EFFICIENT REASONING ALGORITHMS FOR FRAGMENTS OF HORN DESCRIPTION LOGICS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By

DAVID CARRAL
Master of Science, Wright State University, 2012
Bachelor of Science, Universidad Pontificia de Salamanca, 2011

2017
Wright State University

Pascal Hitzler, Ph.D.
Dissertation Director

Mateen Rizki, Ph.D.
Department Chair

Committee on Final Examination

Bernardo Cuenca Grau, Ph.D.

Pascal Hitzler, Ph.D.

Krishnaprasad Thirunarayan, Ph.D.

Michael L. Raymer, Ph.D.

Robert E. W. Fyffe, Ph.D.
Vice President for Research and Dean of the Graduate School
ABSTRACT


We characterize two fragments of Horn Description Logics and we define two specialized reasoning algorithms that effectively solve the standard reasoning tasks over each of such fragments. We believe our work to be of general interest since (1) a rather large proportion of real-world Horn ontologies belong to some of these two fragments and (2) the implementations based on our reasoning approach significantly outperform state-of-the-art reasoners. Claims (1) and (2) are extensively proven via empirically evaluation.
## Contents

1 Introduction ................................................. 1
   1.1 Broader Impact ........................................ 2
   1.2 Problem Statement .................................... 2
   1.3 Structure .............................................. 3

2 An Intuitive Introduction to Description Logics ............ 4
   2.1 An Informal Introduction to DL ....................... 4
      2.1.1 Entities ........................................... 5
      2.1.2 Logical Constructors ............................. 5
      2.1.3 Axioms ............................................. 6
         2.1.3.1 ABox axioms .................................. 7
         2.1.3.2 TBox axioms .................................. 7
      2.1.4 Ontologies ....................................... 7
      2.1.5 Reasoning Tasks .................................. 8

3 Formal Preliminaries ........................................ 9
   3.1 Existential Rules ..................................... 9
   3.2 Horn Description Logics ............................... 10
   3.3 The Relation between DLs and Existential Rules ... 13
   3.4 Axiomatizations of the Equality Predicate .......... 13
      3.4.1 The Standard Equality Axiomatization .......... 13
      3.4.2 Singularization .................................. 14
   3.5 The Chase Algorithm .................................. 15

4 The RCA_n Fragment of Horn DL ............................. 17
   4.1 Model Faithful Acyclicity ............................. 18
   4.2 Restricted Chase Acyclicity ......................... 19
   4.3 Evaluation ............................................ 22
      4.3.1 An Empirical Comparison of RCA_n and MFA' .... 22
      4.3.2 A Materialization Based Reasoner ............... 23
   4.4 Proofs ................................................. 26
      4.4.1 Overchase and Termination ..................... 26
List of Figures

3.1 Syntax and Semantics of Axioms. In the above, $A(i), B \in N_C, R, S, V \in N_R$ and $a, b \in N_I$. .................................................. 11

3.2 Function $\pi$ Mapping DL Axioms to Rules. In the above, $A(i), B \in N_C, R, S, V \in N_R$, $a \in N_I \cap C$ and $x, y, z \in V$. .................................................. 13

