από τον ορισμό του ρ, ισχύει ότι το c ανήκει σε κάποια επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}'$ συνεπώς $\mathcal{T}' \cup \mathcal{Q} \models c$. Από τις δύο ανωτέρω συνθήκες και την ιδιότητα της μονοτονίας προκύπτει ότι $\mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{Q} \models c$. Αυτό ισχύει για κάθε στοιχείο του $\mathcal{R}'$ οπότε τελικά ισχύει ότι $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{R}'$, συνεπώς για κάθε $\mathcal{I}$ οι απαντήσεις στο $\mathcal{R}'$ επί του $\mathcal{I}$ είναι επίσης απαντήσεις του $\mathcal{Q}$ επί του $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{I}$.

Στη συνέχεια θα δείξουμε ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) \subseteq \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$. Αφού το $\mathcal{R}$ είναι κλειστό ως προς τα υποσύνολα, υπάρχει $\mathcal{R}'' \subseteq \mathcal{R}$ τέτοιο ώστε το $\mathcal{R}''$ να είναι μία επαναγραφή με βάση την $\mathcal{T} \setminus \mathcal{T}^-$. Είναι προφανές ότι για κάθε $c \in \mathcal{R}''$ και για κάθε $\mathcal{T}' \in \rho(c)$ έχουμε ότι $\mathcal{T}' \subseteq \mathcal{T} \setminus \mathcal{T}^-$. Συνεπώς, έπεται από την κατασκευή του $\mathcal{R}'$ ότι περιέχει το c και συνεπως ισχύει ότι $\mathcal{R}'' \subseteq \mathcal{R}'$. Άρα, $\mathcal{R}' \models \mathcal{R}''$ και έπεται ότι κάθε απάντηση του $\mathcal{Q}$ επί του $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{I}$ είναι επίσης απάντηση του $\mathcal{R}'$ επί του $\mathcal{I}$.

Τέλος, θα δείξουμε ότι η επαναγραφή $\mathcal{R}'$ είναι κλειστή ως προς τα υποσύνολα. Έστω ένα αυθαίρετο υποσύνολο $\mathcal{T}_s \subseteq \mathcal{T} \setminus \mathcal{T}^-$. Είναι προφανές ότι $\mathcal{T}_s \subseteq \mathcal{T}$ και αφού η επαναγραφή $\mathcal{R}$ είναι κλειστή ως προς τα υποσύνολα, υπάρχει κάποιο σύνολο $\mathcal{R}_s \subseteq \mathcal{R}$ τέτοιο ώστε το $\mathcal{R}_s$ να είναι επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}_s$. Συνεπώς, για κάθε $c_s \in \mathcal{R}_s$ έχουμε ότι $\mathcal{T}_s \cup \mathcal{Q} \models c_s$, οπότε υπάρχει $\mathcal{T}'_s \in \rho(c_s)$ το οποίο είναι ελάχιστο για το c_s. Συνεπως, $c_s \in \mathcal{R}'$. Αφού το c_s είναι μια τυχαία πρόταση, προκύπτει ότι $\mathcal{R}_s \subseteq \mathcal{R}'$. Επίσης δεδομένου ότι και το $\mathcal{T}_s$ είναι τυχαίο, προκύπτει ότι η επαναγραφή $\mathcal{R}'$ είναι κλειστή ως προς τα υποσύνολα. ■

Όπως αποδεικνύεται στο προηγούμενο θεώρημα, αν η αρχική επαναγραφή είναι κλειστή ως προς τα υποσύνολα τότε και το αποτέλεσμα του αλγορίθμου Delete θα είναι επίσης μία επαναγραφή κλειστή ως προς τα υποσύνολα. Συνεπώς, ο αλγόριθμος αυτός μπορεί να χρησιμοποιηθεί και για μελλοντικές αφαιρέσεις αξιωμάτων από την οντολογία.

Όπως φαίνεται από το προηγούμενο παράδειγμα, ο υπολογισμός μίας κλειστής ως προς τα υποσύνολα επαναγραφής απαιτεί την απενεργοποίηση όλων (ή των περισσοτέρων) βελτιστοποιήσεων που βασίζονται στην ιδιότητα της υπαγωγής. Ο υπολογισμός μίας τέτοιας επαναγραφής θεωρητικά είναι εφικτός και υπάρχουν ιδιαίτερα αποδοτικοί αλγόριθμοι [137] οι οποίοι μπορούν να υπολογίσουν τέτοιου τύπου επαναγραφές. Ταυτόχρονα, θεωρούμε ότι αυτός ο υπολογισμός θα πραγματοποιηθεί μία φορά εκτός-σύνδεσης. Ωστόσο, είναι προφανές ότι τίθενται το πρόβλημα της απόδοσης. Τόσο ο χρόνος υπολογισμού τους, όσο και το μέγεθός τους μπορεί να είναι σημαντικά μεγαλύτερο από τον χρόνο υπολογισμού και το μέγεθος του ελαχίστου κορεσμένου συνόλου $\mathcal{S}$ (ελάχιστο ως προς τις σχέσεις υπαγωγής). Αυτό μπορεί να έχει ως αποτέλεσμα την επανάληψη του for-βρόγχου στις γραμμές 2–5 του Αλγορίθμου 2 πολύ περισσότερες φορές. Όπως παρουσιάζεται στο Κεφάλαιο 6, πράγματι ο υπολογισμός της αρχικής, κλειστής ως προς τα υποσύνολα, επαναγραφής δεν είναι πάντα εφικτος αλλά άπαξ και αυτή υπολογιστεί ο Αλγόριθμος 2 είναι πολύ αποδοτικός.

3.5 Επαναγραφή Ερωτημάτων πάνω σε Οντολογίες που Συστέλλονται με τη χρήση Γράφων

Για να αφαιρέσει ο Αλγόριθμος 2 τις προτάσεις που δεν μπορούν να παραχθούν πλέον από το ερώτημα και τη νέα οντολογία, εφαρμόζει τον for-βρόγχο των γραμμών 2–5 στο σύνολο προτάσεων εισόδου. Το σύνολο αυτό, είτε είναι το κορεσμένο σύνολο στο οποίο έχουν εφαρμοστεί όλοι ή κάποιοι έλεγχοι υπαγωγής (Ενότητα 3.3) είτε μία κλειστή ως προς τα υποσύνολα επαναγραφή (Ενότητα 3.4), συνεπώς σε κάθε περίπτωση το μέγεθός του μπορεί να είναι αρκετά μεγαλό. Στην ενότητα αυτή, δείχνουμε ότι υπάρχουν κάποιες συγκεκριμένες συσχετίσεις μεταξύ των στοιχείων των συνόλων αυτών που αν αξιοποιήσουμε, ο αλγόριθμος μπορεί να γίνει πολύ πιο αποδοτικός.

Παράδειγμα 33. Έστω ότι η οντολογία $\mathcal{T}$ αποτελείται από τις εξής προτάσεις:

$$\begin{array}{rcrcl} c_1 & = & R_1(x,y) & \leftarrow & R_2(x,y), \\ c_2 & = & R_2(x,y) & \leftarrow & R_3(x,y), \\ c_3 & = & R_3(x,y) & \leftarrow & R_4(x,y), \\ c_4 & = & R_4(x,f(x)) & \leftarrow & A(x) \end{array}$$

και έστω το ερώτημα $\mathcal{Q} = Q(x) \leftarrow R_1(x,y)$. Υποθέτουμε ότι υπάρχει ένα σύστημα συμπερασμού που εφαρμόζει τους ακόλουθους συμπερασμούς:

$$\begin{array}{ccc} \dfrac{\mathcal{Q} \quad c_1}{\mathcal{Q}_1} & \text{όπου} & \mathcal{Q}_1 = Q(x) \leftarrow R_2(x,y) \\ \dfrac{\mathcal{Q}_1 \quad c_2}{\mathcal{Q}_2} & \text{όπου} & \mathcal{Q}_2 = Q(x) \leftarrow R_3(x,y) \\ \dfrac{\mathcal{Q}_2 \quad c_3}{\mathcal{Q}_3} & \text{όπου} & \mathcal{Q}_3 = Q(x) \leftarrow R_4(x,y) \\ \dfrac{\mathcal{Q}_3 \quad c_4}{\mathcal{Q}_4} & \text{όπου} & \mathcal{Q}_4 = Q(x) \leftarrow A(x) \end{array}$$

Τότε μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$ θα αποτελείται από τις εξής προτάσεις: $\mathcal{R} = \{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3, \mathcal{Q}_4\}$.

Έστω τώρα ότι αφαιρείται η πρόταση c_1 κατασκευάζοντας τη νέα οντολογία $\mathcal{T}' = \{c_2, c_3, c_4\}$. Μία επαναγραφή για τα $\mathcal{Q}$, $\mathcal{T}'$ αποτελείται μόνο από το ερώτημα $\mathcal{Q}$. Ο Αλγόριθμος 2 θα υπολογίσει το σύνολο αυτό ελέγχοντας για κάθε πρόταση $c \in \mathcal{R}$ αν για κάθε $\mathcal{T}_c \in \rho(c)$ (όπου το $\rho(c)$ όπως ορίζεται στον Ορισμό 25) ισχύει ότι $c_1 \in \mathcal{T}_c$, όπου σε αυτήν τη περίπτωση το c θα πρέπει να αφαιρεθεί. Στο τρέχον παράδειγμα, για κάθε $\mathcal{T}_{\mathcal{Q}_i} \in \rho(\mathcal{Q}_i)$, με $i \in \{1,2,3,4\}$, ισχύει ότι $c_1 \in \mathcal{T}_{\mathcal{Q}_i}$. Συνεπώς τα ερωτήματα $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{Q}_3$, $\mathcal{Q}_4$ αφαιρούνται.

Ωστόσο, μπορεί αποφευχθεί ο έλεγχος των ερωτημάτων $\mathcal{Q}_2$, $\mathcal{Q}_3$, $\mathcal{Q}_4$ αν τα διατάξουμε με βάση τη σειρά που προκύπτει από τα ελάχιστα τους σύνολα στο ρ. Πιο συγκεκριμένα, στο τρέχον παράδειγμα, το (μόνο) ελάχιστο σύνολο του $\mathcal{Q}_2$ ($\mathcal{T}_{\mathcal{Q}_2} = \{c_1, c_2\}$) είναι υπερσύνολο του (μόνου) ελάχιστου συνόλου του $\mathcal{Q}_1$ ($\mathcal{T}_{\mathcal{Q}_1} = \{c_1\}$), συνεπώς αν αφαιρεθεί το $\mathcal{Q}_1$ εξαιτίας του c_1 που ανήκει στο σύνολο $\mathcal{T}_{\mathcal{Q}_1}$, τότε δεν χρειάζεται να ελεχθούν όλες οι προτάσεις που παράγονται μετά το $\mathcal{Q}_1$ (δηλαδή, οι $\mathcal{Q}_2$–$\mathcal{Q}_4$). $\diamondsuit$

Για να αξιοποιήσουμε την ιδέα αυτή στον αλγόριθμο που θα παρουσιάσουμε στη συνέχεια, διατάσσουμε τα στοιχεία της αρχικής επαναγραφής με βάση τα ελάχιστα σύνολά τους. Η υλοποίηση αυτή επιτυγχάνεται με τη βοήθεια γράφου παραγωγής:

Ορισμός 34. Έστω ένα ΣΕ $\mathcal{Q}$ και μία οντολογία $\mathcal{T}$. Ένας *γράφος παραγωγής* (*derivation graph*) για τα $\mathcal{Q}, \mathcal{T}$ είναι ένας κατευθυνόμενος γράφος $\mathcal{G} = \langle \mathcal{R}, \mathcal{H} \rangle$, όπου $\mathcal{R} = (\mathcal{Q} \cup \mathcal{T})_{\Gamma, \mathcal{D}}$ (ή η κλειστή ως προς τα υποσύνολα επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T}$) και $\mathcal{H}$ είναι μία δυαδική σχέση πάνω στο $\mathcal{R}$ τέτοια ώστε $\langle c_1, c_2 \rangle \in \mathcal{H}$ αν και μόνο αν υπάρχει ελάχιστο υποσύνολο $\mathcal{U}$ του $\mathcal{Q} \cup \mathcal{R} \cup \mathcal{T}$ τέτοιο ώστε $c_1 \in \mathcal{U}$ και το c_2 παράγεται από το $\mathcal{U}$ με βάση το σύστημα συμπερασμού Γ.

Διαισθητικά, $\langle c_1, c_2 \rangle \in \mathcal{H}$ αν και μόνο αν το c_2 ανήκει στο $\mathcal{R}$ και μπορεί να παραχθεί από κάποιο στοιχείο c_1 του $\mathcal{R}$ με τη χρήση ενός ελάχιστου συνόλου προτάσεων. Η χρήση των ελάχιστων συνόλων στον ορισμό εξασφαλίζει ότι η σχέση $\mathcal{H}$ δεν είναι μεταβατική. Συγκεκριμένα, προκύπτει από τον ορισμό ότι για οποιαδήποτε $c_1, c_2, c_3 \in \mathcal{R}$, αν $\langle c_1, c_2 \rangle, \langle c_2, c_3 \rangle \in \mathcal{H}$ τότε $\langle c_1, c_3 \rangle \in \mathcal{H}$ μόνο αν υπάρχει ελάχιστο υποσύνολο $\mathcal{U}$ του $q \cup \mathcal{R} \cup \mathcal{T}$, που να παράγει το c_3, τέτοιο ώστε $c_1 \in \mathcal{U}$ και $c_2 \notin \mathcal{U}$. Για παράδειγμα, έστω η οντολογία $\mathcal{T} = \{A(x) \leftarrow B(x), B(x) \leftarrow C(x), A(x) \leftarrow C(x)\}$ και το ΣΕ $\mathcal{Q} = Q(x) \leftarrow A(x)$. Τότε το $\mathcal{R}$ θα αποτελείται από τα ΣΕ $\{\mathcal{Q}, \mathcal{Q}_1, \mathcal{Q}_2\}$ με $\mathcal{Q}_1 = Q(x) \leftarrow B(x)$, $\mathcal{Q}_2 = Q(x) \leftarrow C(x)$ και το $\mathcal{H}$ από τις ακμές $\langle \mathcal{Q}, \mathcal{Q}_1 \rangle$, $\langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle$, $\langle \mathcal{Q}, \mathcal{Q}_2 \rangle$. Αν η $\mathcal{T}$ όμως δεν περιείχε το αξίωμα $A(x) \leftarrow C(x)$, το $\mathcal{H}$ δεν θα περιείχε την ακμή $\langle \mathcal{Q}, \mathcal{Q}_2 \rangle$.

Στο Παράδειγμα 33, το $\mathcal{G} = \langle \mathcal{R}, \mathcal{H} \rangle$, με $\mathcal{H} = \{\langle \mathcal{Q}, \mathcal{Q}_1 \rangle, \langle \mathcal{Q}_1, \mathcal{Q}_2 \rangle, \langle \mathcal{Q}_2, \mathcal{Q}_3 \rangle, \langle \mathcal{Q}_3, \mathcal{Q}_4 \rangle\}$, είναι ένας γράφος παραγωγής για τα $\mathcal{Q}, \mathcal{T}$.

Ένας γράφος παραγωγής μπορεί να υπολογιστεί από ένα σημειωμένο σύνολο $\mathcal{R}$ (το οποίο μπορεί να είναι είτε το κορεσμένο σύνολο είτε μία επαναγραφή) συγκρίνοντας τις ετικέτες $\rho(c_1)$ και $\rho(c_2)$ για κάθε c_1, c_2 του $\mathcal{R}$ και κατασκευάζοντας την μεταβατικά μειωμένη (transitively reduced) δομή $\mathcal{H}$. Εναλλακτικά, μπορεί να κατασκευαστεί τροποποιώντας την εσωτερική δομή ενός αλγορίθμου επαναγραφής ώστε να κατασκευάζει σταδιακά τον γράφο καθώς παράγονται οι προτάσεις. Με τον τρόπο αυτό, μπορεί να αξιοποιηθεί η δομή του γράφου μειώνοντας σημαντικά το μέγεθος της αντιστοίχισης ρ. Συγκεκριμένα, για κάθε πρόταση c_2 με $\langle c_1, c_2 \rangle \in \mathcal{H}$ αρκεί να αποθηκεύσουμε τις υποθέσεις που χρειάστηκαν για τον συμπερασμό του c_2 από το c_1, αντί για όλα τα αξιώματα της $\mathcal{T}$ που χρειάζονται για την παραγωγή της c_2.

Ορισμός 35. Ένας *επισημειωμένος γράφος παραγωγής* (*labelled derivation graph*) $\mathcal{G}_{\mathcal{L}}$ για ένα ΣΕ $\mathcal{Q}$ και μία οντολογία $\mathcal{T}$ είναι ένας επισημειωμένος γράφος $\langle \mathcal{R}, \mathcal{H}, \mu \rangle$, όπου $\langle \mathcal{R}, \mathcal{H} \rangle$ ο γράφος παραγωγής και μ μία αντιστοίχιση από το $\mathcal{R}$ στο $2^{\mathcal{Q} \cup \mathcal{R} \cup \mathcal{T}}$ που ορίζεται ως εξής:

- αν δεν υπάρχει c_1 με $\langle c_1, c_2 \rangle \in \mathcal{H}$, τότε το $\mu(c_2)$ αποτελείται από τα ελάχιστα υποσύνολα $\mathcal{U}$ του $\mathcal{T} \cup \mathcal{Q}$ τέτοια ώστε το c_2 να παράγεται από το $\mathcal{U}$ με το σύστημα συμπερασμού Γ.

- αν υπάρχουν $\langle c_1, c \rangle, \ldots, \langle c_n, c \rangle \in \mathcal{H}$ τότε το $\mu(c)$ επιστρέφει όλα τα ελάχιστα υποσύνολα S του $\{c_1, \ldots, c_n\} \cup \mathcal{T}$ τέτοια ώστε το c να παράγεται από το S και, για κάθε $1 \leq i \leq n$, το c_i να μην παράγεται από το $S \setminus \{c_i\}$ με το σύστημα συμπερασμού Γ.

Σύμφωνα με τον Ορισμό 35, η επισημείωση του Παραδείγματος 33 θα μετασχηματιστεί ως εξής:

$$\begin{array}{rcl}
\mu(\mathcal{Q}) & = & \{\{\mathcal{Q}\}\} \\
\mu(\mathcal{Q}_1) & = & \{\{\mathcal{Q}, c_1\}\} \\
\mu(\mathcal{Q}_2) & = & \{\{\mathcal{Q}_1, c_2\}\} \\
\mu(\mathcal{Q}_3) & = & \{\{\mathcal{Q}_2, c_3\}\} \\
\mu(\mathcal{Q}_4) & = & \{\{\mathcal{Q}_3, c_4\}\}
\end{array}$$

3.5.1 Αλγόριθμος

Ο Αλγόριθμος 4 αξιοποιεί τη δομή του γράφου εισόδου και υπολογίζει μία επαναγραφή με βάση τη μειωμένη οντολογία. Ο αλγόριθμος δέχεται την ίδια είσοδο με τον αλγόριθμο $\mathsf{Delete}_{\Gamma,\mathcal{D}}$, αλλά αντί για ένα σημειωμένο σύνολο (κορεσμένο σύνολο ή επαναγραφή) δέχεται έναν σημειωμένο γράφο παραγωγής $\mathcal{G}_{\mathcal{L}}$. Στην αρχή, αρχικοποιεί ένα νέο σύνολο $\mathcal{R}'$ το οποίο είναι η έξοδος. Στη συνέχεια, διατρέχεται το $\mathcal{G}_{\mathcal{L}}$ κατά βάθος (με τη χρήση της στοίβας S) και ελέγχει αν για ένα στοιχείο c του γράφου υπάρχει σύνολο $\mu(c)$ που να είναι υποσύνολο του $\mathcal{Q} \cup \mathcal{R} \cup \mathcal{T} \setminus \mathcal{T}^-$, δηλαδή αν υπάρχει σύνολο υποθέσεων που να συμπεραίνει το c και τα στοιχεία του να μπορούν ακόμα να παραχθούν (γραμμή 5). Αν το c μπορεί ακόμα να παραχθεί, τότε όλοι οι επόμενοι κόμβοι του γράφου που δεν έχουν ήδη προστεθεί στο $\mathcal{R}'$ προστίθενται στη στοίβα (γραμμές 7). Τέλος, όπως στον αλγόριθμο $\mathsf{Delete}_{\Gamma,\mathcal{D}}$, συνεχίζει εφαρμόζοντας τους απαραίτητους συμπερασμούς. Πρέπει να σημειωθεί ότι με την τεχνική αυτή δεν είναι εφικτή η ανανέωση κάθε ετικέτας κατά το τρέξιμο του αλγορίθμου. Ωστόσο, αυτό θεωρούμε ότι δεν αποτελεί ιδιαίτερο πρόβλημα αφού αρκεί να διατρέξουμε μία φορά εκτός σύνδεσης τον νέο γράφο για να ανανεώσουμε τις ετικέτες.

Παράδειγμα 36. Έστω η οντολογία $\mathcal{T}$ και το ΣΕ $\mathcal{Q}$ του Παραδείγματος 33. Τότε, ο επισημειωμένος γράφος παραγωγής για τα $\mathcal{Q}, \mathcal{T}$ θα είναι ο $\mathcal{G}_{\mathcal{L}}$, όπως αυτός εμφανίζεται στο Σχήμα 3.1.

Αν θεωρήσουμε ότι $\mathcal{T}^- = \emptyset$, τότε ο αλγόριθμος $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ θα προσθέσει αρχικά στο S και στη συνέχεια στο $\mathcal{R}$ το $\mathcal{Q}$ και τα αξιώματα c_1, c_2, c_3, c_4 της $\mathcal{T}$. Αφού τα $\mathcal{Q}$, c_1 ανήκουν στο $\mathcal{R}$, θα προστεθεί και το "παιδί" τους $\mathcal{Q}_1$. Ομοίως, θα προστεθούν και τα $\mathcal{Q}_2$, $\mathcal{Q}_3$,

Algorithm 4 G-Delete$_{\Gamma,\mathcal{D}}(\mathcal{G}_\mathcal{L}, \mathcal{T}, \mathcal{T}^-, \lambda_\mathcal{R})$

Input: $\mathcal{G}_\mathcal{L} = \langle \mathcal{R}, \mathcal{H}, \mu \rangle$ σημειωμένος γράφος παραγωγής, $\mathcal{T}$, $\mathcal{T}^-$ σύνολα αξιωμάτων, $\lambda_\mathcal{R}$ αντιστοίχιση
Output: $\mathcal{R}'$ επαναγραφή για την μειωμένη οντολογία

$\mathcal{R}' := \emptyset$
Αρχικοποίηση της στοίβας S ώστε να περιέχει όλες τις προτάσεις c του $\mathcal{R}$ τ.ω. δεν υπάρχει c' με $\langle c', c \rangle \in \mathcal{H}$
while $S \neq \emptyset$ **do**
  Εξάγουμε ένα στοιχείο c από το S
  if υπάρχει $\mathcal{U} \in \mu(c)$ με $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{R}' \cup (\mathcal{T} \setminus \mathcal{T}^-)$ **then**
    $\mathcal{R}' := \mathcal{R}' \cup \{c\}$
    Ώθησε όλα τα c' τ.ω. $\langle c, c' \rangle \in \mathcal{H}$ και $c' \notin \mathcal{R}'$ to S
  end if
end while
$\mathcal{T}^+ := \{c' \mid$ 1. $c' \in \lambda_\mathcal{R}(c)$ με $c \in S \setminus \mathcal{R}'$
    2. υπάρχει $\mathcal{U} \in \mu(c')$ τ.ω. $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{R}' \cup (\mathcal{T} \setminus \mathcal{T}^-)$
    3. δεν υπάρχει $c'' \in \mathcal{R}'$ με $c' \in \lambda_\mathcal{R}(c'')\}$
$\mathcal{R}' :=$ Add$_{\Gamma,\mathcal{D}}(\mathcal{R}', \mathcal{T}^+)$
return $\mathcal{R}'$

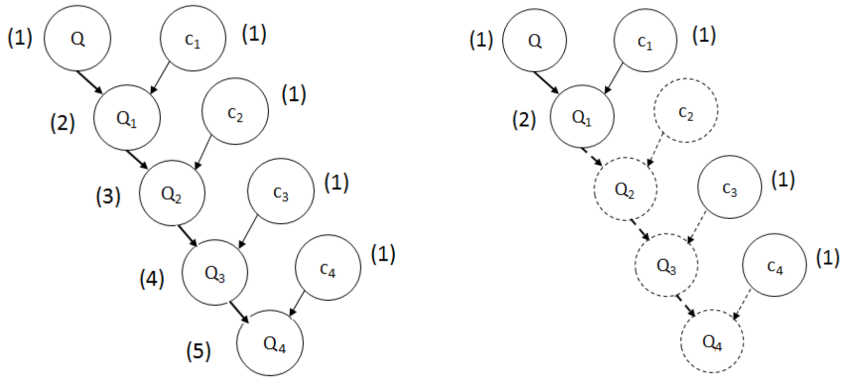
$\mathcal{Q}_4$. Στον αριστερό γράφο του Σχήματος 3.1 αναπαρίσταται με τα αντίστοιχα νούμερα σε παρενθέσεις η σειρά με την οποία θα επεξεργαστεί ο Delete$_{\Gamma,\mathcal{D}}$ τον γράφο $\mathcal{G}_\mathcal{L}$.

Αν υποθέσουμε ότι $\mathcal{T}^- = \{c_2\}$, τότε όπως και προηγουμένως, ο αλγόριθμος θα προσθέσει αρχικά στο S και στη συνέχεια στο $\mathcal{R}$ το $\mathcal{Q}$ και τα αξιώματα c_1, c_2, c_4. Επίσης, θα προσθέσει το $\mathcal{Q}_1$ για τους ίδιους λόγους, αλλά δεν θα συνεχίσει με τη προσθήκη του $\mathcal{Q}_2$, διότι η υπόθεση c_2 δεν ανήκει στο $\mathcal{R}$. Ομοίως, δεν θα προσθέσει το $\mathcal{Q}_3$, διότι η απαραίτητη για την παραγωγή του υπόθεση, $\mathcal{Q}_2$, δεν ανήκει στο $\mathcal{R}$. Ομοίως για το $\mathcal{Q}_4$. Στον δεξία γράφο του Σχήματος 3.1 αναπαρίσταται η σειρά με την οποία θα επεξεργαστεί ο αλγόριθμος τον γράφο $\mathcal{G}_\mathcal{L}$. $\diamondsuit$

Σε περίπτωση που ο αλγόριθμος Add$_{\Gamma,\mathcal{D}}$ δεν κληθεί, τότε απο τη στιγμή που ο γράφος διατρέχεται κατά βάθος συμπεραίνουμε ότι η χείριστη πολυπλοκότητα του Delete είναι $O(|\mathcal{R}|^2)$. Διαφορετικά η πολυπλοκότητα του αλγορίθμου εξαρτάται από την πολυπλοκότητα του συστήματος επαναγραφής που χρησιμοποιεί το σύστημα συμπερασμού Γ. Στη συνέχεια θα δείξουμε την ορθότητα του Αλγορίθμου 4.

3.5.2 Απόδειξη Ορθότητας

Θεώρημα 37. *Έστω $\mathcal{T} \in \mathcal{L}$ μία οντολογία, έστω $\mathcal{Q}$ ένα ΣΕ και έστω $\mathcal{G}_\mathcal{L} = \langle \mathcal{R}, \mathcal{H}, \mu \rangle$ ένας επισημειωμένος γράφος παραγωγής για τα $\mathcal{Q}$, $\mathcal{T}$, όπου το $\mathcal{R}$ είναι είτε το σύνολο $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$ ή μία κλειστή ως προς τα υποσύνολα επαναγραφή, που παράγεται από έναν αλγόριθμο επα-*



Σχήμα 3.1: *Σειρά επεξεργασίας του $\mathcal{G}_\mathcal{L}$ από τον* $\mathsf{Delete}_{\Gamma,\mathcal{D}}$ *για $\mathcal{T}^- = \emptyset$ και $\mathcal{T}^- = \{c_2\}$*

ναγραφής που βασίζεται στη μέθοδο της ανάλυσης για τη γλώσσα $\mathcal{L}$. Έστω επίσης το $\lambda_\mathcal{R}$ όπως ορίζεται στον Ορισμό 27 για κάθε $c \in \mathcal{R}$ και έστω $\mathcal{T}^-$ ένα υποσύνολο της $\mathcal{T}$. Αν ο αλγόριθμος επαναγραφής τερματίζει για κάθε είσοδο, τότε όταν ο αλγόριθμος $\mathsf{G\text{-}Delete}_{\Gamma,\mathcal{D}}$ *εφαρμόζεται στα $\mathcal{G}_\mathcal{L}$, $\mathcal{T}$, $\mathcal{T}^-$ και $\lambda_\mathcal{R}$ τερματίζει. Έστω $\mathcal{R}'$ το σύνολο των προτάσεων που παράγεται από τον αλγόριθμο, τότε το $\mathcal{R}'$ είναι μία επαναγραφή για τα $\mathcal{Q}$, $\mathcal{T} \setminus \mathcal{T}^-$.*

Απόδειξη: Αρχικά θα δείξουμε ότι ο αλγόριθμος τερματίζει. Έστω ότι υπάρχει κύκλος στον $\mathcal{G}_\mathcal{L}$—δηλαδή, έστω ότι υπάρχουν κόμβοι c_1, c_2 τ.ω. το c_2 να είναι προσεγγίσιμο από το c_1 και το c_1 από το c_2. Θα μελετήσουμε τις εξής περιπτώσεις για τα c_1, c_2: (i) και οι δύο προτάσεις c_1, c_2 ικανοποιούν τη συνθήκη της γραμμής 5, (ii) η πρόταση c_1 δεν ικανοποιεί τη συνθήκη της γραμμής 5 ενώ η c_2 την ικανοποιεί. Για όλες τις άλλες περιπτώσεις η απόδειξη είναι προφανής. Αν ισχύει το (i) και η c_1 έχει ήδη προστεθεί στο $\mathcal{R}'$ και προσεγγιστεί από το c_2, τότε από τη γραμμή 7 προκύπτει ότι δεν θα ξαναπροστεθεί στην στοίβα (το ίδιο ισχύει αν αντι για το c_1, το c_2 προστεθεί πρώτα στο $\mathcal{R}'$). Αν ισχύει το (ii) τότε ενώ το c_1 θα προστεθεί τελικά στη στοίβα (εξαιτίας του c_2), το c_2 δεν θα επεξεργαστεί δεύτερη φορά αφού κανένα από τα παιδιά του c_1 δεν θα προστεθεί στη στοίβα.

Στη συνέχεια, θα δείξουμε ότι το $\mathcal{R}'$ όπως υπολογίζεται από τον αλγόριθμο είναι μία επαναγραφή του $\mathcal{Q}$ με βάση την $\mathcal{T} \setminus \mathcal{T}^-$. Αρκεί να δείξουμε ότι για όλα τα στιγμιότυπα $\mathcal{I}$ ισχύει ότι $\mathsf{cert}(\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^- \cup \mathcal{I}) = \mathsf{cert}(\mathcal{R}'_Q, \mathcal{I} \cup \mathcal{R}'_D)$, όπου τα $\mathcal{R}'_Q$, $\mathcal{R}'_D$ είναι τέτοια ώστε $\mathcal{R}' = \mathcal{R}'_Q \cup \mathcal{R}'_D$.

($\supseteq$) Θα δείξουμε τον ισχυρισμό με πλήρη επαγωγή στην σειρά με την οποία προστίθενται οι προτάσεις στο $\mathcal{R}'$. Έστω $\mathcal{R}'_k$ το σύνολο των προτάσεων στο βήμα k του αλγορίθμου.

- *Βασική περίπτωση (i=0):* Η δομή του αλγορίθμου είναι τέτοια ώστε στην αρχή προσθέτει στο $\mathcal{R}'$ τις ρίζες, δηλαδή, όλες τις προτάσεις c για τις οποίες δεν υπάρχει c' τέτοιο ώστε $\langle c', c\rangle$. Μία ρίζα c προστίθεται στο $\mathcal{R}'$ αν υπάρχει ένα σύνολο $\mathcal{U} \in \mu(c)$ τέτοιο ώστε $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$, ή, από τον Ορισμό 35, $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^- \models c$.

- *Επαγωγικό βήμα:* Θα δείξουμε ότι για κάθε $c_k \in \mathcal{R}'$, $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^- \models c_k$. Εφόσον $c_k \in \mathcal{R}'$ συμπεραίνουμε ότι υπάρχει ένα σύνολο $c_1, \ldots, c_{k-1} \in \mathcal{R}'$ με $\langle c_i, c_k\rangle \in \mathcal{H}$ (για κάθε $1 \leq i \leq k-1$) και ένα υποσύνολο $\mathcal{T}'$ του $\mathcal{T} \setminus \mathcal{T}^-$ τέτοιο ώστε

$\{c_1, \ldots, c_{k-1}\} \cup \mathcal{T}' \models c_k$. Από την επαγωγική υπόθεση ισχύει ότι $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^- \models c_i$ (για κάθε $1 \leq i \leq k-1$), συνεπώς $\mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^- \models c_k$. Τέλος, από το Θεώρημα 32, συμπεραίνουμε ότι το ίδιο ισχύει και για τις προτάσεις που προστίθενται στο $\mathcal{R}'$ στη γραμμή 11 του αλγορίθμου.

Συνεπώς για όλα τα στοιχεία του $\mathcal{R}'$ ισχύει ότι $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{Q} \models \mathcal{R}'$, που σημαίνει ότι για κάθε $\mathcal{I}$ οι απαντήσεις του $\mathcal{R}'$ στο $\mathcal{I}$ είναι επίσης απαντήσεις του $\mathcal{Q}$ στο $(\mathcal{T} \setminus \mathcal{T}^-) \cup \mathcal{I}$.

($\subseteq$) Έστω $\mathcal{R}''$ μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T} \setminus \mathcal{T}^-$ όπως παράγεται από έναν αλγόριθμο επαναγραφής που βασίζεται στην μέθοδο της ανάλυσης και που χρησιμοποιεί το σύστημα συμπερασμού Γ. Έστω επίσης $\mathcal{R}''_k$ το σύνολο προτάσεων στο βήμα k. Έστω S_l η στοίβα και $\mathcal{R}'_l$ το σύνολο $\mathcal{R}'$ του αλγορίθμου $\textsf{G-Delete}_{\Gamma,D}$ στην επανάληψη l του βρόγχου στις γραμμές 3–9. Έστω $S_n = \emptyset$ η στοίβα όπως έχει σχηματιστεί στη γραμμή 9. Η απόδειξη προκύπτει από την ακόλουθη πρόταση που θα δείξουμε στη συνέχεια:

Ιδιότητα 38. *Για κάθε $k \geq 0$ και $c \in \mathcal{R}''_k$ ισχύει:*

1. *είτε υπάρχουν $c' \in \mathcal{R}$ και $l \leq n$ τέτοια ώστε το c' να υπάγει το c και να ισχύει ένα από τα ακόλουθα:*

   1.1 *υπάρχει $c' \in S_l$ και ένα σύνολο $\mathcal{U} \in \mu(c')$ τέτοιο ώστε $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{R}'_l \cup \mathcal{T} \setminus \mathcal{T}^-$,*

   1.2 *υπάρχει ένα ελάχιστο σύνολο $\{c_1, \ldots, c_p\} \subseteq \mathcal{R}'_l$ τέτοιο ώστε για κάθε c_j με $1 \leq j \leq p$ υπάρχει μονοπάτι στο $\mathcal{H}$ από το c_j στο c', $\{c_1, \ldots, c_p\} \vdash_\Gamma c'$ και υπάρχει $m \geq l$ τέτοιο ώστε $c' \in S_m$,*

   1.3 *$c' \in \mathcal{R}'_l$,*

2. *είτε υπάρχει μία πρόταση $c' \in \textsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{R}', \mathcal{T}^+)$ με $c' \notin \mathcal{R}$ τέτοια ώστε το c' να υπάγει το c.*

Θα δείξουμε την Ιδιότητα 38 επαγωγικά στην παραγωγή του $\mathcal{R}''_k$ για την περίπτωση που το σύνολο εισόδου είναι το $(\mathcal{Q} \cup \mathcal{T})_{\Gamma,\mathcal{D}}$. Αν το σύνολο εισόδου είναι η κλειστή ως προς τα υποσύνολα επαναγραφή τότε η απόδειξη είναι παρόμοια.

- *Βασική περίπτωση (i=0):* Από τον ορισμό των αλγορίθμων επαναγραφής που βασίζονται στη μέθοδο της ανάλυσης ισχύει ότι $\mathcal{R}''_0 = \mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$. Υπάρχουν δύο περιπτώσεις για το $c \in \mathcal{R}''_0$. Πρώτη περίπτωση: Υπάρχει αντικατάσταση σ τέτοια ώστε $c\sigma \in S_0$. Επίσης, προφανώς υπάρχει σύνολο $\mathcal{U} \in \mu(c)$ τέτοιο ώστε $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$. Συνεπώς, ισχύει η Περίπτωση 1.1 της ιδιότητας. Δεύτερη περίπτωση: υπάρχει μοναδικό $c' \in S_0$ τέτοιο ώστε το c' να υπάγει το c και για κάθε $\mathcal{U}' \in \mu(c'), \mathcal{U}' \cap \mathcal{T}^- \neq \emptyset$. Εφόσον $c \in \mathcal{R}''_0$, ισχύει ότι $c \in \mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$ άρα από τον Ορισμό 34, υπάρχει σύνολο $\mathcal{U} \in \mu(c)$ τέτοιο ώστε $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{T} \setminus \mathcal{T}^-$. Προκύπτει από τη γραμμή 10 του αλγορίθμου ότι $c \in \mathcal{T}^+$, και άρα ισχύει η Περίπτωση 2 της ιδιότητας.

- *Επαγωγικό βήμα:* Έστω ότι στο βήμα k για κάθε στοιχείο του $\mathcal{R}''_k$, ισχύει η Ιδιότητα 38. Τότε, στο βήμα $k+1$ θα παραχθεί κάποια πρόταση c από κάποιον συμπερασμό του οποίου οι υποθέσεις θα ανήκουν στο $\mathcal{R}''_k$. Έστω $\langle c_1, \dots, c_r; c\rangle$ ένας τέτοιος συμπερασμός. Από την επαγωγική υπόθεση θα ισχύουν ένα από τα παρακάτω: a) κάθε c_j, με $1 \leq j \leq r$, ικανοποιεί την περίπτωση 1, b) υπάρχει $m < r$ τέτοιο ώστε τα $c_1, \dots, c_m$ ικανοποιούν την Περίπτωση 1 και τα $c_{m+1}, \dots, c_r$ ικανοποιούν την Περίπτωση 2, c) κάθε c_j, με $1 \leq j \leq r$, ικανοποιεί την Περίπτωση 2.

  Έστω ότι ισχύει το a) και ότι για κάθε c_j, με $1 \leq j \leq r$, ισχύει η Περίπτωση 1.1 της ιδιότητας, δηλαδή, υπάρχουν $c'_j \in S_l$ που υπάγουν κάθε $c_j \in \mathcal{R}''_k$ και είναι τέτοια ώστε να υπάρχει κάποιο σύνολο $\mathcal{U}'_j \in \mu(c'_j)$ με $\mathcal{U}'_j \subseteq \mathcal{Q} \cup \mathcal{R}'_l \cup \mathcal{T} \setminus \mathcal{T}^-$. Συνεπώς, προκύπτει από τη γραμμή 5 ότι υπάρχει ένα $m \geq l$ τέτοιο ώστε $c'_j \in \mathcal{R}'_m$. Επίσης, από τις ιδιότητες της υπαγωγής, υπάρχει ένας συμπερασμός $\langle c'_1, \dots, c'_r; c'\rangle$ τέτοιος ώστε το c' να υπάγει το c. Από τον ορισμό του γράφου παραγωγής, αν το $c' \in \mathcal{R}$ τότε υπάρχουν ακμές $\langle c'_j, c'\rangle \in \mathcal{H}$. Επίσης, προκύπτει από τη γραμμή 7 ότι $c' \in S_m$. Συνεπώς ισχύει η Περίπτωση 1.2 της ιδιότητας. Αν $c' \notin \mathcal{R}$ τότε πρέπει να υπάρχει κάποια πρόταση $c'' \in \mathcal{R}$ που να υπάγει το c'. Αν $c'' \in \mathcal{R}'_n$ τότε ισχύει η Περίπτωση 1.3 της ιδιότητας. Αν $c'' \notin \mathcal{R}'_n$, τότε αφού $\{c'_1, \dots, c'_r\} \in \mu(c')$, ισχύει ότι $c' \in \mathcal{T}^+$, συνεπώς ισχύει η Περίπτωση 2 της ιδιότητας.

  Έστω ότι ισχύει το a) και για κάθε c_j, με $1 \leq j \leq r$, ισχύει η Περίπτωση 1.2 της ιδιότητας, δηλαδή,έστω ότι υπάρχει $c'_j \in \mathcal{R}$ που να υπάγει το c_j και ότι υπάρχει ένα ελάχιστο σύνολο $\{c_{j1}, \dots, c_{jk_j}\} \subseteq \mathcal{R}'_l$, με $l \leq n$, τέτοιο ώστε να υπάρχει ένα μονοπάτι από το c_{jt} στο c'_j (για κάθε $1 \leq t \leq k_j$) και $\{c_{j1}, \dots, c_{jk_j}\} \vdash_\Gamma c'_j$. Από τις ιδιότητες της υπαγωγής, υπάρχει ένας συμπερασμός $\langle c'_1, \dots, c'_p; c'\rangle$ τέτοιος ώστε το c' να υπάγει το c. Επίσης, αφού ισχύει ότι $\{c_{j1}, \dots, c_{jk_j}\} \vdash_\Gamma c'_j$, για κάθε $1 \leq j \leq r$, και $\langle c'_1, \dots, c'_p; c'\rangle$ συμπεραίνουμε ότι $\{c_{11}, \dots, c_{1k_1}, \dots, c_{r1}, \dots, c_{rk_r}\} \vdash_\Gamma c'$. Επιπροσθέτως, ισχύει ότι για κάθε $1 \leq j \leq r$, υπάρχει ένα σύνολο προτάσεων $\mathcal{U}_j \in \mu(c'_j)$ τέτοιο ώστε $\mathcal{U}_j \subseteq \mathcal{Q} \cup \mathcal{R}'_l \cup \mathcal{T} \setminus \mathcal{T}^-$. Συνεπώς, υπάρχει $m \geq l$ τέτοιο ώστε $c'_j \in \mathcal{R}'_m$. Αν $c' \in \mathcal{R}$ τότε από τον ορισμό του γράφου παραγωγής θα υπάρχουν οι ακμές $\langle c'_j, c'\rangle \in \mathcal{H}$ για κάθε $1 \leq j \leq r$. Εφόσον ισχύει ότι $c_{jk_j} \in \mathcal{R}'_l$ και αφού υπάρχει μονοπάτι από το c_{jk_j} στο c'_j, υπάρχει επίσης μονοπάτι από το c_{jk_j} στο c'. Προκύπτει από τη δομή του αλγορίθμου ότι υπάρχει $m' \geq m$ τέτοιο ώστε $c' \in S_{m'}$. Συνεπώς, ισχύει η Περίπτωση 1.2 της ιδιότητας. Αν $c' \notin \mathcal{R}$ τότε προκύπτει όπως προηγουμένως ότι ισχύει είτε η Περίπτωση 1.3 είτε η Περίπτωση 2 (αφού $c'_i \in \mathcal{R}'_n$, υπάρχει ένα σύνολο $\mathcal{U} \in \mu(c')$ τέτοιο ώστε $\mathcal{U} \subseteq \mathcal{Q} \cup \mathcal{R}'_n \cup \mathcal{T} \setminus \mathcal{T}^-$).

  Έστω ότι ισχύει το a) και ότι για κάθε c_j, με $1 \leq j \leq r$, ισχύει η Περίπτωση 1.3, δηλαδή υπάρχει ήδη μία πρόταση c'_j στο $\mathcal{R}'_l$ που να υπάγει το c_j. Αφού οι προτάσεις αυτές ανήκουν στο $\mathcal{R}'_l$, ισχύει για κάθε μία από αυτές η συνθήκη της γραμμής 5. Επίσης από τις ιδιότητες της υπαγωγής ισχύει ότι υπάρχει ένας συμπερασμός $\langle c'_1, \dots, c'_r; c'\rangle$ όπου το c' υπάγει το c ενώ ταυτόχρονα, από τον ορισμό του γράφου παραγωγής ισχύει ότ αν $c' \in \mathcal{R}$ τότε πρέπει να υπάρχουν οι ακμές $\langle c'_j, c'\rangle$ στο $\mathcal{H}$.

Συνεπώς, προκύπτει από τη δομή του αλγορίθμου ότι το c' θα προστεθεί στο S_l στη γραμμή 7 και συνεπώς ισχύει η Περίπτωση 1.2 της ιδιότητας για το c'. Αν $c' \notin \mathcal{R}$ τότε πρέπει να υπάρχει κάποια πρόταση $c'' \in \mathcal{R}$ που να υπάγει την c'. Αν $c'' \in \mathcal{R}'_n$ τότε ισχύει η Περίπτωση 1.3 της ιδιότητας. Αν $c'' \notin \mathcal{R}'_n$ τότε αφού $c'_1, \ldots, c'_r \in \mathcal{R}'_n$ $(\mathcal{R}'_l \subseteq \mathcal{R}'_n)$ και $\{c'_1, \ldots, c'_r\} \in \mu(c')$, $c' \in \mathcal{T}^+$, ισχύει η Περίπτωση 2 της ιδιότητας.

Έστω ότι ισχύει το a) και ότι υπάρχουν r_1, r_2, με $1 \leq r_1 < r_2 < r$, τέτοια ώστε τα $c_1, \ldots, c_{r_1}$ ικανοποιούν την Περίπτωση 1.1, τα $c_{r_1+1}, \ldots, c_{r_2}$ ικανοποιούν την Περίπτωση 1.2 και τα $c_{r_2+1}, \ldots, c_r$ ικανοποιούν την Περίπτωση 1.3. Ακολουθώντας την ίδια διαδικασία όπως στις περιπτώσεις που όλες οι υποθέσεις ικανοποιούν είτε την Περίπτωση 1.1 είτε την Περίπτωση 1.2, συμπεραίνουμε ότι υπάρχουν $c'_1, \ldots, c'_{r_2}$ τέτοια ώστε το c'_j να υπάγει το c_j, για κάθε $1 \leq j \leq r_2$ και υπάρχει $m \geq l$ τέτοιο ώστε $c'_1, \ldots, c'_{r_2} \in \mathcal{R}'_m$. Από τις ιδιότητες της υπαγωγής συμπεραίνουμε ότι υπάρχει $\langle c'_1, \ldots, c'_{r_1}, \ldots, c'_{r_2}, c_{r_2+1} \ldots, c_r; c' \rangle$ τέτοιο ώστε το c' να υπάγει c. Αν $c' \in \mathcal{R}$, συμπεραίνουμε ότι για κάθε $1 \leq j \leq r$ υπάρχει μία ακμή $\langle c'_j, c' \rangle \in \mathcal{H}$. Επίσης, αφού $c'_1, \ldots, c'_{r_2} \in \mathcal{R}'_m$ και $c_{r_2+1}, \ldots, c_r \in \mathcal{R}'_l$, συμπεραίνουμε ότι υπάρχει ένα $m' \geq m, l$ τέτοιο ώστε $c' \in \mathcal{S}_{m'}$. Συνεπώς, από την επαγωγική υπόθεση συμπαιρένουμε ότι ισχύει η Περίπτωση 1.2. Διαφορετικά, συμπεραίνουμε, όπως προηγουμένως, ότι ισχύει είτε η Περιπτωση 1.3 είτε η Περίπτωση 2.

Έστω ότι ισχύει το b). Συγκεκριμένα έστω ότι για τα $c_1, \ldots, c_{r_1}$, $1 \leq r_1 < r$ ισχύει η Περίπτωση 2 και για τα $c_{r_1+1}, \ldots, c_r$ ισχύει η Περίπτωση 1. Τότε, για κάθε c_j, με $1 \leq j \leq r_1$, υπάρχει μία πρόταση c'_j που να υπάγει την c_j και για την οποία ισχύουν ένα από τα εξής: (i) $c'_j \in \mathcal{T}^+$ (γραμμή 10) ή (ii) το c'_j παράγεται από το $\mathcal{T}^+$ και το $\mathcal{R}'$(γραμμή 11). Όπως αποδείξαμε και στις προηγούμενες περιπτώσεις, για κάθε $r_1 < j \leq r$, υπάρχει μία πρόταση c'_j που ναυπάγει την c_j και είτε $c'_j \in \mathcal{R}'_n$ είτε $c'_j \in \mathcal{T}^+$. Συνεπώς, αφού η $\mathsf{Add}_{\Gamma,\mathcal{D}}$ είναι πλήρης και σύμφωνα με την ιδιότητα της υπαγωγής συμπεραίνουμε ότι υπάρχει μια πρόταση $c' \in \mathsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{R}', \mathcal{T}^+)$ που να υπάγει την c.

Έστω ότι ισχύει το c), δηλαδή, για κάθε c_j, με $1 \leq j \leq r$, ισχύει η Περίπτωση 2, δηλαδή υπάρχει πρόταση $c'_j \in \mathsf{Add}_{\Gamma,\mathcal{D}}(\mathcal{R}', \mathcal{T}^+)$, με $c'_j \notin \mathcal{R}$, τέτοια ώστε το c'_j να υπάγει το c_j. Προκύπτει άμεσα από την πληρότητα του $\mathsf{Add}_{\Gamma,\mathcal{D}}$ ότι η Περίπτωση 2 ισχύει επίσης για το c.

Από την απόδειξη της Ιδιότητας 38 συμπεραίνουμε ότι για κάθε $c \in \mathcal{R}''$ υπάρχει μία πρόταση c' η οποία τελικά θα προστεθεί στο $\mathcal{R}'$ και η οποία υπάγει το c. ■

3.5.3 Περαιτέρω Βελτιστοποιήσεις

Η απόδοση του Αλγορίθμου 4 μπορεί να βελτιστοποιηθεί περαιτέρω αν αξιοποιηθεί το $\lambda_{\mathcal{R}}$ ως εξής:

- όταν θα επιλέγει μία νέα πρόταση c στη γραμμή 4 να συνεχίζει με την επεξεργασία της c μόνο αν $\lambda_{\mathcal{R}}(c) = \emptyset$ ή αν καμία από τις προτάσεις του $\lambda_{\mathcal{R}}(c)$ δεν περιέχεται στο $\mathcal{R}'$, διαφορετικά το c και όλες οι προτάσεις που βρίσκονται "μετά" το c στον γράφο μπορούν να απορριφθούν συνεχίζοντας με το επόμενο στοιχείο της στοίβας.

- Στη γραμμή 7, αν υπάρχει πρόταση c_k τέτοια ώστε $\langle c, c_k \rangle \in \mathcal{H}$ και ισχύει $c_k \in \lambda_{\mathcal{R}}(c)$, τότε να ωθείται στην στοίβα μόνο το c_k και όχι τα υπόλοιπα παιδιά του c.

Είναι σημαντικό να σημειωθεί ότι η επαναγραφή που θα προκύψει από αυτές τις βελτιστοποιήσεις δεν θα είναι απαραιτήτως κλειστή ως προς τα υποσύνολα.

Κεφάλαιο 4

Πειραματική Αξιολόγηση

Στην ενότητα αυτή θα παρουσιάσουμε την πειραματική αξιολόγηση των τεχνικών που περιγράφονται στο Κεφάλαιο 3. Για να δείξουμε ότι οι προτεινόμενοι αλγόριθμοί μπορούν να εφαρμοστούν σε περισσότερα από ένα συστήματα επαναγραφής προσπαθήσαμε να δημιουργήσουμε περισσότερες από μία υλοποιήσεις για κάθε αλγόριθμο. Συγκεκριμένα, ο Αλγόριθμος $\mathsf{Add}_{\Gamma,\mathcal{D}}$ υλοποιήθηκε με τη χρήση δύο εργαλειών: του Requiem και του Rapid. Επιλέξαμε το Requiem επειδή εφαρμόζει με έναν σχετικά απλό και σαφή τρόπο την μέθοδο της ανάλυσης με ελάχιστες βελτιστοποιήσεις και αποτελεί τη βάση του εμπορικού συστήματος Blackout καθώς και του συστήματος Kyrie. Επίσης, επιλέχθηκε το Rapid επειδή είναι ένα από τα πιο γρήγορα σύγχρονα συστήματα επαναγραφής [132]. Οι υλοποιήσεις αυτές καλούνται $\mathsf{Add}_{\mathsf{Req}}$ και $\mathsf{Add}_{\mathsf{Rap}}$, αντιστοίχως.

Ο αλγόριθμος που παρουσιάζεται στην Ενότητα 3.3, βελτιστοποιημένος σύμφωνα με την τεχνική που παρουσιάζεται στην Ενότητα 3.5, υλοποιήθηκε μόνο με την χρήση του Requiem (και τον καλούμε $\mathsf{Delete}_{\mathsf{Req}}$). Κατέστη πολύ δύσκολη η εφαρμογή του αλγορίθμου μας στο Rapid καθώς απαιτεί την παρέμβαση στον κώδικα του συστήματος και η δομή του Rapid είναι αρκετά περίπλοκη για να επιτευχθεί αυτο. Επίσης, υλοποιήσαμε τον αλγόριθμο που παρουσιάζεται στην Ενότητα 3.4, ο οποίος δε βασίζεται σε κάποιο σύστημα επαναγραφής αφού δεν εφαρμόζει νέους συμπερασμούς. Ονομάζουμε το σύστημα αυτό Delete.

Για να επιτευχθεί διεξοδική αξιολόγηση των προτεινόμενων συστημάτων, αυτά ελέγχθηκαν με δύο τρόπους. Στην Ενότητα 4.1 χρησιμοποιήθηκε ένα σύνολο από πειραματικές οντολογίες στις οποίες προστίθενται ή αφαιρείται ένα σύνολο αξιωμάτων με τεχνητό και τυχαίο τρόπο. Έτσι, ελέγχεται η αποδοτικότητα των συστημάτων σε ένα πλήθος καταστάσεων, ειδικά σε περιπτώσεις μεγάλων ή μικρών αλλαγών. Στην Ενότητα 4.2 χρησιμοποιήθηκαν διαφορετικές εκδόσεις των ίδιων οντολογιών με στόχο να ελεγχθεί η απόδοση των συστημάτων μας σε πραγματικά σενάρια.

Στην ενότητα αυτή παρουσιάζεται ένα μέρος μόνο των πειραματικών αποτελεσμάτων. Το σύνολο των πειραματικών αποτελεσμάτων παρουσιάζεται στο Παράρτημα Β′.

Πίνακας 4.1: *Στατιστικά στοιχεία των οντολογιών που παρουσιάζονται στην Ενότητα 4.1.*

Οντολογία	#Εννοιών	#Ρόλων	#GCIs	#RIAs
P5X	6	5	18	4
AX	74	31	152	28
SWEET$^{\mathcal{L}}$	4,550	305	5,935	218
PTO$^{\mathcal{L}}$	4,150	22	8,802	15
PRO$^{\mathcal{L}}$	35,351	6	43,351	0
SWEET$^{\mathcal{E}}$	4,572	312	6,873	218
PTO$^{\mathcal{E}}$	4,282	22	10,355	15
PRO$^{\mathcal{E}}$	37,560	6	59,975	0

4.1 Αξιολόγηση με Τεχνητές Τροποποιήσεις

Στην ενότητα αυτή χρησιμοποιήθηκαν οκτώ DL-Lite$_{\mathcal{R}}$-οντολογίες (V, S, U, A, P5, UX, AX, P5X) οι οποίες αρχικά χρησιμοποιήθηκαν από τους Pérez-Urbina et al. [100] για την αξιολόγηση συστημάτων επαναγραφής. Σε κάθε οντολογία αντιστοιχούν πέντε ερωτήματα. Επίσης, κατασκευάστηκαν οι DL-Lite$_{\mathcal{R}}$ και $\mathcal{ELHI}$ εκδόσεις των εξής οντολογιών: SWEET 2.3[1] (Semantic Web for Earth and Environmental Terminology), PTO[2] (Periodic Table Ontology) και PRO[3] (PRotein Ontology). Για μία οντολογία X χρησιμοποιούμε τον συμβολισμό $X^{\mathcal{L}}$ για την DL-Lite$_{\mathcal{R}}$ έκδοσή της και τον συμβολισμό $X^{\mathcal{E}}$ για την $\mathcal{ELHI}$ έκδοσή της. Τα στατιστικά στοιχεία που αφορούν στις οντολογίες που παρουσιάζονται στην Ενότητα αυτή παρουσιάζονται στον Πίνακα 4.1.

Επίσης, για κάθε οντολογία χρησιμοποιήθηκαν τα πέντε ερωτήματα που κατασκευάστηκαν από τους Chortaras et al. [35] για την αξιολόγηση του Rapid. Για περαιτέρω αξιολόγηση κατασκευάσαμε άλλα δύο ερωτήματα για τις οντολογίες PTO και PRO με περισσότερα άτομα. Ο Πίνακας 4.2 περιέχει τα στατιστικά στοιχεία που αφορούν στα ερωτήματα που παρουσιάζονται στην Ενότητα αυτή. Επίσης, περιέχει τα μεγέθη των επαναγραφών τους με βάση τις οντολογίες που παρουσιάζονται στο κεφάλαιο αυτό, όπως αυτές υπολογίζονται από το Rapid. Με $|\mathcal{Q}_i|$ και vars συμβολίζουμε τον αριθμό των ατόμων και των μεταβλητών που εμφανίζονται στο σώμα του ερωτήματος και με ans τον αριθμό των μεταβλητών που εμφανίζονται στην κεφαλή του ερωτήματος. Τέλος, με |rew| συμβολίζουμε το μέγεθος της επαναγραφής. Τα ερωτήματα και τα μεγέθη των επαναγραφών για τις υπόλοιπες οντολογίες βρίσκονται στα Παραρτήματα A′.1,B′. Το ερώτημα q_i μεταξύ δύο οντολογιών $X^{\mathcal{L}}$, $X^{\mathcal{E}}$ είναι το ίδιο. Μερικές αξιοσημείωτες διαφορές στα μεγέθη των επαναγραφών μεταξύ των δύο εκδόσεων της ίδιας οντολογίας οφείλονται στην διαφορετική εκφραστικότητα των οντολογιών. Για παράδειγμα, το μέγεθος της επαναγραφής για τα $\mathcal{Q}_5$, SWEET$^{\mathcal{E}}$ είναι σημαντικά μεγαλύτερο από το αντίστοιχο μέγεθος για την SWEET$^{\mathcal{L}}$

[1] `http://sweet.jpl.nasa.gov/ontology/`
[2] `http://www.cs.man.ac.uk/~stevensr/ontology/`
[3] `http://www.obofoundry.org/`

Πίνακας 4.2: *Στατιστικά στοιχεία των ερωτημάτων και τα μεγέθη των επαναγραφών τους μβτ οντολογίες των οποίων τα αποτελέσματα παρουσιάζονται στις Ενότητες 4.1, 4.2*

Οντολογία	Ερώτημα	$\lvert\mathcal{Q}_i\rvert$	vars	ans	|rew|
P5X	$\mathcal{Q}_5$	5	6	1	37
AX	$\mathcal{Q}_4$	3	2	1	75
SWEET$^{\mathcal{L}}$	$\mathcal{Q}_5$	5	3	2	382
PTO$^{\mathcal{L}}$	$\mathcal{Q}_4$	3	2	1	1,034
	$\mathcal{Q}_5$	3	2	1	1,168
	$\mathcal{Q}_6$	5	3	1	1,171
PRO$^{\mathcal{L}}$	$\mathcal{Q}_2$	2	2	1	1,356
	$\mathcal{Q}_3$	1	1	1	33,919
	$\mathcal{Q}_4$	2	2	1	34,879
	$\mathcal{Q}_5$	4	3	1	27,907
	$\mathcal{Q}_6$	6	4	1	2,458
	$\mathcal{Q}_7$	5	3	1	1,364
SWEET$^{\mathcal{E}}$	$\mathcal{Q}_5$	5	3	2	822
PTO$^{\mathcal{E}}$	$\mathcal{Q}_4$	3	2	1	1,609
	$\mathcal{Q}_5$	3	2	1	1,743
	$\mathcal{Q}_6$	5	3	1	1,746
	$\mathcal{Q}_7$	5	4	1	1,739
PRO$^{\mathcal{E}}$	$\mathcal{Q}_5$	4	3	1	79,427
	$\mathcal{Q}_6$	6	4	1	53,983
	$\mathcal{Q}_7$	5	3	1	52,881

επειδή στο $\mathcal{Q}_5$ υφίσταται το άτομο Region το οποίο συμμετέχει έμμεσα σε οκτώ αξιώματα της μορφής $\mathsf{Region}(x) \leftarrow R(x,y) \wedge B(y)$, τα οποία δεν επιτρέπονται DL-Lite$_{\mathcal{R}}$ και συνεπώς δεν ανήκουν στην SWEET$^{\mathcal{L}}$. Ταυτόχρονα, τα $R(x,y), B(y)$ συμμετέχουν σε πολλά άλλα αξιώματα.

Σε όλα τα πειράματα που εκτελέστηκαν όλα τα συστήματα επέστρεψαν επαναγραφές ίδιου μεγέθους οπότε είναι άσκοπη η παρουσίαση των αριθμών αυτών.

Εφόσον ο στόχος της αξιολόγησης αυτής είναι η μελέτη του τρόπου με τον οποίο οι αλγόριθμοι που προτείνονται βελτιστοποιούν την απόδοση των υπάρχοντων συστημάτων επαναγραφής, συγκρίνουμε τα αποτελέσματα των υλοποιήσεών μας με τα αντίστοιχα αποτελέσματα των συστημάτων επαναγραφής πάνω στα οποία έχουν εφαρμοστεί οι υλοποιήσεις αυτές. Για παράδειγμα, συγκρίνεται η απόδοση του $\mathsf{Add}_{\mathsf{Req}}$ ($\mathsf{Delete}_{\mathsf{Req}}$) σε σχέση με το Requiem και η απόδοση του $\mathsf{Add}_{\mathsf{Rap}}$ σε σχέση με το Rapid. Εφόσον ο αλγόριθμος Delete δεν χρησιμοποιεί κάποιο σύστημα επαναγραφής, η απόδοσή του συγκρίνεται και με τα δύο συστήματα. Σύμφωνα με τη μέθοδο που ακολουθήθηκε, συγκρίνεται ο χρόνος που χρειάστηκαν τα προτεινόμενα συστήματα να υπολογίσουν μία νέα επαναγραφή όταν τροποποιήθηκαν οι οντολογίες δεδομένης μίας προηγούμενης επαναγραφής με τον χρόνο

Πίνακας 4.3: *Σύγκριση χρονικών αποτελεσμάτων (σε sec) των δύο υλοποιήσεων του αλγορίθμου* Add *με το Requiem και το Rapid.*

Σύγκριση των Add_{Req}, Requiem												
Ontology	Query	0.5%	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%	4.5%	5%	100%
P5X	$\mathcal{Q}_5$	18.3	2.6	12.3	8.6	12.2	6.0	15.7	11.0	2.9	13.1	176.5
AX	$\mathcal{Q}_4$	1.6	0.9	1.2	0.8	2.2	3.5	1.3	3.3	2.2	3.0	15.3
$\text{SWEET}^{\mathcal{L}}$	$\mathcal{Q}_5$	20.2	12.5	15.2	21.1	33.5	27.7	28.2	24.8	29.9	27.3	123.6
$\text{PTO}^{\mathcal{L}}$	$\mathcal{Q}_5$	2.0	2.3	2.4	3.7	5.3	10.9	7.2	5.7	7.0	12.0	51.7
$\text{PTO}^{\mathcal{L}}$	$\mathcal{Q}_6$	4.0	6.6	10.4	12.6	44.7	51.8	18.3	29.6	25.2	95.4	309.8
$\text{PRO}^{\mathcal{L}}$	$\mathcal{Q}_4$	435.2	427.5	360.8	355.8	475.2	421.2	549.2	625.2	727.8	987.8	1,178.7
$\text{SWEET}^{\mathcal{E}}$	$\mathcal{Q}_5$	0.1	0.2	0.2	0.8	1.0	0.5	1.4	0.9	1.4	1.4	3.6
$\text{PTO}^{\mathcal{E}}$	$\mathcal{Q}_5$	14.0	173.3	146.1	157.7	416.1	220.8	173.2	203.7	723.4	280.6	1,081.9
Σύγκριση των Add_{Rap}, Rapid												
		0.5%	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%	4.5%	5%	100%
$\text{PTO}^{\mathcal{L}}$	$\mathcal{Q}_6$	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0.1
$\text{PRO}^{\mathcal{L}}$	$\mathcal{Q}_5$	8.3	11.1	28.7	35.6	48.5	52.3	110.7	93.1	110.1	99.8	364.2
$\text{PRO}^{\mathcal{L}}$	$\mathcal{Q}_6$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
$\text{PRO}^{\mathcal{L}}$	$\mathcal{Q}_7$	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	1.3
$\text{PTO}^{\mathcal{E}}$	$\mathcal{Q}_6$	<0.1	<0.1	<0.1	0.1	<0.1	<0.1	<0.1	0.1	0.1	0.1	0.3
$\text{PTO}^{\mathcal{E}}$	$\mathcal{Q}_7$	<0.1	<0.1	<0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
$\text{PRO}^{\mathcal{E}}$	$\mathcal{Q}_5$	2.7	2.6	2.7	5.7	4.9	6.5	7.2	8.2	30.2	94.5	373.0
$\text{PRO}^{\mathcal{E}}$	$\mathcal{Q}_6$	2.1	2.1	2.1	2.1	2.7	2.5	2.3	2.7	2.4	2.0	9.5
$\text{PRO}^{\mathcal{E}}$	$\mathcal{Q}_7$	1.7	1.8	1.8	1.8	1.9	2.0	2.1	2.2	3.0	3.1	6.7

που χρειάζεται το Requiem και το Rapid για τον υπολογισμό μίας νέας επαναγραφής από την αρχή με βάση τη νέα οντολογία.

Όλα τα πειράματα διεξήχθησαν σε υπολογιστή Intel(R) Core (TM) PC σε Windows 7 με επεξεργαστή στα 3.20GHz και μνήμη 8GB RAM.

4.1.1 Αξιολόγηση του Αλγορίθμου Add

Για την αξιολόγηση των εργαλείων Add_{Req}, Add_{Rap} ακολουθήσαμε την προσέγγιση των Cuenca Grau et al. [53]. Πιο συγκεκριμένα, για κάθε μία από τις 14 οντολογίες εφαρμόσαμε τα ακόλουθα βήματα: 1) αφαιρέσαμε p% τυχαία αξιώματα (συμβολίζουμε με $\mathcal{T}^p$ την προκύπτουσα οντολογία) όπου το πεδίο του p είναι μεταξύ 0.5 και 5.0 με βήμα 0.5 (όταν το p% της οντολογίας ήταν μικρότερο του 1 αφαιρούσαμε ένα αξίωμα), 2) χρησιμοποιήσαμε το Requiem (Rapid) για να υπολογίσουμε το κορεσμένο σύνολο (έχοντας εφαρμόσει και τα κριτήρια περιττότητας του αντίστοιχου συστήματος) $\mathcal{S}$ του $\mathcal{Q} \cup \mathcal{T}'$, όπου $\mathcal{T}' = \mathcal{T} \setminus \mathcal{T}^p$, 3) υπολογίσαμε μια επαναγραφή για την οντολογία $\mathcal{T}' \cup \mathcal{T}^p$ με το Add_{Req}, (Add_{Rap}, αντίστοιχα) χρησιμοποιώντας το σύνολο $\mathcal{S}$ που υπολογίστηκε στο βήμα 2). Επαναλάβαμε τα βήματα 1)–3) δέκα φορές για κάθε ερώτημα και οντολογία και υπολογίσαμε τον μέσο όρο των χρονικών αποτελεσμάτων.

Αξιολόγηση του συστήματος $\mathsf{Add_{Req}}$.

Στον Πίνακα 4.3 παρουσιάζονται τα αποτελέσματα για τις οντολογίες και τα ερωτήματα που το $\mathsf{Add_{Req}}$ απαιτεί περισσότερο χρόνο για τον υπολογισμό των επαναγραφών τους και συνεπώς έχει σημασία η διαφορά στην απόδοση μεταξύ του $\mathsf{Add_{Req}}$ και του Requiem. Συγκεκριμένα, στον πίνακα εμφανίζονται τα χρονικά αποτελέσματα που αφορούν στις οντολογίες P5X, $\mathsf{SWEET}^{\mathcal{L}}$, $\mathsf{PTO}^{\mathcal{L}}$, $\mathsf{SWEET}^{\mathcal{E}}$, $\mathsf{PTO}^{\mathcal{E}}$ και στα αντίστοιχα ερωτήματα $\mathcal{Q}_5$ και στις οντολογίες AX, $\mathsf{PRO}^{\mathcal{L}}$ και στα αντίστοιχα ερωτήματα $\mathcal{Q}_4$. Για την οντολογία $\mathsf{PTO}^{\mathcal{L}}$ παρουσιάζεται επίσης το ερώτημα $\mathcal{Q}_6$. Σχετικά με τα ερωτήματα $\mathcal{Q}_5$ που αντιστοιχούν στις οντολογίες AX και $\mathsf{PRO}^{\mathcal{L}}$, το Requiem δεν κατάφερε να τερματίσει σε χρόνο μικρότερο από 2 ώρες και συνεπώς, το αρχικό σύνολο $\mathcal{S}$ δεν ήταν διαθέσιμο. Επίσης, το ίδιο συνέβη με την οντολογία $\mathsf{PTO}^{\mathcal{L}}$ και το αντίστοιχο ερώτημα $\mathcal{Q}_7$, με την $\mathsf{PTO}^{\mathcal{E}}$ και τα αντίστοιχα ερωτήματα $\mathcal{Q}_6, \mathcal{Q}_7$, και με την $\mathsf{PRO}^{\mathcal{E}}$ και οποιοδήποτε αντίστοιχο ερώτημα. Για τις περισσότερες από τις οντολογίες A, P5, S, V, U, UX και οποιοδήποτε αντίστοιχο ερώτημα, το $\mathsf{Add_{Req}}$ υπολογισε μια επαναγραφή σε λιγότερο από 2 δευτερόλεπτα ενώ το Requiem σε πολλές περιπτώσεις χρειάστηκε εώς και 12 δευτερόλεπτα για τον υπολογισμό της ίδιας επαναγραφής.

Οι πρώτες 10 στήλες του πίνακα αντιστοιχούν στον χρόνο που χρειάστηκε το $\mathsf{Add_{Req}}$ να υπολογίσει μία επαναγραφή για το αντίστοιχο ερώτημα και την οντολογία $\mathcal{T}' \cup \mathcal{T}^p$ για τα διάφορα p. Η τελευταία στήλη (με το όνομα “100%”) αντιστοιχεί στον χρόνο που χρειάζεται το Requiem για να υπολογίσει την επαναγραφή για το ίδιο ερώτημα και την ίδια οντολογία $\mathcal{T}' \cup \mathcal{T}^p$ από την αρχή. Όπως φαίνεται στον πίνακα, το σύστημα $\mathsf{Add_{Req}}$ υπολογίζει όλες τις αντίστοιχες επαναγραφές πιο γρήγορα από το Requiem. Πιο συγκεκριμένα, το προτεινόμενο σύστημα παρουσιάζει ιδιαίτερα καλή απόδοση όταν προστίθενται σχετικά λίγα νέα αξιώματα στην οντολογία (τρεις πρώτες στήλες). Τέλος, είναι ενδιαφέρον να παρατηρηθεί ότι η απόδοση του $\mathsf{Add_{Req}}$ αυξάνεται σχεδόν αναλογικά με το p.

Αξιολόγηση του συστήματος $\mathsf{Add_{Rap}}$.

Στις περισσότερες περιπτώσεις και τα δύο συστήματα, Rapid και $\mathsf{Add_{Rap}}$, υπολόγισαν τη νέα επαναγραφή σχεδόν σε αμελητέο χρόνο. Συγκεκριμένα, το Rapid υπολόγισε την επαναγραφή για οποιοδήποτε ερώτημα με βάση τις οκτώ DL-Lite$_{\mathcal{R}}$ οντολογίες και τις δύο εκδόσεις της SWEET σε λιγότερο από 0.08sec και το $\mathsf{Add_{Rap}}$ σε λιγότερο από 0.04sec. Τέτοιες διαφορές σε τόσο μικρούς χρόνους δεν έχουν ιδιαίτερο ενδιαφέρον και συνεπώς δεν θα παρουσιάσουμε αναλυτικά τα αποτελέσματα αυτά. Στο κάτω μέρος του Πίνακα 4.3 παρουσιάζονται τα χρονικά αποτελέσματα που αφορούν στις DL-Lite$_{\mathcal{R}}$ και $\mathcal{ELHI}$ εκδόσεις των οντολογιών PTO και PRO, για τις οποίες τα συστήματα αυτά χρειάστηκαν περισσότερο χρόνο για να υπολογίσουν μία επαναγραφή. Παρατηρούμε ότι σε όλες τις περιπτώσεις το σύστημα $\mathsf{Add_{Rap}}$ είναι πιο γρήγορο από το Rapid. Σε μερικές περιπτώσεις η διαφορά είναι αρκετά σημαντική, ειδικά όταν προστίθενται σχετικά λίγα αξιώματα όπου η διαφορά φτάνει τις δύο τάξεις μεγέθους.

Μια γενική παρατήρηση που αφορά και στα δύο συστήματα Add_{Req} και Add_{Rap} είναι ότι στις περισσότερες περιπτώσεις έχουν πολύ καλύτερη απόδοση από τα αντίστοιχα συστήματα Requiem και Rapid. Επίσης, είναι ενδιαφέρον να σημειωθεί ότι όταν η τιμή του p διακυμαίνεται μεταξύ 0.5 και 3 η απόδοση των συστημάτων μας είναι ιδιαίτερα καλή. Για παράδειγμα, όταν το πολύ $1,810$ προτάσεις προστίθενται στην οντολογία $\mathsf{PRO}^{\mathcal{E}}$ τότε είναι πολύ πιο αποδοτικό να υπολογίσουμε μία νέα επαναγραφή με τα συστήματα που έχουμε υλοποιήσει από το να υπολογιστούν νέες επαναγραφές από την αρχή. Αυτή η διαπίστωση μπορεί εύκολα να επαληθευθεί αν παρατηρήσουμε ότι για παράδειγμα το Rapid χρειάζεται περίπου 6 λεπτά για τον υπολογισμό μίας επαναγραφής του ερωτήματος $\mathcal{Q}_5$ μ.β.τ. $\mathsf{PRO}^{\mathcal{E}}$ ενώ το Add_{Rap} μόλις 5 δευτερόλεπτα.

4.1.2 Αξιολόγηση των αλγορίθμων Delete και Delete_{Γ}

Αξιολογήσαμε τα σύστηματα Delete και $\mathsf{Delete}_{\mathsf{Req}}$ χρησιμοποιώντας τις 14 οντολογίες και τα αντίστοιχα ερωτήματα και προσαρμόζοντας τη μεθοδολογία της προηγούμενης ενότητας ως εξής: 1) Χρησιμοποιήσαμε το Requiem για να υπολογίσουμε το κορεσμένο σύνολο (εφαρμόζοντας τα αντίστοιχα κριτήρια περιττότητας) $\mathcal{S}$ του $\mathcal{Q} \cup \mathcal{T}$ (για το σύστημα Delete επιχειρήσαμε να υπολογίσουμε την κλειστή ως προς τα υποσύνολα επαναγραφή των $\mathcal{Q}$, $\mathcal{T}$) και 2) για $\mathcal{T}^p$ ένα (τυχαίο) p% των αξιωμάτων της $\mathcal{T}$, όπου το p είναι μεταξύ των 0.5 και 5.0 με βήμα 0.5, υπολογίσαμε μία νέα επαναγραφή του $\mathcal{Q}$ μ.β.τ. $\mathcal{T} \setminus \mathcal{T}^p$ χρησιμοποιώντας το Delete και το $\mathsf{Delete}_{\mathsf{Req}}$. Επαναλάβαμε το δεύτερο βήμα δέκα φορές για κάθε ερώτημα και οντολογία και πήραμε τις μέσες τιμές.

Αξιολόγηση του Delete.

Όπως περιγράφεται στην Ενότητα 3.4, για να υπολογίσει το σύστημα Delete την νέα επαναγραφή θα πρέπει η αρχική επαναγραφή να είναι κλειστή ως προς τα υποσύνολα. Για τον λόγο αυτό απενεργοποιήσαμε από το σύστημα Requiem κάποιους από τους ελέγχους υπαγωγής. Εξαιτίας των αλλαγών αυτών στον κώδικα του Requiem το ίδιο το σύστημα δεν κατάφερε να τερματίσει, σε χρόνο λιγότερο των δύο ωρών, για τα ερωτήματα $\mathcal{Q}_3$–$\mathcal{Q}_5$ και τις οντολογίες $\mathsf{PRO}^{\mathcal{L}}$, το ερώτημα $\mathcal{Q}_5$ και την AX, $\mathsf{PTO}^{\mathcal{L}}$, και για όλα τα ερωτήματα και τις οντολογίες $\mathsf{PTO}^{\mathcal{E}}$, $\mathsf{PRO}^{\mathcal{E}}$.

Παρουσιάζουμε στον Πίνακα 4.4 τα πιο σημαντικά αποτελέσματα για τα οποία ο υπολογισμός της κλειστής ως προς τα υποσύνολα επαναγραφής ήταν εφικτός. Συγκεκριμένα, για τις DL-Lite οντολογίες AX και $\mathsf{PTO}^{\mathcal{L}}$ τα αποτελέσματα αφορούν στα αντίστοιχα ερωτήματα $\mathcal{Q}_4$, για την οντολογία $\mathsf{PRO}^{\mathcal{L}}$ στα αντίστοιχα ερωτήματα $\mathcal{Q}_2$ και για τις υπόλοιπες οντολογίες (P5X, $\mathsf{SWEET}^{\mathcal{L}}$) στα αντίστοιχα ερωτήματα $\mathcal{Q}_5$. Για τις $\mathcal{ELHI}$ οντολογίες, το Requiem υπολόγισε την αρχική επαναγραφή μόνο για την οντολογία $\mathsf{SWEET}^{\mathcal{E}}$, για την οποία, παρουσιάζουμε τα αποτελέσματα για τα αντίστοιχα ερωτήματα $\mathcal{Q}_4$, $\mathcal{Q}_5$ στον Πίνακα 4.4.

Παρατηρείται από τον Πίνακα 4.4 ότι σχεδόν σε όλες τις περιπτώσεις το Delete τερμα-

Πίνακας 4.4: *Χρονικά αποτελέσματα (σε sec) του* Delete *όταν οι DL-Lite$_R$ και $\mathcal{ELHI}$ οντολογίες μειώνονται κατά* 5%.

Οντολογία	Ερώτημα	Delete	Rapid	Requiem
P5X	$\mathcal{Q}_5$	<0.1	2.7	137.3
AX	$\mathcal{Q}_4$	<0.1	0.5	18.3
SWEET$^{\mathcal{L}}$	$\mathcal{Q}_5$	0.1	0.8	80.3
PTO$^{\mathcal{L}}$	$\mathcal{Q}_4$	<0.1	0.1	0.9
PRO$^{\mathcal{L}}$	$\mathcal{Q}_2$	0.1	1.5	60.3
SWEET$^{\mathcal{E}}$	$\mathcal{Q}_4$	<0.1	0.1	2.3
SWEET$^{\mathcal{E}}$	$\mathcal{Q}_5$	<0.1	0.1	2.4

Πίνακας 4.5: *Χρονικά αποτελέσματα (σε sec) του* Delete$_{\text{Req}}$ *όταν οι DL-Lite$_R$ και $\mathcal{ELHI}$ οντολογίες μειώνονται κατά* 5%.

Οντολογία	Ερώτημα	Delete$_{\text{Req}}$	Requiem
PTO$^{\mathcal{L}}$	$\mathcal{Q}_5$	0.1	10.9
PRO$^{\mathcal{L}}$	$\mathcal{Q}_3$	150.2	736.6
PRO$^{\mathcal{L}}$	$\mathcal{Q}_4$	150.3	777.1
PTO$^{\mathcal{E}}$	$\mathcal{Q}_4$	1.6	932.9
PTO$^{\mathcal{E}}$	$\mathcal{Q}_5$	1.8	1,222.6
PTO$^{\mathcal{E}}$	$\mathcal{Q}_6$	3.8	1,883.3

τίζει σχεδόν σε 0.1sec, δηλαδή, σημαντικά πιο γρήγορα από τα Requiem και Rapid. Επίσης, στις περισσότερες περιπτώσεις η διαφορά μεταξύ του Delete και του Requiem είναι περισσότερη από 2 τάξεις μεγέθους. Τα αποτελέσματα αυτά είναι σχετικά αναμενόμενα αφού Delete δεν χρησιμοποιεί κάποιο σύστημα επαναγραφής για να εφαρμόσει νέους συμπερασμούς αλλά υπολογίζει την τελική επαναγραφή είτε διαγράφοντας είτε αντιγράφοντας ερωτήματα από την αρχική επαναγραφή. Ωστόσο, όπως έχει ήδη διατυπωθεί, η αρχική επαναγραφή πρέπει να είναι κλειστή ως προς τα υποσύνολα, κάτι που όπως φαίνεται και από την πειραματική αξιολόγηση δεν είναι πάντα εφικτό σε λογικό χρόνο.

Αξιολόγηση του Delete$_{\text{Req}}$.

Αντίθετα, το Delete$_{\text{Req}}$ δεν απαιτεί να είναι η αρχική επαναγραφή κλειστή ως προς τα υποσύνολα. Συνεπώς, στην περίπτωση αυτή, έχουμε περισσότερα αποτελέσματα. Ωστόσο, και πάλι, το Requiem δεν κατάφερε να υπολογίσει το αρχικό κορεσμένο σύνολο για το ερώτημα $\mathcal{Q}_5$ και την οντολογία PRO$^{\mathcal{L}}$ και για όλα τα ερωτήματα με βάση την οντολογία PRO$^{\mathcal{E}}$.

Στον Πίνακα 4.5 παρουσιάζονται τα χρονικά αποτελέσματα για τα ερωτήματα και τις οντολογίες για τα οποία δεν κατάφερε να υπολογίσει το Requiem τις αρχικές τους επαναγραφές. Συγκεκριμένα,παρουσιάζουμε τα αποτελέσματα από τον υπολογισμο των επαναγραφών των $\mathcal{Q}_5$, PTO$^{\mathcal{L}}$, των $\mathcal{Q}_3$, $\mathcal{Q}_4$, PRO$^{\mathcal{L}}$, και των $\mathcal{Q}_4$, $\mathcal{Q}_5$, $\mathcal{Q}_6$, PTO$^{\mathcal{E}}$. Τα υπόλοιπα αποτελέσματα είναι διαθέσιμα στο Παράρτημα Β′. Όπως και στον Πίνακα 4.4, παρουσιάζουμε τα αποτελέσματα μόνο για την περίπτωση που η οντολογία μειώνεται κατά 5%, αφού όπως παρατηρείται και στο Παράρτημα, η διακύμανση στην απόδοση της $\mathsf{Delete}_{\mathsf{Req}}$ εξαρτάται ελάχιστα από την μεταβολή του p%.

Όπως φαίνεται στον Πίνακα 4.5, το $\mathsf{Delete}_{\mathsf{Req}}$ είναι πιο γρήγορο από το Requiem και μάλιστα σε κάποιες περιπτώσεις σχεδόν κατά τρεις τάξεις μεγέθους. Οι νέες επαναγραφές υπολογίζονται πολύ γρήγορα από το $\mathsf{Delete}_{\mathsf{Req}}$ (για παράδειγμα, όταν το 5% της $\mathsf{PTO}^{\mathcal{E}}$ διαγράφεται τότε μία νέα επαναγραφή για το ερώτημα $\mathcal{Q}_6$ υπολογίζεται σε 4 δευτερόλεπτα ενώ το Requiem το υπολογίζει σχεδόν σε 31 λεπτά) και αυτό οφείλεται στο γεγονός ότι σε ελάχιστες περιπτώσεις χρειάστηκαν να εφαρμοστούν νέοι συμπερασμοί. Επίσης, αξίζει να σημειωθεί ότι, όπως φαίνεται και στο Παράρτημα, η απόδοση του $\mathsf{Delete}_{\mathsf{Req}}$ είναι εξίσου καλή και σε όλες τις υπόλοιπες περιπτώσεις.

Τέλος, παρατηρούμε ότι και τα δύο συστήματα, Delete και $\mathsf{Delete}_{\mathsf{Req}}$, έχουν ιδιαίτερα καλή απόδοση. Είναι προφανές από τα πειραματικά αποτελέσματα ότι για κάθε p είναι πολύ πιο γρήγορο να υπολογιστεί μία νέα επαναγραφή για τη νέα οντολογία με τα συστήματα αυτά παρά να υπολογιστεί από την αρχή με την χρήση είτε του Requiem είτε του Rapid.