3.3 Rules Employed in the Axiomatization of Equality. .................................................. 14

3.4 Ontology $O = \langle T, A \rangle$, Program $P(O)$ and the (Restricted and Oblivious) Chase of $P(O)$. 16

4.1 Expansion rules for the construction of $V_T$. .................................................. 21

4.2 Queries for Reactome, Uniprot, LUBM and UOBM. .................................................. 39

4.3 Expansion rules for the construction of $W_T$. .................................................. 40

5.1 Example ontology $O$. .................................................. 42

5.2 Checking acyclicity of our example ontology $O$. .................................................. 44

5.3 Running Example: Reasoning. .................................................. 46
List of Tables

4.1 Results for the ORE and ODP Repositories. .................................................. 24
4.2 Results for Reactome, Uniprot, LUBM and UOBM (sorted from top to bottom in the above table). ................................................................. 25
5.1 Acyclicity evaluation results for ontologies outside the OWL 2 profiles. ............ 50
Amongst the many people I have had the pleasure to work with throughout these past years, there are three without whom I would not have been able to successfully finish this doctorate degree. The first of them is Dr. Bernardo Cuenca Grau, who has on many occasions attempted to teach me how to clearly structure and organize ideas, an instrumental skill I still hope to pick up some day. The second person I would like to recognize is Dr. Cristina Feier, my closest and most esteemed collaborator thus far! Many-a-deadlines I would have missed, had you not been there to help me out! And, last but not least, I would like to dearly thank my advisor Dr. Pascal Hitzler, who was unfortunate enough to have to mentor me during these most melodramatic early years of my adult life and, nevertheless, made an outstanding job out of it.

Finally, I would like to mention the other members of my dissertation committee, Dr. Krishnaprasad Thirunarayan and Dr. Michael L. Raymer. Thank you very much for seeing me through these last steps of my PhD.

On the more formal side, this work was partially supported by the National Science Foundation under award 1017225 “III: Small: TROn – Tractable Reasoning with Ontologies.” Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Furthermore, the author was also partially funded by the “La Caixa” foundation fellowship for postgraduate studies.
1 Introduction

On many applications, machines are primarily used to store, transmit and display information for human consumption, but often this content is “incomprehensible” to the machines themselves. Amongst other reasons, this is because the machines lack the necessary (and basic) necessary background to “understand” such content.

For example, if you ask the server hosting the Wikipedia page of King Hamlet “What is the name of the uncle of Prince Hamlet?,” even if it could somehow correctly interpret the question, it still could not produce the correct answer. This is because, such an answer is not explicitly included in the webpage. Nevertheless, the webpage does contain (at the time of this write-up) the following relevant pieces of information: (a) “Prince Hamlet is the son of King Hamlet” and (b) “Claudius the brother of King Hamlet.” Upon reading (a) and (b), any human would readily ascertain the answer of the previously poised question; namely, that (c) “Claudius is the uncle of Prince Hamlet.” The server, lacking an “understanding” of the entities uncle, brother and father and their relation, is unable to come up with such a conclusion.

The previous issue can be addressed making use of Description Logics (DL); a knowledge representation family of languages which may be employed to encode information about the relationships between a set of entities within a given domain. Using DL languages, we can encode knowledge about a particular domain into axioms which, in turn, can be grouped into larger structures, referred to as ontologies. After crafting such ontologies, we can employ automated reasoning algorithms to derive implicit knowledge that follows from the explicitly stated assertions.

For example, we may use DL axioms (1.1) and (1.2) to convey facts (a) and (b), respectively.

\[
\text{HasFather(princeHamlet, kingHamlet)} \quad (1.1)
\]

\[
\text{HasBrother(kingHamlet, claudius)} \quad (1.2)
\]

Furthermore, to provide the machine with some kind of “understanding” of the relation between the entities uncle, brother and father, we may employ the following axiom.

\[
\text{HasFather} \circ \text{HasBrother} \sqsubseteq \text{HasUncle} \quad (1.3)
\]

Intuitively the previous axiom indicates that, if somebody has a father who has a brother, then this brother is the uncle of that “somebody” (many more such examples will be presented in Section 2).
1.1 Broader Impact

With the example provided in the previous section, we intend to instruct the reader about the potential of knowledge representation languages. Nevertheless, this is only a toy example which does not accurately showcase the usefulness of these technologies. We proceed thus, to elaborate more broadly about the use of these languages in different domain applications.

The use of DL ontologies has been steadily gaining in popularity and is used in countless application areas, such as social network analysis [Fan 2012], city data processing [Lécué et al. 2014] and network traffic analysis [Barrett et al. 2000]. Perhaps, the most relevant and widespread application of DL takes place within the context of the Semantic Web [Hitzler et al. 2009], an extension of the World Wide Web where structure and meaning are provided in the hopes of making the information in web pages understandable.

The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium which promote common data formats and exchange protocols on the Web. One of such standards is the Web Ontology Language (OWL) which provides a computer based syntax for DL languages. Using OWL, ontologies can be encoded in a way and manner that the information such as the one presented in the previous section can be encoded in the content of webpages. Alas, the use of OWL would allow use to solve the issue presented in the previous example!

As with DLs, the use of OWL allows for the automated inference of implicit information from explicit statements. It is thus, of the utmost importance to develop efficient reasoning algorithms which are scalable over large ontologies. In this thesis we propose the use of two of such reasoning algorithms which have the potential of improving the performance of many existing real-world applications.

1.2 Problem Statement

Hopefully by now, the reader is somewhat convinced of the usefulness of knowledge representation languages and automated reasoning techniques. We proceed in this section to argue why the development of reasoning algorithms is non-trivial and challenging endeavor.

First of, note that, reasoning is over many expressive DL languages theoretically hard. In fact, for all DL languages which which cannot be characterized within the tractable fragments; namely, $\mathcal{EL}^+\!+$ [Baader et al. 2005a], DL-LiteR [Calvanese et al. 2007] and DLP [Grosif et al. 2003]; reasoning is at least ExpTime-hard.

On this thesis, we focus on the theoretical definition and implementation of very efficient reasoning
1.3 Structure

We proceed to briefly describe the structure of this thesis which includes five more further sections.

- **Section 2**: In this section we informally introduce DL languages in an attempt to convey an intuitive understanding of this standards aimed at the unexperienced reader.

- **Section 3**: In this section we formally define DL languages and introduce several notions used in the subsequent sections of the thesis.

- **Sections 4**: This section formally presents the RCA fragment of Horn-\textit{SROIQ} and an algorithm to reason over RCA ontologies. Furthermore, this section includes an extensive evaluation of the RCA fragment plus a comparison of the implementation of the reasoning algorithm against other reasoners.

- **Section 5**: This section is similar in quite similar to the previous one in terms of content, but it considers the RSA fragment of Horn-\textit{SROIQ} instead of RCA.

- **Section 6**: This section lists the conclusions of this thesis and elaborates about possible further work.
2

An Intuitive Introduction to Description Logics

Description logics (DL) is a family of knowledge representation formalisms widely used in ontological modeling. All DL languages are fragments of first-order logic and as such, they are equipped with a formal semantics. Thus, the use of DL enables a precise specification of the understanding of DL logical formulas (also referred as axioms) which, apart from allowing the exchange of knowledge without ambiguity also enable formal deduction. Given a set of DL formulas (also referred to as an ontology), we can employ logical deduction to infer additional information from the facts explicitly included in such ontology. The derivation of implicitly entailed formulas from a given set of explicitly stated axioms is often referred to as reasoning, which is a challenging and relevant task employed in many real-world applications.

The main contribution of this thesis is the definition of two fragments of Horn DL (i.e., the subset of DL which does not allow for the use of non-deterministic logical constructors such as disjunction) and the presentation of two corresponding algorithms to solve reasoning tasks over such fragments. As our main research hypothesis, we aim to empirically verify that our algorithms significantly outperform state-of-the-art systems and that our newly defined fragments characterize a large proportion of real-world Horn DL ontologies.

The remainder of the thesis is structured as follows. In the rest of this chapter we present an informal introduction to DL. In the subsequent chapter we present a formal definition of DL which also includes the definition of many preliminary notions that will be used in further sections. Chapters 4 and 5 describe the two aforementioned fragments of Horn DL and the algorithms to reason over each of such fragments, respectively. On the last chapter of this thesis we list our conclusions and elaborate about possible further work.

2.1 An Informal Introduction to DL

Along this section, we provide an informal and intuitive introduction to DL. See the following section for a formal and precise definition of the syntax and semantics of this family of knowledge represen-
2.1. AN INFORMAL INTRODUCTION TO DL

tation languages.