4.2 Αξιολόγηση σε Πραγματικά Σενάρια

Στην ενότητα αυτή παρουσιάζουμε τα αποτελέσματα των πειραμάτων που εφαρμόστηκαν σε πραγματικά σενάρια. Για τον λόγο αυτό, χρησιμοποιούμε τις οντολογίες NCI[4] και SWEET. Συγκεκριμένα, χρησιμοποιήσαμε το DL-Lite$_{\mathcal{R}}$ τμήμα των εκδόσεων 10j, 12a και 12e της NCI οντολογίας (για όλες τις εκδόσεις το DL-Lite$_{\mathcal{R}}$-τμήμα τους ταυτίζεται με το $\mathcal{ELHI}$-τμήμα τους, συνεπώς ήταν άσκοπο να πειραματιστούμε και με το $\mathcal{ELHI}$-τμήμα). Επίσης, χρησιμοποιήσαμε το $\mathcal{ELHI}$-τμήμα των διαδοχικών εκδόσεων 2.1, 2.2 και 2.3 της οντολογίας SWEET. Για το σύνολο των πειραμάτων, χρησιμοποιήσαμε τα ερωτήματα που κατασκευάστηκαν για την αξιολόγηση του Rapid. Τα στατιστικά στοιχεία που αφορούν στις οντολογίες αυτές παρουσιάζονται στον Πίνακα 4.6. Επίσης, στον Πίνακα 4.7 παρουσιάζεται ο αριθμός των GCIs και RIAs που αφαιρέθηκαν ($\mathcal{T}^-_{x,y}$) και προστέθηκαν ($\mathcal{T}^+_{x,y}$) μεταξύ δύο διαδοχικών εκδόσεων $\mathcal{T}_x$ και $\mathcal{T}_y$, δηλαδή $\mathcal{T}^-_{x,y} = \mathcal{T}_x \setminus \mathcal{T}_y$ και $\mathcal{T}^+_{x,y} = \mathcal{T}_y \setminus \mathcal{T}_x$, συνεπώ $\mathcal{T}_y = (\mathcal{T}_x \setminus \mathcal{T}^-_{x,y}) \cup \mathcal{T}^+_{x,y}$.

Δεδομένων δύο διαδοχικών εκδόσεων $\mathcal{T}_x, \mathcal{T}_y$ μίας οντολογίας και δεδομένου ενός ερωτήματος $\mathcal{Q}$ ακολουθήσαμε τα παρακάτω βήματα για τα πειράματά μας: Αρχικά, υπολογίστηκε το αρχικό κορεσμένο σύνολο $\mathcal{S}$ του $\mathcal{Q} \cup \mathcal{T}_x$ με το Requiem, στη συνέχεια, υπολογίστηκε με το $\mathsf{Delete}_{\mathsf{Req}}$ το κορεσμένο σύνολο $\mathcal{S}'$ του $\mathcal{Q} \cup (\mathcal{T}_x \setminus \mathcal{T}^-_{x,y})$ και τέλος, εφαρμόστηκε το $\mathsf{Add}_{\mathsf{Req}}$ στα $\mathcal{S}', \mathcal{T}^+_{x,y}$ για τον υπολογισμό της τελικής επαναγραφής του $\mathcal{Q}$ με βάση το $\mathcal{T}_y$. Σχετικά με την αξιολόγηση του $\mathsf{Add}_{\mathsf{Rap}}$, αφού δεν υπάρχει κάποια υλοποίηση του $\mathsf{Delete}_{\mathsf{Rap}}$, εφαρμόσαμε τα πειράματά μας χρησιμοποιώντας μόνο τα προστιθέμενα αξιώ-

[4] ftp://ftp1.nci.nih.gov/pub/cacore/EVS/NCI_Thesaurus/archive

Πίνακας 4.6: *Στατιστικά στοιχεία ρεαλιστικών οντολογιών.*

Οντολογία	#Εννοιών	#Ρόλων	#GCIs	#RIAs
NCI_{10j}	28,758	67	50,933	0
NCI_{12a}	29,058	67	52,571	0
NCI_{12e}	29,173	66	53,341	0
$\mathsf{SWEET}_{2.1}$	4,451	543	6,028	434
$\mathsf{SWEET}_{2.2}$	4,445	543	6,020	429
$\mathsf{SWEET}_{2.3}$	4,572	312	6,873	218

Πίνακας 4.7: *Αριθμός αξιωμάτων που έχουν προστεθεί και αφαιρεθεί.*

Αξιώματα	#GCIs	#RIAs
$\mathsf{NCI}^{-}_{10j,12_a}$	610	0
$\mathsf{NCI}^{+}_{10j,12_a}$	2,248	0
$\mathsf{NCI}^{-}_{12_a,12_e}$	10,087	0
$\mathsf{NCI}^{+}_{12_a,12_e}$	10,857	0
$\mathsf{SWEET}^{-}_{2.1,2.2}$	1,314	58
$\mathsf{SWEET}^{+}_{2.1,2.2}$	1,306	53
$\mathsf{SWEET}^{-}_{2.2,2.3}$	2,228	429
$\mathsf{SWEET}^{+}_{2.2,2.3}$	2,403	218

ματα, δηλαδή αρχικά υπολογίσαμε το κορεσμένο σύνολο $\mathcal{Q} \cup \mathcal{T}_x \setminus \mathcal{T}^{-}_{x,y}$ με το Rapid και στη συνέχεια εφαρμόσαμε το $\mathsf{Add}_{\mathsf{Rap}}$ στο σύνολο $\mathcal{T}^{+}_{x,y}$ για τον υπολογισμό μίας επαναγραφής του $\mathcal{Q}$ με βάση το $\mathcal{T}_y$.

Στον Πίνακα 4.8 παρουσιάζονται τα χρονικά αποτελέσματα των συστημάτων $\mathsf{Delete}_{\mathsf{Req}}$, $\mathsf{Add}_{\mathsf{Req}}$, $\mathsf{Add}_{\mathsf{Rap}}$ και των συστημάτων Requiem και Rapid για τον υπολογισμό μίας επαναγραφής του $\mathcal{Q}$, με βάση το $\mathcal{T}_y$ από την αρχή. Ο υπολογισμός των κορεσμένων συνόλων των $\mathcal{Q}_3 \cup \mathsf{NCI}_{10j}$ και $\mathcal{Q}_3 \cup \mathsf{NCI}_{12a}$ δεν κατέστη εφικτός από το Requiem αφού εξαντλήθηκε η μνήμη του. Όπως φαίνεται από τον πίνακα, ο συνολικός χρόνος $(t(\mathsf{Delete}_{\mathsf{Req}}) + t(\mathsf{Add}_{\mathsf{Req}}))$ που χρειάστηκαν και τα δύο συστήματα, $\mathsf{Delete}_{\mathsf{Req}}$ και $\mathsf{Add}_{\mathsf{Req}}$, για να υπολογίσουν μία νέα επαναγραφή με βάση την $\mathcal{T}_y$, είναι στις περισσότερες περιπτώσεις ο μισός σε σχέση με τον χρόνο που χρειάζεται το Requiem να υπολογίσει την ίδια επαναγραφή από την αρχή. Επίσης, η απόδοση του $\mathsf{Add}_{\mathsf{Rap}}$ είναι αντίστοιχα καλή, δηλαδή σχεδόν σε όλες τις περιπτώσει το $\mathsf{Add}_{\mathsf{Rap}}$ είναι πιο γρήγορο από το Rapid (στις μόνες δύο περιπτώσεις που Requiem δεν είναι πιο γρήγορο από το Rapid είναι όταν και τα δύο συστήματα υπολογίζουν τη νέα επαναγραφή σε 0,02 δευτερόλεπτα).

Αξίζει να σημειωθεί ότι το $\mathsf{Delete}_{\mathsf{Req}}$ εφάρμοσε νέους συμπερασμούς μόνο όταν ελέγχθηκε με τα ερωτήματα $\mathcal{Q}_4$, $\mathcal{Q}_5$ και την οντολογία $\mathsf{SWEET}_{2.3}$. Η χαμηλή απόδοση που σημείωσε στα ερωτήματα $\mathcal{Q}_2$, $\mathcal{Q}_4$ και τις οντολογίες NCI_{12a} και NCI_{12e}, οφείλεται στο μεγάλο μέγεθος του γράφου του κορεσμένου συνόλου. Για παράδειγμα, ο αρχικός γράφος $\mathcal{Q}_4 \cup \mathsf{NCI}_{12e}$ περιείχε 16.749 κόμβους. Η σχετικά κακή απόδοση του $\mathsf{Add}_{\mathsf{Req}}$ στην περί-

Πίνακας 4.8: *Χρονικά αποτελέσματα (σε sec) σε πραγματικά σενάρια.*

Οντολογία	Ερώτημα	$\text{Delete}_{\text{Req}}$	Add_{Req}	Σύνολο	Requiem	Add_{Rap}	Rapid
NCI_{12a}	$\mathcal{Q}_1$	0.01	0.17	0.18	7.33	0.02	0.04
	$\mathcal{Q}_2$	0.73	32.75	33.48	67.06	0.07	0.13
	$\mathcal{Q}_3$	-	-	-	-	0.32	0.55
	$\mathcal{Q}_4$	1.39	29.10	30.49	77.26	0.07	0.15
	$\mathcal{Q}_5$	0.35	6.92	7.27	22.57	0.01	0.02
NCI_{12e}	$\mathcal{Q}_1$	0.02	0.13	0.15	67.84	0.01	0.04
	$\mathcal{Q}_2$	0.59	55.72	56.31	68.37	0.09	0.14
	$\mathcal{Q}_3$	-	-	-	-	0.31	0.57
	$\mathcal{Q}_4$	0.94	965.31	966.25	713.18	0.08	0.17
	$\mathcal{Q}_5$	0.06	168.41	168.47	189.45	0.03	0.09
$\text{SWEET}_{2.2}$	$\mathcal{Q}_1$	0.02	0.11	0.13	0.97	0.01	0.02
	$\mathcal{Q}_2$	0.03	0.07	0.10	0.96	0.01	0.04
	$\mathcal{Q}_3$	0.05	0.12	0.17	1.19	0.02	0.06
	$\mathcal{Q}_4$	0.10	0.52	0.62	2.65	0.06	0.13
	$\mathcal{Q}_5$	0.10	0.54	0.64	2.70	0.06	0.14
$\text{SWEET}_{2.3}$	$\mathcal{Q}_1$	0.04	0.45	0.49	1.16	0.02	0.02
	$\mathcal{Q}_2$	0.05	0.27	0.32	0.99	0.02	0.02
	$\mathcal{Q}_3$	0.06	0.60	0.66	1.36	0.03	0.04
	$\mathcal{Q}_4$	0.16	1.97	2.13	3.50	0.05	0.12
	$\mathcal{Q}_5$	0.15	1.90	2.05	3.57	0.05	0.12

πτωση των $\mathcal{Q}_4$, NCI_{12e} οφείλεται στο μεγάλο μέγεθος του συνόλου των προστιθέμενων αξιωμάτων που σχετίζοταν με το $\mathcal{Q}_4$. Συγκεκριμένα, το σύνολο αυτό περιείχε 8.491 ενώ το αρχικό κορεσμένο σύνολο 1.882 προτάσεις. Θα ήταν αναμενόμενο όμως, η απόδοση του $\mathsf{Add}_{\mathsf{Req}}$ να είναι τουλάχιστον η ίδια με του Requiem. Ωστόσο, όπως είναι γνωστό, επειδή το Requiem εφαρμόζει εμπρόσθιο έλεγχο υπαγωγής, έχει πολύ μεγάλη σημασία για την απόδοσή του η σειρά με την οποία επεξεργάζεται τις προτάσεις.

Μέρος ΙΙΙ

Απάντηση Ερωτημάτων σε Ασυνεπείς Βάσεις Γνώσης

Κεφάλαιο 5

Απάντηση Ερωτημάτων σε Ασυνεπείς Βάσεις Γνώσης

Η συχνή τροποποίηση των οντολογιών, ειδικά όταν αυτή εφαρμόζεται από πολλούς διαφορετικούς φορείς, μπορεί να προκαλέσει ασυνέπεια στην τελική βάση γνώσης. Μπορεί, για παράδειγμα, να έχει οριστεί στην οντολογία ότι δύο έννοιες είναι ξένες μεταξύ τους, αλλά στο σώμα ισχυρισμών να μοιράζονται κοινά άτομα. Για την αντιμετώπιση του προβλήματος αυτού προτείνονται στη βιβλιογραφία κυρίως τρεις λύσεις. Η πρώτη λύση είναι να διορθωθεί η οντολογία, η δεύτερη να διορθωθεί το σώμα των ισχυρισμών και η τρίτη να εξάγουμε συνεπή συμπεράσματα πάρα την ασυνέπεια της βάσης γνώσης. Επειδή συνήθως το πρόβλημα προέρχεται από το σώμα ισχυρισμών και όχι από την οντολογία, στην παρούσα διατριβή θεωρούμε ότι η οντολογία είναι συνεπής και μελετάμε τρόπους αποδοτικού υπολογισμού απαντήσεων σε συζευκτικά ερωτήματα με τις δύο άλλες λύσεις.

Το πρόβλημα συνεπούς απάντησης ερωτημάτων είναι δισεπίλυτο σε γλώσσες υψηλής εκφραστικότητας. Μπορεί, όμως, να μετατραπεί σε πολυωνυμικό για γλώσσες χαμηλής εκφραστικότητας, όπως είναι η DL-Lite$_{\mathcal{R}}$, με τη χρήση δύο ειδών διορθώσεων, της *Τομής ABox Διορθώσεων (Intersection ABox Repair-IAR)* και της *Τομής Διορθώσεων του Κλειστού ABox (Intersection Closed ABox Repair-ICAR)*, που ορίζουν τις αντίστοιχες σημασιολογίες. Με την IAR διόρθωση πρακτικά αφαιρούνται όλοι οι ισχυρισμοί που *εμφανίζονται* στη βάση γνώσης και είναι ασυνέπεις σε σχέση με την οντολογία. Ένα μειονέκτημα της διόρθωσης αυτής είναι ότι εξαρτάται από το συντακτικό της βάσης γνώσης, δηλαδή τους ισχυρισμούς που περιέχονται σε αυτήν και όχι από τους ισχυρισμούς που συμπεραίνονται από αυτήν. Αντίθετα, με την ICAR διόρθωση, υπολογίζονται πρώτα όλοι οι ισχυρισμοί που μπορούν να *συμπεραχθούν* από τη βάση γνώσης και στη συνέχεια αφαιρούνται οι ισχυρισμοί που προκαλούν ασυνέπεια. Οι τυπικοί ορισμοί των διορθώσεων αυτών και των αντίστοιχων σημασιολογιών παρουσιάζονται αναλυτικά στο Κεφάλαιο 2.

Στο κεφάλαιο αυτό, προτείνουμε, αρχικά, έναν αποδοτικό αλγόριθμο υπολογισμού απαντήσεων υπό την ICAR σημασιολογία για DL-Lite$_{\mathcal{R}}$ οντολογίες. Εξ όσων γνωρίζουμε δεν υπάρχει κάποιο άλλο αποδοτικό σύστημα υπολογισμού των ICAR απαντήσεων. Η αποδοτικότητα του αλγορίθμου μας βασίζεται στη χρήση συστημάτων κορεσμού δεδομένων, τα οποία είναι ιδιαιτερα αποδοτικά για χαμηλές εκφραστικότητες. Ένα ακόμα χαρακτηριστικό τους είναι ότι υπολογίζουν όλους τους ισχυρισμούς που παράγονται από μία (χαμηλής εκφραστικότητας) ΒΓ. Αξιοποιώντας την ιδιότητα αυτή, στη συνέχεια, προτείνουμε μία βελτιστοποίηση του πρώτου αλγοριθμου. Στην επόμενη ενότητα, προτείνουμε έναν αλγόριθμο συνεπούς απάντησης ερωτημάτων υπό την ICAR σημασιολογία για πιο εκφραστικές Περιγραφικές Λογικές. Τέλος, προτείνουμε έναν αποδοτικό τρόπο υπολογισμού των IAR-απαντήσεων για DL-Lite$_\mathcal{R}$ οντολογίες. Για κάθε προτεινόμενο αλγόριθμο παρουσιάζεται και η αντίστοιχη απόδειξη της ορθότητάς του. Η πειραματική αξιολόγηση των αλγορίθμων παρουσιάζεται στο Κεφάλαιο 6.

5.1 Υπολογισμός ICAR-απαντήσεων στη DL-Lite$_\mathcal{R}$

Σύμφωνα με τη βιβλιογραφία, υπάρχουν δύο κυρίως τρόποι υπολογισμού των ICAR-απαντήσεων. Ο ένας τρόπος είναι, πρακτικά, η εφαρμογή του ορισμού του ICAR, δηλαδή ο υπολογισμός όλων των ισχυρισμών που μπορεί να παραχθούν από τη βάση γνώσης και στη συνέχεια η αφαίρεση των ισχυρισμών που οδηγούν σε ασυνέπεια. Συνήθως όμως επειδή στις πραγματικές ΒΓ το σώμα των ισχυρισμών είναι εξαρχής αρκετά μεγάλο, είναι προφάνες ότι η προσέγγιση αυτή θα έχει αρκετά χαμηλή αποδοση. Άλλωστε, η πρώτη προσπάθεια που σημειώθηκε από τους Masotti et al. [84] περιορίστηκε σε πολύ μικρές ΒΓ ενώ δεν συνεχίστηκε η έρευνα τους προς την ίδια κατεύθυνση.

Ο άλλος τρόπος είναι με την τροποποίηση της επαναγραφής του τιθέμενου ερωτήματος και των θετικών αξιωμάτων της οντολογίας, έτσι ώστε να αποκλειστούν οι ασυνεπείς απαντήσεις. Η συνάρτηση $\mathsf{ref}^{\mathsf{ic}}$ όπως έχει οριστεί από τους Lembo et al [77] εφαρμόζει ακριβώς αυτήν την τροποποίηση. Ακολουθεί ένα παράδειγμα που περιγράφει τον υπολογισμό των ICAR-απαντήσεων τον τρόπο υπολογισμού της $\mathsf{ref}^{\mathsf{ic}}$ αλλά και της ref, την οποία θα αξιοποιήσουμε εκτενώς στη συνέχεια. Ο τυπικός ορισμός των ref, $\mathsf{ref}^{\mathsf{ic}}$ παρουσιάζεται στην Ενότητα 2.4.3.

Παράδειγμα 39. Έστω τα ακόλουθα TBox και ABox:

$$\begin{aligned}\mathcal{T} &= \{\exists P \sqsubseteq B,\ P \sqsubseteq \neg P', B \sqsubseteq \neg B'\},\\ \mathcal{A} &= \{P(a,b), P'(a,b), P(c,d), B(e), B'(e)\}\end{aligned}$$

και έστω το ερώτημα $\mathcal{Q} = Q(x) \leftarrow B(x)$. Όπως είναι προφανές, το $\mathcal{A}$ είναι ασυνεπές σε σχέση με το $\mathcal{T}$ εξαιτίας των συνόλων $\{B(e), B'(e), B \sqsubseteq \neg B'\}$ και $\{P(a,b), P'(a,b), P \sqsubseteq \neg P'\}$. Υπάρχουν οι εξής AR-διορθώσεις για τα $\mathcal{T}, \mathcal{A}$:

$$\begin{aligned}r_1 &= \{P(a,b), B(e), P(c,d)\}, & r_2 &= \{P(a,b), B'(e), P(c,d)\},\\ r_3 &= \{P'(a,b), B(e), P(c,d)\}, & r_4 &= \{P'(a,b), B'(e), P(c,d)\}\end{aligned}$$

Συνεπώς, η τομή των ABox διοθώσεων, δηλαδή το IAR, θα είναι $\mathcal{A}_I = r_1 \cap r_2 \cap r_3 \cap r_4 = \{P(c,d)\}$, επομένως προκύπτει ότι $\mathsf{cert}^{\mathsf{ia}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}_I) = \{c\}$. Επίσης, ισχύει ότι:

$$clc(\mathcal{T}, \mathcal{A}) = \mathcal{A} \cup \{B(a), B(c)\}$$

Επομένως, οι διορθώσεις για το $clc(\mathcal{T}, \mathcal{A})$ είναι:

$$\begin{aligned} r'_1 &= \{P(a,b), B(e), P(c,d), B(a), B(c)\}, \\ r'_2 &= \{P(a,b), B'(e), P(c,d), B(a), B(c)\}, \\ r'_3 &= \{P'(a,b), B(e), P(c,d), B(a), B(c)\}, \\ r'_4 &= \{P'(a,b), B'(e), P(c,d), B(a), B(c)\} \end{aligned}$$

και η τομή των κλειστών διορθωμένων ABox, δηλαδή το ICAR, θα είναι $\mathcal{A}_C = r'_1 \cap r'_2 \cap r'_3 \cap r'_4 = \{P(c,d), B(a), B(c)\}$. Συνεπώς, $\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{cert}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}_C) = \{a, c\}$.

Για τον υπολογισμό των ref, $\mathsf{ref}^{\mathsf{ic}}$, αρχικά πρέπει να υπολογίσουμε την επαναγραφή του $\mathcal{Q}$ μβτ $\mathcal{T}$, που είναι το σύνολο $\mathcal{R} = \{Q(x) \leftarrow B(x), Q(x) \leftarrow P(x,y)\}$. Σύμφωνα με την Ενότητα 2.4.3, με τη συνάρτηση $\mathsf{ref}(\mathcal{R}_Q, \mathcal{T})$, για κάθε άτομο $A(\vec{x})$ που εμφανίζεται σε ένα ερώτημα $\mathcal{Q}_i$ του $\mathcal{R}$, η ref επεκτείνει το $\mathcal{Q}_i$ με τα άτομα με τα οποία είναι ξένο το $A(\vec{x})$. Με τον τρόπο αυτό, υπολογίζεται το σύνολο:

$$\mathcal{R}^{\mathsf{ia}} = \{Q(x) \leftarrow B(x) \wedge \neg B'(x), Q(x) \leftarrow P(x,y) \wedge \neg P'(x,y) \wedge \neg B'(x)\}$$

Είναι προφανές, ότι αν αποτιμήσουμε τη $\mathsf{ref}(\mathcal{R}_Q, \mathcal{T})$ στο $\mathcal{A}$, θα επιστραφούν οι IAR-απαντήσεις, δηλαδή το σύνολο $\{c\}$. Στο σημείο αυτό είναι ενδιαφέρον να παρατηρήσουμε πως εξαιτίας του αρνητικού ατόμου $\neg B'(x)$, το "e" δεν επιστρέφεται στο σύνολο των απαντήσεων. Τέλος, για να υπολογιστούν οι ICAR-απαντήσεις, η συνάρτηση $\mathsf{ref}^{\mathsf{ic}}(\mathcal{R}_Q, \mathcal{T})$ επαναγράφει *περαιτέρω* το σύνολο $\mathcal{R}^{\mathsf{ia}}$, αναλύοντας το $Q(x) \leftarrow B(x) \wedge \neg B'(x)$ ως εξής:

$$\frac{Q(x) \leftarrow B(x) \wedge \neg B'(x) \quad B(x) \leftarrow P(x,y)}{Q(x) \leftarrow P(x,y) \wedge \neg B'(x)} \tag{5.1}$$

δηλαδή, επαναγράφει μόνο τα θετικά άτομα που εμφανίζονται στα ερωτήματα που ανήκουν στο $\mathsf{ref}(\mathcal{R}_Q, \mathcal{T})$, χωρίς όμως να λαμβάνει υπόψη αξιώματα του $\mathcal{T}$ που έχουν υπαρξιακούς ποσοδείκτες στα δεξιά, και αντιγράφοντας τα αρνητικά άτομα όπως είναι. Επομένως, το σύνολο $\mathcal{R}^{\mathsf{ic}}$ σχηματίζεται ως εξής:

$$\mathcal{R}^{\mathsf{ic}} = \mathcal{R}^{\mathsf{ia}} \cup \{Q(x) \leftarrow P(x,y) \wedge \neg B'(x)\}$$

Αν αποτιμήσουμε το $\mathcal{R}^{\mathsf{ic}}$ στο $\mathcal{A}$, τότε πράγματι θα ανακτήσουμε και τις απαντήσεις "$\{a, c\}$". $\diamondsuit$

Μία πρώτη ενδιαφέρουσα παρατήρηση στο σημείο αυτό είναι ότι εφόσον κάθε θετικό άτομο του $\mathcal{R}^{\mathsf{ia}}$ ξανά-επαναγράφεται, το σύνολο των ερωτημάτων που τελικά υπολογίζει η $\mathsf{ref}^{\mathsf{ic}}$ είναι τόσο μεγάλο που πρακτικά η εκτέλεσή του στη βάση δεδομένων είναι πρακτικά ανέφικτη. Άλλωστε, για αυτόν τον λόγο οι Lembo et al. [76] που προτείνουν τη μεθοδολογία αυτή, δεν παρουσιάζουν σχετικά πειραματικά αποτελέσματα. Όμως, αν μελετήσουμε περαιτέρω το σύνολο που υπολογίζει η $\mathsf{ref}^{\mathsf{ic}}$ παρατηρούμε ότι υπάρχουν πολλά ερωτήματα που είναι περιττά. Στο Παράδειγμα 39, το ερώτημα $Q(x) \leftarrow P(x,y) \wedge \neg P'(x,y) \wedge \neg B'(x)$ είναι περιττό, αφού στο σύνολο $\mathcal{R}^{\mathsf{ic}}$ υπάρχει το ερώτημα $Q(x) \leftarrow P(x,y) \wedge \neg B'(x)$, που η εκτέλεσή του θα επιστρέψει περισσότερες απαντήσεις.

Μια δεύτερη ενδιαφέρουσα παρατήρηση είναι ότι αν εφαρμοστεί, στο Παράδειγμα 39 η ref *μόνο* στο ερώτημα $\mathcal{Q}$ και αποτιμηθεί στο σύνολο clc, τότε θα επιστραφούν οι ICAR-απαντήσεις. Συνεπώς, αρκεί να βρεθεί μία μέθοδος που να υπολογίζει αποδοτικά το clc. Όπως φαίνεται στο παράδειγμα που ακολουθεί, είναι πιθανό ο κορεσμός του $\mathcal{A}$ με βάση το $\mathcal{T}$, όπως υπολογίζεται από συστήματα κορεσμού δεδομένων, να μπορεί να προσεγγίσει το σύνολο $clc(\mathcal{T}, \mathcal{A})$.

Παράδειγμα 40. Έστω $\mathcal{T}$ η οντολογία, $\mathcal{A}$ το ABox και $\mathcal{Q}$ το ερώτημα του Παραδείγματος 39 και έστω κάποιο $\mathcal{RL}$ σύστημα κορεσμού δεδομένων ans. Δεδομένου ότι το $\mathcal{T}$ είναι εκφρασμένο στην $\mathcal{RL}$, το σύστημα ans θα υπολογίσει τον κορεσμό των $\mathcal{T}$, $\mathcal{A}$, δηλαδή το σύνολο $\mathcal{A}_s = \mathcal{A} \cup \{B(a), B(c)\}$. Επίσης, αν το σύνολο ref$(\{\mathcal{Q}\}, \mathcal{T}) = \{Q(x) \leftarrow B(x) \wedge \neg B'(x)\}$ εκτελεστεί στο $\mathcal{A}_s$ θα επιστραφεί το σύνολο $\{a, c\}$, δηλαδή, ακριβώς το σύνολο των ICAR απαντήσεων. ♢

Παρόλο που τα $\mathcal{RL}$ συστήματα κορεσμού δεδομένων είναι ιδιαίτερα αποδοτικά για $\mathcal{RL}$-οντολογίες, για υψηλότερες εκφραστικότητες αδυνατούν υπολογίσουν όλο το σύνολο των απαντήσεων. Για να αντιμετωπιστεί το πρόβλημα αυτό οι Stoilos et al. [129, 128] εισήγαγαν την έννοια του $\mathcal{T}$-συμπληρώματος για μία οντολογία $\mathcal{T}$ και ένα σύστημα ans. Πρακτικά, το $\mathcal{T}$-συμπλήρωμα αποτελείται από το ελάχιστο σύνολο αξιωμάτων που συμπεραίνονται από την οντολογία τέτοιο ώστε ένα $\mathcal{RL}$-σύστημα να δύναται να υπολογίσει όλες τις απαντήσεις ακόμα και για οντολογίες υψηλότερης εκφραστικότητας.

Παράδειγμα 41. Έστω το TBox $\mathcal{T}$, το ABox $\mathcal{A}$:

$$\begin{aligned}\mathcal{T} &= \{A \sqsubseteq \exists P,\ \exists P \sqsubseteq B\}\\ \mathcal{A} &= \{A(a)\}\end{aligned}$$

και το ερώτημα $\mathcal{Q} = Q(x) \leftarrow B(x)$. Όπως περιγράφεται στην Ενότητα 2.4.2, η $\mathcal{RL}$ δεν υποστηρίζει αξιώματα με υπαρξιακό ποσοδείκτη στα δεξιά, επομένως οποιοδήποτε $\mathcal{RL}$-ans σύστημα θα υπολογίσει το σύνολο ans$(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \emptyset$. Αν όμως προσθέσουμε στο $\mathcal{T}$ το αξίωμα $c = A \sqsubseteq B$, το οποίο συμπεραίνεται από το $\mathcal{T}$, τότε ans$(\mathcal{Q}, \mathcal{T} \cup \{c\} \cup \mathcal{A}) = \{a\}$. Το σύνολο $\{c\}$ αποτελεί ένα $\mathcal{T}$-συμπλήρωμα για το σύστημα ans. ♢

Στη συνέχεια, παραθέτουμε τον τυπικό ορισμό του $\mathcal{T}$-συμπληρώματος.

Ορισμός 42 ([129, 128])**.** Έστω $\mathcal{T}$ ένα TBox και έστω ans ένα $\mathcal{RL}$ ABox-σύστημα κορεσμού. Το *συμπλήρωμα* του $\mathcal{T}$ (ή $\mathcal{T}$-συμπλήρωμα) για το ans είναι ένα σύνολο $\mathcal{C}$ τέτοιο ώστε $\mathcal{T} \models \mathcal{C}$ και για κάθε ABox $\mathcal{A}$ και ΣΕ $\mathcal{Q}$ που δεν περιέχει υπαρξιακούς ποσοδείκτες ισχύει ότι: cert$(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) \subseteq$ ans$(\mathcal{Q}, \mathcal{T} \cup \mathcal{C} \cup \mathcal{A})$.

Επίσης, έστω $\mathcal{Q}$ ένα οποιοδήποτε ΣΕ. Αν υπάρχει μία επαναγραφή $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ για τα $\mathcal{Q}, \mathcal{T}$, τότε ισχύει ότι: cert$(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) \subseteq$ ans$(\mathcal{R}_Q, \mathcal{T} \cup \mathcal{C} \cup \mathcal{A})$.

Όπως φαίνεται στο επόμενο παράδειγμα, αν χρησιμοποιηθεί ένα $\mathcal{RL}$ σύστημα κορεσμού δεδομένων ans και το TBox-συμπλήρωμα μίας DL-Lite$_\mathcal{R}$ οντολογίας για το ans, τότε μπορούν να υπολογιστούν αποδοτικά οι ICAR-απαντήσεις.

Παράδειγμα 43. Έστω τα $\mathcal{T}$, $\mathcal{A}$ όπως έχουν οριστεί στο Παράδειγμα 39. Έστω επίσης τα $\mathcal{T}'$, $\mathcal{A}'$ που ορίζονται ως εξής:

$$\mathcal{T}' = \{C \sqsubseteq \exists P\} \cup \mathcal{T} \qquad \mathcal{A}' = \{C(f)\} \cup \mathcal{A}$$

Είναι εύκολο να παρατηρηθεί ότι $\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T}' \cup \mathcal{A}') = \mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) \cup \{f\}$. Έστω επίσης ένα $\mathcal{RL}$ ABox σύστημα κορεσμού ans. Το τμήμα του $\mathcal{T}'$ που θα λάβει το ans υπόψη είναι το $\mathcal{T}|_{\mathsf{rl}} = \{\exists P \sqsubseteq B\}$. Συμπεραίνουμε, έτσι, ότι το ans δε θα επιστρέψει την απάντηση f, δηλαδή, $f \notin \mathsf{ans}(\mathcal{Q}, \mathcal{T}' \cup \{C(f)\})$ αφού $\mathcal{T}|_{\mathsf{rl}} \cup \{C(f)\} \not\models B(f)$. Συνεπως, η διαδικασία που περιγράφεται στο Παράδειγμα 40 δεν θα επιστρέψει την απάντηση "f".

Αν, όμως, προσθέσουμε το $\mathcal{T}$-συμπλήρωμα $\mathcal{C} = \{C \sqsubseteq B\}$, τότε το τμήμα του $\mathcal{T} \cup \mathcal{C}$ το οποίο το ans θα λάβει υπόψη είναι το $\mathcal{T}|_{\mathsf{rl}} \cup \mathcal{C}$. Συνεπώς, θα υπολογίσει το σύνολο $clc(\mathcal{T}|_{\mathsf{rl}} \cup \mathcal{C}, \mathcal{A}')$, το οποίο είναι ίσο με το σύνολο $\mathcal{A}_C = \mathcal{A} \cup \{B(f)\}$. Τέλος, εκτελώντας το $Q(x) \leftarrow B(x) \wedge \neg B'(x)$ στο $\mathcal{A}_C$ θα επιστραφούν οι επιθυμητές ICAR-απαντήσεις, δηλαδή το σύνολο $\{a, c, f\}$. $\diamondsuit$

Στο σημείο αυτό πρέπει να σημειωθεί ότι πριν την εκτέλεση του ans θα πρέπει να αφαιρεθούν οι ισχυρισμοί α για τους οποίους το σύνολο $\mathcal{T} \cup \{\alpha\}$ είναι ασυνεπές.

Παράδειγμα 44. Έστω η οντολογία $\mathcal{T}$ του Παραδείγματος 39 και έστω το $\mathcal{T}''$:

$$\mathcal{T}'' = \{P \sqsubseteq P'\} \cup \mathcal{T}$$

Έστω επίσης το ABox $\mathcal{A}$ και το ερώτημα $\mathcal{Q}$ του ίδιου παραδείγματος. Στην περίπτωση αυτή, το P είναι μη-ικανοποιήσιμο, οπότε, $P(a, b)$, $P(c, d) \notin clc(\mathcal{T}, \mathcal{A})$ και άρα

$$\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T}'' \cup \mathcal{A}) = \emptyset$$

Αν όμως χρησιμοποιήσουμε κάποιο $\mathcal{RL}$ σύστημα κορεσμού δεδομένων, όπως στο Παράδειγμα 40, τότε θα επιστραφεί το σύνολο των απαντήσεων $\{a, c\}$, το οποίο είναι λάθος. Επίσης, αξίζει να σημειωθεί ότι οι ίδιες απαντήσεις θα επιστραφούν αν αποτιμήσουμε τον $\mathsf{ref}^{\mathsf{ic}}$ μετασχηματισμό στο $\mathcal{A}$:

$$\mathsf{ref}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T}'') = \{Q(x) \leftarrow B(x) \wedge \neg B'(x), Q(x) \leftarrow P(x, z) \wedge \neg B'(x)\}$$

Οι σωστές απαντήσεις θα ληφθούν αν πρώτα αφαιρέσουμε από το ABox τους ισχυρισμούς που αφορούν στο P, δηλαδη τους ισχυρισμούς $P(a, b)$, $P(c, d)$. Αν στην συνέχεια αποτιμήσουμε το $\mathcal{Q}(x) \leftarrow B(x) \wedge \neg B'(x)$ στο $A_s = \{P'(a, b), B(e), B'(e)\}$ θα πάρουμε το κενό σύνολο, το οποίο στην περίπτωση αυτή είναι και το σωστό σύνολο απαντήσεων. $\diamondsuit$

Από το προηγούμενο παράδειγμα αποδεικνύεται ότι ο αλγόριθμος που προτείνεται από τους Lembo et. al [76] δεν είναι ορθός, αφού μετά το δεύτερο βήμα της επαναγραφής θα πρέπει να ξανα-ελεγχθούν τα άτομα ως προς την ικανοποιησιμότητά τους. Συγκεκριμένα, αν στο Παράδειγμα 44 είχε εντοπιστεί ότι το άτομο $P(x, z)$, το οποίο δημιουργείται από την επαναγραφή του $B(x)$, είναι μη-ικανοποιήσιμο, τότε δεν θα αποτιμόταν το ερώτημα $Q(x) \leftarrow P(x, z) \wedge \neg B'(x)$ και επομένως δεν θα επιστρέφονταν οι απαντήσεις $\{a, c\}$.

Ορισμός 45. Έστω $\mathcal{T}$ ένα $\mathcal{DL}$-TBox και $\mathcal{A}$ ένα ABox. Συμβολίζουμε με $cr(\mathcal{A}, \mathcal{T})$ το σύνολο που προκύπτει από το $\mathcal{A}$ αφαιρώντας όλους τους ισχυρισμούς α για τους οποίους ισχύει ότι το σύνολο $\mathcal{T} \cup \{\alpha\}$ είναι ασυνεπές.

Στη συνέχεια, αποδεικνύουμε ότι η μέθοδος που προτείνουμε είναι ορθή.

Θεώρημα 46. *Έστω $\mathcal{Q}$ ένα ΣΕ, $\mathcal{T}$ ένα DL-Lite$_{\mathcal{R}}$ TBox, $\mathcal{A}$ ένα ABox και έστω $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T}$. Έστω* ans *ένα $\mathcal{RL}$ σύστημα κορεσμού δεδομένων και έστω $\mathcal{C} = \mathcal{T}^+ \cup \mathcal{C}$, όπου $\mathcal{C}$ το $\mathcal{T}$-συμπλήρωμα για το* ans*. Τότε,*

$$\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{ans}(\mathsf{ref}(\mathcal{R}_Q, \mathcal{T}), \mathcal{C} \cup cr(\mathcal{A}, \mathcal{T}))$$

Απόδειξη: Εφεξής θα συμβολίζουμε με $\mathcal{Q}(\vec{a})$ την φόρμουλα $\mathcal{Q}\sigma$, όπου $\sigma = \{\vec{x} \mapsto \vec{a}\}$, για ένα ερώτημα $\mathcal{Q}$ με διακεκριμένες μεταβλητές $\vec{x}$.

Έστω μια πλειάδα σταθερών $\vec{a}$ τέτοια ώστε $\vec{a} \in \mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A})$, δηλαδή, $\mathcal{T} \cup \mathcal{A} \models_{ICAR} \mathcal{Q}(\vec{a})$. Από τους Ορισμούς 10,12 προκύπτει ότι $\mathcal{T} \cup \mathcal{A} \models_{ICAR} \mathcal{Q}(\vec{a})$ ανν $\mathcal{T} \cup clc(\mathcal{T}, \mathcal{A}) \models_{IAR} \mathcal{Q}(\vec{a})$, το οποίο, σύμφωνα με το Λήμμα 18, ισχύει ανν υπάρχει UCQ επαναγραφή $\mathcal{R}$ για τα $\mathcal{Q}$, $\mathcal{T}$ τέτοια ώστε $clc(\mathcal{T}, \mathcal{A}) \models \mathsf{ref}(\mathcal{R}, \mathcal{T})(\vec{a})$. Δεδομένου ότι από την DL-Lite$_{\mathcal{R}}$ απουσιάζει ο τελεστής της ένωσης, το τελευταίο ισχύει ανν υπάρχει κάποιο $\mathcal{Q}' \in \mathsf{ref}(\mathcal{R}, \mathcal{T})$ τέτοιο ώστε $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$.

Έστω κάποιο $\mathcal{RL}$ ABox-σύστημα κορεσμού ans, ένα συμπλήρωμα $\mathcal{C}$ του $\mathcal{T}$ για το ans και το κορεσμένο σύνολο $\mathcal{A}_s$ του $cr(\mathcal{A}, \mathcal{T})$ υπολογισμένο από το ans χρησιμοποιώντας κάποιο $\mathcal{C}$.Έστω επίσης μία *οποιαδήποτε* επαναγραφή $\mathcal{R}' = \langle \mathcal{R}'_Q, \mathcal{R}'_D \rangle$ για τα $\mathcal{Q}, \mathcal{T}$ πιθανόν διαφορετική από το $\mathcal{R}$. Τότε, σύμφωνα με τον Ορισμό 5, $\mathsf{ans}(\mathsf{ref}(\mathcal{R}_Q, \mathcal{T}), \mathcal{C} \cup cr(\mathcal{A}, \mathcal{T})) = \mathsf{cert}(\mathsf{ref}(\mathcal{R}_Q, \mathcal{T}), \mathcal{A}_s)$. Συνεπώς, για να δειχθεί το θεώρημα αρκεί να δείξουμε ότι για κάθε $\vec{a}$ και για κάθε $\mathcal{Q}' \in \mathsf{ref}(\mathcal{R}, \mathcal{T})$, ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν υπάρχει $\mathcal{Q}'' \in \mathsf{ref}(\mathcal{R}'_Q, \mathcal{T})$ τέτοιο ώστε $\mathcal{A}_s \models \mathcal{Q}''(\vec{a})$. Αρχικά θα δείξουμε ότι για κάθε ισχυρισμό $\alpha \in HB(\mathcal{T} \cup \mathcal{A})$ ισχύει ότι $\alpha \in clc(\mathcal{T}, \mathcal{A})$ ανν $\alpha \in \mathcal{A}_s$, από το οποίο συμπεραίνεται ότι $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν $\mathcal{A}_s \models \mathcal{Q}'(\vec{a})$.

Είναι προφανές ότι $clc(\mathcal{T}, \mathcal{A}) \subseteq \mathcal{A}_s$, αφού το $\mathcal{C}$ είναι ένα συμπλήρωμα του $\mathcal{T}$ για το ans και αφού φορτώνουμε μόνο θετικά GCIs του $\mathcal{T}$ στο ans. Συνεπώς πρέπει να δείξουμε επίσης ότι $\mathcal{A}_s \subseteq clc(\mathcal{T}, \mathcal{A})$. Αντί αυτού, θα δείξουμε κάτι πιο ισχυρό—δηλαδή, ότι αν $\alpha \in \mathcal{A}_s$ τότε $\alpha \in clc(\mathcal{T}, \mathcal{A})$ και υπάρχει κάποιο συνεπές $\{\alpha'\} \subseteq \mathcal{A}$ τέτοιο ώστε $\{\alpha'\} \cup \mathcal{T} \models \alpha$.

Έστω $\mathcal{A}^s_i$ το κορεσμένο σύνολο όπως υπολογίζεται από το ans στο βήμα i, όπου $\mathcal{A}^1_s = cr(\mathcal{A}, \mathcal{T})$ και για κάποιο πεπερασμένο n ισχύει ότι $\mathcal{A}^n_s = \mathcal{A}_s$. Επίσης, μπορούμε να υποθέσουμε ότι το σύστημα ans δεν αφαιρεί ποτέ ισχυρισμούς από το $\mathcal{A}^i_s$ οπότε η ακολουθία είναι καλώς διατεταγμένη. Θα δείξουμε τον ισχυρισμό μας επαγωγικά στην κατασκευή του $\mathcal{A}_s$.

- *Βασικό Βήμα (i=1)*: Τότε, $\mathcal{A}^1_s = cr(\mathcal{A}, \mathcal{T})$. Το σύνολο $clc(\mathcal{T}, \mathcal{A})$ περιέχει όλους τους ισχυρισμούς $A(o)$, όπου το A είναι ικανοποιήσιμο, αφού για $S = \{A(o)\}$ ισχύει ότι $\mathcal{T} \cup S \models A(o)$, άρα από τον ορισμό του $cr(\cdot)$ ισχύει ότι $\mathcal{A}^1_s \subseteq clc(\mathcal{T}, \mathcal{A})$.

- *Επαγωγικό Βήμα.* Έστω ότι για κάθε ισχυρισμό $\alpha \in \mathcal{A}_s^i$, με $1 \leq i < n$, ισχύει ότι $\alpha \in clc(\mathcal{T}, \mathcal{A})$ και υπάρχει κάποιο συνεπές $\{\alpha'\} \subseteq \mathcal{A}$ τέτοιο ώστε $\{\alpha'\} \cup \mathcal{T} \models \alpha$. Έστω ότι στο επόμενο βήμα το σύστημα ans υπολογίζει και προσθέτει έναν νέο ισχυρισμό α''—δηλαδή, $\mathcal{A}_s^{i+1} = \mathcal{A}_s^i \cup \{\alpha''\}$. Υπάρχουν οι εξής περιπτώσεις σε σχέση με την μορφή του α:
  - Αν το α είναι της μορφής $A(a)$, τότε από το συντακτικό της DL-Lite$_\mathcal{R}$ και δεδομένου ότι το ans χρησιμοποιεί μόνο $\mathcal{RL}$ τμήμα του $\mathcal{C}$, ο ισχυρισμός αυτός μπορεί να υπολογιστεί από το ans αν υπάρχει: (i) κάποιος ισχυρισμός της μορφής $B(a) \in \mathcal{A}_s^i$ και κάποιο αξίωμα $B \sqsubseteq A \in \mathcal{C}$, ή (ii) κάποιος ισχυρισμός της μορφής $R(a, c) \in \mathcal{A}_s^i$ και κάποιο αξιώμα $\exists R \sqsubseteq A \in \mathcal{C}$, ή (iii) κάποιος ισχυρισμός της μορφής $R(c, a) \in \mathcal{A}_s^i$ και κάποιο αξίωμα $\exists R^- \sqsubseteq A \in \mathcal{C}$. Θα δείξουμε μόνο την περίπτωση (i) καθώς οι υπόλοιπες περιπτώσεις μπορούν να δειχθούν με παρόμοιο τρόπο. Από την επαγωγική υπόθεση ισχύει ότι θα υπάρχει κάποιος ισχυρισμος $\{\alpha'\} \subseteq \mathcal{A}$ τέτοιος ώστε $\{\alpha'\} \cup \mathcal{T} \models B(a)$. Από τον Ορισμό 42 προκύπτει ότι αφού $\mathcal{T} \models \mathcal{C}$, ισχύει ότι $\mathcal{T} \models B \sqsubseteq A$ και συνεπώς από το $\{\alpha'\} \cup \mathcal{T} \models B(a)$ προκύπτει ότι $\{\alpha'\} \cup \mathcal{T} \models A(a)$. Επίσης αφού το $\{\alpha'\}$ είναι συνεπές σε σχέση με το $\mathcal{T}$, ισχύει ότι $A(a) \in clc(\mathcal{T}, \mathcal{A})$.
  - Αν το α είναι της μορφής $R(a, b)$ τότε λόγω του συντακτικού της DL-Lite$_\mathcal{R}$ και του τμήματος της που υποστηρίζει το ans, ο ισχυρισμός αυτός μπορεί να υπολογιστεί από το ans είτε από έναν ισχυρισμό $P(a, b) \in \mathcal{A}_s^i$ και ένα αξίωμα $P \sqsubseteq R \in \mathcal{C}$ ή από έναν ισχυρισμό $P(b, a) \in \mathcal{A}_s^i$ και ένα αξίωμα $P^- \sqsubseteq R \in \mathcal{C}$. Όπως προηγουμένως μπορούμε να δείξουμε τον ισχυρισμό μας χρησιμοποιώντας την επαγωγική υπόθεση και την ορθότητα του $\mathcal{T}$-συμπληρώματος.

Από τα παραπάνω προκύπτει ότι για κάθε $\vec{\alpha}$ και για κάθε $\mathcal{Q}' \in \mathsf{ref}(\mathcal{R}, \mathcal{T})$, ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν $\mathcal{A}_s \models \mathcal{Q}'(\vec{a})$. Συνεπώς, αρκεί να δείξουμε ότι για *οποιαδήποτε* επαναγραφή $\mathcal{R}' = \langle \mathcal{R}'_Q, \mathcal{R}'_D \rangle$ υπάρχει κάποιο $\mathcal{Q}'' \in \mathsf{ref}(\mathcal{R}'_Q, \mathcal{T})$ τέτοιο ώστε $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν $\mathcal{A}_s \models \mathcal{Q}''(\vec{a})$.

Καταρχήν, ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν $[clc(\mathcal{T}, \mathcal{A})]^{ia} \models \mathcal{Q}^-(\vec{a})$, όπου το $\mathcal{Q}^-$ προκύπτει από το $\mathcal{Q}'$ αν αφαιρέσουμε τα αρνητικά του άτομα και το $[clc(\mathcal{T}, \mathcal{A})]^{ia}$ είναι η IAR διόρθωση του $clc(\mathcal{T}, \mathcal{A})$. Αφού το $\mathcal{R}'$ είναι επίσης μία επαναγραφή, υπάρχει κάποιο $\mathcal{Q}'' \in \mathcal{R}'_Q$ τέτοιο ώστε $clc(\mathcal{T}, \mathcal{A}) \models \mathcal{Q}'(\vec{a})$ ανν $\mathcal{R}'_D \cup [clc(\mathcal{T}, \mathcal{A})]^{ia} \models \mathcal{Q}''(\vec{a})$, το οποίο, σύμφωνα με τα προηγούμενα ισχύει ανν $\mathcal{R}'_D \cup [\mathcal{A}_s]^{ia} \models \mathcal{Q}''(\vec{a})$ το οποίο ισχύει ανν $\mathcal{R}'_D \cup \mathcal{A}_s \models \mathsf{ref}(\mathcal{Q}''(\vec{a}), \mathcal{T})$. Όμως, αφου το $\mathcal{A}_s$ έχει υπολογιστεί με την χρήση του TBox-συμπληρώματος, θα είναι πλήρες και αφού το $\mathcal{R}'_D$ είναι datalog πρόγραμμα προκύπτει ότι $\mathcal{A}_s \models \mathsf{ref}(\mathcal{Q}''(\vec{a}), \mathcal{T})$.

■

Αξίζει να σημειωθεί ότι για DL-Lite$_\mathcal{R}$ οντολογίες, πάντα υπάρχουν TBox-συμπληρώματα για $\mathcal{RL}$ συστήματα καθώς και επαναγραφές για οποιοδήποτε ΣΕ. Συνεπώς το αποτελέσμά μας είναι ιδιαίτερης πρακτικής σημασίας.

5.1.1 Επιπλέον Βελτιστοποιήσεις

Η βελτιστοποίηση που προτείνουμε στην ενότητα αυτή βασίζεται στην ιδιότητα που έχουν οι ABox αλγόριθμοι κορεσμού να υπολογίζουν όλους τους δυνατούς ισχυρισμούς που συνεπάγονται από τη βάση γνώσης. Όπως φαίνεται στο παρακάτω παράδειγμα η ιδιότητα αυτή μας επιτρέπει να απλοποιήσουμε περαιτέρω το ερώτημα όπως σχηματίζεται από τον μετασχηματισμό ref.

Παράδειγμα 47. Έστω τα ακόλουθα TBox και ABox:

$$\begin{aligned}\mathcal{T} &= \{B \sqsubseteq A, B' \sqsubseteq A', A \sqsubseteq \neg A'\}\\ \mathcal{A} &= \{B(a), A'(a), B(b), B'(b)\}\end{aligned}$$

και έστω επίσης το ερώτημα $\mathcal{Q} = Q(x) \leftarrow B(x)$. Τότε ισχύει ότι $\mathsf{ref}(\mathcal{Q},\mathcal{T}) = \{\mathcal{Q}'\}$, όπου $\mathcal{Q}' = Q(x) \leftarrow B(x) \wedge \neg A'(x) \wedge \neg B'(x)$. Όταν το ref αποτιμηθεί στο $\mathcal{A}$ θα επιστραφεί το $\emptyset$ σύνολο. Είναι σημαντικό να παρατηρηθεί ότι η παρουσία και των δύο αρνητικών ατόμων απαιτείται στο ερώτημα. Το άτομο $\neg A'(x)$ χρειάζεται για να μην επιστραφεί η απάντηση "a" και το $\neg B'(x)$ για να μην επιστραφεί η απάντηση "b".

Έστω τώρα ένα $\mathcal{RL}$ ABox σύστημα κορεσμού ans. Το $\mathcal{T}$-συμπλήρωμα για το ans είναι το σύνολο $\mathcal{C} = \emptyset$. Συνεπώς, ο κορεσμός του $\mathcal{A} \cup \mathcal{T}$ όπως υπολογίζεται από το ans είναι το σύνολο $\mathcal{A}_s = \mathcal{A} \cup \{A(a), A(b), A'(b)\}$. Όπως είναι φανερό, θα μπορούσε το $\mathcal{Q}'$ να μην περιέχει το άτομο $\neg B'(x)$, εφόσον περιέχει το άτομο $\neg A'(x)$ και $A'(b) \in \mathcal{A}_s$. Πράγματι, αποτιμώντας το $\mathcal{Q}'' = Q(x) \leftarrow B(x) \wedge \neg A'(x)$ στο $\mathcal{A}_s$ λαμβάνουμε το $\emptyset$ σύνολο, το οποίο και είναι το σύνολο των απαντήσεων που πρέπει να λάβουμε. $\diamondsuit$

Όπως υποδεικνύει το προηγούμενο παράδειγμα, η αξιοποίηση συστημάτων κορεσμού επιτρέπει στον μετασχηματισμό του $\mathcal{Q}$ να γίνει με τέτοιο τρόπο ώστε να περιέχει λιγότερα αρνητικά άτομα απο αυτά που προκύπτουν χρησιμοποιώντας τον μετασχηματισμό ref. Διαισθητικά, από ένα ερώτημα μπορούν να αφαιρεθούν τα αρνητικά άτομα εκείνα για τα οποία υπάρχει κάποιο άλλο αρνητικό άτομο στο ερώτημα το οποίο τα υπάγει. Επομένως, όποια στιγμιότυπα συμπεραίνεται από τη ΒΓ ότι ανήκουν στα αφαιρούμενα άτομα συμπεραίνεται επίσης ότι ανήκουν και σε αυτά που τα υπάγουν.

Ορισμός 48. Έστω $\mathcal{T}$ ένα $\mathcal{DL}$-TBox και έστω $\mathcal{Q}$ ένα ΣΕ. Ένα άτομο $\neg A(x)$ που εμφανίζεται στο $\mathcal{Q}$ ονομάζεται *καλυμμένο* αν υπάρχει κάποιο άλλο άτομο $\neg B(x)$ στο $\mathcal{Q}$ τέτοιο ώστε $\mathcal{T} \models A \sqsubseteq B$. Ένα άτομο $\neg R(x,y)$ που εμφανίζεται στο $\mathcal{Q}$ ονομάζεται *καλυμμένο* αν υπάρχει κάποιο άλλο άτομο $\neg A(x)$, $\neg A(y)$, $\neg S(x,y)$, ή $\neg S(y,x)$ στο $\mathcal{Q}$ τέτοιο ώστε $\mathcal{T} \models \exists R \sqsubseteq A$, $\exists R^- \sqsubseteq A$, $R \sqsubseteq S$, ή $R \sqsubseteq S^-$, αντιστοίχως. Συμβολίζουμε με $\mathsf{min}(\mathcal{Q},\mathcal{T})$ το ερώτημα που προκύπτει από το $\mathcal{Q}$ αφαιρώντας όλα τα καλυμμένα άτομα. Επίσης, για ένα UCQ $\mathcal{R}$ ορίζουμε: $\mathsf{ref}_{\mathsf{min}}(\mathcal{R},\mathcal{T}) = \bigvee_{\mathcal{Q}' \in \mathsf{ref}(\mathcal{R},\mathcal{T})} \mathsf{min}(\mathcal{Q}',\mathcal{T})$.

Στη συνέχεια, θα αποδείξουμε την ορθότητα του αλγορίθμου.

Θεώρημα 49. *Έστω $\mathcal{Q}$ ένα ΣΕ, $\mathcal{T}$ ένα DL-Lite$_{\mathcal{R}}$ TBox, $\mathcal{A}$ ένα ABox και έστω $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ μια επαναγραφή για τα $\mathcal{Q}, \mathcal{T}$. Έστω επίσης* ans *ένα $\mathcal{RL}$ ABox-σύστημα κορεσμού και έστω $\mathcal{C} = \mathcal{T}^+ \cup C$, όπου C είναι ένα $\mathcal{T}$-συμπλήρωμα για το* ans. *Τότε,*

$$\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{ans}(\mathsf{ref}_{\mathsf{min}}(\mathcal{R}_Q, \mathcal{T}), \mathcal{C} \cup cr(\mathcal{A}, \mathcal{T}))$$

Απόδειξη: Έστω ότι το $\mathcal{A}_s$ είναι ο κορεσμός του $\mathcal{A}$ υπολογισμένος από το ans χρησιμοποιώντας το σύνολο $\mathcal{C}$. Έστω ένα ΣΕ $\mathcal{Q}' \in \mathsf{ref}(\mathcal{R}_Q, \mathcal{T})$ και έστω $\mathcal{Q}_m = \mathsf{min}(\mathcal{Q}', \mathcal{T})$. Θα δειξουμε ότι αν για κάποιο $\vec{a}$ ισχύει ότι $\mathcal{A}_s \models \mathcal{Q}_m(\vec{a})$ τότε ισχύει επίσης ότι $\mathcal{A}_s \models \mathcal{Q}'(\vec{a})$ (το αντίστροφο είναι προφανές αφού το $\mathsf{min}(\mathcal{Q}', \mathcal{T})$ περιέχει ακριβώς τα ίδια θετικά άτομα με το $\mathcal{Q}$ και λιγότερα αρνητικά).

Η απόδειξη θα γίνει με εις άτοπο απαγωγή. Ας υποθέσουμε ότι για κάποιο $\vec{a}$ ισχύει ότι $\mathcal{A}_s \models \mathcal{Q}_m(\vec{a})$ αλλά $\mathcal{A}_s \not\models \mathcal{Q}'(\vec{a})$. Από τον ορισμό του min προκύπτει ότι τα $\mathcal{Q}'$, $\mathcal{Q}_m$ έχουν τα ίδια θετικά άτομα, έστω $\mathcal{B}(\vec{x}, \vec{y})$, οπότε από την υπόθεση ισχύει ότι $\mathcal{A}_s \models \mathcal{B}(\vec{a}, \vec{c})$. Επίσης, κανένας ισχυρισμός από τα αρνητικά άτομα που εμφανίζονται στο $\mathcal{Q}_m(\vec{a})$ δεν εμφανίζεται στο $\mathcal{A}_s$, αφού $\mathcal{A}_s \models \mathcal{Q}_m(\vec{a})$. Συνεπώς, ισχύει το $\mathcal{A}_s \not\models \mathcal{Q}'(\vec{a})$ μόνο αν υπάρχει κάποιο αρνητικό άτομο $\neg At(\vec{z}) \in \mathcal{Q}'$ και κάποιος ισχυρισμός $At(\vec{d})$ στο $\mathcal{A}_s$ και το $\neg At(\vec{z})$ δεν εμφανίζεται στο $\mathcal{Q}_m$. Αφού το $\mathcal{Q}_m$ παράγεται από το $\mathcal{Q}'$ το άτομο $\neg At(\vec{z})$ θα απομακρυνθεί από το $\mathcal{Q}'$ μόνο αν ισχύει μία από τις περιπτώσεις που περιγράφονται στον Ορισμό 48. Για παράδειγμα, το $At(\vec{z})$ μπορεί να είναι της μορφής $A(x)$ και το $\mathcal{A}_s$ να περιέχει τον ισχυρισμό $A'(d)$. Τότε όμως, σύμφωνα με τον ορισμό, το $\mathcal{Q}_m$ πρέπει να περιέχει το $\neg A'(x)$ και θα πρέπει να ισχύει ότι $\mathcal{T} \models A \sqsubseteq A'$ και αφού το C είναι το $\mathcal{T}$-συμπλήρωμα για το ans, το $\mathcal{A}_s$ θα πρέπει επίσης να περιέχει το $A'(d)$, το οποίο είναι άτοπο, αφού $\mathcal{A}_s \models \mathcal{Q}_m(\vec{a})$. Η απόδειξη είναι παρόμοια για όλες τις άλλες περιπτώσεις και μορφές του $\neg At(\vec{z})$. ■

Όπως φαίνεται στην Ενότητα 6 που παρουσιάζονται τα πειραματικά αποτελέσματα, με τον τρόπο αυτό, το μέγεθος του ερωτήματος μπορεί να μειωθεί έως και δύο τάξεις μεγέθους, αυξάνοντας σημαντικά την απόδοση του συστήματος.

5.2 Υπολογισμός ICAR-απαντήσεων σε εκφραστικές ΠΛ

Όπως αναφέρεται στην Ενότητα 2.4.3, το πρόβλημα απάντησης ερωτημάτων με βάση την ICAR σημασιολογία, ακόμα και για Περιγραφικές Λογικές με πολυνωμιακή πολυπλοκότητα ως προς τα δεδομένα, όπως είναι η $\mathcal{EL}_\perp$, έχει αποδειχθεί ότι είναι coNP-hard [115] ενώ το πρόβλημα δεν μπορεί να γίνει βατό ακόμα και αν εφαρμοστούν συντακτικοί περιορισμοί [115].

Ο αλγόριθμος που προτείναμε στην προηγούμενη ενότητα μπορεί να αποτελέσει τη βάση προσεγγιστικών αλγορίθμων για τον υπολογισμό ICAR απαντήσεων ακόμα και για πιο εκφραστικές ΠΛ. Ωστόσο, όπως φαίνεται στο παράδειγμα που ακολουθεί, η απευθείας εφαρμογή του αλγορίθμου αυτού μπορεί να επιφέρει μη αναμενόμενα αποτελέσματα.

Παράδειγμα 50. Έστω το $\mathcal{EL}_\perp$-TBox $\mathcal{T} = \{A \sqcap B \sqsubseteq C,\ A \sqcap B \sqsubseteq \perp\}$, το ABox $\mathcal{A} =$

$\{A(a), B(a)\}$, και το ΣΕ $\mathcal{Q} = Q(x) \leftarrow C(x)$. Προφανώς, ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) = \mathcal{A}$ και οπότε $\mathsf{cert}^{\mathsf{ic}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \emptyset$.

Έστω ένα $\mathcal{RL}$ ABox-σύστημα κορεσμού ans. Αν φορτώσουμε το $\mathcal{T} \cup \mathcal{A}$ στο ans, τότε το ans θα υπολογίσει τον κορεσμό $\mathcal{A}_s = \mathcal{A} \cup \{C(a)\}$, αφού το $A \sqcap B \sqsubseteq C$ αντιστοιχεί στον datalog κανόνα $A(x) \wedge B(x) \to C(x)$. Τέλος, $\mathsf{ref}(\mathcal{Q}, \mathcal{T}) = \{\mathcal{Q}\}$, οπότε προκύπτει ότι $\mathsf{cert}(\mathsf{ref}(\mathcal{Q}, \mathcal{T}), \mathcal{A}_s) = \{a\}$. ♢

Όπως φαίνεται στο προηγούμενο παράδειγμα, για πιο εκφραστικές γλώσσες από την DL-Lite$_\mathcal{R}$, τα ABox-συστήματα κορεσμού υπολογίζουν περισσότερους ισχυρισμούς και επομένως περισσότερες απαντήσεις. Το πρόβλημα αυτό μπορεί να επιλυθεί με δύο τρόπους. Ο ένας τρόπος είναι να επεκταθεί ο μετασχηματισμός ref με τέτοιο τρόπο ώστε να προστεθούν περισσότερα αρνητικά άτομα (δηλαδή, στο προηγούμενο παράδειγμα, το $\mathsf{ref}(\mathcal{Q}, \mathcal{T})$ να επεκταθεί με το $\neg A(x) \wedge \neg B(x)$). Με αυτόν τον τρόπο όμως θα περιοριστούμε στην συγκεκριμένη γλώσσα που μελετάμε. Ο άλλος τρόπος, κάπως πιο γενικός, είναι να μορφοποιηθεί το σύστημα ans με τέτοιο τρόπο ώστε να μην παραχθούν εξαρχής κάποιοι ισχυρισμοί (στο προηγούμενο παράδειγμα, ο ισχυρισμός $C(a)$).

Η μέθοδος που ακολουθήθηκε στην εργασία αυτή, βασίζεται στον δεύτερο τρόπο. Στο πλαίσιο αυτό, προτείνεται ο Αλγόριθμος 5, ο οποίος υπολογίζει τον κορεσμό ABox με βάση το $\mathcal{RL}$ τμήμα του TBox λαμβάνοντας υπόψη τα αρνητικά γενικευμένα αξιώματα. Συγκεκριμένα, ο αλγόριθμος ApproxAns αφού δεχτεί στην είσοδο ένα TBox $\mathcal{T}$ και ένα ABox $\mathcal{A}$, προσθέτει στο $\mathcal{A}_s$ το $\mathcal{A}$ αφαιρώντας τους ισχυρισμούς που αντιστοιχούν σε μηικανοποιήσιμες έννοιες/ρόλους (γραμμή 2) και μετασχηματίζει τα αξιώματα του $\mathcal{T}^+|_{\mathsf{rl}}$ σε κανόνες datalog (γραμμή 3). Στη συνέχεια, προσθέτει στο $\mathcal{A}_s$ τους ισχυρισμούς που παράγονται από το $\mathcal{T}^+|_{\mathsf{rl}} \cup \mathcal{A}$, ελέγχοντας αν για ένα αξίωμα $\mathcal{B} \to H$, που ανήκει στο $\mathcal{T}^+|_{\mathsf{rl}}$, υπάρχει κάποιο σύνολο ισχυρισμών σ, που ανήκει στο $\mathcal{A}_s$ (όπως αυτό έχει σχηματιστεί την στιγμή του ελέγχου), τέτοιο ώστε το $\mathcal{B}\sigma$ να είναι συνεπές με βάση το $\mathcal{T}$. Στην περίπτωση αυτή, συμπεραίνεται ότι και το $H\sigma$ είναι συνεπές και άρα προστίθενται στο $\mathcal{A}_s$. Δηλαδή, πρακτικά, προσθέτει όλους τους νέους ισχυρισμούς που μπορούν να παραχθούν από όλα τα συνεπή υποσύνολα του $\mathcal{A}$ και την οντολογία και τα προσθέτει στο $\mathcal{A}_s$. Στη συνέχεια, υπολογίζει όλους τους νέους ισχυρισμούς που μπορούν να παραχθούν από όλα τα συνεπή υποσύνολα του $\mathcal{A}_s$ κοκ. Ο αλγόριθμος τερματίζει όταν σταματήσει να παράγει νέους ισχυρισμούς.

Είναι σημαντικό να σημειωθεί ότι ο έλεγχος της συνέπειας που εκτελείται στη γραμμή 7 μπορεί να εφαρμοστεί σε πολυωνυμικό χρόνο για πολλές ΠΛ.

Όπως αποδεικνύεται στην παρακάτω πρόταση, για DL-Lite$_\mathcal{R}$ οντολογίες $\mathcal{T}$ ο Αλγόριθμος 5 υπολογίζει το σύνολο $clc(\mathcal{T}, \mathcal{A})$ για οποιοδήποτε $\mathcal{A}$.

Πρόταση 51. *Έστω $\mathcal{T}$ ένα DL-Lite$_\mathcal{R}$ TBox, $\mathcal{A}$ ένα ABox και $\mathcal{Q}$ ένα ΣΕ. Έστω επίσης* ans *κάποιο $\mathcal{RL}$ ABox-σύστημα κορεσμού και $\mathcal{C}$ το $\mathcal{T}$-συμπλήρωμα για το* ans*. Τότε,* $clc(\mathcal{T}, \mathcal{A}) = \mathsf{ApproxAns}(\mathcal{T} \cup \mathcal{C}, \mathcal{A})$.

Απόδειξη: Σύμφωνα με την γραμμή 7, ένας ισχυρισμός $\mathcal{H}\sigma$ που συμπεραίνεται από το $\mathcal{B}\sigma$ και το $\mathcal{B} \to \mathcal{H}$ θα προστεθεί στο $\mathcal{A}_s$ μόνο αν το $\mathcal{T} \cup \mathcal{B}\sigma$ είναι συνεπές και ισχύει

Algorithm 5 ApproxAns($\mathcal{T}, \mathcal{A}$)

Είσοδος: TBox $\mathcal{T}$, ABox $\mathcal{A}$

1: **Έξοδος:** Το κορεσμένο σύνολο $\mathcal{A}_s$

2: $\mathcal{A}_s := cr(\mathcal{A}, \mathcal{T})$

3: $\mathcal{P} :=$ toRules($\mathcal{T}^+|_{\mathsf{rl}}$)

4: **while** Δεν προστίθενται νέοι ισχυρισμοί στο $\mathcal{A}_s$ **do**

5:     $\mathcal{A}_t := \emptyset$

6:     **for all** $\mathcal{B} \to H \in \mathcal{P}$ τέτοιο ώστε $\sigma \subseteq \mathcal{A}_s$ για κάποιο σ **do**

7:         **if** $\mathcal{B}\sigma$ είναι συνεπές με βάση το $\mathcal{T}$ **do**

8:             $\mathcal{A}_t := \mathcal{A}_t \cup \{H\sigma\}$

9:     $\mathcal{A}_s := \mathcal{A}_s \cup \mathcal{A}_t$

10: **return** $\mathcal{A}_s$

$\mathcal{T} \cup \mathcal{B}\sigma \models \mathcal{H}\sigma$. Λόγω της δομής των DL-Lite$_\mathcal{R}$ αξιωμάτων, προκύπτει ότι το $\mathcal{B}$ είναι ένα μόνο άτομο (έννοια ή ρόλος) $A(x)$ ($R(x, y)$) και το $\mathcal{B}\sigma$ έχει την μορφή $A(a)$ ($R(a, b)$). Από την απόδειξη του Θεωρήματος 46 ισχύει ότι είτε $A(a) \in \mathcal{A}$ ή υπάρχει κάποιο $\alpha \in \mathcal{A}$ τέτοιο ώστε $\{\alpha\} \cup \mathcal{T} \models A(a)$, οπότε το $A(a)$ πρέπει να είναι συνεπές με βάση το $\mathcal{T}$, διαφορετικά το $\{\alpha\}$ θα ήταν ασυνεπές, όμως αφού ο αλγόριθμος στη γραμμή 2 αφαιρεί όλα τους ισχυρισμούς α για τους οποίους το $\mathcal{T} \cup \{\alpha\}$ είναι ασυνεπές, το $\mathcal{A}$ δεν περιέχει ισχυρισμούς τέτοιων ατόμων. Συνεπώς, προκύπτει άμεσα από την δομή του αλγορίθμου ότι το σύνολο $\mathcal{A}_s$ είναι ο κορεσμός του $cr(\mathcal{A}, \mathcal{T})$ με βάση το σύνολο $\mathcal{T} \cup \mathcal{C}$, οπότε με βάση την απόδειξη του Θεωρήματος 46 συμπεραίνεται ότι είναι το σύνολο $clc(\mathcal{T}, \mathcal{A})$. ■

Ωστόσο, για πιο εκφραστικές ΠΛ ο αλγόριθμος προσεγγίζει άνωθεν το $clc(\mathcal{T}, \mathcal{A})$.

Θεώρημα 52. *Έστω $\mathcal{T}$ ένα $\mathcal{DL}$-TBox,* ans *ένα $\mathcal{RL}$ ABox-σύστημα κορεσμού, και $\mathcal{C}$ ένα $\mathcal{T}$-συμπλήρωμα για το* ans. *Τότε, ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) \subseteq$* ApproxAns($\mathcal{T} \cup \mathcal{C}, \mathcal{A}$)

Απόδειξη: Έστω κάποιο $\alpha \in clc(\mathcal{T}, \mathcal{A})$. Από τον ορισμό του clc, προκύπτει ότι υπάρχει κάποιο υποσύνολο $S \subseteq \mathcal{A}$ συνεπές με βάση το $\mathcal{T}$ τέτοιο ώστε $\mathcal{T} \cup S \models \alpha$. Αφού το $\mathcal{C}$ είναι ένα $\mathcal{T}$-συμπλήρωμα για το ans, ισχύει ότι $\mathcal{T}|_{\mathsf{rl}} \cup \mathcal{C} \cup S \models \alpha$. Επίσης από τον ορισμό της γλώσσας $\mathcal{RL}$, ισχύει ότι το $\mathcal{T}|_{\mathsf{rl}} \cup \mathcal{C}$ μπορεί να μετασχηστιστεί σε ένα datalog πρόγραμμα $\mathcal{P}$. Συνεπώς, $\alpha \in MM(\mathcal{P} \cup S)$, όπου $MM(\mathcal{P} \cup S)$ είναι το ελάχιστο μοντέλο του $\mathcal{P}$ στο S. Το ελάχιστο μοντέλο μπορεί να υπολογιστεί υπολογίζοντας τον κορεσμό των ισχυρισμών του S με εμπρόσθιο τρόπο (forward-style manner). Συγκεκριμένα, για $S^0 = S$ και για κάποιο S^i υπολογισμένο στο βήμα i, αν υπάρχει κάποιο $\mathcal{B} \to H \in \mathcal{P}$ και κάποια αντιστοίχιση σ από τις μεταβλητές του $\mathcal{B}$ στις σταθερές του S^i τέτοια ώστε $\mathcal{B}\sigma \subseteq S^i$, τότε το S^{i+1} περιέχει το $H\sigma$. Στο βήμα 0 κάθε τέτοιο $\mathcal{B}\sigma$ είναι συνεπές (αφού είναι υποσύνολο του συνεπούς υποσυνόλου S). Επίσης, το $S^0 \cup \{\alpha\}$ είναι επίσης συνεπές (αφού το α έχει συμπεραχθεί από ένα συνεπές σύνολο από ισχυρισμούς και λόγω της αρχής της μονοτονίας της λογικής πρώτης τάξης). Συνεπώς όλα τα S^i και τα υποσύνολα $\mathcal{B}\sigma$ είναι συνεπή, άρα, το $\mathcal{A}_s$ όπως υπολογίζεται από τον Αλγόριθμο 5 είναι το ελάχιστο μοντέλο του $\mathcal{P}$ στο

S και $\alpha \in \mathcal{A}_s$. ■

5.2.1 Η σημασιολογία ICAR$^+$

Για να υπολογίσει ο Αλγόριθμος 5 το σύνολο $clc(\mathcal{T}, \mathcal{A})$, για κάθε ισχυρισμό $H\sigma$ που συμπεραίνεται στη γραμμή 6 θα πρέπει να ελέγχεται αν υπάρχει κάποιο υποσύνολο $\mathcal{A}'$ του *αρχικού* ABox $\mathcal{A}$ συνεπές με βάση το $\mathcal{T}$ τέτοιο ώστε $\mathcal{T} \cup \mathcal{A}' \models \mathcal{B}\sigma$ (και συνεπώς $\mathcal{T} \cup \mathcal{A}' \models H\sigma$). Αντίθετα, ο αλγόριθμος που προτείνεται εφαρμόζει ενός είδους "τοπικό" έλεγχο πριν προσθέσει το $H\sigma$ στο $\mathcal{A}_s$, δηλαδή, ελέγχει αν το $\mathcal{B}\sigma$ είναι συνεπές με βάση το *παραγόμενο* σύνολο ισχυρισμών (το κορεσμένο σύνολο) $\mathcal{A}_s$. Στο παρακάτω παράδειγμα διαφαίνεται η διαφορά μεταξύ του συνόλου που υπολογίζει ο Αλγόριθμος 5 και του συνόλου $clc(\cdot)$, για πιο εκφραστικές ΠΛ.

Παράδειγμα 53. Έστω τα ακόλουθα TBox και ABox:

$$\begin{aligned}\mathcal{T} &= \{A \sqcap B \sqsubseteq C,\ D \sqsubseteq B,\ A \sqsubseteq \neg D\}\\ \mathcal{A} &= \{A(a), D(a)\}\end{aligned}$$

Ισχύει ότι $clc(\mathcal{T}, \mathcal{A}) = \mathcal{A} \cup \{B(a)\}$. Ο Αλγόριθμος 5 όμως θα υπολογίσει και θα *προσθέσει* το $B(a)$ στο σύνολο $\mathcal{A}_s$ και στη συνέχεια θα υπολογίσει και θα προσθέσει επίσης το $C(a)$ εξαιτίας του συνεπούς υποσυνόλου $\{A(a), B(a)\} \subseteq \mathcal{A}_s$. $\diamondsuit$

Είναι φανερό ότι αν το $B(a)$ εμφανιζόταν ήδη στο $\mathcal{A}$, τότε το $C(a)$ θα άνηκε στο σύνολο $clc(\mathcal{T}, \mathcal{A})$. Δηλαδή, προσθέτοντας ισχυρισμούς που συμπεραίνονται με συνεπή τρόπο από τη βάση γνώσης μπορεί να αυξηθεί το σύνολο clc (οπότε και οι ICAR απαντήσεις), κάτι το οποίο δε συμβαίνει με τις συνήθεις σημασιολογίες πρώτης τάξης, όπου για κάθε Σ και ϕ τέτοια ώστε $\Sigma \models \phi$, τα Σ και $\Sigma \cup \{\phi\}$ είναι ισοδύναμα. Δηλαδή, παρατηρείται ότι για πιο εκφραστικές ΠΛ η ICAR σημασιολογία εμφανίζει την ίδια "δυσλειτουργία" με την IAR σημασιολογία, εξαρτάται από το συντακτικό της ΒΓ.

Σύμφωνα με τα προηγούμενα, εισάγουμε μια προέκταση της συνάρτησης clc, η οποία πρακτικά υπολογίζεται από τον Αλγόριθμο 5.

Ορισμός 54. Έστω $\mathcal{T}$ ένα $\mathcal{DL}$ TBox και έστω $\mathcal{A}$ ένα ABox. Ορίζουμε το σύνολο $clc^+(\mathcal{T}, \mathcal{A})$ ως το ελάχιστο σύνολο ισχυρισμών που ικανοποιεί τις παρακάτω συνθήκες:

- $cr(\mathcal{A}, T) \subseteq clc^+(\mathcal{T}, \mathcal{A})$
- Αν υπάρχει κάποιο $\mathcal{A}' \subseteq clc^+(\mathcal{T}, \mathcal{A})$ συνεπές με βάση το $\mathcal{T}$ τέτοιο ώστε $\mathcal{T} \cup \mathcal{A}' \models \alpha$, τότε το $clc^+(\mathcal{T}, \mathcal{A})$ περιέχει το α.

Σε αντίθεση με το clc, στο clc^+ οποιοσδήποτε ισχυρισμός που συμπεραίνεται από τη βάση γνώσης με συνεπή τρόπο μπορεί να χρησιμοποιηθεί να για συμπεραχθούν νέοι ισχυρισμοί. Συνεπώς, η παρακάτω πρόταση είναι προφανής.

Πρόταση 55. *Για ένα $\mathcal{DL}$ TBox $\mathcal{T}$ και ένα ABox $\mathcal{A}$ ισχύει:* $clc(\mathcal{T}, \mathcal{A}) \subseteq clc^+(\mathcal{T}, \mathcal{A})$.

Επίσης, σε αντίθεση με το clc, για πολλές ΠΛ το clc^+ μπορεί να υπολογιστεί σε πολυωνυμικό χρόνο.

Θεώρημα 56. *Έστω $\mathcal{DL}$ κάποια ΠΛ τέτοια ώστε το πρόβλημα της συνέπειας ενός ABox μπορεί να επιλυθεί σε πολυωνυμικό χρόνο και έστω ότι η $\mathcal{DL}$ είναι datalog επαναγράψιμη. Τότε, για κάθε $\mathcal{DL}$-TBox $\mathcal{T}$ και ABox $\mathcal{A}$, το $clc^+(\mathcal{T}, \mathcal{A})$ μπορεί να υπολογιστεί σε πολυωνυμικό χρόνο ως προς το μέγεθος των δεδομένων.*

Απόδειξη: Από την υπόθεση, η $\mathcal{DL}$ μπορεί να επαναγραφεί σε datalog, συνεπώς για κάθε $\mathcal{T} \in \mathcal{DL}$ υπάρχει κάποιο datalog πρόγραμμα $\mathcal{R}_D$ τέτοιο ώστε για κάθε ερώτημα $\mathcal{Q}$ που δεν περιέχει υπαρξιακές μεταβλητές το $\mathcal{R} = \langle\{\mathcal{Q}\}, \mathcal{R}_D\rangle$ να είναι μία datalog επαναγραφή για τα $\mathcal{Q}, \mathcal{T}$. Δηλαδή, από τα $\mathcal{T}$, $\mathcal{R}_D$ συμπεραίνονται οι ίδιοι ισχυρισμοί, δηλαδή, για κάθε ισχυρισμό α, $\mathcal{T} \cup \mathcal{A} \models \alpha$ ανν $\mathcal{R}_D \cup \mathcal{A} \models \alpha$, το οποίο ισχύει ανν $MM(\mathcal{R}_D \cup \mathcal{A}) \models \alpha$, όπου $MM(\mathcal{R}_D \cup \mathcal{A})$ το ελάχιστο μοντέλο του $\mathcal{R}_D$ στο $\mathcal{A}$. Το ελάχιστο μοντέλο μπορεί να υπολογιστεί εξαντλητικά εφαρμόζοντας όλους τους κανόνες του $\mathcal{R}_D$ στο $\mathcal{A}$ και σε κάθε νέο ισχυρισμό που προκύπτει με εμπρόσθιο τρόπο (forward-chaining style) μέχρι να σταματήσουν να συμπεραίνονται νέοι ισχυρισμοί. Κατά τη διαδικασία αυτή, πριν προστεθεί το συμπέρασμα ενός κανόνα στο σύνολο των νέων ισχυρισμών, μπορεί να ελέγχεται αν οι ισχυρισμοί, πάνω στους οποίους εφαρμόζεται το σώμα του κανόνα, είναι συνεπείς με βάση το $\mathcal{T}$. Είναι προφανές ότι με τη διαδικασία αυτή υπολογίζεται το $clc^+(\mathcal{T}, \mathcal{A})$.

Επιπροσθέτως, από τον ορισμό της datalog επαναγραψιμότητας μίας ΠΛ είναι προφανές ότι ο υπολογισμός του $\mathcal{R}_D$ είναι ανεξάρτητος από τα δεδομένα, ενώ ο υπολογισμός του κορεσμού ενός datalog προγράμματος γίνεται σε πολυωνυμικό χρόνο ως προς τα δεδομένα. Τέλος, από την υπόθεση ισχύει ότι ο έλεγχος της συνέπειας των ισχυρισμών πάνω στους οποίους εφαρμόζονται οι κανόνες μπορεί να γίνει σε πολυωνυμικό χρόνο. Συνεπώς, συνολικά ο αλγόριθμος είναι πολυωνυμικός ως προς το $\mathcal{A}$. ■

Το Πόρισμα που ακολουθεί είναι άμεση συνέπεια του Θεωρήματος 56 και των αποτελεσμάτων που αφορούν στην datalog επαναγραψιμότητα βατών Horn-ΠΛ [30, 103, 63] καθώς και ιδιαίτερα εκφραστικών μη-Horn ΠΛ [67].

Πόρισμα 57. *Έστω ότι $\mathcal{DL}$ είναι μία από τις παρακάτω ΠΛ:*

- *DL-Lite$_\mathcal{R}$, $\mathcal{ELHI}^\neg$, ή Horn-$\mathcal{SHIQ}$; ή*
- markable-*$\mathcal{SHI}$ ή $\mathcal{ALCHI}$ που είναι απλή σε σχέση με τα διαζευκτικά κατηγορήματα.*

Τότε, για κάθε $\mathcal{DL}$-TBox $\mathcal{T}$ και ABox $\mathcal{A}$, το $clc^+(\mathcal{T}, \mathcal{A})$ μπορεί να υπολογιστεί σε πολυωνυμικό χρόνο ως προς το μέγεθος του $\mathcal{A}$.

Με βάση το clc^+ ορίζουμε την ICAR$^+$ σημασιολογία ως εξής:

Ορισμός 58. Έστω η DL βάση γνώσης $\mathcal{K} = \langle\mathcal{T}, \mathcal{A}\rangle$ και έστω το clc^+ όπως ορίζεται στον Ορισμό 54. Η *Τομή Διορθώσεων του Κλειστού$^+$ ABox* (*Intersection Closed$^+$ ABox Repair (ICAR$^+$)*) είναι ένα σύνολο $\mathcal{A}'$ ισχυρισμών τέτοιο ώστε:

- $\mathcal{A}' \subseteq clc^+(\mathcal{T}, \mathcal{A})$
- $Mod(\langle \mathcal{T}, \mathcal{A}' \rangle) \neq \emptyset$
- δεν υπάρχει $\mathcal{A}''$ τ.ω. $\mathcal{A}' \subsetneq \mathcal{A}'' \subseteq clc^+(\mathcal{T}, \mathcal{A})$ και $Mod(\langle \mathcal{T}, \mathcal{A}'' \rangle) \neq \emptyset$

Συμβολίζουμε με $\mathcal{A}_{c^+}(\mathcal{K})$ την τομή διορθώσεων του κλειστού ABox του $\mathcal{K}$.

Η παρακάτω πρόταση προκύπτει άμεσα από την Πρόταση 55.

Πρόταση 59. *Έστω η DL βάση γνώσης $\mathcal{K}$. Ισχύει ότι $\mathcal{A}_C(\mathcal{K}) \subseteq \mathcal{A}_{c^+}(\mathcal{K})$.*

Τέλος, αξίζει να σημειωθεί ότι το αποτέλεσμα του Αλγορίθμου 5 ενδεχομένως να μπορούσε να χρησιμοποιηθεί για τον υπολογισμό των ICAR$^+$-απαντήσεων, δηλαδή, τον άνωθεν προσεγγιστικό υπολογισμό των ICAR-απαντήσεων. Συγκεκριμένα, για κάποιο ΣΕ $\mathcal{Q}$ μπορεί να υπολογιστεί μία επαναγραφή $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ για τα $\mathcal{Q}, \mathcal{T}$ (εφόσον υπάρχει) και μετά να υπολογιστεί το σύνολο των απαντήσεων $\mathsf{cert}(\mathsf{ref}(\mathcal{R}_Q, \mathcal{T}), \mathcal{A}_s)$. Ωστόσο, θα πρέπει να οριστεί το ref για πιο εκφραστικές ΠΛ, αφού έχει οριστεί μόνο για την DL-Lite$_\mathcal{R}$. Η ύπαρξη, όμως, πιθανών αναδρομικών αξιωμάτων που εμφανίζονται στις λογικές αυτές επιτρέπει μόνο έναν προσεγγιστικό υπολογισμό του αντίστοιχου ref.