2.1.1 Entities

DL languages provide means to model the relations between entities in a domain of interest. In DL there are three kinds of entities: concepts, roles and individual names. Concepts represent sets of individuals, roles represent binary relations between individuals and individual names represent single individuals in the domain. Readers familiar with first-order logic will recognize these entities as unary predicates, binary predicates and constants.

For example, if we decide to model the domain of their family relationships, we may use concepts such as Female to represent the set of all female individuals, respectively. Furthermore, we may use roles such as IsParentOf to represent the (binary) relationship between parents and their children. Finally, we may use individual names such as joe to represent the domain individual Joe (note the distinction between “individual names;” i.e., DL entities which represent objects in the domain of interest; and “domain individuals;” i.e., the objects in the domain of interest which these DL entities represent).

2.1.2 Logical Constructors

Combining the use of entities with the available DL logical constructors we can create more involved (and more useful) sets of individuals, which are referred as concept expressions. For example, we can employ the intersection (also called conjunction) constructor to define a concept expressions such as the following.

\[
\text{Female} \sqcap \text{Parent}
\] (2.1)

The previous concept expression subsumes all the domain individuals that are in the intersection of the concepts Female and Parent; i.e., concept expression (2.1) includes a named individual if and only if it belongs to both the Female and Parent concepts. Intuitively, concept expression (2.1) coincides with our understanding of the idea motherhood.

Another way of constructing concept expressions is the use of role restrictions, which link concepts and roles together. For example, making use of role restrictions we can declare concept expressions such as the following.

\[
\exists \text{hasChild} \cdot \text{Person}
\] (2.2)

The previous concept expression, referred to as an existential restriction, characterizes every domain individual related to another domain individual in the class Person via the property hasChild. Intuitively, this expression coincides with the idea of parenthood.

To represent the set of individuals all of whose children are female, we may employ universal restrictions such as the following.

\[
\forall \text{IsParentOf} \cdot \text{Female}
\] (2.3)
The previous concept expression characterizes every individual in domain such that, if this individual is connected to some other individual via role \texttt{IsParentOf}, then the latter is subsumed by the class \texttt{Female}. Note that, such a concept also subsumes those individuals that have no children at all. If this meaning is not intended, one can describe the individuals with some children and with all their children being female with the following concept.

\[ \exists \text{IsParentOf} \top \land \forall \text{IsParentOf} \text{.Female} \]  

The first operand in the conjunction indicates that this concept expression may only subsume individuals with at least some children. Furthermore, the second operand indicates that concept expression only subsumes individuals who only have daughters.

Functional restrictions allow us to restrict the number of individuals that be reached via a given role.

\[ \leq 1 \text{IsParentOf} \top \]  

The previous concept expression, referred to as an at-least restriction, is employed to characterize domain individuals with at most one children.

Finally, nominals may be employed to denote classes containing a single individual.

\[ \{ \text{joe} \} \]  

The previous class expression denotes the set only containing the domain individual Joe (represented by the name individual \texttt{joe}).

Apart from boolean constructors, role restrictions and nominals, which are used to build concept expressions, we can also make use role constructors to construct role expressions. For example, we can make use of the role inverse constructor to define a role expression such as the following.

\[ \text{IsParentOf}^- \]  

The previous role expression comprehends the inverse of the relation subsumed by the role \texttt{IsParentOf}; i.e., if some individual is connected to another via \texttt{IsParentOf}, then the latter is connected to the former via the role expression \texttt{IsParentOf}^-.

Finally, we may use role chains to define relations that result from the composition of two or more existing roles.

\[ \text{hasParent} \circ \text{hasBrother} \]  

The previous role expression subsumes pairs of individuals for which the first individual in the pair is connected to a third individual via role \texttt{hasParent} which, in turn, is connected to the second individual via role \texttt{hasBrother}.