5.3 Υπολογισμός IAR-απαντήσεων στη DL-Lite$_\mathcal{R}$

Στην ενότητα αυτή θα δείξουμε ότι το πλαίσιο που προτείνουμε, δηλαδή η αξιοποίηση συστημάτων κορεσμού δεδομένων, μπορεί να εφαρμοστεί και για τον υπολογισμό των IAR-απαντήσεων. Η μέθοδος που προτείνουμε είναι η εξής: Έστω ένα ΣΕ $\mathcal{Q}$, μία οντολογία $\mathcal{T}$ και έστω $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T}$. Αρχικά, υπολογίζεται η διόρθωση του ABox με βάση την IAR σημασιολογία, στη συνέχεια, η οντολογία με το διορθωμένο ABox καταχωρούνται σε ένα σύστημα κορεσμού δεδομένων και, τέλος, εκτελείται το $\mathcal{R}_Q$ στο σύστημα αυτό. Από τα συμπεράσματα της Ενότητας 5.1 καθώς και τα συμπεράσματα των Rosati et al. [115] προκύπτει η ακόλουθη πρόταση.

Πρόταση 60. *Έστω $\mathcal{T}$ ένα $\mathcal{DL}$-TBox, $\mathcal{A}$ ένα ABox, $\mathcal{Q}$ ένα ΣΕ και έστω ότι το σύνολο $\mathcal{A}^{\mathsf{ia}}$ είναι η IAR-διόρθωση του $\mathcal{A}$ για το $\langle \mathcal{T}, \mathcal{A} \rangle$. Έστω επίσης ότι το* ans *είναι ένα $\mathcal{RL}$ ABox-κορεσμένο σύστημα, $\mathcal{C}$ ένα $\mathcal{T}$-συμπλήρωμα για το* ans *και έστω $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ μία επαναγραφή για τα $\mathcal{Q}, \mathcal{T}$. Τότε, ισχύει:* $\mathsf{cert}^{\mathsf{ia}}(\mathcal{Q}, \mathcal{T} \cup \mathcal{A}) = \mathsf{ans}(\mathcal{R}_Q, \mathcal{T} \cup \mathcal{C} \cup \mathcal{A}^{\mathsf{ia}})$.

Σύμφωνα με τους Stoilos et al. [129], για βάσεις γνώσης εκφρασμένες στις γλώσσες DL-Lite$_\mathcal{R}$ και $\mathcal{EL}$ πάντα υπάρχουν συμπληρώματα για $\mathcal{RL}$ ABox-συστήματα κορεσμού. Επίσης, ο Rosati [115] έδειξε ότι για την DL-Lite$_\mathcal{R}$ και την $\mathcal{EL}_{\bot nr}$ το μέγεθος όλων των ελάχιστων $\mathcal{T}$-συνεπή υποσυνόλων ενός ABox είναι πολυωνυμικά φραγμένο. Συνεπώς, με την προσέγγιση που προτείνεται στην ενότητα αυτή μπορούν να υπολογιστούν αποδοτικά οι IAR-απαντήσεις για τις γλώσσες αυτές αξιοποιώντας κάποιο $\mathcal{RL}$ σύστημα. Δυστυχώς,

Algorithm 6 ABoxIARRepair$(\mathcal{T}, \mathcal{A})$

Input: TBox $\mathcal{T}$, ABox $\mathcal{A}$
1: **Output:** IAR-διόρθωση του $\mathcal{T} \cup \mathcal{A}$
2: $\Phi := \emptyset$
3: **for all** $C \sqsubseteq \neg D \in \mathcal{T}$ **do**
4: Πρόσθεσε το ΣΕ $\pi(C) \wedge \pi(D)$ στο Φ
5: **for all** $\mathcal{Q}_\phi \in \Phi$ **do**
6:     Υπολόγισε μία UCQ επαναγραφή $\mathcal{R}_\phi$ για τα $\mathcal{Q}_\phi, \mathcal{T}$
7:     **for all** $\mathcal{Q}'_\phi \in \mathcal{R}_\phi$ και σ τέτοια ώστε $\mathcal{Q}'_\phi\sigma \subseteq \mathcal{A}$ **do**
8:         Αφαίρεσε το $\mathcal{Q}'_\phi\sigma$ από το $\mathcal{A}$
9: **return** $\mathcal{A}$

ακόμα και για αυτές τις γλώσσες ο υπολογισμός του $\mathcal{A}^{\mathsf{ia}}$ μπορεί να είναι αρκετά δύσκολος και χρονοβόρος αν το $\mathcal{A}$ είναι αρκετά μεγάλο.

Οι Rosati et al. [117] πρότειναν και υλοποίησαν ένα αλγόριθμο για τον υπολογισμό του συνόλου $\mathcal{A}^{\mathsf{ia}}$ για DL-Lite$_\mathcal{R}$ TBoxes. Ο αλγόριθμος αυτός αρχικά διατρέχει όλο το $\mathcal{A}$ "επισημειώνοντας" κάθε ισχυρισμό. Στη συνέχεια, διατρέχει για δεύτερη φορά το $\mathcal{A}$ και αν εντοπίσει κάποιο υποσύνολο $\mathcal{A}' \subseteq \mathcal{A}$ που να είναι ασυνεπές με το $\mathcal{T}$, ανανεώνει την "επισημείωση" των ισχυρισμών που ανήκουν στο $\mathcal{A}'$. Τέλος, διατρέχει για τρίτη φορά το $\mathcal{A}$ διαγράφοντας τους ισχυρισμούς των οποίων η επισημείωση έχει ανανεωθεί. Είναι φανερό, ότι σε ρεαλιστικά σενάρια, που το σώμα ισχυρισμών είναι πολύ μεγάλο η διαδικασία αυτή είναι αρκετά χρονοβόρα, όπως άλλωστε επαληθεύεται και από την αντίστοιχη πειραματική αξιολόγηση που εφαρμόσαμε στο Κεφάλαιο 6.

Από την τεχνική που προτείνουν οι Lembo et al. [77] για τον υπολογισμό του ref εξάγεται και ένας ακόμα αλγόριθμος υπολογισμού του $\mathcal{A}^{\mathsf{ia}}$ που βασίζεται στην επαναγραφή των ερωτημάτων και απεικονίζεται στον Αλγόριθμο 6. Η κεντρική ιδέα βασίζεται στον αλγόριθμο UnsatQuery [77], με τον οποίο εντοπίζονται οι ασυνεπείς ισχυρισμοί. Ο τυπικός ορισμός του UnsatQuery αναφέρεται στην Ενότητα 2.4.3. Συνοπτικά, αρχικά κάθε αξίωμα ξένων εννοιών ή ρόλων που εμφανίζεται στην οντολογία μετασχηματίζεται σε μορφή συζευκτικού ερωτήματος (γραμμές 2, 4), όπου το π αντιστοιχίζει μία έννοια A στο $A(x)$ και μία έννοια $\exists R$ στο $\exists y.R(x,y)$. Για παράδειγμα, αν η οντολογία περιέχει το αξίωμα $A \sqsubseteq \neg B$, τότε σχηματίζεται το ερώτημα $Q(x) \leftarrow A(x) \wedge B(x)$. Όλα τα ερωτήματα που έχουν σχηματιστεί αποθηκεύονται στο Φ και στη συνέχεια υπολογίζεται η UCQ επαναγραφή τους. Με τον τρόπο αυτό, εντοπίζονται όλα τα ζεύγη εννοιών και ρόλων για τα οποία συμπεραίνεται από την οντολογία ότι είναι ξένα μεταξύ τους. Επομένως, η αποτίμηση των αντίστοιχων ερωτημάτων επιστρέφει τους ισχυρισμούς της βάσης γνώσης που δεν είναι συνεπείς ως προς την οντολογία. Τέλος, αφαιρούνται οι ασυνεπείς ισχυρισμοί. Η διαδικασία επαναλαμβάνεται για κάθε ερώτημα που εμφανίζεται στο Φ.

Η παρακάτω πρόταση προκύπτει άμεσα από τον ορισμό της σημασιολογίας IAR.

Πρόταση 61. *Για ένα DL-Lite$_\mathcal{R}$ TBox $\mathcal{T}$ και ένα $\mathcal{A}$ ABox, ο αλγόριθμος* ABoxIARRepair$(\mathcal{T}, \mathcal{A})$

Algorithm 7 ABoxIARRepairOpt($\mathcal{T}, \mathcal{A}$)

Input: TBox $\mathcal{T}$, ABox $\mathcal{A}$
Output: IAR-repair of $\mathcal{T} \cup \mathcal{A}$
$\Phi := \emptyset, \mathcal{H} := \emptyset$
for all $C \sqsubseteq \neg D \in \mathcal{T}$ **do**
  Πρόσθεσε το ΣΕ $\pi(C) \wedge \pi(D)$ στο Φ
for all $\mathcal{Q}_\phi \in \Phi$ **do**
  Υπολόγισε την UCQ επαναγραφή $\mathcal{R}_\phi$ του $\mathcal{Q}_\phi, \mathcal{T}$
  $\Phi = \Phi \cup \mathcal{R}_\phi$
for all άτομο $C(\vec{x})$ που εμφανίζεται στο Φ και σ τ.ω. $\pi(C)\sigma \in \mathcal{A}$
  $\mathcal{H} := \mathcal{H} \cup \langle \mathsf{nonFresh}(\sigma), \bigcup_\sigma \{\pi(C)\sigma\} \rangle$
for all $\pi(C) \wedge \pi(D) \in \Phi$ **do**
  for all σ_1, σ_2 τ.ω. $\langle \vec{a}, \pi(C)\sigma_1 \rangle, \langle \vec{a}, \pi(D)\sigma_2 \rangle \in \mathcal{H}$
    Αφαίρεσε το $\{\pi(C)\sigma_1, \pi(D)\sigma_2\}$ από το $\mathcal{A}$
return $\mathcal{A}$

υπολογίζει την IAR-διόρθωση των $\mathcal{T} \cup \mathcal{A}$.

Παρά το γεγονός ότι συνήθως τα αρνητικά αξιώματα υπαγωγής είναι λίγα και οι επαναγραφές τους επίσης δεν είναι τόσο μεγάλες, η αποτίμηση κάθε ζεύγους ατόμων απαιτεί πρακτικά αρκετό χρόνο, επομένως ο Αλγόριθμος 6 δεν είναι αρκετά αποδοτικός. Ο αλγόριθμος μπορεί να βελτιστοποιηθεί σημαντικά αν παρατηρήσουμε ότι τα άτομα που εμφανίζονται στα ερωτήματα μιας επαναγραφής επαναλαμβάνονται. Αντί να υπολογίζονται όλες οι αντιστοιχίσεις σ τέτοιες ώστε $\mathcal{Q}'_\phi\sigma \subseteq \mathcal{A}$, δηλαδή, όλες οι απαντήσεις του $\mathcal{Q}'_\phi$ στο $\mathcal{A}$, μπορούν να αποτιμηθούν όλα τα άτομα $A(x)$, $R(x, y)$ που συμμετέχουν στην επαναγραφή και να αποθηκευτούν οι απαντήσεις καθώς και οι αντίστοιχοι ισχυρισμοί. Στη συνέχεια, για να υπολογιστούν οι απαντήσεις των $\mathcal{Q}'_\phi$ στο $\mathcal{A}$ αρκεί να υπολογιστεί η τομή των απαντήσεων που αντιστοιχούν στους ισχυρισμούς που εμφανίζονται στο $\mathcal{Q}'_\phi$. Αν η τομή δεν είναι κενή, αφαιρούνται απευθείας οι αντίστοιχοι ισχυρισμοί.

Η βελτιστοποιημένη έκδοση του Αλγορίθμου 6 παρουσιάζεται στον Αλγόριθμο 7. Αρχικά, ο αλγόριθμος ABoxIARRepairOpt κατασκευάζει το Φ, όπως και στον Αλγόριθμο 6, και στη συνέχεια, υπολογίζει την επαναγραφή κάθε ερωτήματος που ανήκει στο Φ αποθηκεύοντάς την επίσης στο Φ. Στη συνέχεια, αποτιμάει κάθε άτομο που εμφανίζεται στο Φ ανάλογα με το αν έχει ελεύθερες μεταβλητές ως προς το Φ. Αν το άτομο αυτό δεν έχει ελεύθερες μεταβλητές στο Φ, τότε αποτιμάται στο $\mathcal{A}$ και αποθηκεύεται το ζεύγος $\langle$στιγμιότυπο, ισχυρισμός$\rangle$. Αν το άτομο έχει ελεύθερες μεταβλητές στο Φ, δηλαδή είναι ρόλος, τότε αποθηκεύεται το ζεύγος $\langle$στιγμιότυπο δεσμευμένης μεταβλητής, σύνολο ισχυρισμών$\rangle$, όπου στο σύνολο ισχυρισμών περιέχονται όλοι οι ισχυρισμοί του ρόλου αυτού που στη δεσμευμένη μεταβλητή έχουν το συγκεκριμένο στιγμιότυπο. Τέλος, για κάθε ερώτημα $\mathcal{Q}$ που ανήκει στο Φ, και κάθε κοινό στιγμιότυπο που έχει το ζεύγος ατόμων του $\mathcal{Q}$, αφαιρείται το αντίστοιχο ζεύγος ισχυρισμών.

Παράδειγμα 62. Έστω το TBox $\mathcal{T}$ και το ABox $\mathcal{A}$:

$$\begin{aligned} \mathcal{T} &= \{\exists P \sqsubseteq B,\ B \sqsubseteq \neg A\}, \\ \mathcal{A} &= \{P(a,b), P(a,c), P(c,d), B(b), A(b), A(a)\} \end{aligned}$$

Ο Αλγόριθμος 7 αρχικά θα φτιάξει από το αξίωμα $B \sqsubseteq \neg A$ το ερώτημα $\mathcal{Q} = Q(x) \leftarrow A(x) \land B(x)$, το οποίο θα προσθέσει στο Φ. Στη συνέχεια, θα υπολογίσει την επαναγραφή του $\mathcal{Q}$ με βάση το $\mathcal{T}$, δηλαδή, θα προσθέσει το ερώτημα $Q(x) \leftarrow A(x) \land P(x,y)$ στο Φ. Στο επόμενο βήμα θα αποτιμήσει τα άτομα $\{A(x), B(x), P(x,y)\}$. Μετά την αποτίμηση των $A(x)$, $B(x)$ θα αποθηκεύσει στο $\mathcal{H}$ τα ζεύγη $\langle a, \{A(a)\}\rangle$, $\langle b, \{A(b)\}\rangle$, $\langle b, \{B(b)\}\rangle$. Επίσης, επειδή το $P(x,y)$ έχει την μεταβλητή y που είναι ελεύθερη στο Φ, θα αποθηκευτούν στο $\mathcal{H}$ τα εξής ζεύγη: $\langle a, \{P(a,b), P(a,c)\}\rangle$, $\langle c, \{P(c,d)\}\rangle$. Συνεπώς, τώρα από το ερώτημα $Q(x) \leftarrow A(x) \land B(x)$ και τα ζεύγη $\langle b, \{A(b)\}\rangle$, $\langle b, \{B(b)\}\rangle$, προκύπτει ότι πρέπει να αφαιρεθούν οι ισχυρισμοί $A(b)$, $B(b)$ και από το ερώτημα $Q(x) \leftarrow A(x) \land P(x,y)$ και τα ζεύγη $\langle a, \{P(a,b), P(a,c)\}\rangle$, $\langle a, \{A(a)\}\rangle$ προκύπτει ότι πρέπει να αφαιρεθούν οι ισχυρισμοί $P(a,b)$, $P(a,c)$, $A(a)$. Οπότε τελικά $\mathcal{A}_I = \{P(c,d)\}$. $\diamondsuit$

Όπως φαίνεται και στην πειραματική αξιολόγηση που ακολουθεί στο επόμενο κεφάλαιο, η μέθοδος αυτή είναι ιδιαίτερα αποδοτική.

Κεφάλαιο 6

Πειραματική Αξιολόγηση

Στο κεφάλαιο αυτό, παρουσιάζεται η πειραματική αξιολόγηση των εργαλείων που υλοποιήθηκαν για την απάντηση ερωτημάτων σε ασυνεπείς βάσεις γνώσης με βάση την ICAR και την IAR σημασιολογία. Οι υλοποιήσεις βασίζονται στους αλγορίθμους που παρουσιάστηκαν στο Κεφάλαιο 5 και ονομάζονται HdIC, HdIA, αντίστοιχα. Για την ICAR σημασιολογία υλοποιήθηκε επίσης η βελτιστοποιημένη μορφή του αλγορίθμου, όπως αυτή περιγράφεται στην Ενότητα 5.1 και το αντίστοιχο σύστημα ονομάζεται $\mathsf{HdIC}_{\mathsf{op}}$. Δεν ήταν εφικτή η υλοποίηση του Αλγορίθμου 5, αφού κάτι τέτοιο απαιτεί την παρέμβαση στον κώδικα ενός ABox-συστήματος κορεσμού ή την κατασκευή ενός τέτοιου συστήματος από την αρχή. Τα συστήματα που κατασκευάσαμε χρησιμοποιούν το $\mathcal{RL}$-σύστημα κορεσμού δεδομένων GraphDB [69], το Hydrowl [128] για τον υπολογισμό των συμπληρωμάτων και το Rapid [131] για τον υπολογισμό των επαναγραφών.

Μετά από μία σύντομη προεπισκόπηση των πειραματικών δεδομένων που χρησιμοποιήθηκαν, παρουσιάζονται τα πειραματικά αποτελέσματα των συστημάτων HdIC, $\mathsf{HdIC}_{\mathsf{op}}$ και HdIA. Επίσης, συγκρίνονται όλα μας τα συστήματα με το CQApri [18] και το QuID [117] που υπολογίζουν τις IAR-απαντήσεις. Τέλος, καθώς το εργαλείο των Masotti et al. [84] για τον υπολογισμό των ICAR απαντήσεων δεν είναι διαθέσιμο, δεν συγκρίναμε την αποδόση των συστημάτων μας με το εργαλείο αυτό.

Όλα τα πειράματα πραγματοποιήθηκαν σε Windows 7 υπολογιστή με επεξεργαστή στα 3.20 GHz με 8GB μνήμης RAM.

6.1 Πειραματικά Δεδομένα

Για την αξιολόγηση των συστημάτων χρησιμοποιήσαμε τα ίδια πειραματικά δεδομένα που χρησιμοποιήθηκαν για την αξιολόγηση του CQApri [18]: μία DL-Lite έκδοση της οντολογίας $\text{LUBM}^{\exists}_{20}$ [81] στην οποία έχουν προστεθεί αρνητικά γενικευμένα αξιώματα, ένα σύνολο ερωτημάτων και ένα σύνολο από ABoxes ασυνεπή ως προς την οντολογία. Η οντολογία περιέχει συνολικά 1081 αξιώματα, εκ των οποίων τα 875 είναι αρνητικά αξιώματα υπαγωγής.

Για τα ABoxes και τα ερωτήματα χρησιμοποιούμε τα ίδια ονόματα με τους Bienvenu et

Πίνακας 6.1: *Αριθμός ατόμων και μεταβλητών των ερωτημάτων.*

Ερώτημα	#ατόμων	#μεταβλητών
req2	3	2
req3	6	3
g2	1	1
g3	1	1
q1	2	2
q2	2	2
q4	7	6
Lutz1	8	4
Lutz5	5	3

Πίνακας 6.2: *Μεγέθη των ABoxes, ποσοστό ασυνεπών ισχυρισμών, μέσος αριθμός ισχυρισμών ανά ασυνέπεια.*

ABox	# ABox	% ασυνεπείς ισχ.	μέσος αρ. ασυνεπών ισχ.
$\mathcal{A}^{1}_{15e\text{-}4}$	75.708	2,05	0,04
$\mathcal{A}^{1}_{5e\text{-}2}$	76.959	30,97	1,03
$\mathcal{A}^{1}_{2e\text{-}1}$	80.454	57,99	3,96
$\mathcal{A}^{5}_{15e\text{-}4}$	499.674	1,70	0,03
$\mathcal{A}^{5}_{5e\text{-}2}$	507.713	33,12	1,21
$\mathcal{A}^{5}_{2e\text{-}1}$	531.607	58,29	4,29
$\mathcal{A}^{10}_{15e\text{-}4}$	930.729	2,37	0,05
$\mathcal{A}^{10}_{5e\text{-}2}$	945.450	33,92	1,31
$\mathcal{A}^{10}_{2e\text{-}1}$	988.882	58,89	4,86
$\mathcal{A}^{20}_{15e\text{-}4}$	1.982.922	2,64	0,05
$\mathcal{A}^{20}_{5e\text{-}2}$	2.014.129	33,91	1,6
$\mathcal{A}^{20}_{2e\text{-}1}$	2.103.366	58,78	5,49

al. Για παράδειγμα, με $\mathcal{A}^{m}_{Xe-Y}$ συμβολίζουμε ένα ABox που περιέχει δεδομένα για m πανεπιστήμια και οι προστιθέμενες ασυνέπειες είναι με πιθανότητα X. Στους Πίνακες 6.1, 6.2 παρουσιάζονται τα στατιστικά στοιχεία που αφορούν στη δομή των ερωτημάτων και των ABoxes, αντίστοιχα. Τα ερωτήματα που χρησιμοποιήθηκαν είναι από ένα πλήθος πειραμάτων που αναφέρεται στη βιβλιογραφία και βασίζεται στην LUBM-οντολογία: τα ερωτήματα req2, req3 είναι από τα πειράματα των Pérez-Urbina et al. [100], τα ερωτήματα g2, g3, q1, q2, q4 κατασκευάστηκαν για την αξιολόγηση του CQApri και τα ερωτήματα Lutz1, Lutz5 κατασκευάστηκαν για τα πειράματα των Lutz et al. [81]. Όλα τα ερωτήματα παρατίθενται στο Παράρτημα Α′.2.

6.2 Γενικά Αποτελέσματα

Ο υπολογισμός του $\text{LUBM}^{\exists}_{20}$-συμπληρώματος για το GraphDB έγινε σε χιλιοστά του δευτερολέπτου. Ο υπολογισμός αυτός γίνεται μόνο μία φορά αφού εξαρτάται μόνο από την οντολογία και το σύστημα κορεσμού. Επίσης, χρειάστηκε λιγότερο από ένα δευτερόλεπτο να διαπιστωθεί ότι η συγκεκριμένη οντολογία δεν περιέχει μη-ικανοποιήσιμες έννοιες/ρόλους .

Στο κεφάλαιο αυτό παρουσιάζονται τα χρονικά αποτελέσματα που αφορούν στα σώματα ισχυρισμών που περιέχουν δεδομένα για 20 πανεπιστήμια, ενώ τα χρονικά αποτελέσματα που αφορούν στα υπόλοιπα σώματα ισχυρισμών παρουσιάζονται στο Παράρτημα Β′. Συγκεκριμένα, στον Πίνακα 6.3 παρουσιάζονται τα αποτελέσματα για τα ABoxes $\mathcal{A}^{20}_{15e\text{-}4}$, $\mathcal{A}^{20}_{5e\text{-}2}$, $\mathcal{A}^{20}_{2e\text{-}1}$ για όλα τα συστήματα. Στη στήλη Load εμφανίζονται οι χρόνοι φόρτωσης της οντολογίας, του συμπληρώματός της και του ABox στο GraphDB καθώς και ο χρόνος υπολογισμού του κορεσμού του ABox με βάση την οντολογία. Για τα συστήματα υπολογισμού IAR-απαντήσεων διαχωρίζεται ο χρόνος προεπεξεργασίας από τον χρόνο αποτίμησης. Για τα συστήματα HdIA, QuID ο χρόνος προεπεξεργασίας περιλαμβάνει τον χρόνο που απαιτείται για τον υπολογισμό της IAR-διόρθωσης ενώ για το CQApri τον χρόνο που απαιτείται για να εντοπιστούν όλα τα σύνολα των ισχυρισμών που δημιουργούν ασυνέπεια.

Μία γενική παρατήρηση που προκύπτει από τα χρονικά αποτελέσματα είναι ότι όσο αυξάνεται ο όγκος των δεδομένων αυξάνεται και ο χρόνος που χρειάζονται όλα τα συστήματα για να ανταποκριθούν. Επίσης είναι ενδιαφέρον να παρατηρήσουμε ότι επιβεβαιώνεται και πειραματικά το Λήμμα 18, δηλαδή ο αριθμός των ICAR-απαντήσεων είναι μεγαλύτερος ή ίσος από τον αριθμό των IAR-απαντήσεων. Συγκεκριμένα, παρατηρείται από τους πίνακες ότι σε πολλά ερωτήματα οι ICAR-απαντήσεις που υπολογίζουν τα HdIC, HdIC_{op} είναι πολύ περισσότερες από τις αντίστοιχες IAR-απαντήσεις, όπως αυτές υπολογίζονται από το CQApri και το HdIA (που μεταξύ τους υπολογίζουν τον ίδιο αριθμό απαντήσεων). Αξίζει να σημειωθεί ότι το QuID υπολόγισε διαφορετικό αριθμό IAR-απαντήσεων, κάτι που ενδεχομένως να υπονοεί ότι υπάρχει κάποιο λάθος στην υλοποίησή του.

6.3 Αξιολόγηση των ICAR-αλγορίθμων

Μία πρώτη παρατήρηση είναι ότι ενώ το HdIC υπολογίζει ICAR-απαντήσεις, δηλαδή επιστρέφει περισσότερα αποτελέσματα, είναι, σχεδόν για όλα τα ερωτήματα, σημαντικά πιο γρήγορο σε σχέση με το QuID και το CQApri. Συγκεκριμένα, σχεδόν σε όλες τις περιπτώσεις ισχύει ότι $t(\text{load}) + t(\text{HdIC}) > t(\text{Προεπεξ.}(\text{CQA})) + t(\text{Αποτίμηση}(\text{CQA})) > t(\text{Διόρθωση.}(\text{QuID})) + t(\text{Αποτίμηση}(\text{QuID}))$. Σε κάθε περίπτωση, το HdIC είναι σημαντικά πιο γρήγορο από το QuID ενώ υπάρχει μία μόνο αξιοσημείωτη περίπτωση που το CQApri είναι πιο γρήγορο από το HdIC. Η διαφορά αυτή εμφανίζεται στο ερώτημα q_4, όμως αυτό μπορεί να δικαιολογηθεί από το γεγονός ότι ο αριθμός των ICAR-απαντήσεων είναι ση-

Πίνακας 6.3: *Αποτελέσματα των συστημάτων* HdIC, $\mathsf{HdIC_{op}}$ *και CQApri* (CQA) *για το* $\mathcal{A}_i^{20}$.

		Αριθμός Απαντήσεων			Χρόνοι Φόρτωσης, Προεπεξεργασίας και Αποτίμησης (σε sec)								
					ICAR Απάντηση Ερωτ.			*IAR Απάντηση Ερωτ.*					
								Διόρθωση/Προεπεξ.			Αποτίμηση		
$\mathcal{A}$	$\mathcal{Q}$	HdIC	HdIA	QuID	Load	HdIC	$\mathsf{HdIC_{op}}$	HdIA	QuID	CQA	HdIA	QuID	CQA
	q_1	544725	535341	539683		120.1	13.8				0.4	38.5	31.2
	q_2	189527	189209	189881		55.0	4.6				0.1	30.1	17.3
	q_4	2033569	1355978	1444799		2701.9	286.6				1.8	31.8	49.8
	g_2	7393	7159	7222		0.9	0.1				<0.1	2.5	2.6
$\mathcal{A}^{20}_{15e\text{-}4}$	g_3	28998	28889	28944	59.7	2.4	0.2	114.2	3184.4	124.6	<0.1	1.7	1.4
	req_2	40428	40428	40755		16.1	1.6				0.2	0.1	0.8
	req_3	4851	4823	4874		4.8	0.9				0.4	7.0	4.5
	lutz_1	414600	414600	418882		514.1	60.3				2.5	892.5	136.2
	lutz_5	81996	81996	82323		84.1	6.7				0.6	172.8	118.4
	q_1	532587	309625	396146		146.9	15.1				0.3	40.6	44.1
	q_2	172121	156361	182504		56.7	4.5				0.1	35.0	25.5
	q_4	1126433	0	0		1920.9	178.4				0.3	1.3	146.9
	g_2	9005	5727	6397		1.2	0.1				<0.1	5.0	5.6
$\mathcal{A}^{20}_{5e\text{-}2}$	g_3	28460	24976	27218	61.2	2.5	0.2	114.0	5246.1	136.9	<0.1	4.5	2.1
	req_2	24183	24183	30635		14.7	1.4				0.1	0.1	1.3
	req_3	2569	2125	2815		4.9	0.8				0.2	5.3	9.2
	lutz_1	15539	15539	21480		336.8	33.4				1.0	202.9	238.7
	lutz_5	53430	53420	60178		82.9	6.1				0.4	149.8	123.6
	q_1	497164	73611	-		149.1	16.7				0.3	-	43.0
	q_2	131228	77475	-		42.1	3.6				<0.1	-	23.9
	q_4	174226	0	-		418.5	38.8				<0.1	-	oom
	g_2	15427	4803	-		2.1	0.1				<0.1	-	3.4
$\mathcal{A}^{20}_{2e\text{-}1}$	g_3	27393	14845	-	66.4	6.9	0.3	122.1	t/o	205.7	<0.1	-	2.7
	req_2	6336	6336	-		10.3	1.2				0.1	-	1.7
	req_3	407	184	-		3.2	0.7				0.1	-	22.1
	lutz_1	0	0	-		175.5	19.7				0.4	-	266.3
	lutz_5	14671	14355	-		58.7	4.6				0.2	-	134.0

μαντικά μεγαλύτερος από τις IAR-απαντήσεις. Επίσης, η διαφορά μειώθηκε σημαντικά με την βελτιστοποιήση που παρουσιάζεται στην Ενότητα 5.1.1, στην οποία βασίζεται η $\mathsf{HdIC_{op}}$. Γενικότερα, η απόδοση του $\mathsf{HdIC_{op}}$ είναι κατά μία τάξη μεγέθους καλύτερη από την απόδοση του HdIC. Αυτό συμβαίνει γιατί, στις περισσότερες περιπτώσεις η $\mathsf{ref_{min}}$, στην οποία βασίζεται η $\mathsf{HdIC_{op}}$, υπολογίζει ερωτήματα των οποίων ο αριθμός των αρνητικών ατόμων έχει επίσης μειωθεί κατά μία τάξη μεγέθους. Στον Πίνακα 6.4 παρουσιάζεται ο αριθμός των αρνητικών ατόμων κάθε ερωτήματος όπως υπολογίζεται από την HdIC και $\mathsf{HdIC_{op}}$ αντίστοιχα.

Τέλος, αξίζει να σημειωθεί ότι για το ABox $\mathcal{A}^{20}_{2e\text{-}1}$ στο CQApri εξαντλήθηκε η μνήμη πριν επιστραφεί κάποιο αποτέλεσμα, ενώ ο χρόνος υπολογισμού της δίορθωσης του από το QuID ξεπέρασε το χρονικό όριο της 1 ώρας 30 λεπτών που είχαμε θέσει.

6.4 Αξιολόγηση του IAR-αλγορίθμου

Όπως φαίνεται από τους ίδιους πίνακες το σύστημα HdIA είναι σημαντικά πιο αποδοτικό από τα συστήματα QuID και CQApri. Καταρχήν, η απόδοση του συστήματός μας για τον υπολογισμό της IAR-διόρθωσης ήταν κατά αρκετές τάξεις μεγέθους καλύτερη από

Πίνακας 6.4: *Αριθμός αρνητικών ατόμων όπως υπολογίζονται από τα* HdIC, HdIC_{op}

Ερώτημα	HdIC	HdIC_{op}
q_1	353	32
q_2	1001	33
q_4	2001	223
g_2	148	4
g_3	116	5
req_2	649	44
req_3	1308	113
$lutz_1$	1010	270
$lutz_5$	1155	74

Πίνακας 6.5: *Αριθμός αρνητικών ατόμων όπως υπολογίζονται από την* HdIC_{dp}

Ερώτημα	HdIC_{dp}
q_1	11
q_2	13
q_4	78
g_2	4
g_3	5
req_2	24
req_3	51
$lutz_1$	104
$lutz_5$	34

την αντίστοιχη απόδοση του QuID, ενώ ήταν σημαντικά καλύτερη από την απόδοση του CQApri για το βήμα της προεπεξεργασίας. Αξίζει να σημειωθεί ότι στην ουσία το CQApri κατά την προεπεξεργασία πραγματοποιεί την ίδια εργασία με τα HdIA και QuID, δηλαδή εντοπίζει όλα τα υποσύνολα της βάσης δεδομένων που είναι ασυνεπή ως προς την οντολογία, με τη μόνη διαφορά ότι δεν τα αφαιρεί από τη βάση αλλά τα αφαιρεί πριν επιστρέψει τις IAR-απαντήσεις. Σε αντίθεση λοιπόν με την περίπτωση της ICAR-σημασιολογίας, το σύστημά μας έχει σε όλες τις περιπτώσεις καλύτερη απόδοση. Επιπλέον, λόγω της Πρότασης 60, μετά από τον βήμα υπολογισμού της IAR-διόρθωσης, το σύστημά μας υπολογίζει τις IAR-απαντήσεις με τον συνήθη τρόπο αποτίμησης ερωτημάτων με τη χρήση του GraphDB. Για κάθε ερώτημα και ABox, η διαδικασία αυτή πραγματοποιείται σε μερικά χιλιοστά του δευτερολέπτου ενώ για την ίδια διαδικασία το QuID και το CQApri, που υπολογίζουν την επαναγραφή του ερωτήματος με βάση την οντολογία και στην συνέχεια την αποτιμούν στη βάση δεδομένων με τη χρήση της PostgreSQL, χρειάζονται μερικά δευτερόλεπτα.

6.5 Βελτιστοποίηση

Παρά το γεγονός ότι η ref_{min} μειώνει σημαντικά τον αριθμό των αρνητικών ατόμων, παρατηρούμε από τον Πίνακα 6.4 ότι υπάρχουν περιπτώσεις, όπως για παράδειγμα το ερώτημα q_4 που αριθμός αυτός παραμένει μεγάλος, επηρεάζοντας αρνητικά τη συνολική απόδοση του συστήματος. Από περαιτέρω ανάλυση του ερωτήματος αυτού, διαπιστώσαμε ότι το μεγαλύτερο μέρος των αρνητικών ατόμων που περιέχει οφείλεται στα αρνητικά αξιώματα υπαγωγής μεταξύ ρόλων. Ωστόσο, πρακτικά είναι κάπως απίθανο να παραβιαστούν τα αξιώματα αυτά από ισχυρισμούς. Για τον λόγο αυτόν προσθέσαμε ένα ακόμα βήμα προεπεξεργασίας κατά το οποίο εντοπίζονται ποια από τα αξιώματα υπαγωγής μεταξύ ρόλων παραβιάζονται και λαμβάνουμε υπόψη μόνο αυτά. Δηλαδή, στην περίπτωση των εργαλείων HdIC, HdIC_{op}, απο τους αρνητικούς ρόλους που θεωρητικά πρέπει να προ-

Πίνακας 6.6: *Αποτελέσματα των συστημάτων* HdIC$_{dp}$, HdIA$_{dp}$ *για τα ABox* $\mathcal{A}_i^{20}$.

$\mathcal{A}$	$\mathcal{Q}$	Preproc.	HdIC$_{dp}$	HdIA$_{dp}$
$\mathcal{A}_{15e\text{-}4}^{20}$	q$_1$	11,5	4,1	92,2
	q$_2$		1,5	
	q$_4$		104,8	
	g$_2$		<0,1	
	g$_3$		0,2	
	req$_2$		0,7	
	req$_3$		0,6	
	lutz$_1$		19,2	
	lutz$_5$		2,8	
$\mathcal{A}_{5e\text{-}2}^{20}$	q$_1$	11,7	4,8	98,2
	q$_2$		1,6	
	q$_4$		79,6	
	g$_2$		0,1	
	g$_3$		0,2	
	req$_2$		1,2	
	req$_3$		0,6	
	lutz$_1$		13,3	
	lutz$_5$		2,5	
$\mathcal{A}_{2e\text{-}1}^{20}$	q$_1$	12,9	5,5	99,3
	q$_2$		1,5	
	q$_4$		19,3	
	g$_2$		0,1	
	g$_3$		0,2	
	req$_2$		0,6	
	req$_3$		0,5	
	lutz$_1$		9,3	
	lutz$_5$		1,9	

στεθούν κατά τον σχηματισμό του τελικού ερωτήματος, προσθέτουμε μόνο εκείνους που συμμετέχουν στα αξιώματα που έχουμε εντοπίσει κατά το βήμα προεπεξεργασίας. Την διαδικασία αυτή την ονομάζουμε dp. Όπως φαίνεται από τον Πίνακα 6.5, με τον τρόπο αυτό, ο αριθμός αρνητικών ατόμων που εμφανίζεται στα ερωτήματα μειώθηκε στο μισό. Στην περίπτωση του HdIA, από τα αρνητικά αξιώματα υπαγωγής μεταξύ ρόλων που εμφανίζονται στην οντολογία λαμβάνουμε υπόψη μόνο τα αξιώματα που έχουμε εντοπίσει και φτιάχνουμε μόνο τα αντίστοιχα ερωτήματα. Ονομάζουμε τα συστήματα HdIC$_{op}$, HdIA που αξιοποιούν την dp, HdIC$_{dp}$, HdIA$_{dp}$.

Στον Πίνακα 6.6 παρουσιάζονται τα χρονικά αποτελέσματα των συστημάτων HdIC$_{dp}$, HdIA$_{dp}$ για τις περιπτώσεις που το σώμα των ισχυρισμών περιέχει δεδομένα για 20 πανεπιστήμια. Τα χρονικά αποτελέσματα που αντιστοιχούν στις υπόλοιπες περιπτώσεις βρίσκονται στο Παράρτημα Β′. Στη στήλη "Preproc." παρουσιάζεται ο χρόνος που χρειάζεται για το βήμα προεπεξεργασίας της dp, στη στήλη HdIC$_{dp}$ ο χρόνος για να υπολογιστούν οι ICAR-απαντήσεις και στη στήλη HdIA$_{dp}$ ο χρόνος για τον υπολογισμό της IAR-διόρθωσης. Όπως φαίνεται από τους αντίστοιχους πίνακες η βελτιστοποίηση αυτή επηρέασε σημαντικά την απόδοση των συστημάτων και σε μερικές περιπτώσεις τη διπλασίασε. Τέλος, για τις λίγες περιπτώσεις που στο ερώτημα q_4 το CQApri ήταν σημαντικά πιο γρήγορο από το $HdIC_{dp}$ είτε η διαφορά τους εκμηδενίστηκε είτε ελαχιστοποιήθηκε στα 1-2 δευτερόλεπτα (όταν ο συνολικός χρόνος υπολογισμού απαντήσεων φτάνει τα 3 λεπτά).

Μέρος IV

Επίλογος

Κεφάλαιο 7

Σχετικές Εργασίες

Στην ενότητα αυτή θα παρουσιάσουμε συνοπτικά τις εργασίες που έχουν σημειωθεί στη βιβλιογραφία και σχετίζονται με το περιεχόμενο της διατριβής αυτής. Θα μας απασχολήσουν κυρίως οι εργασίες που σχετίζονται με την εξέλιξη οντολογιών, την επαναγραφή ερωτημάτων και την συνεπή απάντηση ερωτημάτων.

7.1 Απάντηση Ερωτημάτων και Εξέλιξη Οντολογιών

Τα τελευταία χρόνια έχει σημειωθεί ένας σημαντικός όγκος εργασιών που σχετίζονται με την εξέλιξη των οντολογιών, τόσο θεωρητικό όσο και πρακτικό επίπεδο (μερικές ενδεικτικές εργασίες είναι οι εξής: [68, 53, 127, 22, 36, 32, 37, 45, 111]). Ωστόσο, ως προς την απάντηση ερωτημάτων, εξ όσων γνωρίζουμε, στο πεδίο των οντολογιών υπάρχουν μόνο οι εργασίες των Kondylakis et al. [70, 71], οι οποίοι εστιάζουν στο τμήμα που αφορά στις συσχετίσεις των οντολογιών με τις βάσεις δεδομένων. Ο στόχος της εργασίας τους είναι να μην επαναπροσδιορίζονται οι συσχετίσεις κάθε φορά που αλλάζει η οντολογία. Αυτό επιτυγχάνεται υπολογίζοντας την επαναγραφή του ερωτήματος με βάση κάθε νέα οντολογία, η οποία εκτελείται στο σύστημα ολοκλήρωσης δεδομένων (data integration system). Θεωρούμε ότι η εργασία των Kondylakis et al. [71] μπορεί να βελτιστοποιηθεί περαιτέρω αν αξιοποιηθούν οι αλγόριθμοι που προτείνουμε στο Κεφάλαιο 3 ώστε να μην υπολογίζονται οι επαναγραφές μεταξύ των διαφορετικών εκδόσεων της οντολογίας κάθε φορά απο την αρχή. Στο πεδίο των βάσεων δεδομένων, μία σχετική εργασία είναι των Curino et al. [38], στην οποία όμως τα προβλήματα που αντιμετωπίζονται είναι θεμελιωδώς διαφορετικά. Συγκεκριμένα, ο στόχος της εργασίας είναι η μελέτη της ομαλής και ασφαλούς συγχώνευσης των δεδομένων στο νέο σχήμα και, επομένως, ο μετασχηματισμός των επαναγραφών ώστε να συνεχίσει να υπάρχει αντιστοιχία με αυτό. Όμως, η δομή των βάσεων δεδομένων είναι αρκετά διαφορετική από τις οντολογίες, διότι, σε αντίθεση με τις βάσεις δεδομένων που είναι πλήρεις ως προς τα δεδομένα, τα αξιώματα στις οντολογίες χρησιμοποιούνται για να παραχθεί νέα πληροφορία.

Από την άλλη πλευρά, στην ερευνητική περιοχή της επαναγραφής ερωτημάτων με βάση σταθερές οντολογίες έχει προταθεί ένα ευρύ σύνολο συστημάτων επαναγραφής. Ο πρώτος αλγόριθμος επαναγραφής, με το όνομα PerfectRef, προτάθηκε από τους Calvanese et al. [31] για την οικογένεια γλωσσών της DL-Lite. Το σύστημα QuOnto [2] αποτελεί υλοποίηση του PerfectRef. Ο αλγόριθμος επαναγράφει το τιθέμενο ερώτημα με βάση τα αξιώματα της οντολογίας σε ένα σύνολο ερωτημάτων, εφαρμόζοντας ένα σύνολο κανόνων μετασχηματισμού. Ο στόχος είναι να υποβληθεί η επαναγραφή σε μία βάση δεδομένων. Με τον τρόπο αυτό, το πρόβλημα απάντησης ερωτημάτων σε μία βάση γνώσης ανάγεται στο πρόβλημα απάντησης ερωτημάτων σε ένα στιγμιότυπο μίας βάσης δεδομένων, χωρίς πλέον να είναι απαραίτητη η ύπαρξη της οντολογίας. Η προσέγγιση αυτή, έδωσε το έναυσμα για ένα καινούριο πεδίο έρευνας στην απάντηση ερωτημάτων.

Στη συνέχεια, οι Pérez-Urbina et al. παρουσίασαν τον πρώτο αλγόριθμο επαναγραφής βασισμένο στη μέθοδο της ανάλυσης για την DL-Lite [102]. Βασικό κίνητρό τους ήταν να ελαχιστοποιήσουν το μέγεθος της επαναγραφής ώστε η εκτέλεσή της στη βάση δεδομένων να πραγματοποιείται σε εύλογο χρονικό διάστημα. Στο πλαίσιο αυτό, εισήγαγαν κριτήρια περιττότητας μειώνοντας έτσι και άλλο το μέγεθος της τελικής επαναγραφής. Από την υλοποίηση του αλγορίθμου δημιουργήθηκε το σύστημα Requiem, το οποίο όπως έδειξε πειραματική αξιολόγηση, είναι πιο αποδοτικό από το QuOnto. Στη συνέχεια, οι Pérez-Urbina et al. εξέλιξαν τον αλγόριθμο αυτό και για τις γλώσσες DL-Lite$^+$ [101] και $\mathcal{ELHIO}^\neg$ [103]. Επίσης, στις ίδιες εργασίες έδειξαν ότι στη γλώσσα DL-Lite$_\mathcal{R}$ η επαναγραφή στη χειρότερη περίπτωση είναι ένα ΣΣΕ, στη DL-Lite$^+$ η επαναγραφή είναι στη χειρότερη περίπτωση ένα ΣΣΕ και ένα γραμμικό πρόγραμμα Datalog και τέλος στην $\mathcal{ELHI}$ ένα μη-γραμμικό πρόγραμμα Datalog. Με βάση τα αποτελέσματα αυτά, υπολόγισαν και τις αντίστοιχες πολυπλοκότητες αποτίμησης των επαναγραφών, όπως αυτές παρουσιάζονται στο Κεφάλαιο 2. Μετά το Requiem ακολούθησαν οι βελτιστοποιήσεις του, Kyrie [89] και Blackout [104].

Οι Rosati και Almatelli πρότειναν το σύστημα το Presto [116] που επίσης βασίζεται στη μέθοδο της ανάλυσης και υπολογίζει μία επαναγραφή ενός ΣΕ με βάση μία DL-Lite οντολογία. Στο Presto εφαρμόζεται ένα σύνολο από τεχνικές που μειώνουν σημαντικά το μέγεθος της παραγώμενης επαναγραφής. Για να βελτιστοποιηθεί περαιτέρω ο χρόνος αποτίμησης της επαναγραφής στη βάση, αντί για ένα ΣΣΕ, επιστρέφει ένα μη-αναδρομικό πρόγραμμα Datalog το οποίο εκτελείται πολύ πιο γρήγορα από μία SQL βάση. Στη συνέχεια, οι Di Giacomo et al. αξιοποίησαν το σύστημα επαναγραφής Presto στο σύστημα απάντησης ερωτημάτων Mastro [28], στο οποίο επαναγράφουν το μη-αναδρομικό πρόγραμμα Datalog σε ένα ΣΣΕ αφαιρώντας ταυτόχρονα τις περιττές προτάσεις.

Ακολουθώντας την ίδια λογική, οι Trivela et al. πρότειναν το σύστημα Rapid του οποίου ο αλγόριθμος εφαρμόζει τον κανόνα της ανάλυσης επιλεκτικά και με τέτοιο τρόπο ώστε να μην παράγονται περιττές προτάσεις. Με τον τρόπο αυτό, οι τέλικοι έλεγχοι υπαγωγής μειώνονται σημαντικά παράγοντας έτσι την τελική επαναγραφή πολύ πιο σύντομα από άλλα αντίστοιχα συστήματα. Το Rapid μπορεί να υπολογίσει επαναγραφές για τις γλώσσες DL-Lite, $\mathcal{ELHI}$ [131], Horn-$\mathcal{SHIQ}$ [133]. Οι Gottlob et al. ακολουθώντας την ίδια προσέγγιση με τα άλλα συστήματα, ανέπτυξαν το σύστημα Nyaya [52] που υποστηρίζει τη γλώσσα Linear-Datalog$^{\pm}$. Το Nyaya, δηλαδή, εφαρμόζει τον κανόνα της ανάλυσης και χρησιμοποιεί τις προτάσεις της οντολογίας με τέτοιο τρόπο ώστε με βάση τους κανόνες υπαγωγής να ελαχιστοποιήσει το μέγεθος της επαναγραφής. Ένα ακόμα σύστημα που βασίζεται στον κανόνα της ανάλυσης είναι το Clipper [41] το οποίο είναι το πρώτο σύστημα που υποστηρίζει Horn-$\mathcal{SHIQ}$ οντολογίες με μεταβατικούς ρόλους.

Ωστόσο, υπάρχουν συστήματα, όπως είναι το Ontop [114], και το IQAROS [137] των οποίων οι αλγόριθμοι δεν μπορούν να αναχθούν στην μέθοδο της ανάλυσης. Για παράδειγμα, ο αλγόριθμος του IQAROS υπολογίζει μία νέα επαναγραφή για DL-Lite οντολογίες προοδευτικά, δηλαδή υπολογίζει μία επαναγραφή για κάθε άτομο του ερωτήματος και συνδυάζει τα αποτελέσματα με αποδοτικό τρόπο.

Οι Venetis et al. [137, 136] επίσης μελέτησαν το πρόβλημα υπολογισμού μίας επαναγραφής για την περίπτωση που άτομα ή μεταβλητές αφαιρούνται ή προστίθενται στο τιθέμενο ερώτημα, αξιοποιώντας την επαναγραφή του αρχικού ερωτήματος. Θεωρητικά θα μπορούσε το πρόβλημα υπολογισμού νέας επαναγραφής όταν προστίθενται νέα αξιώματα να αναχθεί στο πρόβλημα υπολογισμού μίας επαναγραφής όταν προστίθενται νέα άτομα. Ωστόσο, η τεχνική των Venetis et al. εφαρμόζεται μόνο για DL-Lite οντολογίες ενώ η δική μας τεχνική εφαρμόζεται σε οποιαδήποτε γλώσσα που είναι επαναγράψιμη. Ως προς την αφαίρεση αξιωμάτων (ατόμων) οι προσεγγίσεις μας είναι αρκετά διαφορετικές. Είναι ενδιαφέρον να σημειώσουμε ότι οι εργασίες των Venetis et al. και η δική μας είναι συμπληρωματικές, διότι μπορούν να καλύψουν όλο το φάσμα των πιθανών αλλαγών, στην αρχική οντολογία ή στο τιθέμενο ερώτημα, αποφεύγοντας υπολογισμούς που έχουν ήδη εφαρμοστεί.

7.2 Συνεπής Απάντηση Ερωτημάτων

Το πρόβλημα της ασυνέπειας έχει αποσχολήσει διάφορα πεδία έρευνας και ειδικότερα το πεδίο της Λογικής, των Βάσεων Δεδομένων και της Τεχνητής Νοημοσύνης. Στη βιβλιογραφία έχει σημειωθεί πλήθος δημοσιεύσεων (κάποιες ενδεικτικές εργασίες είναι οι [98, 82, 83, 130]) που προέρχεται από το πεδίο της παρασυνεπούς λογικής (paraconsistent logic), δηλαδή, της λογικής που επιτρέπει την εφαρμογή λογικών συμπερασμών από ασυνεπείς θεωρίες. Στην ενότητα αυτή, αρχικά θα παρουσιάσουμε συνοπτικά τις προσεγγίσεις που σχετίζονται περισσότερο με την διαχείριση της ασυνέπειας στις οντολογίες, στην αναθεώρηση πεποιθήσεων (revision of belief) και στις βάσεις δεδομένων, και στη συνέχεια θα εξετάσουμε πιο αναλυτικά τις εργασίες που σχετίζονται με την IAR και την ICAR σημασιολογία.

Οι Huang et al. [61] προσεγγίζουν το πρόβλημα της συλλογιστικής σε ασυνεπείς ΒΓ ορίζοντας μία συνάρτηση επιλογής, η οποία επιλέγει ένα συνεπές υποσύνολο της οντολογίας. Στη συνέχεια, εφαρμόζονται οι συνήθεις κανόνες της συλλογιστικής για το επιλεγμένο υποσύνολο. Ένα πιο διευρυμένο πλαίσιο προτείνεται από τους Haase et al. [58], στο οποίο παρουσιάζονται τέσσερα προβλήματα που σχετίζονται με την ασυνέπεια. Συγκεκριμένα, μελετάται η συνεπής επέκταση μίας ΒΓ, η διόρθωση της ασυνέπειας, συλλογιστική σε ασυνεπείς ΒΓ και τέλος η συλλογιστική σε διαφορετικές εκδόσεις μίας ΒΓ που μεταξύ τους μπορεί να είναι ασυνεπείς. Η λύση που προτείνουν για τη διόρθωση της ασυνέπειας περιορίζεται στην εύρεση και αφαίρεση του ελαχίστου αριθμού αξιωμάτων ώστε να εξασφαλστεί η συνέπεια. Οσον αφορά τη συλλογιστική σε ασυνεπείς ΒΓ ακολουθούν τη προσέγγιση των Huang et al.. Υπάρχουν, επίσης, αρκετές εργασίες που επικεντρώνονται στον εντοπισμό των ασυνεπειών σε μία ΒΓ. Στο πλαίσιο αυτό, οι Parsia et al [97, 65, 66] παρουσιάζουν τρόπους εντοπισμού μη-ικανοποιήσιμων εννοιών και διόρθωσης τους. Ταυτόχρονα, οι Horridge et al. [59] εξελίσσοντας τις εργασίες αυτές, εντοπίζουν το σύνολο των ελάχιστων υποσυνόλων (justification) της ΒΓ που δημιουργούν την ασύνεπεια και προτείνουν τρόπους επίλυσης του προβλήματος. Οι Baclawski et al. [10] παρουσιάζουν ένα οπτικοποιημένο εργαλείο με το οποίο ο χρήστης μπορεί να ελέγξει τη ΒΓ ως προς τη συνέπειά της. Τέλος, στο ίδιο πλαίσιο, μία από τις δυνατότητες που προσφέρει το σύστημα Mastro [28] είναι ο έλεγχος μίας DL-Lite ΒΓ ως προς την συνέπεια. Η πλειοψηφία των εργασιών αυτών ανάγει το πρόβλημα της ασυνέπειας στην αποτίμηση ερωτημάτων λογικής πρώτης τάξης και κυρίως επικεντρώνονται σε ασυνέπειες που εμφανίζονται στο σώμα ορολογίας.

Οι σημασιολογίες IAR και ICAR με τις οποίες ασχοληθήκαμε στη διατριβή αυτή, βασίζονται στην αρχή WIDTO (when in doubt throw it out) [46] που προέρχεται από τον χώρο αναθεώρησης πεποιθήσεων, αλλά οι σημασιολογίες που έχουν οριστεί στο πλαίσιο αυτό δεν διαχειρίζονται το πρόβλημα της ασυνέπειας. Από την άλλη πλευρά, στην ίδια ερευνητική περιοχή οι Qi et al. [107], διαχειρίζονται το πρόβλημα της ασυνέπειας με την χρήση ενός τελεστή αναθεώρησης, με το οποίο τροποποιείται η αρχική ΒΓ τροποιώντας τόσο το σώμα ισχυρισμών όσο και το σώμα ορολογίας. Ενώ οι ίδιοι συγγραφείς σε επόμενες εργασίες τους [93, 92, 91] πρότειναν την επίλυση του προβλήματος της ασυνέπειας μέσω μίας "βαθμολόγησης" των υπεύθυνων αξιωμάτων. Επίσης, οι Meyer et al. [88] προσέγγισαν το πρόβλημα τροποποιώντας τα ίδια τα αξιώματα.

Η προσέγγιση που ακολουθούμε στην διατριβή αυτή προέρχεται κυρίως από τον χώρο των Βάσεων Δεδομένων, όπου αντιμετωπίζεται επίσης το πρόβλημα δημιουργίας βάσεων από πολλούς διαφορετικούς φορείς ή το πρόβλημα διαχείρισης των δεδομένων μέσω περίπλοκων και μακροχρόνιων διεργασιών, που είναι πιθανό να οδηγήσει σε ασυνέπειες. Μία από τις κύριες προσεγγίσεις βασίζεται στην ελάχιστη δυνατή αλλαγή των δεδομένων της βάσης (minimal-change integrity maintenance) έτσι ώστε να διασφαλιστεί η συνέπεια της [42, 34]. Στο ίδιο πλαίσιο, οι Arenas et al. [4] εισήγαγαν την έννοια της "διόρθωσης" ως το στιγμιότυπο της βάσης που ικανοποιεί τους περιορισμούς ακεραιότητας (integrity constraints) και διαφέρει ελάχιστα από την αρχική οντολογία. Μια διαφορετική προσέγγιση που προέρχεται επίσης από τον χώρο των βάσεων δεδομένων είναι η συνεπής απάντηση ερωτημάτων [25, 33] σύμφωνα με την οποία, ο στόχος δεν είναι η αλλαγή της βάσης αλλά η εξαγωγή συνεπούς πληροφορίας, παρά την ασυνέπεια της. Την προσέγγιση αυτή ακολουθήσαμε στις Ενότητες 5.1,5.2.

7.2.1 Οι σημασιολογίες IAR και ICAR

Πρώτη φορά τόσο η έννοια της διόρθωσης για μία ΒΓ, με τον τρόπο χρησιμοποιείται στη διατριβή αυτή, όσο και η AR-σημασιολογία (ABox-Repair semantics) εισήχθη από τους Lembo et al. [78] το 2007. Στη συνέχεια, Lembo et al. [75], αφού απέδειξαν ότι για τις ΠΛ που ανήκουν στη DL-Lite οικογένεια, η συνεπής απάντηση ερωτημάτων υπό την AR-σημασιολογία είναι coNP-complete, εισήγαγαν την IAR-σημασιολογία, που αποτελεί προσέγγισή της. Επιπλέον, έδειξαν ότι το πρόβλημα μετατρέπεται σε πολυωνυμικό για την DL-Lite$_{\mathcal{A}}$, που είναι στην ουσία μία επέκταση της DL-Lite$_{\mathcal{R}}$ τέτοια ώστε να μπορούν να οριστούν και ιδιότητες τύπων δεδομένων (datatype properties) καθώς και η δυνατότητα προσδιορισμού ενός ρόλου ως συναρτησιακού. Επίσης, εισήγαγαν την CAR-σημασιολογία (Closed ABox Repair-Διόρθωση Κλειστών ABox) σύμφωνα με την οποία οι απαντήσεις υπολογίζονται με βάση τις διορθώσεις του κλειστού ABox, δηλαδή του ABox στο οποίο έχουν εφαρμοστεί όλοι οι συμπερασμοί με βάση το TBox. Ωστόσο, όπως έδειξαν στην ίδια εργασία η απάντηση ερωτημάτων υπό την CAR-σημασιολογία είναι επίσης coNP-complete πρόβλημα για την DL-Lite$_{core}$, που είναι η απλούστερη γλώσσα της οικογένειας DL-Lite. Όποτε με την ίδια λογική, εισήγαγαν την ICAR-σημασιολογία και έδειξαν ότι έχει πολυωνυμική πολυπλοκότητα για την DL-Lite$_{\mathcal{A}}$.

Το 2011, η Bienvenu [17] μελέτησε τις περιπτώσεις όπου η συνεπής απάντηση ερωτημάτων είναι εφικτή για την DL-Lite υπό την AR-σημασιολογία, και συγκεκριμένα με τη χρήση της τεχνικής της επαναγραφής για την DL-Lite$_{core}$. Ταυτόχρονα, οι Lembo et al. [76] έδειξαν ότι το πρόβλημα απάντησης ΣΕ στην DL-Lite$_{A}$, και στην συνέχεια στην DL-Lite$_{A,id,den}$ [77], τόσο υπό την IAR όσο και υπό την ICAR σημασιολογία μπορεί να αναχθεί στο πρόβλημα απάντησης ενός ερωτήματος της λογικής πρώτης τάξης σε μία βάση δεδομένων και συνεπώς ελαχιστοποίησαν την πολυπλοκότητα του προβλήματος από πολυωνυμική σε AC^0. Στην ίδια εργασία, πρότειναν τις αντίστοιχες μεθόδους για τον υπολογισμό της IAR και της ICAR-επαναγραφής (που αντιστοιχούν στους μετασχηματισμούς ref, ref^{ic} που παρουσιάζουμε αναλυτικά στην Ενότητα 2.4.3).

Το ίδιο έτος ο Rosati [115] έδειξε ότι το προβλήμα της απάντησης ερωτημάτων στις σημασιολογίες AR, IAR, CAR, ICAR, για ένα μεγάλο εύρος ΠΛ, από την $\mathcal{EL}_{\bot}$ έως την $\mathcal{SHIQ}$ είναι δυσεπίλυτο (intractable). Επίσης, πρότεινε το υποσύνολο $\mathcal{EL}_{\bot nr}$ της γλώσσας $\mathcal{EL}_{\bot}$, στο οποίο το πρόβλημα για την IAR-σημασιολογία έχει πολυωνυμική πολυπλοκότητα, αλλά για την ICAR-σημασιολογία το πρόβλημα παραμένει δυσεπίλυτο. Τα αποτελέσματα της εργασίας αυτής παρουσιάζονται στην Ενότητα 2.4.3. Επίσης, οι Lukasiewicz et al. [79] έδειξαν ότι το ίδιο πρόβλημα για την IAR-σημασιολογία στη φυλαγμένη (guarded) Datalog είναι coNP-complete ενώ για τη γραμμική Datalog πρότειναν μία μεθοδολογία βασισμένη στην έννοια της επαναγραφής, αντίστοιχη της εργασίας [76]. Με τη μεθοδολογία αυτή, έδειξαν ότι το πρόβλημα μπορεί να αναχθεί στο πρόβλημα απάντησης ενός ερωτήματος της λογικής πρώτης τάξης σε μία βάση δεδομένων.

Συστήματα Υπολογισμού Συνεπών Απαντήσεων

Μια πρώτη προσπάθεια υπολογισμού των IAR και ICAR απαντήσεων έγινε από την Massoti [84] με το σύστημα QuAC, με το οποίο κατασκευάζονται οι IAR και ICAR-διορθώσεις της ΒΓ. Το σύστημα αυτό ήταν μία πρώτη πρακτική προσέγγιση προς αυτήν την κατεύθυνση και η πειραματική αξιολόγησή του αρκετά περιορισμένη. Σε επόμενη εργασία, οι Rosati et al. [118] παρουσίασαν το σύστημα QuID, ως εξέλιξη του QuAC, με το οποίο μπορούν να υπολογιστούν οι IAR-απαντήσεις στη γλώσσα DL-Lite$_{\mathcal{A}}$ είτε με τον υπολογισμό της IAR-διόρθωσης της ΒΓ και κατόπιν αποτίμησης του τιθέμενου ερωτήματος σε αυτήν είτε με τον υπολογισμό της IAR-επαναγραφής του ερωτήματος και αποτίμησής του στην αρχική ΒΓ. Σε κάθε περίπτωση, η αποτίμηση γίνεται με τη χρήση του PostgreSQL. Τέλος, οι Bienvenu et al. [18] μελέτησαν τέσσερις παραλλαγές των σημασιολογιών AR και IAR, βασισμένες στην έννοια της προτίμησης (preference), και αφού έδειξαν ότι για την DL-Lite$_{\mathcal{R}}$ το πρόβλημα γίνεται coNP-complete ή δυσκολότερο, υλοποίησαν το εργαλείο CQApri με το οποίο υπολογίζουν τις απαντήσεις AR, IAR, και τις αντίστοιχες απαντήσεις $\subseteq_p$-AR, $\subseteq_p$-IAR με βάση την πιο εύκολα υπολογίσιμη κατηγορία προτιμήσεων p. Συνοπτικά, για τον υπολογισμό των IAR-απαντήσεων, το CQApri αρχικά εντοπίζει και αποθηκεύει τα υποσύνολα του σώματος των ισχυρισμών που προκαλούν ασυνέπεια, στη συνέχεια, υπολογίζει με το Rapid και αποτιμάει με την PostgreSQL την επαναγραφή του τιθέμενου ερωτήματος και, τέλος, επιστρέφει μόνο τις απαντήσεις που δεν εμπλέκονται σε κάποια ασυνέπεια.

Στο Κεφάλαιο 6, παρουσιάζεται η σύγκριση των συστημάτων μας με το QuID και το CQApri και δείχνουμε ότι το σύστημά μας παρουσιάζει σαφώς καλύτερη απόδοση.

Προσεγγίσεις

Το 2013 οι Bienvenu et al. [20] όρισαν δύο νέες σημασιολογίες, τις k-support και k-defeater, που έχουν τις καλές υπολογιστικές ιδιότητες της IAR-σημασιολογίας, αλλά ταυτόχρονα δίνουν τόσες απαντήσεις όσες η AR-σημασιολογία. Ωστόσο, δεν παρουσιάζουν κάποια υλοποιήση που να υπολογίζει απαντήσεις υπό αυτές τις σημασιολογίες. Επίσης, όρισαν την *θαρραλέα* (*brave semantics*), σύμφωνα με την οποία, ένα boolean ερώτημα $\mathcal{Q}$ συνεπάγεται από μία βάση γνώσης $\langle \mathcal{T}, \mathcal{A} \rangle$ αν υπάρχει μία AR-διόρθωση $\mathcal{A}' \subseteq \mathcal{A}$ τέτοια ώστε $\mathcal{T}, \mathcal{A}' \models \mathcal{Q}$. Ωστόσο, όπως οι ίδιοι αναφέρουν, από την σημασιολογία αυτή εξάγονται μη-συνεπή συμπεράσματα. Μια άλλη προσπάθεια διεύρυνσης των IAR-απαντήσεων έχει σημειωθεί από τους Lukasiewicz et al. [79], με την k-χαλαρή (k-lazy) σημασιολογία, η οποία όμως έχει τις ίδιες υπολογιστικές ιδιότητες με την IAR-σημασιολογία μόνο αν το ερώτημα περιέχει μόνο ένα άτομο, διαφορετικά το πρόβλημα γίνεται coNP-hard ως προς τα δεδομένα.

Κεφάλαιο 8

Συνεισφορά και Θέματα προς Έρευνα

Τα τελευταία χρόνια το πεδίο απάντησης συζευκτικών ερωτημάτων σε μεγάλα σύνολα δεδομένων έχει καταστεί αντικείμενο συνεχούς έρευνας. Μία από τις πιο διαδεδομένες προσεγγίσεις στο πρόβλημα αυτό βασίζεται στη τεχνική επαναγραφής ερωτημάτων κατά την οποία το αρχικό ερώτημα επαναγράφεται με βάση τα αξιώματα της οντολογίας. Η επαναγραφή εκτελείται με τέτοιο τρόπο ώστε να περιλαμβάνει όλη τη σχετική πληροφορία από την οντολογία για την απάντηση του ερωτήματος πάνω σε ένα οποιοδήποτε σύνολο δεδομένων.

Τα υπάρχοντα συστήματα επαναγραφής ερωτημάτων δέχονται στην είσοδο ένα συζευκτικό ερώτημα και μία οντολογία και υπολογίζουν μία επαναγραφή του ερωτήματος με βάση την οντολογία. Ωστόσο, τα συστήματα αύτα είναι έτσι σχεδιασμένα ώστε όταν η οντολογία τροποποιείται -δηλαδή, επεκτείνεται ή μειώνεται κατά ένα σύνολο αξιωμάτων- υπολογίζουν την νέα επαναγραφή από την αρχή. Κατά αυτόν τον τρόπο δεν αξιοποιούν την πληροφορία που έχει παραχθεί από τις προηγούμενες επαναγραφές. Καθώς, όμως, οι οντολογίες που χρησιμοποιούνται για να μοντελοποιήσουν την επιστημονική γνώση σε πραγματικά πεδία συνεχώς αλλάζουν, με τα υπάρχοντα συστήματα θα πρέπει κάθε φορά η επαναγραφή του ερωτήματος να εκτελείται από την αρχή. Τελικά, με τον τρόπο αυτόν προκαλούνται καθυστερήσεις στην τελική απάντηση των ερωτημάτων.

Επιπλέον, στη προσέγγιση απάντησης ερωτημάτων μέσω επαναγραφής θεωρείται πάντα ότι η ΒΓ είναι συνεπής. Ωστόσο, η τροποποίηση των οντολογιών μπορεί να προκαλέσει ασυνέπεια στη ΒΓ, δηλαδή να υπάρχουν ισχυρισμοί που να παραβιάζουν τα αξιώματα της οντολογίας. Το πρόβλημα της ασυνέπειας μπορεί να επιλυθεί είτε αφαιρώντας από το σύνολο των δεδομένων τους ισχυρισμούς που προκαλούν ασυνέπεια είτε υπολογίζοντας απαντήσεις με νόημα, παρά την ασυνέπεια της ΒΓ.

Στην παρούσα διατριβή, αρχικά μελετάμε το πρόβλημα υπολογισμού μίας επαναγραφής ενός ερωτήματος πάνω σε μία οντολογία που έχει τροποποιηθεί, αξιοποιώντας τη πληροφορία που έχει παραχθεί από τον υπολογισμό μίας επαναγραφής πάνω σε προηγούμενη έκδοση της οντολογίας. Στη συνέχεια, μελετάμε το πρόβλημα υπολογισμού απαντήσεων σε ασυνεπείς ΒΓ. Πιο συγκεκριμένα, η συνεισφορά της εργασίας μας συνοψίζεται στα εξής:

- *Υπολογισμός επαναγραφής όταν η οντολογία επεκτείνεται*: Μελετήσαμε το πρόβλημα θεωρητικά και προτείναμε έναν αλγόριθμο, του οποίου αποδείξαμε την ορθότητα. Ο αλγόριθμος εστιάζει στην υλοποιήση των συμπερασμών που οφείλονται μόνο στην εισαγωγή των νέων αξιωμάτων. Ένα σημαντικό θεωρητικό αποτέλεσμα είναι ότι για τον αποδοτικό υπολογισμό της νέας επαναγραφής δεν αρκεί η αρχική επαναγραφή $\mathcal{R}$, αλλά απαιτείται το ενδιάμεσο σύνολο που παράγεται κατά τη διαδικασία υπολογισμού του $\mathcal{R}$. Με τον τρόπο αυτό, ο αλγόριθμος εκτελεί *μόνο* την πρόσθετη εργασία που είναι απαραίτητη για τον υπολογισμό μιας επανεγγραφής για την επεκτεταμένη οντολογία. Επίσης, μπορεί να εφαρμοστεί σε οποιοδήποτε σύστημα που βασίζεται στη μέθοδο της ανάλυσης. Για την πειραματική αξιολόγησή του τον εφαρμόσαμε στα συστήματα Requiem και Rapid. Διαπιστώσαμε ότι όταν προστίθενται στην οντολογία ένα σύνολο αξιωμάτων που αντιστοιχεί στο 5% του μεγέθους της, είναι πολύ πιο αποδοτικό να χρησιμοποιηθεί το σύστημά μας από το αντίστοιχο σύστημα επαναγραφής.

- *Υπολογισμός επαναγραφής όταν η οντολογία συστέλλεται*: Δεδομένης μίας επαναγραφής με βάση μία αρχική οντολογία, μελετήσαμε το πρόβλημα υπολογισμού μίας νέας επαναγραφής όταν από την οντολογία αφαιρείται ένα σύνολο αξιωμάτων. Μελετήσαμε το πρόβλημα θεωρητικά και παρουσιάσαμε έναν γενικό αλγόριθμο που υπολογίζει μία νέα επαναγραφή. Σύμφωνα με τον αλγόριθμο αυτόν, αρχικά αφαιρούνται οι προτάσεις που δεν συμπεραίνονται από το ερώτημα και τη νέα οντολογία. Για να επιτευχθεί αυτό, θεωρούμε ότι κάθε πρόταση της αρχικής επαναγραφής είναι επισημειωμένη με το σύνολο των υποσυνόλων της οντολογίας που είναι υπεύθυνα για την παραγωγή της. Με τον τρόπο αυτό, για να παραμείνει μια πρόταση στη νέα επαναγραφή αρκεί στην επισημείωσή της να υπάρχει κάποιο υποσύνολο της οντολογίας που να μην περιέχει κάποιο απο τα αξιώματα που έχουν αφαιρεθεί. Ωστόσο, η λύση αυτή δεν είναι πλήρης στην περίπτωση που από την αρχική επαναγραφή έχουν αφαιρεθεί οι προτάσεις που είναι περιττές, διότι υπάρχει περίπτωση μία πρόταση που στην αρχική επαναγραφή ήταν περιττή να μην είναι περιττή στην νέα επαναγραφή. Συνεπώς, πρέπει να προστεθεί η πρόταση αυτή καθώς και να εφαρμοστούν νέοι συμπερασμοί εξαιτίας αυτής της προσθήκης. Για να εφαρμοστούν μόνο οι απαραίτητοι νέοι συμπερασμοί καλείται ο αλγόριθμος υπολογισμού επαναγραφής όταν η οντολογία επεκτείνεται, άρα όπως και προηγουμένως, δεν αρκεί η αρχική επαναγραφή αλλά απαιτείται το ενδιάμεσο σύνολο που προκύπτει κατά τον υπολογισμό της αρχικής επαναγραφής.

- *Υπολογισμός επαναγραφής όταν η οντολογία συστέλλεται, χωρίς την εφαρμογή νέων συμπερασμών*: Είναι προφανές ότι αν κατα τον υπολογισμό της αρχικής επαναγραφής δεν έχουν εφαρμοστεί κριτήρια περιττότητας τότε για τον υπολογισμό της νέας επαναγραφής δεν χρειάζεται να εφαρμοστούν νέοι συμπερασμοί. Όπως είναι φυσικό όμως, υπό αυτές τις συνθήκες, αν η οντολογία είναι αρκετά μεγάλη τότε η επαναγραφή δεν είναι πρακτικά υπολογίσιμη. Βελτιστοποιήσαμε το αποτέλεσμα αυτό, αποδεικνύοντας ότι αρκεί η αρχική επαναγραφή να είναι κλειστή ως προς τα υποσύνολα, δηλαδή μπορεί να έχουν εφαρμοστεί κάποιοι έλεγχοι περιττότητας ώστε να περιοριστεί το πρόβλημα αυτό.

- *Υπολογισμός επαναγραφής όταν η οντολογία συστέλλεται, με τη χρήση γράφων*: Ένα μειονέκτημα των προηγούμενων αλγορίθμων που αφορούν στη συστολή των οντολογιών είναι ότι απαιτούν την αποθήκευση μεγάλου όγκου πληροφορίας. Αυτό συμβαίνει γιατί *κάθε* πρόταση του συνόλου εισόδου, που μπορεί να είναι είτε υπερσύνολο της αρχικής επαναγραφής είτε μία αρχική επαναγραφή κλειστή ως προς τα υποσύνολα, πρέπει να είναι επισημειώμενη με *όλα* τα υποσύνολα της οντολογίας από τα οποία μπορεί να παραχθεί. Επίσης, οι αλγόριθμοι θα πρέπει να διατρέξουν *όλες* τις προτάσεις της αρχικής επαναγραφής και τις επισημειώσεις τους. Αξιοποιώντας όμως τις συσχετίσεις που υπάρχουν μεταξύ των προτάσεων του συνόλου εισόδου, τότε οι προηγούμενοι αλγόριθμοι μπορούν να βελτιστοποιηθούν τόσο χωρικά όσο και χρονικά. Μελετήσαμε το πρόβλημα θεωρητικά και στη συνέχεια παρουσιάσαμε έναν αποδοτικό αλγόριθμο. Καταρχήν, ορίσαμε τον επισημειωμένο γράφο παραγωγής κάθε κόμβος του οποίου είναι μία πρόταση του συνόλου, σημειωμένος με τις προτάσεις του συνόλου που την συμπεραίνουν άμεσα. Ο αλγόριθμος διατρέχει τον γράφο κατά βάθος, προσθέτει στο σύνολο εξόδου κάθε κόμβο του οποίου η επισημείωση περιέχει προτάσεις που ανήκουν στο σύνολο εξόδου ή στη νέα οντολογία και συνεχίζει με τα παιδιά του. Με τον τρόπο αυτό δεν διατρέχεται ολόκληρος ο γράφος αλλά μόνο οι κόμβοι του που αντιστοιχούν σε προτάσεις που μπορούν να παραχθούν από τη νέα οντολογία. Αποδείξαμε την ορθότητά του αλγορίθμου και τον υλοποιήσαμε. Για την περίπτωση που το αρχικό σύνολο είναι το ενδιάμεσο σύνολο, χρησιμοποιήσαμε το σύστημα Requiem. Για την περίπτωση που το αρχικό σύνολο είναι η επαναγραφή κλειστή ως προς τα υποσύνολα, δεν βασιστήκαμε σε κάποιο σύστημα επαναγραφής διότι σε αυτή την περίπτωση δεν εφαρμόζονται νέοι συμπερασμοί. Τα συστήματα μας ελέγχθηκαν σε σενάρια συστολής της οντολογίας από 0,5% έως και 5% και αποδείχθηκαν σε όλες τις περιπτώσεις πιο αποδοτικά από τα συστήματα Rapid και Requiem.

Από όσο γνωρίζουμε, αυτή είναι η πρώτη θεωρητική και πρακτική μελέτη του προβλήματος αποδοτικού υπολογισμού μίας νέας επαναγραφής σε εξελισσόμενες οντολογίες.

- *Συνεπής Απάντηση Ερωτημάτων με την ICAR σημασιολογία για DL-Lite$_{\mathcal{R}}$ οντολογίες*: Μελετήσαμε το πρόβλημα υπολογισμού όλων των απαντήσεων σε ένα ερώτημα που μπορούν να *παραχθούν* από τη ΒΓ, παρά την ασυνέπειά της. Σύμφωνα με τη προσέγγιση που προτείναμε, αρχικά υπολογίζεται μία επαναγραφή $\langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ του ερωτήματος με βάση τα θετικά αξιώματα της οντολογίας. Στη συνέχεια, επεκτείνεται κάθε ερώτημα του $\mathcal{R}_Q$ με τα κατάλληλα αρνητικά άτομα, τέτοια ώστε να εξασφαλιστεί ότι δε θα επιστραφούν ασυνεπείς απαντήσεις. Τέλος, για να υπολογιστούν οι απαντήσεις, αξιοποιείται ένα $\mathcal{RL}$ σύστημα κορεσμού δεδομένων το οποίο δέχεται το νέο $\mathcal{R}_Q$, την οντολογία μαζί με το συμπλήρωμά της για την $\mathcal{RL}$ και το σώμα των ισχυρισμών. Ο αλγόριθμος υλοποιήθηκε σε ένα πρωτότυπο σύστημα και η σύγκρισή του με τα συστήματα CQApri και QuID που υπολογίζουν απαντήσεις με βάση την IAR σημασιολογία (και επομένως υπολογίζουν λιγότερες απαντήσεις) έδειξε ότι είχε σχετικά καλή απόδοση.

- *Βελτιστοποίηση Αλγορίθμου Συνεπούς Απάντησης Ερωτημάτων με την ICAR σημασιολογία για DL-Lite$_\mathcal{R}$ οντολογίες*: Αξιοποιήσαμε περαιτέρω τις ιδιότητες των συστημάτων κορεσμού δεδομένων βελτιστοποιώντας σημαντικά τον αλγόριθμό μας. Συγκεκριμένα, λαμβάνοντας υπόψη ότι τα συστήματα αυτά υπολογίζουν αρχικά όλους τους ισχυρισμούς που μπορεί να συμπεραχθούν από τη ΒΓ, μειώσαμε σημαντικά τον αριθμό των αρνητικών ατόμων που εμφανίζονται στα ερωτήματα του $\mathcal{R}_Q$. Η υλοποίηση της βελτιστοποίησης αυτής αύξησε σημαντικά την απόδοση του συστήματός μας· στις περισσότερες περιπτώσεις ανταποκρίθηκε πιο γρήγορα από το CQApri και το QuID. Τέλος, ελέγχοντας κατά το βήμα της προεπεξεργασίας ποια αρνητικά αξιώματα μεταξύ ρόλων παραβιάζονται από το σώμα ισχυρισμών, βελτιστοποιήσαμε περαιτέρω το σύστημά μας. Από όσο γνωρίζουμε, αυτό είναι το πρώτο αποδοτικό σύστημα υπολογισμού των ICAR απαντήσεων.

- *Υπολογισμός ICAR-απαντήσεων σε εκφραστικές ΠΛ*: Εισάγαμε μία νέα ICAR-τύπου σημασιολογία, που προσεγγίζει την ICAR σημασιολογία άνωθεν και παρουσιάσαμε έναν αλγόριθμο υπολογισμού απαντήσεων με βάση τη σημασιολογία αυτή. Επίσης, αποδείξαμε ότι το πρόβλημα συνεπούς απάντησης ερωτημάτων με τη νέα σημασιολογία μπορεί να επιλυθεί σε πολυωνυμικό χρόνο για τις ΠΛ: DL-Lite$_\mathcal{R}$, $\mathcal{ELHI}^\neg$, Horn-$\mathcal{SHIQ}$, *markable*-$\mathcal{SHI}$, $\mathcal{ALCHI}$.

- *Υπολογισμός IAR-απαντήσεων για DL-Lite$_\mathcal{R}$ οντολογίες*: Παρουσιάσαμε έναν αποδοτικό αλγόριθμο υπολογισμού συνεπών απαντήσεων με βαση την IAR σημασιολογία. Ο αλγόριθμος αυτός αρχικά υπολογίζει την IAR διόρθωση του σώματος των ισχυρισμών με αποδοτικό τρόπο και στη συνέχεια χρησιμοποιεί ένα σύστημα κορεσμού δεδομένων για τον υπολογισμό των απαντήσεων. Η πειραματική ανάλυση της υλοποίησης του αλγορίθμου έδειξε ότι είναι πιο αποδοτικός από τα συστήματα CQApri και QuId, τα οποία είναι μοναδικά συστήματα που υπολογίζουν τις IAR απαντήσεις.

Στο υπόλοιπο του κεφαλαίου αυτού παρουσιάζουμε κάποια σχετικά ανοικτά θέματα για περαιτέρω έρευνα:

- *Συνεπής απάντηση ερωτημάτων με την ICAR σημασιολογία σε ΒΓ που εξελίσσονται*: Ο αλγόριθμος υπολογισμού συνεπών απαντήσεων με την ICAR σημασιολογία για DL-Lite βάσεις γνώσεις που προτείναμε στη διατριβή αυτή αποτελείται τέσσερα βήματα, αρχικά υπολογίζει: 1. την επαναγραφή $\mathcal{R} = \langle \mathcal{R}_Q, \mathcal{R}_D \rangle$ του ερωτήματος με βάση τα θετικά αξιώματα της οντολογίας, 2. όλα τα αρνητικά αξιώματα υπαγωγής που παράγονται από την οντολογία, 3. το συμπλήρωμα της οντολογίας για το $\mathcal{RL}$ σύστημα κορεσμού και 4. χρησιμοποιεί το σύστημα κορεσμού για τον υπολογισμό των απαντήσεων. Σε περίπτωση που η οντολογία συστέλλεται ή επεκτείνεται τότε μπορούν κάποιοι υπολογισμοί να μην επαναληφθούν αν για τον υπολογισμό της νέας επαναγραφής $\mathcal{R}$ αξιοποιηθούν οι αλγόριθμοι που παρουσιάσαμε στη διατριβή αυτή και αφορούν στις εξελισσόμενες ΒΓ. Έχει, όμως, επίσης ερευνητικό και πρακτικό ενδιαφέρον η βελτιστοποίηση των υπόλοιπων βημάτων με τέτοιο τρόπο ώστε να εκτελεστούν μόνο οι απαραίτητοι υπολογισμοί.

- *Υπολογισμός εξηγήσεων (explanations) για την ICAR σημασιολογία*: Οι Bienvenu et al [19] εισήγαγαν την έννοια της *εξήγησης* για τις σημασιολογίες AR και IAR τόσο για τις απαντήσεις που επιστρέφονται όσο και για τις απαντήσεις που *δεν* επιστρέφονται. Διαισθητικά, μια εξήγηση είναι ένα υποσύνολο του σώματος των ισχυρισμών που υποδεικνύει τα αίτια της απάντησης ή μη-απάντησης. Στο πλαίσιο αυτό, θα είχε ιδιαίτερο ενδιαφέρον, το σύστημά μας για την ICAR σημασιολογία να υπολογίζει τις αντίστοιχες εξηγήσεις ώστε ο χρήστης να έχει τη δυνατότητα να διορθώσει τη ΒΓ.

- *Κάτωθεν προσέγγιση της ICAR σημασιολογίας για εκφραστικές ΠΛ*: Στη διατριβή αυτή παρουσιάσαμε μία άνωθεν προσέγγιση της ICAR σημασιολογίας για εκφραστικές ΠΛ. Θα ήταν πολύ σημαντικό να μελετηθεί και μία κάτωθεν προσέγγιση της ICAR σημασιολογίας για εκφραστικές ΠΛ που να είναι επίσης πρακτικά υπολογίσιμη.