### 2.1.3 Axioms

Concept and role expressions may be employed in the construction of axioms, which are formulas that capture partial knowledge about the domain of interest. As customary, we separate DL axioms into two groups: assertional or ABox axioms, and terminological or TBox axioms.
2.1.3.1 ABox axioms

Some ABox axioms, such as the following, capture knowledge the concepts to which named individuals belong to.

\[
\begin{align*}
\text{Mother(mary)} & \quad (2.9) \\
\text{Male(joe)} & \quad (2.10)
\end{align*}
\]

The previous axioms indicate that Mary—the domain individual represented by the named individual \text{mary}—is a mother and Joe—the domain individual represented by the named individual \text{joe}—is male, respectively. ABox axioms such as (2.9) and (2.10) are referred to as concept assertions.

ABox axioms may also be used to describe relations between named individuals.

\[
\begin{align*}
\text{IsParentOf(mary, joe)} & \quad (2.11) \\
\text{SisterOf(mary, mike)} & \quad (2.12)
\end{align*}
\]

The previous ABox axioms are referred to role assertions.

Although common sense may indicate that Mary and Joe are different individuals, this fact does not logically follow from the previous axioms. Note that, DLs do not make the unique name assumption (UNA), and so different named individuals may refer to the same domain individual.

2.1.3.2 TBox axioms

TBox (terminological box) axioms allow us to define relations at a terminological level; i.e., relationships between concept and role expressions. For example, we may use TBox axioms such as the following to indicate that Mother is a subclass of Parent.

\[
\text{Mother} \sqsubseteq \text{Parent} \quad (2.13)
\]

Axioms such as (2.13) are often referred to as subclass axioms.

In a similar manner, we may also use TBox axioms to enforce subclass and equivalence relations over role expressions.

\[
\begin{align*}
\text{hasParent} \circ \text{hasBrother} & \sqsubseteq \text{hasUncle} \quad (2.14) \\
\text{IsMotherOf} & \sqsubseteq \text{IsParentOf} \quad (2.15)
\end{align*}
\]

Axioms (2.14) and (2.15) intuitively indicate the following: the composition relation resulting from concatenating the roles \text{hasParent} and \text{hasBrother} is a subrelation of \text{hasUncle} and the role \text{IsMotherOf} is a subrelation of the rule \text{IsParentOf}.

2.1.4 Ontologies

An ontology is a set of axioms which often satisfies some additional syntactic restrictions. Such restrictions are dependent on the particular DL language employed and will be described at length in the following section.
2.1.5 Reasoning Tasks

Given some ontology, the task of computing the axioms and/or answers to queries entailed by such ontology is referred to as reasoning. There exist several different types of reasoning tasks which are grouped as follows.

1. The standard reasoning tasks: Amongst others, these include ontology satisfiability, i.e., checking whether an ontology entails a logical inconsistency; classification, i.e., computing all axioms of the form $A \sqsubseteq B$, where $A$ and $B$ are concepts, entailed by an ontology; and ABox retrieval, i.e., computing all ABox assertions entailed by an ontology.

2. Conjunctive query answering: This is the task of determining whether a sequence of individuals is an answer of a conjunctive query with respect to a given ontology. A CQ is a formula such as $\exists x, z : \text{attendsCourse}(x, y) \land \text{Course}(y) \land \text{taughtBy}(y, z) \land \text{supervisedBy}(x, z)$ (intuitively, the answer of the previous query would include all pairs of students and supervisors such that the student attends a course taught by the supervisor).
Formal Preliminaries

3.1 Existential Rules

Even though reasoning over programs with existential rules is not the main topic of this thesis, some of the reasoning algorithms presented in later sections are based on this paradigm and thus, we formally introduce it in this chapter.

Let $\mathbf{C}, \mathbf{F}, \mathbf{V}$ and $\mathbf{P}$ be pairwise disjoint and infinite countable sets of constants, function symbols, variables and predicates, respectively, where every function symbol and predicate $s$ is associated with some arity $\text{ar}(s) \geq 1$. Furthermore, $\top, \bot, \approx \in \mathbf{P}$ with $\text{ar}(\top) = \text{ar}(\bot) = 1$ and $\text{ar}(\approx) = 2$. As customary, we often refer to the special predicate $\approx$ as the equality predicate.