Μέρος V

Παραρτήματα

Παράρτημα Α′

Ερωτήματα Αξιολόγησης

Α′.1 Ερωτήματα Αξιολόγησης Κεφαλαίου 4

V:

$$
\begin{aligned}
\mathsf{q}_1: &\quad Q(x) &\leftarrow&\quad \mathsf{Location}(x)\\
\mathsf{q}_2: &\quad Q(x,y) &\leftarrow&\quad \mathsf{MilitaryPerson}(x) \wedge \mathsf{hasRole}(y,x) \wedge \mathsf{related}(x,z)\\
\mathsf{q}_3: &\quad Q(x,y) &\leftarrow&\quad \mathsf{TimeDependantRelation}(x) \wedge \mathsf{hasRelationMember}(x,y) \wedge \mathsf{Event}(y)\\
\mathsf{q}_4: &\quad Q(x,y) &\leftarrow&\quad \mathsf{Object}(x) \wedge \mathsf{hasRole}(x,y) \wedge \mathsf{Symbol}(y)\\
\mathsf{q}_5: &\quad Q(x) &\leftarrow&\quad \mathsf{Individual}(x) \wedge \mathsf{hasRole}(x,y) \wedge \mathsf{Scientist}(y) \wedge\\
&&&\quad \mathsf{hasRole}(x,z) \wedge \mathsf{Discoverer}(z) \wedge \mathsf{hasRole}(x,w) \wedge \mathsf{Inventor}(w)
\end{aligned}
$$

P5, P5X:

$$
\begin{aligned}
\mathsf{q}_1: &\quad Q(x) &\leftarrow&\quad \mathsf{edge}(x,y)\\
\mathsf{q}_2: &\quad Q(x) &\leftarrow&\quad \mathsf{edge}(x,y) \wedge \mathsf{edge}(y,z)\\
\mathsf{q}_3: &\quad Q(x) &\leftarrow&\quad \mathsf{edge}(x,y) \wedge \mathsf{edge}(y,z) \wedge \mathsf{edge}(z,w)\\
\mathsf{q}_4: &\quad Q(x) &\leftarrow&\quad \mathsf{edge}(x,y) \wedge \mathsf{edge}(y,z) \wedge \mathsf{edge}(z,w)\mathsf{edge}(w,z)\\
\mathsf{q}_5: &\quad Q(x) &\leftarrow&\quad \mathsf{edge}(x,y) \wedge \mathsf{edge}(y,z) \wedge \mathsf{edge}(z,w) \wedge\\
&&&\quad \mathsf{edge}(w,z) \wedge \mathsf{edge}(z,w)
\end{aligned}
$$

A, AX:

$$
\begin{aligned}
\mathsf{q}_1: &\quad Q(x) &\leftarrow&\quad \mathsf{Device}(x) \wedge \mathsf{assistsWith}(x,y)\\
\mathsf{q}_2: &\quad Q(x) &\leftarrow&\quad \mathsf{Device}(x) \wedge \mathsf{assistsWith}(x,y) \wedge \mathsf{UpperLimbMobility}(y)\\
\mathsf{q}_3: &\quad Q(x) &\leftarrow&\quad \mathsf{Device}(x) \wedge \mathsf{assistsWith}(x,y) \wedge \mathsf{Hear}(y) \wedge \mathsf{affects}(z,y) \wedge \mathsf{Autism}(z)\\
\mathsf{q}_4: &\quad Q(x) &\leftarrow&\quad \mathsf{Device}(x) \wedge \mathsf{assistsWith}(x,y) \wedge \mathsf{PhysicalAbility}(y)\\
\mathsf{q}_5: &\quad Q(x) &\leftarrow&\quad \mathsf{Device}(x) \wedge \mathsf{assistsWith}(x,y) \wedge \mathsf{PhysicalAbility}(y) \wedge \mathsf{affects}(z,y) \wedge\\
&&&\quad \mathsf{Quadriplegia}(z)
\end{aligned}
$$

U,UX:

$$\begin{array}{rrcl}
\mathsf{q}_1: & Q(x) & \leftarrow & \mathsf{worksFor}(x,y) \wedge \mathsf{affiliatedOrganizationOf}(y,z) \\
\mathsf{q}_2: & Q(x,y) & \leftarrow & \mathsf{Person}(x) \wedge \mathsf{teacherOf}(x,y) \wedge \mathsf{Course}(y) \\
\mathsf{q}_3: & Q(x,y,z) & \leftarrow & \mathsf{Student}(x) \wedge \mathsf{advisor}(x,y) \wedge \mathsf{FacultyStaff}(y) \wedge \\
 & & & \mathsf{takesCourse}(x,z) \wedge \mathsf{teacherOf}(y,z) \wedge \mathsf{Course}(z) \\
\mathsf{q}_4: & Q(x,y) & \leftarrow & \mathsf{Person}(x) \wedge \mathsf{worksFor}(x,y) \wedge \mathsf{Organization}(y) \\
\mathsf{q}_5: & Q(x) & \leftarrow & \mathsf{Person}(x) \wedge \mathsf{worksFor}(x,y) \wedge \mathsf{University}(y) \wedge \mathsf{hasAlumnus}(y,x)
\end{array}$$

SWEET:

$$\begin{array}{rrcl}
\mathsf{q}_1: & Q(x) & \leftarrow & \mathsf{Ocean}(x) \wedge \mathsf{contains}(x,y) \wedge \mathsf{Volcano}(y) \\
\mathsf{q}_2: & Q(x,y) & \leftarrow & \mathsf{City}(x) \wedge \mathsf{inside}(x,y) \wedge \mathsf{Country}(y) \wedge \mathsf{isAdjacentTo}(x,z) \wedge \mathsf{Desert}(z) \\
\mathsf{q}_3: & Q(x) & \leftarrow & \mathsf{Country}(x) \wedge \mathsf{partOf}(x,y) \wedge \mathsf{Continent}(y) \wedge \mathsf{surroundedBy}(y,z) \wedge \mathsf{Ocean}(z) \\
\mathsf{q}_4: & Q(x) & \leftarrow & \mathsf{Region}(x) \\
\mathsf{q}_5: & Q(x,y) & \leftarrow & \mathsf{Island}(x) \wedge \mathsf{isPart}(x,y) \wedge \mathsf{Region}(y) \wedge \mathsf{contains}(y,z) \wedge \mathsf{Volcano}(z)
\end{array}$$

PRO:

$$\begin{array}{rrcl}
\mathsf{q}_1: & Q(x) & \leftarrow & \mathsf{haspart}(x,y) \wedge \mathsf{CHEBI23367}(y) \\
\mathsf{q}_2: & Q(x) & \leftarrow & \mathsf{lackspart}(x,y) \wedge \mathsf{SO0000418}(y) \\
\mathsf{q}_3: & Q(x) & \leftarrow & \mathsf{PR000000001}(x) \\
\mathsf{q}_4: & Q(x) & \leftarrow & \mathsf{derivesfrom}(x,y) \wedge \mathsf{PR000000001}(y) \\
\mathsf{q}_5: & Q(x) & \leftarrow & \mathsf{haspart}(x,y) \wedge \mathsf{CHEBI23367}(y), \mathsf{lackspart}(x,z) \wedge \mathsf{SO0000418}(z) \\
\mathsf{q}_6: & Q(x) & \leftarrow & \mathsf{haspart}(x,y) \wedge \mathsf{CHEBI23367}(y) \wedge \mathsf{lackspart}(x,z) \wedge \\
 & & & \mathsf{SO0000418}(z) \wedge \mathsf{onlyintaxon}(z,w) \wedge \mathsf{OBI0100026}(w) \\
\mathsf{q}_7: & Q(x) & \leftarrow & \mathsf{haspart}(x,y) \wedge \mathsf{CHEBI23367}(y) \wedge \mathsf{lackspart}(x,z) \wedge \\
 & & & \mathsf{SO0000418}(z) \wedge \mathsf{MOD00033}(y)
\end{array}$$

PTO:

$$\begin{array}{rrcl}
\mathsf{q}_1: & Q(x) & \leftarrow & \mathsf{Molecule}(x) \\
\mathsf{q}_2: & Q(x) & \leftarrow & \mathsf{Ion}(x) \\
\mathsf{q}_3: & Q(x) & \leftarrow & \mathsf{Atom}(x) \\
\mathsf{q}_4: & Q(x) & \leftarrow & \mathsf{MoleOfChemical}(x) \wedge \mathsf{hasColor}(x,z) \wedge \mathsf{WhiteColor}(z) \\
\mathsf{q}_5: & Q(x) & \leftarrow & \mathsf{MoleOfChemical}(x) \wedge \mathsf{hasState}(x,y) \wedge \mathsf{SolidState}(y) \\
\mathsf{q}_6: & Q(x) & \leftarrow & \mathsf{AmountOfMatter}(x) \wedge \mathsf{hasState}(x,y) \wedge \mathsf{SolidState}(y) \wedge \\
 & & & \mathsf{hasColour}(x,z) \wedge \mathsf{PaleYellowColour}(z) \\
\mathsf{q}_7: & Q(x) & \leftarrow & \mathsf{hasPart}(x,y) \wedge \mathsf{DiscretePhysicalObject}(y) \wedge \mathsf{hasCharge}(y,z) \wedge \\
 & & & \mathsf{NegativeCharge}(z) \wedge \mathsf{ionisedFrom}(x,w)
\end{array}$$

NCI:

$$
\begin{aligned}
\mathsf{q}_1: \quad Q(x) &\leftarrow \mathsf{Techniques}(x) \wedge \mathsf{rTechniqueHasPurpose}(x,y) \wedge \mathsf{Detection}(y)\\
\mathsf{q}_2: \quad Q(x) &\leftarrow \mathsf{Gene}(x)\\
\mathsf{q}_3: \quad Q(x) &\leftarrow \mathsf{Gene}(x) \wedge \mathsf{rGeneAssociatedWithDisease}(x,y) \wedge \mathsf{Carcinoma}(y)\\
\mathsf{q}_4: \quad Q(x) &\leftarrow \mathsf{Gene}(x) \wedge \mathsf{rGenePlaysRoleinProcess}(x,y) \wedge \mathsf{Pathogenesis}(y)\\
\mathsf{q}_5: \quad Q(x) &\leftarrow \mathsf{ImmunoproteinGene}(x) \wedge \mathsf{rGenePlaysRoleinProcess}(x,y) \wedge\\
&\qquad \mathsf{ImmuneResponse}(y) \wedge \mathsf{rGeneisBiomarkerof}(x,z) \wedge \mathsf{BreastCarcinoma}(z)
\end{aligned}
$$

Α′.2 Ερωτήματα Αξιολόγησης Κεφαλαίου 6

$$
\begin{aligned}
\mathsf{q}_1: \quad Q(x,y) &\leftarrow \mathsf{Person}(x) \wedge \mathsf{takesCourse}(x,y)\\
\mathsf{q}_2: \quad Q(x,y) &\leftarrow \mathsf{Employee}(x) \wedge \mathsf{publicationAuthor}(y,x)\\
\mathsf{q}_4: \quad Q(x,y,z,w,u,v) &\leftarrow \mathsf{FullProfessor}(x) \wedge \mathsf{publicationAuthor}(y,x) \wedge\\
&\qquad \mathsf{teacherOf}(x,z) \wedge \mathsf{advisor}(w,x) \wedge \mathsf{GraduateStudent}(w) \wedge\\
&\qquad \mathsf{degreeFrom}(x,u) \wedge \mathsf{degreeFrom}(w,v)\\
\mathsf{g}_2: \quad Q(x) &\leftarrow \mathsf{Organization}(x)\\
\mathsf{g}_3: \quad Q(x) &\leftarrow \mathsf{Employee}(x)\\
\mathsf{req}_2: \quad Q(x,y) &\leftarrow \mathsf{Person}(x) \wedge \mathsf{teacherOf}(x,y) \wedge \mathsf{Course}(y)\\
\mathsf{req}_3: \quad Q(x,y,z) &\leftarrow \mathsf{Student}(x) \wedge \mathsf{advisor}(x,y) \wedge \mathsf{Faculty}(y) \wedge\\
&\qquad \mathsf{takesCourse}(x,z) \wedge \mathsf{teacherOf}(y,z) \wedge \mathsf{Course}(z)\\
\mathsf{lutz}_1: \quad Q(x,y) &\leftarrow \mathsf{Student}(x) \wedge \mathsf{takesCourse}(x,y) \wedge \mathsf{Course}(y) \wedge\\
&\qquad \mathsf{teacherOf}(z,y) \wedge \mathsf{Faculty}(z) \wedge \mathsf{worksFor}(z,w) \wedge\\
&\qquad \mathsf{Department}(w) \wedge \mathsf{memberOf}(x,w)\\
\mathsf{lutz}_5: \quad Q(x) &\leftarrow \mathsf{Publication}(x) \wedge \mathsf{publicationAuthor}(x,y) \wedge\\
&\qquad \mathsf{Professor}(y) \wedge \mathsf{publicationAuthor}(x,z) \wedge \mathsf{Student}(z)
\end{aligned}
$$

Παράρτημα Β′

Πειραματικά Αποτελέσματα

Β′.1 Πειραματικά Αποτελέσματα Κεφαλαίου 4

Πίνακας Β′.1: *Μεγέθη datalog επαναγραφών όπως υπολογίστηκαν από το Rapid*

	Q_1	Q_2	Q_3	Q_4	Q_5
UX	5	1	12	5	25
P5X	23	31	34	36	37
AX	74	53	57	75	82
SWEET$^{\mathcal{L}}$	12	10	2	372	382
SWEET$^{\mathcal{E}}$	182	192	356	812	822

Πίνακας Β′.2: *Μεγέθη datalog επαναγραφών όπως υπολογίστηκαν από το Rapid*

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
PTO$^{\mathcal{L}}$	22	315	195	1034	1168	1171	616
PTO$^{\mathcal{E}}$	29	1356	33919	34879	27907	2458	1364
PRO$^{\mathcal{L}}$	1103	879	1653	1609	1743	1746	1739
PRO$^{\mathcal{E}}$	51641	52877	51614	52407	79427	53983	52881

Πίνακας Β′.3: *Αποτελέσματα* $\mathsf{Add}_{\mathsf{Req}}$, *Requiem (σε msec) για τον υπολογισμό UCQ επαναγραφής*

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_1	Q_2	Q_3	Q_4	Q_5
	P5					**P5X**				
0.5%	0	0	2	50	1058	0	2	15	344	18355
1.0%	0	0	3	52	1484	0	2	7	47	2614
1.5%	0	0	4	76	2889	0	2	8	321	12299
2.0%	0	0	0	43	1399	0	1	16	273	8608
2.5%	0	0	5	89	350	0	2	7	46	12195
3.0%	0	0	2	15	568	0	1	21	185	5998
3.5%	0	0	3	48	349	0	2	8	302	15662
4.0%	0	0	1	93	1028	1	1	20	254	11021
4.5%	0	0	2	14	454	0	1	6	0	2846
5.0%	0	0	3	45	1940	0	0	2	100	13059
Requiem (100%)	10	20	51	472	12300	10	44	314	5360	176490
	S					**V**				
0.5%	0	1	0	0	2	0	0	0	5	0
1.0%	0	1	0	0	0	0	0	0	7	0
1.5%	0	1	0	0	0	0	0	0	10	0
2.0%	0	1	0	0	0	0	0	2	10	0
2.5%	0	1	0	0	1	0	0	1	10	0
3.0%	0	1	1	1	0	0	0	2	23	1
3.5%	0	1	0	0	1	0	0	3	21	0
4.0%	0	0	0	0	2	0	0	2	27	4
4.5%	0	2	1	1	3	0	0	3	22	0
5.0%	0	1	1	1	2	0	0	3	27	1
Requiem (100%)	10	51	515	895	15256	0	7	30	145	22
	A					**AX**				
0.5%	3	4	12	8	117	6	121	1425	1574	-
1.0%	3	8	13	14	118	3	128	290	855	-
1.5%	5	9	19	21	106	9	417	2215	1233	-
2.0%	4	5	18	31	113	9	577	2589	833	-
2.5%	5	8	16	23	102	9	394	5131	2180	-
3.0%	6	8	17	23	173	9	286	4146	3490	-
3.5%	5	8	19	24	120	8	522	3896	1331	-
4.0%	5	8	20	36	199	10	661	3042	3310	-
4.5%	5	8	20	28	168	12	677	2811	2218	-
5.0%	6	7	19	33	156	11	509	4437	3037	-
Requiem (100%)	69	57	78	124	438	74	2062	20282	15325	t.o.

Πίνακας Β′.4: *Αποτελέσματα* $\mathsf{Add}_{\mathsf{Req}}$, *Requiem (σε msec) για τον υπολογισμό UCQ επαναγραφής*

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_1	Q_2	Q_3	Q_4	Q_5
			U					**UX**		
0.5%	0	1	0	2	0	1	3	0	1	0
1.0%	0	1	0	0	0	0	1	0	2	0
1.5%	0	1	0	2	2	0	2	1	1	3
2.0%	0	1	0	1	1	0	0	0	2	0
2.5%	0	2	0	2	2	0	2	0	4	0
3.0%	0	0	0	2	2	0	2	0	2	0
3.5%	0	1	0	3	2	0	2	0	2	6
4.0%	0	1	0	2	2	0	2	1	3	0
4.5%	0	1	0	1	2	0	2	0	8	0
5.0%	0	1	0	2	2	0	1	0	10	3
Requiem (100%)	0	61	118	1467	5232	22	106	943	13123	41435
			SWEET$^{\mathcal{L}}$					**PTO**$^{\mathcal{L}}$		
0.5%	0	0	0	170	20179	14	25	16	47	1990
1.0%	0	0	0	169	12543	18	37	19	95	2483
1.5%	0	0	0	166	15231	20	59	23	64	2424
2.0%	1	0	0	179	21099	116	54	39	87	3743
2.5%	0	0	0	194	33477	30	93	43	104	5298
3.0%	3	2	2	194	27709	84	63	37	90	10906
3.5%	0	1	2	198	28216	54	100	52	128	7161
4.0%	2	1	1	178	24800	104	104	60	149	5733
4.5%	3	1	0	185	29891	133	134	59	276	6975
5.0%	1	0	2	184	27332	132	86	61	130	12986
Requiem (100%)	22	17	14	343	123572	562	711	435	1029	51656
			PRO$^{\mathcal{L}}$							
0.5%	199	1217	414212	435178						
1.0%	286	1848	332112	427479						
1.5%	421	2546	325722	360840						
2.0%	593	3357	334979	355799						
2.5%	657	3797	393472	475168						
3.0%	755	4590	368566	421230						
3.5%	1024	5568	540383	549273						
4.0%	917	6257	596061	625165						
4.5%	1108	6706	647301	727802						
5.0%	1212	7388	847767	987835						
Requiem (100%)	12175	71456	1142559	1178659						

Πίνακας Β′.5: *Αποτελέσματα* Add$_{\text{Req}}$, *Requiem (σε msec) για τον υπολογισμό datalog επαναγραφής*

	Q_1	Q_2	Q_3	Q_4	Q_5	Q_1	Q_2	Q_3	Q_4	Q_5
	SWEET$^{\mathcal{E}}$					**PTO$^{\mathcal{E}}$**				
0.5%	16	14	36	74	82	20470	17392	24329	12579	13985
1.0%	48	14	62	241	167	40112	37330	29814	54034	173321
1.5%	36	28	126	508	220	46736	50409	145210	142882	146132
2.0%	50	32	110	747	843	63612	126662	118069	78644	157664
2.5%	112	58	202	656	1015	149617	165681	89864	107146	416131
3.0%	88	80	143	788	479	110131	116209	115506	104115	220839
3.5%	102	80	178	794	1413	114490	137445	208385	209683	173227
4.0%	114	46	160	894	937	152805	163930	322345	217995	203729
4.5%	136	84	182	1190	1395	222355	196104	336689	249451	723430
5.0%	116	106	264	1092	1411	194479	196604	286359	201245	280624
Requiem(100%)	1160	990	1360	3500	3570	291214	781780	782106	850778	1081858

Πίνακας Β′.6: *Αποτελέσματα* $\mathsf{Add}_{\mathsf{Rap}}$, *Rapid (σε msec) για τον υπολογισμό datalog επαναγραφής*

	DL-Lite							$\mathcal{ELHI}$						
	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
	SWEET							**SWEET**						
0.5%	0	1	1	12	7	-	-	4	4	9	10	32	-	-
1.0%	0	0	0	5	11	-	-	5	3	9	24	16	-	-
1.5%	0	1	0	5	10	-	-	5	3	6	21	20	-	-
2.0%	1	1	0	6	6	-	-	6	2	7	40	30	-	-
2.5%	0	1	0	4	6	-	-	3	2	7	21	34	-	-
3.0%	1	0	0	6	10	-	-	3	2	8	31	20	-	-
3.5%	0	0	1	5	10	-	-	4	3	8	24	10	-	-
4.0%	0	0	0	5	10	-	-	4	3	7	25	17	-	-
4.5%	0	0	0	4	2	-	-	4	5	6	30	30	-	-
5.0%	0	0	0	6	6	-	-	4	3	6	39	21	-	-
Rapid(100%)	2	2	1	22	18	t.o.	t.o.	14	14	21	49	84	t.o.	t.o.
	PRO							**PRO**						
0.5%	2	15	431	418	8345	70	528	1922	1358	1593	1398	2712	2135	1732
1.0%	5	21	504	436	11109	76	531	2312	1448	1621	1565	2610	2106	1757
1.5%	4	19	541	467	28712	68	538	1923	1502	1630	1577	2740	2156	1813
2.0%	1	18	537	474	35635	70	543	2728	1529	1828	1629	5682	2146	1821
2.5%	3	21	581	513	48475	70	548	1963	1595	1756	1674	4863	2731	1932
3.0%	1	27	654	546	52287	72	544	2307	1567	1895	1731	6452	2545	2028
3.5%	0	44	611	556	110730	76	563	2021	1644	1909	1695	7227	2248	2113
4.0%	4	39	661	600	93132	76	566	2789	1688	1926	1815	8170	2752	2176
4.5%	1	29	749	624	110064	82	584	3634	1782	2017	1859	30175	2370	3014
5.0%	3	40	671	655	99763	92	589	3029	1783	1965	1880	94525	2009	3121
Rapid(100%)	28	561	3768	4254	364155	210	1307	5741	6103	6416	6436	373033	9450	6664
	PTO							**PTO**						
0.5%	1	8	1	12	15	16		40	25	59	52	71	52	57
1.0%	0	7	0	10	15	16		35	31	61	54	73	52	52
1.5%	0	8	1	15	18	16		38	33	66	58	79	50	53
2.0%	1	8	2	16	23	13		42	35	77	61	86	55	58
2.5%	0	7	3	18	22	20		43	36	73	62	90	52	58
3.0%	0	12	6	17	26	18		50	40	89	67	101	52	58
3.5%	1	10	2	25	28	19		53	42	83	70	99	52	61
4.0%	1	6	4	24	27	20		59	42	88	73	111	57	59
4.5%	1	6	3	27	32	17		56	48	95	74	108	61	66
5.0%	0	6	4	26	37	20		57	42	92	83	113	61	64
Rapid(100%)	9	11	7	23	45	122		63	50	110	96	122	297	153

Πίνακας Β′.7: *Αποτελέσματα* $\mathsf{Delete}_{\mathsf{Req}}$, *Requiem (σε msec) για τον υπολογισμό datalog επαναγραφής*

DL-Lite$_{\mathcal{R}}$

	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}
	Q_1		Q_2		Q_3		Q_4		Q_5	
P5										
0.5%	0	0	10	0	10	0	51	10	642	30
1.0%	0	0	10	0	10	0	180	10	3573	132
1.5%		0	0	0	0	0	60	0	3500	101
2.0%	0	0	0	0	0	0	50	0	3762	112
2.5%	10	0	0	10	10	0	190	10	3881	101
3.0%	0	0	0	0	10	10	50	0	650	15
3.5%	0	0	0	0	0	0	190	0	606	21
4.0%	0	0	1	0	10	0	62	0	690	30
4.5%	0	0	0	0	20	0	50	20	3743	111
5.0%	0	0	0	1	20	0	50	10	650	30
P5X										
0.5%	0	10	0	0	10	0	51	10	3513	91
1.0%	0	0	0	0	10	0	60	10	3459	91
1.5%	0	0	0	0	20	0	50	10	651	20
2.0%	0	0	10	0	10	0	50	10	3520	101
2.5%	0	0	0	0	20	0	60	0	631	20
3.0%	0	0	0	0	10	0	190	10	651	21
3.5%	0	0	0	0	20	0	180	10	600	20
4.0%	0	0	0	0	10	0	199	10	620	21
4.5%	0	0	0	0	20	0	50	0	672	20
5.0%	10	0	0	0	20	0	70	0	3742	91
A										
0.5%	10	0	20	0	10	0	30	0	20	0
1.0%	10	0	10	10	10	0	10	10	10	0
1.5%	10	0	10	0	10	0	10	0	20	0
2.0%	20	0	0	0	10	0	10	0	10	0
2.5%	10	0	0	0	0	0	10	0	10	0
3.0%	10	0	10	0	0	0	10	0	10	0
3.5%	10	0	0	0	10	0	0	0	10	0
4.0%	10	0	0	0	10	0	10	0	20	0
4.5%	0	0	10	0	0	0	0	0	10	0
5.0%	0	0	0	0	0	0	10	0	10	0
AX										
0.5%	10	10	20	0	20	0	20	10	20	0
1.0%	10	0	10	0	1	0	10	0	10	10
1.5%	10	0	20	0	26	0	10	0	10	0
2.0%	10	0	10	0	10	0	10	0	10	0
2.5%	10	0	10	0	10	0	10	0	10	0
3.0%	0	0	0	0	10	0	10	0	10	0
3.5%	0	0	0	0	10	0	10	0	30	0
4.0%	10	0	10	0	10	0	10	0	10	0
4.5%	10	0	10	0	10	0	10	0	10	0
5.0%	10	0	0	0	10	0	10	0	20	0

Πίνακας Β′.8: *Αποτελέσματα* $\mathsf{Delete}_{\mathsf{Req}}$, *Requiem (σε msec) για τον υπολογισμό datalog επαναγραφής*

Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}		
Q_1		Q_2		Q_3		Q_4		Q_5		Q_6	
SWEET$^{\mathcal{L}}$											
1	0	0	0	2	0	23	3	23	3		
1	0	1	0	1	0	18	3	20	4		
1	0	1	0	1	0	18	3	19	3		
0	0	0	1	1	0	16	3	22	3		
1	0	0	0	1	0	8	2	17	3		
0	0	0	0	0	0	15	2	15	3		
1	0	0	0	1	0	9	2	14	2		
1	0	1	0	1	0	12	3	1	0		
0	0	1	0	0	0	14	3	14	3		
0	0	1	0	1	0	7	1	13	2		
PTO$^{\mathcal{L}}$											
440	5	693	13	414	8	731	10	18221	100	2935134	85571
427	4	636	24	410	0	722	10	17914	100	t.o.	-
424	3	623	11	383	10	715	10	17221	100	t.o.	-
421	2	628	10	414	10	675	10	18302	90	t.o.	-
442	2	650	10	373	0	745	10	16295	90	t.o.	-
420	3	657	10	373	0	705	10	13548	80	t.o.	-
412	2	620	10	373	3	705	10	15730	80	t.o.	-
411	3	617	10	383	0	705	10	11649	70	t.o.	-
409	2	613	9	383	0	705	10	15887	90	t.o.	-
409	2	614	10	363	0	695	20	10924	80	t.o.	-
PRO$^{\mathcal{L}}$											
12149	30	91986	270	814688	174775	839047	171913	t.o.	-	t.o.	-
12185	30	92652	250	810616	174308	826422	170041	t.o.	-	t.o.	-
12069	30	90480	220	659771	172596	824544	167289	t.o.	-	t.o.	-
12069	30	90840	220	797398	171159	816812	165871	t.o.	-	t.o.	-
12101	30	90306	220	783928	166038	807072	163300	t.o.	-	t.o.	-
12133	30	88790	214	771638	163431	800553	161344	t.o.	-	t.o.	-
12068	30	88329	220	757310	161587	793537	156123	t.o.	-	t.o.	-
12035	30	87803	210	744442	155839	784304	155215	t.o.	-	t.o.	-
12061	30	87387	227	743335	154551	778846	152589	t.o.	-	t.o.	-
12019	20	85812	210	736605	150163	777029	150287	t.o.	-	t.o.	-
SWEET$^{\mathcal{E}}$											
1020	20	927	7	1220	17	2576	73	2468	87		
1060	10	923	12	1181	20	2291	50	1038	44		
1040	10	974	12	1228	18	1193	37	2165	43		
1100	10	901	9	1188	18	2555	81	2043	47		
1140	30	852	5	1213	16	999	30	2086	46		
1110	0	928	9	1159	16	2463	52	2677	53		
1059	10	972	13	1169	16	2945	59	2238	57		
1040	10	938	17	1200	16	2879	60	2553	60		
1010	20	926	9	1179	20	1145	36	2077	61		
PTO$^{\mathcal{E}}$											
537973	3045	493852	854	526849	1559	1516651	1736	1495290	1801	1368474	3759
492413	1930	571771	1540	510428	1736	1090124	1541	1473445	1703	1815393	3757
540316	1325	619737	2042	404889	1541	1081453	1601	1255766	1740	1372746	2944
378637	1256	391288	1433	457399	1601	1014268	1551	1281720	1736	1623899	3751
531932	1830	373997	2007	501167	1551	1047134	1541	1354339	1640	1677866	3808
506860	1541	577360	2423	398463	1541	943522	1301	1289245	1705	1685082	3482
412356	1642	526917	2220	489236	1301	863402	1601	1352420	1431	1526158	4206
484171	1962	505700	2120	395185	1601	987207	1641	1047415	1642	1755134	2682
376558	1957	580765	2122	483367	1641	932933	1591	1222630	1768	1883275	3847

Πίνακας Β′.9: *Αποτελέσματα* Delete, *Requiem (σε msec) για τον υπολογισμό UCQ επαναγραφής*

DL-Lite$_{\mathcal{R}}$

	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$
	Q_1		Q_2		Q_3		Q_4		Q_5	
	SWEET$^{\mathcal{L}}$									
0.5%	10	0	10	2	12	0	223	2	122515	188
1.0%	12	2	0	2	10	0	214	4	120422	174
1.5%	20	0	2	0	10	0	200	8	117436	169
2.0%	14	0	0	0	10	0	194	0	112044	188
2.5%	10	0	2	0	12	0	182	2	113140	162
3.0%	12	0	4	0	10	0	216	4	86724	132
3.5%	14	0	0	0	10	0	206	4	71527	105
4.0%	12	0	0	0	10	2	210	6	91486	138
4.5%	10	0	2	0	10	0	186	4	101588	147
5.0%	10	0	4	0	10	0	192	0	80367	128
	PTO$^{\mathcal{L}}$									
0.5%	412	0	699	0	439	0	1101	2	t.o.	-
1.0%	416	0	692	1	435	0	1046	2	t.o.	-
1.5%	410	0	691	1	436	0	1022	3	t.o.	-
2.0%	407	0	689	1	432	0	1019	3	t.o.	-
2.5%	407	0	685	1	428	0	1022	3	t.o.	-
3.0%	405	0	681	1	430	0	990	4	t.o.	-
3.5%	385	0	677	2	432	1	1019	4	t.o.	-
4.0%	403	0	689	1	427	1	994	5	t.o.	-
4.5%	382	0	678	2	428	1	928	5	t.o.	-
5.0%	380	0	645	2	420	1	945	6	t.o.	-
	PRO$^{\mathcal{L}}$									
0.5%	12019	0	66097	10	t.o.	t.o.	t.o.	t.o.	t.o.	-
1.0%	12002	0	66669	16	t.o.	t.o.	t.o.	t.o.	t.o.	-
1.5%	11984	0	64506	23	t.o.	t.o.	t.o.	t.o.	t.o.	-
2.0%	11973	0	64316	30	t.o.	t.o.	t.o.	t.o.	t.o.	-
2.5%	11919	0	62535	37	t.o.	t.o.	t.o.	t.o.	t.o.	-
3.0%	11942	0	62675	45	t.o.	t.o.	t.o.	t.o.	t.o.	-
3.5%	11936	0	61287	54	t.o.	t.o.	t.o.	t.o.	t.o.	-
4.0%	11919	0	61473	64	t.o.	t.o.	t.o.	t.o.	t.o.	-
4.5%	11909	0	60787	75	t.o.	t.o.	t.o.	t.o.	t.o.	-
5.0%	11887	0	60260	88	t.o.	t.o.	t.o.	t.o.	t.o.	-

Πίνακας Β′.10: *Αποτελέσματα* Delete*, Requiem (σε msec) για τον υπολογισμό UCQ επαναγραφής*

DL-Lite$_{\mathcal{R}}$

	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$
	Q_1		Q_2		Q_3		Q_4		Q_5	
	A									
0.5%	21	8	16	0	46	0	119	7	595	1
1.0%	18	8	16	0	45	0	111	6	689	1
1.5%	18	8	16	0	50	0	118	5	648	1
2.0%	24	8	13	0	49	1	108	4	653	1
2.5%	13	6	14	0	43	0	112	5	680	2
3.0%	19	8	14	0	50	0	78	4	574	2
3.5%	17	8	14	0	50	0	96	4	499	2
4.0%	18	6	14	0	44	0	93	5	620	2
4.5%	20	6	15	0	44	0	92	4	503	2
5.0%	19	8	11	0	38	0	64	3	260	1
	AX									
0.5%	48	2	1597	1	17240	41	19175	17	t.o.	-
1.0%	44	2	1922	1	20742	41	20590	17	t.o.	-
1.5%	45	1	1873	0	14671	41	19178	17	t.o.	-
2.0%	47	2	1842	1	16328	42	17237	15	t.o.	-
2.5%	34	2	1985	31	18619	41	18184	17	t.o.	-
3.0%	33	2	1531	10	15041	40	12933	14	t.o.	-
3.5%	28	2	1423	0	16811	41	17160	17	t.o.	-
4.0%	58	2	1683	0	19793	41	15802	17	t.o.	-
4.5%	31	2	1883	0	10046	21	18113	18	t.o.	-
5.0%	47	2	1749	1	16501	31	18291	21	t.o.	-
	P5									
0.5%	1	0	2	0	15	0	306	0	5894	0
1.0%	1	0	3	0	14	0	185	0	6191	0
1.5%	0	0	2	0	17	0	310	0	6352	0
2.0%	1	0	2	0	17	0	265	0	9914	0
2.5%	0	0	2	0	18	0	303	0	2175	0
3.0%	0	0	2	0	12	0	249	0	6016	0
3.5%	0	0	1	0	13	0	324	0	2188	0
4.0%	1	0	1	0	10	0	199	0	5941	0
4.5%	0	0	1	0	17	0	301	0	9898	0
5.0%	0	0	1	0	15	0	308	0	6066	0
	P5X									
0.5%	2	0	10	0	118	1	3873	7	71457	37
1.0%	1	0	8	0	134	1	3176	6	117219	45
1.5%	2	0	7	0	143	1	4428	6	88049	39
2.0%	1	0	7	0	159	1	2894	5	84950	37
2.5%	1	0	8	0	161	1	4655	6	140693	48
3.0%	1	0	8	0	155	1	2944	6	119734	43
3.5%	1	0	8	0	145	1	3922	6	129999	45
4.0%	1	0	7	0	166	1	3220	5	136500	47
4.5%	1	0	8	0	115	1	3779	6	90984	38
5.0%	0	0	7	0	136	1	3709	6	137324	46

Πίνακας Β′.11: *Αποτελέσματα* Delete, *Requiem (σε msec) για τον υπολογισμό UCQ επαναγραφής*

DL-Lite$_{\mathcal{R}}$

	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$	Requiem	Delete$_{Req}$
	Q_1		Q_2		Q_3		Q_4		Q_5	
	U									
0.5%	1	0	31	0	66	0	1376	5	5404	12
1.0%	0	0	30	0	59	0	1507	4	4387	10
1.5%	1	0	30	0	57	0	1329	4	4003	9
2.0%	1	0	29	0	50	0	1286	3	3918	9
2.5%	0	0	30	1	61	0	1466	3	4207	8
3.0%	1	0	30	0	63	0	1335	4	4913	9
3.5%	1	0	29	0	61	0	1114	4	5090	9
4.0%	0	0	31	0	52	0	1336	4	4367	9
4.5%	1	0	26	0	66	0	1351	4	4908	10
5.0%	0	0	29	0	61	0	858	4	5186	10
	UX									
0.5%	2	0	42	1	593	2	10553	14	33392	30
1.0%	1	0	43	0	777	1	13009	14	38246	28
1.5%	1	0	42	0	585	1	12926	14	31967	25
2.0%	0	0	39	0	673	1	11111	13	36013	28
2.5%	1	0	44	0	616	1	11924	13	30660	26
3.0%	1	0	41	0	713	1	8658	12	33587	28
3.5%	0	0	36	0	588	1	11754	14	31523	27
4.0%	1	0	40	0	701	1	10426	13	30548	30
4.5%	2	0	35	0	548	1	10849	15	17139	23
5.0%	1	0	34	0	601	1	11578	16	33836	31
	S									
0.5%	1	0	27	3	229	2	727	4	13160	27
1.0%	1	0	26	2	323	3	841	3	12855	25
1.5%	1	0	26	3	387	2	800	3	12909	22
2.0%	1	0	26	2	328	2	835	3	12057	21
2.5%	1	0	26	2	314	3	755	3	15578	26
3.0%	1	0	30	2	326	2	697	2	14977	24
3.5%	1	0	20	1	340	3	711	3	12868	21
4.0%	1	0	27	2	277	2	840	3	15161	27
4.5%	0	0	25	2	353	2	725	3	15176	25
5.0%	0	0	24	2	357	2	563	2	14499	25
	V									
0.5%	1	0	1	0	10	0	50	0	8	0
1.0%	2	0	1	0	10	0	46	0	9	0
1.5%	0	0	1	0	12	0	51	0	8	0
2.0%	1	0	1	0	10	0	40	0	8	0
2.5%	1	0	1	0	9	0	50	0	11	0
3.0%	0	0	1	0	8	0	48	0	9	0
3.5%	0	0	1	0	8	0	45	0	9	0
4.0%	0	0	1	0	9	0	49	0	7	0
4.5%	1	0	0	0	9	0	51	0	8	0
5.0%	1	0	1	0	9	0	46	0	7	0

Πίνακας Β′.12: *Αποτελέσματα* Delete, *Requiem (σε msec) για τον υπολογισμό datalog επαναγραφής*

	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}	Requiem	Delete_{Req}
	Q_1		Q_2		Q_3		Q_4		Q_5	
	SWEET$^{\mathcal{E}}$									
0.5%	1129	6	968	0	1286	8	3236	19	3382	23
1.0%	1106	7	956	0	1286	16	2941	18	3106	24
1.5%	1096	6	954	4	1241	6	2472	15	3140	22
2.0%	1106	6	940	8	1189	4	2756	17	3158	30
2.5%	1074	6	941	4	1205	14	2705	22	2873	20
3.0%	1102	4	912	0	1194	4	2590	18	2701	16
3.5%	1057	4	916	0	1202	10	2196	14	2804	24
4.0%	1030	2	902	2	1124	8	2319	17	2470	20
4.5%	1037	6	935	2	1130	8	2344	16	2420	25
5.0%	1038	5	932	4	1096	6	2299	19	2383	12

B′.2 Πειραματικά Αποτελέσματα Κεφαλαίου 6

Πίνακας B′.13: *Αποτελέσματα των συστημάτων* HdIC, HdIC_{op} *και CQApri* (CQA) *για τα ABox* $\mathcal{A}_i^1$.

		Αριθμός Απαντήσεων			**Χρόνοι Φόρτωσης, Προεπεξεργασίας και Αποτίμησης (σε sec)**								
					ICAR Απάντηση Ερωτ.			*IAR Απάντηση Ερωτ.*					
								Διόρθωση/Προεπεξ.			Αποτίμηση		
$\mathcal{A}$	$\mathcal{Q}$	HdIC	HdIA	QuID	Load	HdIC	HdIC_{op}	HdIA	QuID	CQA	HdIA	QuID	CQA
	q_1	20386	20119	20269		4.2	0.7				<0.1	5.0	1.9
	q_2	7214	7200	7219		1.5	0.1				<0.1	3.2	1.1
	q_4	82273	79907	80537		73.7	6.6				0.1	1.8	2.3
	g_2	1195	1184	1186		0.1	<0.1				<0.1	0.9	0.5
$\mathcal{A}^1_{15e\text{-}4}$	g_3	1069	1066	1067	1.8	0.1	<0.1	8.6	133.3	8.2	<0.1	0.2	0.4
	req_2	1519	1519	1533		0.6	0.1				<0.1	0.1	0.3
	req_3	191	191	194		0.4	0.1				<0.1	0.4	0.4
	$lutz_1$	15081	15081	15175		14.8	1.4				0.1	43.7	22.4
	$lutz_5$	3136	3136	3147		2.9	0.2				<0.1	5.1	3.6
	q_1	19919	11430	14363		4.0	0.7				<0.1	0.5	1.4
	q_2	6624	5851	6862		1.6	0.2				<0.1	2.4	1.1
	q_4	37824	13520	16986		38.0	3.5				<0.1	1.1	3.0
	g_2	1252	946	989		0.1	<0.1				<0.1	0.8	0.5
$\mathcal{A}^1_{5e\text{-}2}$	g_3	1067	908	995	1.9	0.1	<0.1	8.9	143.5	9.1	<0.1	0.2	2.0
	req_2	883	883	1126		0.5	0.1				<0.1	0.1	0.3
	req_3	96	71	94		0.3	<0.1				<0.1	0.4	0.5
	$lutz_1$	495	495	754		8.6	0.8				<0.1	37.4	22.7
	$lutz_5$	2010	2010	5924		2.1	0.2				<0.1	7.9	3.6
	q_1	18660	2894	5380		4.2	0.5				<0.1	0.7	1.4
	q_2	4605	2826	5999		1.2	0.2				<0.1	2.0	1.2
	q_4	8772	292	670		12.7	1.2				<0.1	1.0	3.1
	g_2	1420	546	635		0.1	<0.1				<0.1	1.2	0.5
$\mathcal{A}^1_{2e\text{-}1}$	g_3	984	534	827	1.9	0.1	<0.1	8.1	192.4	10.6	<0.1	0.4	0.4
	req_2	243	243	554		0.3	0.1				<0.1	0.1	0.3
	req_3	14	4	12		0.3	<0.1				<0.1	0.2	0.5
	$lutz_1$	0	0	0		5.2	0.5				<0.1	24.1	24.9
	$lutz_5$	554	543	5883		1.9	0.1				<0.1	7.8	4.2

Πίνακας Β′.14: *Αποτελέσματα των συστημάτων* HdIC, HdIC_{op} *και CQApri* (CQA) *για τα ABox* $\mathcal{A}_i^5$.

		Αριθμός Απαντήσεων			Χρόνοι Φόρτωσης, Προεπεξεργασίας και Αποτίμησης (σε sec)								
					ICAR Απάντηση Ερωτ.			*IAR Απάντηση Ερωτ.*					
								Διόρθωση/Προεπεξ.			Αποτίμηση		
$\mathcal{A}$	$\mathcal{Q}$	HdIC	HdIA	QuID	Load	HdIC	HdIC_{op}	HdIA	QuID	CQA	HdIA	QuID	CQA
	q_1	135746	133525	134487		24.6	4.5				0.2	12.5	5.4
	q_2	47280	47204	47317		11.0	0.8				<0.1	8.8	4.3
	q_4	521096	460955	465330		542.0	48.0				0.6	8.8	10.7
	g_2	2597	2531	2541		0.3	<0.1				<0.1	1.9	0.9
$\mathcal{A}^5_{15e\text{-}4}$	g_3	7363	7341	7349	8.5	0.5	0.1	29.7	669.1	23.7	<0.1	0.6	0.7
	req_2	10082	10082	10146		3.6	0.3				<0.1	0.1	0.4
	req_3	1173	1166	1174		1.2	0.2				0.1	1.7	1.1
	lutz_1	105389	105389	106340		105.8	9.9				0.5	213.9	42.3
	lutz_5	20333	20333	38257		16.5	1.2				0.1	76.5	22.3
	q_1	132704	77289	95109		26.7	2.5				0.1	11.3	7.3
	q_2	43385	39174	45472		13.3	1.0				<0.1	7.3	5.0
	q_4	289817	7844	11805		332.5	34.7				0.1	1.1	13.6
	g_2	2951	1643	1825		0.3	<0.1				<0.1	0.6	0.9
$\mathcal{A}^5_{5e\text{-}2}$	g_3	7260	6340	6896	8.8	0.5	0.1	28.7	824.3	26.2	<0.1	0.2	0.8
	req_2	6014	6014	7556		3.0	0.3				<0.1	0.1	0.4
	req_3	635	506	663		1.1	0.2				<0.1	1.3	1.1
	lutz_1	2104	2104	2816		59.5	5.4				0.2	140.3	57.7
	lutz_5	13193	13191	38173		14.5	1.1				0.1	67.5	25.6
	q_1	124131	18780	35473		29.3	2.7				0.1	8.1	8.5
	q_2	33280	19883	40348		10.4	0.7				<0.1	5.1	6.5
	q_4	44493	0	0		75.6	7.7				<0.1	0.7	14.7
	g_2	4499	1189	1612		0.5	<0.1				<0.1	0.6	1.3
$\mathcal{A}^5_{2e\text{-}1}$	g_3	7096	3895	5816	9.8	0.6	0.1	29.8	1280.2	36.2	<0.1	0.2	0.9
	req_2	1592	1592	3385		2.3	0.4				<0.1	0.1	0.7
	req_3	106	55	123		0.8	0.1				<0.1	0.8	1.4
	lutz_1	0	0	0		35.0	3.4				0.1	83.4	72.6
	lutz_5	3733	3656	37938		12.3	0.9				0.1	57.1	29.5

Πίνακας Β′.15: *Αποτελέσματα των συστημάτων* HdIC, HdIC$_{op}$ *και CQApri* (CQA) *για τα ABox* $\mathcal{A}_i^{10}$.