The set of terms $\mathbf{T} \supseteq \mathbf{C} \cup \mathbf{V}$ is the minimal set such that, for every function symbol $f$ with $\text{ar}(f) = n$ and every sequence of terms $t_1, \ldots, t_n, f(t_1, \ldots, t_n) \in \mathbf{T}$. The depth $\text{dep}(t)$ of a term $t$ is defined as 0 if $t \in \mathbf{C} \cup \mathbf{V}$, and $\text{dep}(t) = t + \max_{i=1}^n \text{dep}(t_i)$ if $t$ is of the form $f(t_1, \ldots, t_n)$. We often abbreviate a sequence of terms $t_1, \ldots, t_n$ as $\vec{t}$ and identify such a sequence with the set $\{\vec{t}\}$. A term $t'$ is a subterm of another term $t$ if and only if $t' = t$, or $t = f(\vec{t})$ and $t'$ is a subterm of some $s \in \vec{t}$; if additionally $t' \neq t$, then $t'$ is a proper subterm of $t$. A term $t$ is $n$-cyclic if and only if there exists a sequence of terms of the form $f(s_1^\rightarrow), \ldots, f(s_{n+1}^\rightarrow)$ such that $f(s_{n+1}^\rightarrow)$ is a subterm of $t$ and, for every $i = 1, \ldots, n$, $f(s_i^\rightarrow)$ is a proper subterm of $f(s_{i+1}^\rightarrow)$. We simply refer to 1-cyclic terms as cyclic.

An atom is a formula of the form $p(\vec{t})$ where $p$ is some predicate with $\text{ar}(p) = |\vec{t}|$. A fact is a ground atom; i.e., an atom without occurrences of variables.

A substitution is a function mapping terms to terms. Given a substitution $\sigma$ and an atom $\alpha = p(t_1, \ldots, t_n)$, the application of $\sigma$ on $\alpha$ is the atom $\alpha \sigma = p(\sigma(t_1), \ldots, \sigma(t_n))$. The result of the application of a substitution on a sequence of variables, a conjunction of atoms or a rule is defined in the obvious manner. Given a sequence of terms $\vec{t}$, we define $\sigma_{\vec{t}} \subseteq \sigma$ as the maximal substitution only defined for the terms in $\vec{t}$. As customary, we represent the substitution mapping $t_i$ to $u_i$ for all $i = 1, \ldots, n$ with the expression $[t_1/u_1, \ldots, t_n/u_n]$.

Let $t$ be some ground term and $c$ some constant. Let $t_c$ be the term obtained from $t$ by replacing every occurrence of a constant by $c$, i.e., $f(d, g(c))_c = f(c, g(c))$. The notation is analogously extended to facts and sets of facts.
3.2 Horn Description Logics

With \( \phi(\bar{x}) \) we stress that \( \bar{x} = x_1, \ldots, x_n \) are the free variables occurring in the formula \( \phi \). Furthermore, we often identify a conjunction of atoms \( \phi_1 \land \ldots \land \phi_n \) with the set \( \{\phi_1, \ldots, \phi_n\} \).

A rule is a first-order logic (FOL) formula of one of the following forms.

\[
\forall \bar{x} \forall \bar{z} (\beta(\bar{x}, \bar{z}) \rightarrow \exists \bar{y} \eta(\bar{x}, \bar{y})) \tag{3.1}
\]

\[
\forall \bar{x} (\beta(\bar{x}) \rightarrow t \approx u) \tag{3.2}
\]

In the above, \( \beta \) and \( \eta \) are non-empty conjunctions of atoms which do not contain occurrences of function symbols or equality; \( \bar{x}, \bar{y} \) and \( \bar{z} \) are pairwise disjoint; \( \bar{x} \) is non-empty; and \( t, u \in \bar{x} \cup C \). To simplify the notation, we frequently omit the universal quantifiers from rules. As customary, we refer to rules of the forms (3.1) and (3.2) as tuple generating dependencies (TGDs) and equality generating dependencies (EGDs), respectively.

Given a set of rules \( R \), we define \( R^3 \) and \( R^\forall \) as the sets of all the TGDs in \( R \) which do and do not contain existentially quantified variables, respectively. Moreover, let \( R^\approx \) be the set of all EGDs in \( R \). A program is a tuple \( \langle R, I \rangle \) where \( R \) is a set of rules and \( I \) is an instance; i.e., a finite set of equality and function free facts. Given a program \( \langle R, I \rangle \), we assume that, without loss of generality, every predicate in \( R \) occurs in \( I \) and every constant in \( R \) occurs in \( I \). Often, we abuse notation and identify the set of rules and facts \( R \cup I \) with the program \( \langle R, I \rangle \).

For the remainder of the paper, we assume that \( \top \) and \( \bot \) are treated as ordinary unary predicates when dealing with programs and that the semantics of \( \top \) is captured explicitly in any program \( P = \langle R, I \rangle \) by including the rule \( p(x_1, \ldots, x_n) \rightarrow \top(x_1) \land \ldots \land \top(x_n) \) in \( R \) for every predicate \( p \) with arity \( n \) occurring in \( P \).

We will later employ skolemization to define the consequences of a TGD over a set of facts. The skolemization \( sk(\rho) \) of some TGD \( \rho = \beta(\bar{x}, \bar{z}) \rightarrow \exists \bar{y} \eta(\bar{x}, \bar{y}) \) is the rule \( \beta(\bar{x}, \bar{z}) \rightarrow \eta(\bar{x}, \bar{y}) \sigma_{sk} \) where \( \sigma_{sk} \) is a substitution mapping every \( y \in \bar{y} \) into \( f_y^\rho(\bar{x}) \) where \( f_y^\rho \) is a fresh function unique for every variable \( y \) and TGD \( \rho \).

3.2 Horn Description Logics

In this section, we formally define the syntax and semantics of several Horn DL languages considered across this paper. More precisely, we define Horn-SROIQ and Horn-SHOTIQ \cite{Kroetzsch2013}, \( \mathcal{E} \mathcal{L}^{++} \) \cite{Baader2005b}, DL-LiteR \cite{Calvanese2007} and DLP \cite{Grosof2003}. Readers familiar with the Web Ontology Language (OWL) \cite{Hitzler2009} may recognize these languages as the logics that underpin the DL, EL, QL and RL standards, respectively. We assume basic familiarity with the topic and otherwise refer to the literature for further details: for a thorough theoretical introduction see \cite{Baader2007}; an extended introduction to DL and Semantic Web technologies is provided in \cite{Hitzler2009}, where also the relationships between DL and the different OWL standards are explained in detail. Without loss of generality, we restrict our attention to ontologies in a normal form close to those in \cite{Kroetzsch2007, Ortiz2010}.

A DL signature is a tuple \( \langle NC, NR, NI \rangle \) where \( NC, NR \) and \( NI \) are countable and mutually disjoint
3.2. HORN DESCRIPTION LOGICS

\[ A \subseteq B \quad \Rightarrow \quad \bigcap_{i=1}^{n} A_i \subseteq B \]
\[ A \subseteq B \quad \Rightarrow \quad x \in A^T \quad \exists y((x, y) \in R^T \land y \in B^T) \]  
\[ A \subseteq B \quad \Rightarrow \quad (x, y) \in R^T \land y \in A \rightarrow x \in B \]  
\[ \top \subseteq \forall R.B \quad \Rightarrow \quad (x, y) \in R^T \rightarrow y \in B^T \]  
\[ A \subseteq \exists R.a \quad \Rightarrow \quad x \in A^T \rightarrow (x, a^T) \in R^T \]  
\[ R \subseteq S \quad \Rightarrow \quad R^T \subseteq S^T \]  
\[ R^{-} \subseteq S \quad \Rightarrow \quad (x, y) \in R^T \rightarrow (y, x) \in S^T \]  
\[ R \circ S \subseteq V \quad \Rightarrow \quad (x, y) \in R^T \land (y, z) \in S^T \rightarrow (x, z) \in S^T \]  
\[ A(a) \quad \Rightarrow \quad a^T \in A^T \]
\[ R(a, b) \quad \Rightarrow \quad (a^T, b^T) \in R^T \]