		Αριθμός Απαντήσεων			**Χρόνοι Φόρτωσης, Προεπεξεργασίας και Αποτίμησης (σε sec)**								
					ICAR Απάντηση Ερωτ.			*IAR Απάντηση Ερωτ.*					
								Διόρθωση/Προεπεξ.			Αποτίμηση		
$\mathcal{A}$	$\mathcal{Q}$	HdIC	HdIA	QuID	Load	HdIC	HdIC$_{op}$	HdIA	QuID	CQA	HdIA	QuID	CQA
	q$_1$	255839	251991	253600		47.5	4.6				0.2	17.6	11.5
	q$_2$	88994	88816	89110		22.3	1.7				0.1	12.3	6.9
	q$_4$	966856	769786	794576		1060.3	99.4				1.1	14.1	17.9
	g$_2$	4019	3893	3926		0.4	<0.1				<0.1	2.8	1.3
$\mathcal{A}_{15e\text{-}4}^{10}$	g$_3$	13562	13513	13536	19.7	1.0	0.1	52.9	1437.4	58.1	<0.1	2.8	0.9
	req$_2$	19008	19008	19142		6.9	0.6				0.1	0.1	0.5
	req$_3$	2242	2228	2253		2.4	0.3				0.1	2.8	2.1
	lutz$_1$	189519	189519	191141		227.6	21.2				0.9	191.0	61.3
	lutz$_5$	38244	38244	38374		33.6	2.3				0.2	84.0	48.7
	q$_1$	250320	145488	185489		51.9	5.0				0.1	31.0	14.0
	q$_2$	80803	73453	85643		20.2	1.6				0.1	8.9	7.8
	q$_4$	540237	1001	1505		683.1	65.4				0.1	1.1	26.5
	g$_2$	4739	2744	3086		0.6	<0.1				<0.1	3.3	1.5
$\mathcal{A}_{5e\text{-}2}^{10}$	g$_3$	13279	11669	12733	20.2	1.0	0.1	53.8	2072.3	65.6	<0.1	3.3	2.7
	req$_2$	11334	11334	14338		5.9	0.5				0.1	0.1	0.7
	req$_3$	1221	1012	1300		1.8	0.3				0.1	2.3	4.1
	lutz$_1$	5249	5249	7237		118.5	11.7				0.4	98.2	83.0
	lutz$_5$	25168	25163	28325		29.4	2.1				0.2	59.0	49.7
	q$_1$	233153	35086	78806		62.5	10.4				0.1	14.0	17.1
	q$_2$	61171	35762	75896		17.0	1.4				<0.1	6.6	13.7
	q$_4$	88323	0	0		161.2	15.6				<0.1	0.7	29.2
	g$_2$	7528	2269	3040		1.0	0.1				<0.1	3.3	4.9
$\mathcal{A}_{2e\text{-}1}^{10}$	g$_3$	12751	6870	10649	21.3	1.0	0.1	55.2	3205.0	84.6	<0.1	1.7	4.3
	req$_2$	2997	2997	6395		4.3	0.5				<0.1	0.1	0.8
	req$_3$	197	98	248		1.5	0.3				<0.1	1.2	4.1
	lutz$_1$	0	0	0		70.7	7.0				0.1	78.7	132.4
	lutz$_5$	6927	6797	10736		24.2	1.7				0.1	60.8	62.0

Πίνακας Β′.16: *Αποτελέσματα των συστημάτων* HdIC$_{dp}$, HdIA$_{dp}$ *για τα ABox* $\mathcal{A}_i^1$.

$\mathcal{A}$	$\mathcal{Q}$	Preproc.	HdIC$_{dp}$	HdIA$_{dp}$
$\mathcal{A}^1_{15e\text{-}4}$	q$_1$	0,5	0,2	7,0
	q$_2$		0,1	
	q$_4$		3,0	
	g$_2$		<0,1	
	g$_3$		<0,1	
	req$_2$		<0,1	
	req$_3$		<0,1	
	lutz$_1$		0,6	
	lutz$_5$		0,1	
$\mathcal{A}^1_{5e\text{-}2}$	q$_1$	0,5	0,2	7,0
	q$_2$		0,1	
	q$_4$		1,7	
	g$_2$		<0,1	
	g$_3$		<0,1	
	req$_2$		<0,1	
	req$_3$		<0,1	
	lutz$_1$		0,4	
	lutz$_5$		0,1	
$\mathcal{A}^1_{2e\text{-}1}$	q$_1$	0,6	0,2	6,9
	q$_2$		0,1	
	q$_4$		0,7	
	g$_2$		<0,1	
	g$_3$		<0,1	
	req$_2$		<0,1	
	req$_3$		<0,1	
	lutz$_1$		0,2	
	lutz$_5$		0,1	

Πίνακας Β′.17: *Αποτελέσματα των συστημάτων* HdIC$_{dp}$, HdIA$_{dp}$ *για τα ABox* $\mathcal{A}_i^5$.

$\mathcal{A}$	$\mathcal{Q}$	Preproc.	HdIC$_{dp}$	HdIA$_{dp}$
$\mathcal{A}^5_{15e\text{-}4}$	q$_1$	2,2	1,0	24,2
	q$_2$		0,5	
	q$_4$		22,0	
	g$_2$		<0,1	
	g$_3$		<0,1	
	req$_2$		0,2	
	req$_3$		0,1	
	lutz$_1$		5,4	
	lutz$_5$		0,6	
$\mathcal{A}^5_{5e\text{-}2}$	q$_1$	2,3	1,0	23,8
	q$_2$		0,4	
	q$_4$		14,8	
	g$_2$		<0,1	
	g$_3$		<0,1	
	req$_2$		0,2	
	req$_3$		0,1	
	lutz$_1$		2,8	
	lutz$_5$		0,5	
$\mathcal{A}^5_{2e\text{-}1}$	q$_1$	2,5	1,1	24,6
	q$_2$		0,3	
	q$_4$		4,4	
	g$_2$		<0,1	
	g$_3$		0,1	
	req$_2$		0,1	
	req$_3$		0,1	
	lutz$_1$		2,0	
	lutz$_5$		0,4	

Πίνακας Β′.18: *Αποτελέσματα των συστημάτων* $\mathsf{HdlC_{dp}}$, $\mathsf{HdlA_{dp}}$ *για τα ABox* $\mathcal{A}_i^{10}$.

$\mathcal{A}$	$\mathcal{Q}$	Preproc.	$\mathsf{HdlC_{dp}}$	$\mathsf{HdlA_{dp}}$
	$\mathsf{q_1}$		1,7	
	$\mathsf{q_2}$		0,7	
	$\mathsf{q_4}$		38,5	
	$\mathsf{g_2}$		<0,1	
$\mathcal{A}^{10}_{15e\text{-}4}$	$\mathsf{g_3}$	4,5	0,1	44,0
	$\mathsf{req_2}$		0,4	
	$\mathsf{req_3}$		0,2	
	$\mathsf{lutz_1}$		7,7	
	$\mathsf{lutz_5}$		1,2	
	$\mathsf{q_1}$		1,9	
	$\mathsf{q_2}$		0,7	
	$\mathsf{q_4}$		28,8	
	$\mathsf{g_2}$		<0,1	
$\mathcal{A}^{10}_{5e\text{-}2}$	$\mathsf{g_3}$	4,7	0,1	43,2
	$\mathsf{req_2}$		0,3	
	$\mathsf{req_3}$		0,2	
	$\mathsf{lutz_1}$		5,2	
	$\mathsf{lutz_5}$		1,0	
	$\mathsf{q_1}$		2,6	
	$\mathsf{q_2}$		0,6	
	$\mathsf{q_4}$		12,5	
	$\mathsf{g_2}$		0,1	
$\mathcal{A}^{10}_{2e\text{-}1}$	$\mathsf{g_3}$	5,0	0,1	45,1
	$\mathsf{req_2}$		0,3	
	$\mathsf{req_3}$		0,2	
	$\mathsf{lutz_1}$		3,7	
	$\mathsf{lutz_5}$		0,8	

Παράρτημα Γ′

Αποδόσεις Ξένων Όρων

Abox saturation system	:	σύστημα κορεσμού δεδομένων
approximation	:	προσέγγιση
answer variable	:	μεταβλητή απάντησης
atomic	:	ατομικός
assertion	:	ισχυρισμός
backward subsumption checking	:	οπισθοδρομικός έλεγχος υπαγωγής
basic	:	βασικός
binary resolution	:	δυαδική ανάλυση
compliment	:	συμπλήρωμα
concept	:	έννοια
concjunctive query	:	συζευκτικό ερώτημα
constructor	:	κατασκευαστής
clause	:	πρόταση
data integration	:	ολοκλήρωση δεδομένων
database	:	βάση δεδομένων
darivation	:	παραγωγή
forward subsumption checking	:	εμπρόσθιος έλεγχος υπαγωγής
fragment	:	υποσύνολο
robustly decidable	:	εύρωστα αποφασίσιμος
deductive	:	επαγωγικός
description logic	:	περιγραφική λογική
disjunction	:	διάζευξη
distinguished variable	:	διακεκριμένη μεταβλητή
domain	:	χώρος, πεδίο
existential restriction	:	υπαρξιακός περιορισμός
general	:	γενικός
individual	:	άτομο
inclusion axiom	:	αξίωμα υπαγωγής
inference	:	συμπερασμός
instance relation	:	σχέση στιγμιοτύπου
intractable	:	δισεπίλυτο

inverse	:	αντίστροφος
knowledge	:	γνώση
knowledge base	:	βάση γνώσης
knowledge representation	:	αναπαράσταση γνώσης
label	:	επισημείωση
logic	:	λογική
nominal	:	ονοματική έννοια
number restriction	:	περιορισμός πληθυκότητας
object	:	αντικείμενο
ontology	:	οντολογία
quantifier	:	ποσοδείκτης
query	:	ερώτημα
query answering	:	απάντηση ερωτημάτων
query rewriting	:	επαναγραφή ερωτημάτων
reasoner	:	μηχανή συλλογιστικής
reasoning	:	συλλογιστική
reduction	:	μείωση, αναγωγή
redundant	:	περιττός
resolution	:	ανάλυση
restriction	:	περιορισμός
rewritable	:	επαναγράψιμος
role	:	ρόλος
satisfiable	:	ικανοποιήσιμος
saturation	:	κορεσμός
semantics	:	σημασιολογία
subsumption	:	υπαγωγή
syntax	:	συντακτικό
top concept	:	καθολική έννοια
tractable	:	βατός
transitive	:	μεταβατικός
unqualified existential restriction	:	απροσδιόριστος υπαρξιακός περιορισμός
upper approximation	:	άνωθεν προσέγγιση

Παράρτημα Δ′

Γλωσσάριο Συμβόλων

C	:	Το σύνολο των ατομικών εννοιών μιας ΠΛ γλώσσας
R	:	Το σύνολο των ατομικών ρόλων μιας ΠΛ γλώσσας
I	:	Το σύνολο των ατόμων μιας ΠΛ γλώσσας
$\exists$	:	Ο υπαρξιακός προσοδείκτης που στις ΠΛ χρησιμοποιείται στους υπαρξιακούς περιορισμούς
$\forall$	:	Ο καθολικός προσοδείκτης που στις ΠΛ χρησιμοποιείται στους περιορισμούς τιμής
$\sqsubseteq$	:	Το σύμβολο της υπαγωγής δυο εννοιών ή ρόλων
$\subseteq$	:	Το σύμβολο του υποσυνόλου
$\neg$	:	Το σύμβολο της άρνησης μιας έννοιας, ή γενικότερα ενός στοιχείου
$\bot$	:	Η κενή έννοια
$\top$	:	Η καθολική έννοια
P^-	:	Ο αντίστροφος του ρόλου P
$\mathcal{O}$	:	Ένα σώμα ορολογίας
$\mathcal{A}$	:	Ένα σώμα ισχυρισμών
$\mathcal{I}$	:	Μια ερμηνεία (interpretation)
$\Delta^{\mathcal{I}}$	:	Ένας χώρος ερμηνείας (domain of interpretation) ο οποίος αποτελείται από ένα σύνολο αντικειμένων
$.^{\mathcal{I}}$	:	Μια συνάρτηση ερμηνείας που ερμηνεύει τα στοιχεία μιας ΠΛ γλώσσας
$\models$	:	Το σύμβολο της λογικής συνεπαγωγής (entailment)
mgu	:	Η συνάρτηση εύρεσης του μέγιστου κοινού ενοποιητή (most general unifier)

$\square$

Βιβλιογραφία

[1] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison-Wesley, 1995.

[2] Andrea Acciarri, Diego Calvanese, Domenico Lembo, Maurizio Lenzerini, Mattia Palmieri, and Riccardo Rosati. Quonto: Querying ontologies. In *In Proc. of AAAI 2005*, pages 1670–1671, 2005.

[3] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. Consistent query answers in inconsistent databases. In *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA*, pages 68–79, 1999.

[4] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. Answer sets for consistent query answering in inconsistent databases. *TPLP*, 3(4-5):393–424, 2003.

[5] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev. The dl-lite family and relations. *CoRR*, abs/1401.3487, 2014.

[6] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 325–330. Morgan Kaufmann, 2003.

[7] Franz Baader. Terminological cycles in a description logic with existential restrictions. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, pages 325–330, 2003.

[8] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[9] Leo Bachmair and Harald Ganzinger. Resolution theorem proving. In *Handbook of Automated Reasoning*, pages 19–99. Elsevier and MIT Press, 2001.

[10] Kenneth Baclawski, Mieczyslaw M. Kokar, Richard J. Waldinger, and Paul A. Kogut. Consistency checking of semantic web ontologies. In *The Semantic Web - ISWC*

2002, First International Semantic Web Conference, Sardinia, Italy, June 9-12, 2002, Proceedings, pages 454–459, 2002.

[11] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On rules with existential variables: Walking the decidability line. *Artificial Intelligence*, 175(9–10):1620–1654, 2011.

[12] Sean Bechhofer and Carole Goble. Description Logics and Multimedia - Applying Lessons Learnt from the GALEN Project. In *KRIMS 96 Workshop on Knowledge Representation for Interactive Multimedia Systems, ECAI 96*, Budapest, 1996.

[13] Sean Bechhofer, Frank van Harmelen, James Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein eds. OWL Web Ontology Language Reference. URL `http://www.w3.org/TR/owl-ref/`, Feb 2004.

[14] J. I. Berman, H. H. Moore IV, and J. R. Wright. In *CLASSIC and PROSE stories: Enabling technologies for knowledge based systems*, pages 69–78. AT and T Technical Journal, 1994.

[15] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, 2001.

[16] Leopoldo E. Bertossi, Anthony Hunter, and Torsten Schaub. Introduction to inconsistency tolerance. In *Inconsistency Tolerance [result from a Dagstuhl seminar]*, pages 1–14, 2005.

[17] Meghyn Bienvenu. First-Order Expressibility Results for Queries over Inconsistent DL-Lite Knowledge Bases. In *24th International Workshop on Description Logics (DL 2011)*, Espagne, 2011-07.

[18] Meghyn Bienvenu, Camille Bourgaux, and François Goasdoué. Querying inconsistent description logic knowledge bases under preferred repair semantics. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pages 996–1002, 2014.

[19] Meghyn Bienvenu, Camille Bourgaux, and François Goasdoué. Explaining query answers under inconsistency-tolerant semantics over description logic knowledge bases (extended abstract). In *Proceedings of the 28th International Workshop on Description Logics, Athens,Greece, June 7-10, 2015.*, 2015.

[20] Meghyn Bienvenu and Riccardo Rosati. New inconsistency-tolerant semantics for robust ontology-based data access. In *Proceedings of the 26th International Workshop on Description Logics*, 2013.

[21] S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, S. Handschuh, Y. Kompatsiaris, S. Staab, and M. G. Strintzis. *Semantic Annotation of Images and Videos for Multimedia Analysis*. Lecture Notes in Computer Science – The Semantic Web: Research and Applications, Springer, Vol. 3532, 2005, pp. 592-607, 2005.

[22] Richard Booth, Thomas Meyer, and Ivan José Varzinczak. First steps in EL contraction. In *Proceedings of the Workshop on Automated Reasoning about Context and Ontology Evolution (ARCOE 2009)*, 2009.

[23] A. Borgida. On the relative expressiveness of description logics and predicate logics. *Artificial Intelligence*, 82:353–367, 1996.

[24] R. Brachman and H. Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann, 2004.

[25] François Bry. Query answering in information systems with integrity constraints. In *Integrity and Internal Control in Information Systems, IFIP TC11 Working Group 11.5, First Working Conference on Integrity and Internal Control in Information Systems: Increasing the confidence in Information Systems, Zurich, Switzerland, December 4-5, 1997*, pages 113–130, 1997.

[26] Andrea Calì, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, and Andreas Pieris. Datalog+/-: A family of logical knowledge representation and query languages for new applications. In *Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science (LICS 2010)*, pages 228–242, 2010.

[27] Andrea Calì, Domenico Lembo, and Riccardo Rosati. On the decidability and complexity of query answering over inconsistent and incomplete databases. In Frank Neven, Catriel Beeri, and Tova Milo, editors, *PODS*, pages 260–271. ACM, 2003.

[28] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. The Mastro system for ontology-based data access. *Semantic Web Journal*, 2(1):43–53, 2011.

[29] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning*, 39(3):385–429, 2007.

[30] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in

description logics: The DL-Lite family. *Journal of Automated Reasoning*, 39(3):385–429, 2007.

[31] Diego Calvanese, Giuseppe Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. Autom. Reason.*, 39(3):385–429, October 2007.

[32] Diego Calvanese, Evgeny Kharlamov, Werner Nutt, and Dmitriy Zheleznyakov. Evolution of DL-Lite knowledge bases. In *International Semantic Web Conference (1)*, pages 112–128, 2010.

[33] Jan Chomicki. Consistent query answering: Five easy pieces. In *Database Theory - ICDT 2007, 11th International Conference, Barcelona, Spain, January 10-12, 2007, Proceedings*, pages 1–17, 2007.

[34] Jan Chomicki and Jerzy Marcinkowski. On the computational complexity of minimal-change integrity maintenance in relational databases. In *Inconsistency Tolerance [result from a Dagstuhl seminar]*, pages 119–150, 2005.

[35] Alexandros Chortaras, Despoina Trivela, and Giorgos Stamou. Optimized query rewriting in OWL 2 QL. In *Proceedings of the 23rd International Conference on Automated Deduction (CADE 23), Polland*, pages 192–206, 2011.

[36] Bernardo Cuenca Grau, Ernesto Jimenez-Ruiz, Evgeny Kharlamov, and Dmitriy Zheleznyakov. Ontology evolution under semantic constraints. In *13th International Conference on Principles of Knowledge Representation and Reasoning*, March 2012.

[37] Bernardo Cuenca Grau, Evgeny Kharlamov, and Dmitriy Zheleznyakov. Ontology contraction: Beyond propositional paradise. In *Alberto Mendelzon International Workshop on Foundations of Data Management (AMW)*, Ouro Preto, Brazil, June 2012.

[38] Carlo Curino, Hyun Jin Moon, Alin Deutsch, and Carlo Zaniolo. Automating the database schema evolution process. *The VLDB Journal*, 22(1):73–98, 2013.

[39] P. T. Devanbu, R. J. Brachman, P. G. Selfridge, and B. W. Ballard. Lassienowledge-based software information system. In *Proceedings of the 12th international conference on Software engineering*, ICSE '90, pages 249–261, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[40] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, and Andrea Schaerf. Deduction in concept languages: From subsumption to instance checking. *Journal of Logic and Computation*, 4(4):423–452, 1994.

[41] Thomas Eiter, Magdalena Ortiz, Mantas Simkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for Horn-SHIQ plus rules. In *AAAI*, 2012.

[42] Suzanne M. Embury, Sue M. Brandt, John S. Robinson, Iain Sutherland, Frank A. Bisby, W. A. Gray, Andrew C. Jones, and Richard J. White. Adapting integrity enforcement techniques for data reconciliation. *Inf. Syst.*, 26(8):657–689, 2001.

[43] Herbert B. Enderton. *A mathematical introduction to logic*. Academic Press, 1972.

[44] FaCT++. `http://owl.man.ac.uk/factplusplus/`, 2003.

[45] Giorgos Flouris, Dimitris Manakanatas, Haridimos Kondylakis, Dimitris Plexousakis, and Grigoris Antoniou. Ontology change: Classification and survey. *Knowl. Eng. Rev.*, 23(2):117–152, June 2008.

[46] M I Ginsberg and M L Ginsberg. Counterfactuals. *Artif. Intell.*, 30(1):35–80, October 1986.

[47] Christine Golbreich, Songmao Zhang, and Olivier Bodenreider. The foundational model of anatomy in owl: Experience and perspectives. In *Web Semantics: Science, Services and Agents on the World Wide Web*, pages 181–195.

[48] Rafael S. Gonçalves, Bijan Parsia, and Ulrike Sattler. Analysing the evolution of the nci thesaurus. In *CBMS*, pages 1–6. IEEE, 2011.

[49] Rafael S. Gonçalves, Bijan Parsia, and Ulrike Sattler. Analysing multiple versions of an ontology: A study of the nci thesaurus. In Riccardo Rosati, Sebastian Rudolph, and Michael Zakharyaschev, editors, *Description Logics*, volume 745 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2011.

[50] John Goodwin. Experiences of using owl at the ordnance survey. In *Proceedings of OWL: Experiences and Directions*, 2005.

[51] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proceedings of the 27th International Conference on Data Engineering, ICDE 2011*, 2011.

[52] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *Proceedings of the 2011 IEEE 27th International Conference on Data Engineering*, ICDE '11, pages 2–13, Washington, DC, USA, 2011. IEEE Computer Society.

[53] Bernardo Cuenca Grau, Christian Halaschek-Wiener, Yevgeny Kazakov, and Boontawee Suntisrivaraporn. Incremental classification of description logics ontologies. *JAR*, 44(4):337–369, 2010.

[54] Benjamin N. Grosof, Ian Horrocks, Raphael Volz, and Stefan Decker. Description logic programs: combining logic programs with description logic. In *Proceedings of the Twelfth International World Wide Web Conference, WWW 2003, Budapest, Hungary, May 20-24, 2003*, pages 48–57, 2003.

[55] Volker Haarslev and Ralf Möller. RACER System Description. In *International Joint Conference on Automated Reasoning-01*, volume 2083, 2001.

[56] Volker Haarslev, Ralf Moller, Ragnhild Van Der Straeten, and Michael Wessel. Extended query facilities for racer and an application to software-engineering problems. In *Proceedings of the International Workshop on Description Logics (DL-2004)*, 2004.

[57] P. Haase and L. Stojanovic. Consistent evolution of OWL ontologies. In *2nd European Semantic Web Conf.*, 2005.

[58] Peter Haase, Frank van Harmelen, Zhisheng Huang, Heiner Stuckenschmidt, and York Sure. A framework for handling inconsistency in changing ontologies. In *The Semantic Web - ISWC 2005, 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005, Proceedings*, pages 353–367, 2005.

[59] Matthew Horridge, Bijan Parsia, and Ulrike Sattler. Explaining inconsistencies in OWL ontologies. In *Scalable Uncertainty Management, Third International Conference, SUM 2009, Washington, DC, USA, September 28-30, 2009. Proceedings*, pages 124–137, 2009.

[60] Ian Horrocks and Ulrike Sattler. A description logic with transitive and inverse roles and role hierarchies. *Journal of Logic and Computation*, 9(3):385–410, 1999.

[61] Zhisheng Huang, Frank van Harmelen, and Annette ten Teije. Reasoning with inconsistent ontologies. In *BNAIC 2005 - Proceedings of the Seventeenth Belgium-Netherlands Conference on Artificial Intelligence, Brussels, Belgium, October 17-18, 2005*, pages 349–350, 2005.

[62] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Data complexity of reasoning in very expressive description logics. In *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005*, pages 466–471, 2005.

[63] Ullrich Hustadt, Boris Motik, and Ulrike Sattler. Data complexity of reasoning in very expressive description logics. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *Proc. of the 19th Int. Joint Conference on Artificial Intelligence (IJCAI 2005)*, pages 466–471, 2005.

[64] E. Jiménez Ruiz, B. Cuenca Grau, I. Horrocks, and R. Berlanga. Supporting concurrent ontology development: Framework, algorithms and tool. *Data Knowl. Eng.*, 70(1):146–164, January 2011.

[65] Aditya Kalyanpur, Bijan Parsia, and Evren Sirin. Black box techniques for debugging unsatisfiable concepts. In *Proceedings of the 2005 International Workshop on Description Logics (DL2005), Edinburgh, Scotland, UK, July 26-28, 2005*, 2005.

[66] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, and Bernardo Cuenca Grau. Repairing unsatisfiable concepts in OWL ontologies. In *The Semantic Web: Research and Applications, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings*, pages 170–184, 2006.

[67] Mark Kaminski, Yavor Nenov, and Bernardo Cuenca Grau. Computing datalog rewritings for disjunctive datalog programs and description logic ontologies. In *Proceedings of the 8th International Conference on Web Reasoning and Rule Systems (RR)*, 2014.

[68] Yevgeny Kazakov and Pavel Klinov. Incremental reasoning in EL+ without bookkeeping. In *Description Logics*, pages 294–315, 2013.

[69] Atanas Kiryakov, Barry Bishop, Damyan Ognyanoff, Ivan Peikov, Zdravko Tashev, and Ruslan Velkov. The features of BigOWLIM that enabled the BBC's World Cup website. In *Workshop on Semantic Data Management (SemData)*, 2010.

[70] Haridimos Kondylakis and Dimitris Plexousakis. Ontology evolution in data integration: Query rewriting to the rescue. In *Proceedings of the 30th International Conference on Conceptual Modeling*, ER'11, pages 393–401, Berlin, Heidelberg, 2011. Springer-Verlag.

[71] Haridimos Kondylakis and Dimitris Plexousakis. Ontology evolution without tears. *Web Semantics: Science, Services and Agents on the World Wide Web*, 19(0), 2013.

[72] Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. A sound and complete backward chaining algorithm for existential rules. In *Proceedings of the 6th International Conference on Web Reasoning and Rule Systems, Vienna, Austria, 10-12 September*, pages 122–138. Springer-Verlag, Berlin, 2012.

[73] Markus Krötzsch. OWL 2 profiles: An introduction to lightweight ontology languages. In *Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School 2012, Vienna, Austria, September 3-8, 2012. Proceedings*, pages 112–183, 2012.

[74] Markus Krötzsch, Sebastian Rudolph, and Pascal Hitzler. Complexity boundaries for horn description logics. In *Proceedings of the Twenty-Second AAAI Conference on*

Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada, pages 452–457, 2007.

[75] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Inconsistency-tolerant semantics for description logics. In *Proceedings of the 4th International Conference on Web Reasoning and Rule Systems (RR)*, pages 103–117, 2010.

[76] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Query rewriting for inconsistent dl-lite ontologies. In *RR*, pages 155–169, 2011.

[77] Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo. Inconsistency-tolerant query answering in ontology-based data access. *J. of Web Semantics*, 33:3–29, 2015.

[78] Domenico Lembo and Marco Ruzzi. Consistent query answering over description logic ontologies. In *Description Logics*, 2007.

[79] Thomas Lukasiewicz, Maria Vanina Martinez, and Gerardo I. Simari. Inconsistency handling in datalog+/- ontologies. In *ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012*, pages 558–563, 2012.

[80] Carsten Lutz. The complexity of conjunctive query answering in expressive description logics. In Alessandro Armando, Peter Baumgartner, and Gilles Dowek, editors, *Proceedings of the 4th International Joint Conference on Automated Reasoning (IJCAR2008)*, number 5195 in LNAI, pages 179–193. Springer, 2008.

[81] Carsten Lutz, Inanç Seylan, David Toman, and Frank Wolter. The combined approach to OBDA: taming role hierarchies using filters. In *Proceedings of the 12th International Semantic Web Conference (ISWC 2013)*, pages 314–330, 2013.

[82] Yue Ma, Pascal Hitzler, and Zuoquan Lin. Algorithms for paraconsistent reasoning with OWL. In *The Semantic Web: Research and Applications, 4th European Semantic Web Conference, ESWC 2007, Innsbruck, Austria, June 3-7, 2007, Proceedings*, pages 399–413, 2007.

[83] Yue Ma, Guilin Qi, and Pascal Hitzler. Computing inconsistency measure based on paraconsistent semantics. *J. Log. Comput.*, 21(6):1257–1281, 2011.

[84] Giulia Masotti, Riccardo Rosati, and Marco Ruzzi. Practical abox cleaning in dl-lite (progress report). In *Proceedings of the 24th International Workshop on Description Logics (DL)*, 2011.

[85] Deborah L. McGuinness and Jon R. Wright. Conceptual modelling for configuration: A description logic-based approach. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 12(4):333–344, September 1998.

[86] Carlo Meghini, Fabrizio Sebastiani, and Umberto Straccia. A model of multimedia information retrieval. *Journal of the ACM*, 48(5):909–970, 2001.

[87] E. Mendelson. *Introduction to Mathematical Logic, Fourth Edition*. Discrete Mathematics and Its Applications. Taylor & Francis, 1997.

[88] Thomas Andreas Meyer, Kevin Lee, and Richard Booth. Knowledge integration for description logics. In *Proceedings, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, July 9-13, 2005, Pittsburgh, Pennsylvania, USA*, pages 645–650, 2005.

[89] José Mora and Óscar Corcho. Towards a systematic benchmarking of ontology-based query rewriting systems. In *Proceedings of the 12th International Semantic Web Conference*, pages 376–391, 2013.

[90] Boris Motik, Rob Shearer, and Ian Horrocks. A Hypertableau Calculus for SHIQ. In Diego Calvanese, Enriso Franconi, Volker Haarslev, Domenico Lembo, Boris Motik, Sergio Tessaris, and Anny-Yasmin Turhan, editors, *Proceedings of the 20th Int. Workshop on Description Logics (DL 2007)*, pages 419–426, Brixen/Bressanone, Italy, June 8–10 2007. Bozen/Bolzano University Press.

[91] Kedian Mu, Zhi Jin, Weiru Liu, Didar Zowghi, and Bo Wei. Measuring the significance of inconsistency in the viewpoints framework. *Sci. Comput. Program.*, 78(9):1572–1599, 2013.

[92] Kedian Mu, Weiru Liu, and Zhi Jin. A blame-based approach to generating proposals for handling inconsistency in software requirements. *IJKSS*, 3(1):1–17, 2012.

[93] Kedian Mu, Weiru Liu, and Zhi Jin. Measuring the blame of each formula for inconsistent prioritized knowledge bases. *J. Log. Comput.*, 22(3):481–516, 2012.

[94] Natalya F. Noy, Sandhya Kunnatur, Michel Klein, and Mark A. Musen. Tracking changes during ontology evolution. In Sheila A. Mcilraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *Third International Semantic Web Conference*, pages 259+, Hiroshima, Japan, November 2004. Springer Berlin.

[95] Giorgio Orsi and Andreas Pieris. Optimizing query answering under ontological constraints. *Proceedings of the VLDB Endowment*, 4(11):1004–1015, 2011.

[96] Christos H. Papadimitriou. *Computational complexity*. Academic Internet Publ., 2007.

[97] Bijan Parsia, Evren Sirin, and Aditya Kalyanpur. Debugging OWL ontologies. In *Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*, pages 633–640, 2005.

[98] Peter F. Patel-Schneider. Adding number restrictions to a four-valued terminological logic. In *Proceedings of the 7th National Conference on Artificial Intelligence. St. Paul, MN, August 21-26, 1988.*, pages 485–490, 1988.

[99] Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. Technical report, W3C, Feb. 2004. W3C Recommendation, URL `http://www.w3.org/TR/2004/REC-owl-semantics-20040210/`.

[100] Héctor Pérez-Urbina, Ian Horrocks, and Boris Motik. Efficient query answering for OWL 2. In *Proc. of the Int. Semantic Web Conference (ISWC2009)*, Chantilly, VA, USA., October 2009.

[101] Héctor Pérez-Urbina, Boris Motik, and Ian Horrocks. Rewriting conjunctive queries over description logic knowledge bases. In *Semantics in Data and Knowledge Bases, Third International Workshop, SDKB 2008, Nantes, France, March 29, 2008, Revised Selected Papers*, pages 199–214, 2008.

[102] Héctor Pérez-Urbina, Boris Motik, and Ian Horrocks. Rewriting conjunctive queries under description logic constraints. Technical report, University of Oxford, 2008.

[103] Héctor Pérez-Urbina, Boris Motik, and Ian Horrocks. Tractable query answering and rewriting under description logic constraints. *Journal of Applied Logic*, 8(2):186–209, 2010.

[104] Héctor Pérez-Urbina, Edgar Rodríguez-Díaz, Michael Grove, George Konstantinidis, and Evren Sirin. Evaluation of query rewriting approaches for owl 2. In *TJoint Workshop on Scalable and High-Performance Semantic Web Systems*, 2012.

[105] K. Petridis, S. Bloehdorn, C. Saathoff, N. Simou, S. Dasiopoulou, V. Tzouvaras, S. Handschuh, Y. Avrithis, I. Kompatsiaris, and S. Staab. Knowledge representation and semantic annotation of multimedia content. 2006.

[106] Guilin Qi and Jianfeng Du. Model-based revision operators for terminologies in description logics, 2009.

[107] Guilin Qi, Weiru Liu, and David A. Bell. A revision-based approach to handling inconsistency in description logics. *Artif. Intell. Rev.*, 26(1-2):115–128, 2006.

[108] M. R. Quillian. Semantic memory. In M. Minsky, editor, *Semantic Information Processing*, pages 216–270. MIT Press, Cambridge, MA, 1968.

[109] Rob Raskin and Michael Pan. Semantic web for earth and environmental terminology (sweet. In *Proceedings of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data*, 2003.

[110] A. L. Rector and I. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *Proceedings Workshop on Ontological Engineering, AAAI Spring Symposium, Stanford CA.*, pages 100–107. Hanley and Belfus, Inc., Philadelphia, PA, 1997.

[111] Marcio Ribeiro, Renata Wassermann, Grigoris Antoniou, Giorgos Flouris, and Jeff Pan. Belief contraction in web-ontology languages. In *Proceedings of the 3rd International Workshop on Ontology Dynamics (IWOD-09)*, 2009.

[112] John Alan Robinson. A machine-oriented logic based on the resolution principle. *J. ACM*, 12(1):23–41, 1965.

[113] Mariano Rodriguez-Muro and Diego Calvanese. High performance query answering over DL-Lite ontologies. In *Proc. of the 13th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 308–318. AAAI Press, 2012.

[114] Mariano Rodriguez-Muro, Roman Kontchakov, and Michael Zakharyaschev. Ontology-based data access: Ontop of databases. In *Proceedings of the 12th International Semantic Web Conference, Part I, Sydney, NSW, Australia, 21-25 October*, pages 558–573. Springer, Berlin, 2013.

[115] Riccardo Rosati. On the complexity of dealing with inconsistency in description logic ontologies. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, IJCAI'11, pages 1057–1062. AAAI Press, 2011.

[116] Riccardo Rosati and Alessandro Almatelli. Improving query answering over DL-Lite ontologies. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR 2010)*, 2010.

[117] Riccardo Rosati, Marco Ruzzi, Mirko Graziosi, and Giulia Masotti. Evaluation of techniques for inconsistency handling in OWL 2 QL ontologies. In *Proceedings of the 11th International Semantic Web Conference (ISWC)*, pages 337–349, 2012.

[118] Riccardo Rosati, Marco Ruzzi, Mirko Graziosi, and Giulia Masotti. Evaluation of techniques for inconsistency handling in owl 2 ql ontologies. In *Proceedings of the 11th International Conference on The Semantic Web - Volume Part II*, ISWC'12, pages 337–349, Berlin, Heidelberg, 2012. Springer-Verlag.

[119] Ulrike Sattler. A concept language extended with different kinds of transitive roles. In *Proceedings of the 20th Annual German Conference on Artificial Intelligence: Advances in Artificial Intelligence*, KI '96, pages 333–345, London, UK, UK, 1996. Springer-Verlag.

[120] Domenico Fabio Savo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro, Vittorio Romagnoli, Marco Ruzzi, and Gabriele Stella. Experimenting ontology-based data access with mastro (extended abstract). In *Proceedings of the Eighteenth Italian Symposium on Advanced Database Systems, SEBD 2010, Rimini, Italy, June 20-23, 2010*, pages 326–333, 2010.

[121] Andrea Schaerf. Reasoning with individuals in concept languages. *Data and Knowledge Engineering*, 13(2):141–176, 1994.

[122] Manfred Schmidt-Schaubßand Gert Smolka. Attributive concept descriptions with complements. *Artif. Intell.*, 48(1):1–26, February 1991.

[123] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, May 2006.

[124] Amandeep S. Sidhu, Tharam S. Dillon, Elizabeth Chang, and Baldev S. Sidhu. Protein ontology development using owl. In Bernardo Cuenca Grau, Ian Horrocks, Bijan Parsia, and Peter F. Patel-Schneider, editors, *OWLED*, volume 188 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2005.

[125] N. Simou, Th. Athanasiadis, G. Stoilos, and S. Kollias. Image indexing and retrieval using expressive fuzzy description logics. 2008.

[126] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics*, 5:51–53, 2007.

[127] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semant.*, 5(2):51–53, June 2007.

[128] Giorgos Stoilos. Ontology-based data access using rewriting, OWL 2 RL systems and repairing. In *Proceedings of the 11th European Semantic Web Conference ESWC*, pages 317–332, 2014.

[129] Giorgos Stoilos, Bernardo Cuenca Grau, Boris Motik, and Ian Horrocks. Repairing ontologies for incomplete reasoners. In *Proceedings of the 10th International Semantic Web Conference (ISWC-11), Bonn, Germany*, pages 681–696, 2011.

[130] Umberto Straccia. A sequent calculus for reasoning in four-valued description logics. In *Automated Reasoning with Analytic Tableaux and Related Methods,*

International Conference, TABLEAUX '97, Pont-à-Mousson, France, May 13-16, 1997, Proceedings, pages 343–357, 1997.

[131] Despoina Trivela, Giorgos Stoilos, Alexandros Chortaras, and Giorgos Stamou. Optimising resolution-based rewriting algorithms for owl ontologies. *Journal of Web Semantics*, 33:30–49, 2015.

[132] Despoina Trivela, Giorgos Stoilos, Alexandros Chortaras, and Giorgos B. Stamou. Optimising resolution-based rewriting algorithms for DL ontologies. In *Description Logics*, pages 464–476, 2013.

[133] Despoina Trivela, Giorgos Stoilos, Alexandros Chortaras, and Giorgos B. Stamou. Query rewriting in Horn-SHIQ. In *Proceedings of the 28th International Workshop on Description Logics (DL 2015), Athens, Greece, 7–10 June*. CEUR-WS.org, 2015.

[134] Moshe Y. Vardi. The complexity of relational query languages (extended abstract). In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, STOC '82, pages 137–146, New York, NY, USA, 1982. ACM.

[135] Moshe Y. Vardi. On the complexity of bounded-variable queries (extended abstract). In *Proceedings of the Fourteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS '95, pages 266–276, New York, NY, USA, 1995. ACM.

[136] Tassos Venetis, Giorgos Stoilos, and Giorgos Stamou. Query rewriting under query extensions for OWL 2 QL ontologies. In *Proceedings of the 7th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2011), Bonn, Germany*, 2011.

[137] Tassos Venetis, Giorgos Stoilos, and Giorgos Stamou. Incremental query rewriting for OWL 2 QL. In *Proceedings of the 25th International Workshop on Description Logics (DL 2012), Rome, Italy*, 2012.

[138] Zhe Wang, Kewen Wang, and Rodney Topor. A new approach to knowledge base revision in DL-Lite. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010)*, 2010.

Ευρετήριο Όρων

Κατάλογος Δημοσιεύσεων

E. Tsalapati, G. Stoilos, G. Stamou, G. Koletsos. : Efficient Query Answering over Expressive Inconsistent Description Logics. In *International Joint Conference on Artificial Intelligence (IJCAI)*, New York, 2016. (υπό έκδοση)

E. Tsalapati, G. Stoilos, A. Chortaras, G. B. Stamou, G. Koletsos. Query Rewriting under Ontology Change. In *The Computer Science Journal*, (υπό έκδοση).

C. Hondrou, E. Tsalapati, A. Raouzaiou, G. Marandianos, K. Karpouzis, S. D. Kollias. Player-Specific Conflict Handling Ontology. In *International Journan of Serious Games 1(3)*, 2014.

E. Tsalapati, G. Stoilos, G. B. Stamou, G. Koletsos. Query Rewriting under Ontology Evolution. In *Proceedings of the Description Logics Workshop (DL 13), Ulm, Germany*, pp. 975-987, 2013.

C. Hondrou, E. Tsalapati, A. Raouzaiou, K. Karpouzis, S. Kollias. Player Specific Conflict Handling Ontology. *GALA, Paris, France*,pp. 304-315, 2013.

E. Tsalapati, G. Stoilos, G. B. Stamou, G. Koletsos. Query Rewriting under Ontology Contraction. In *Proceedings of the International Conference On Web Reasoning And Rule, Vienna, Austria*, pp. 172-187, 2012.

E. Tsalapati, N. Simou, N. Drosopoulos and R. Stein. Evolving LIDO based aggregations into Linked Data. *CIDOC2012 - Enriching Cultural Heritage Helsinki, Finland*, 2012.

E. Tsalapati, G. Stamou, G. Koletsos. : A Method for Approximation to Ontology Reuse Problem. In *Proceedings of theKEOD - International Conference on Knowledge Engineering and Ontology Development, Madeira, Portugal*, pp. 416-419, 2009.

□