Figure 3.1: Syntax and Semantics of Axioms. In the above, \( A_i, B \in N_C \), \( R, S, V \in N_R \) and \( a, b \in N_I \). [86x788]

sets of concepts, roles and individual names, respectively, where we additionally assume that \( \top, \bot \in N_C \). As customary, we refer to \( \top \) and \( \bot \) as the top and bottom concepts, respectively. For the rest of this paper we assume that a DL signature has been fixed and so omit further references to it when possible without introducing ambiguity.

An axiom is a formula of one of the forms given in the left hand side of Figure 3.1. Some of the axioms such figure may appear redundant, as they can be normalized away using other axioms in such figure; nevertheless, these extra axioms are useful for defining the different types of DL languages considered in this document (see Definition 3.2.1).

Axioms of the form \( 3.3 \text{a,b,c,d,e,f,g,h,i} \text{ a,b,c,d,e,f,g,h,i} \) are referred to terminological or TBox axioms. Axioms of the form \( 3.13 \text{ and } 3.14 \text{ a,b,c,d,e,f,g,h,i} \) are referred to assertional or ABox axioms (or simply assertions). A TBox (resp. an ABox) is a set of TBox (resp. ABox) axioms.

Given some TBox \( T \), we define the relation \( \sqsubseteq_T \) over \( N_R \cup \{ R^{-} \mid R \in N_R \} \) as the minimal transitive and reflexive relation such that \( R \sqsubseteq_T S \) and \( R^{-} \sqsubseteq_T S^{-} \) if \( R \subseteq S \in T \), and \( R^{-} \sqsubseteq_T S^{-} \) if \( R^{-} \subseteq S \in T \), where \( R, S \in N_R \).

Definition 3.2.1 We proceed with the definition of all the different types of TBoxes considered across this document.

1. A Horn-SROIQ TBox is a TBox.
2. A Horn-SHOTIQ TBox is a TBox such that, for every axiom of the form \( R \circ S \subseteq V \in T \), \( R = S = V \).
3. A Horn-SRIQ TBox is a TBox which does not contain axioms of the form \( 3.8 \text{ or } 3.9 \text{ a,b,c,d,e,f,g,h,i} \text{ a,b,c,d,e,f,g,h,i} \).
4. An $\mathcal{EL}^{++}$ TBox $\mathcal{T}$ is a TBox which does not contain axioms of the form (3.7) or (3.11). Furthermore, if $\mathcal{T}$ contains axioms the form $R \circ S \sqsubseteq V$ and $\top \sqsubseteq \forall V'.B$ in $\mathcal{T}$ with $V \sqsubseteq \top V'$, then $\top \sqsubseteq \forall S.B \in \mathcal{T}$.

5. A DL-Lite$_R$ TBox $\mathcal{T}$ is a TBox which does not contain axioms of the form (3.3), (3.7), (3.8), (3.9) or (3.12). Furthermore, for every axiom in $\mathcal{T}$ of the form (3.3), $n = 1$.

6. An DLP TBox is a TBox which does not contain axioms of the form (3.4).

An $\mathcal{L}$ ontology, where $\mathcal{L}$ is one of the previously discussed DL languages, is a tuple $\langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ is an $\mathcal{L}$ TBox.

As with programs, we often abuse notation and simply identify the set of axioms $\mathcal{T} \cup \mathcal{A}$ with the ontology $\langle \mathcal{T}, \mathcal{A} \rangle$.

Note that, we disregard certain syntactic restrictions, such as simplicity of roles occurring with numeric quantifiers or role regularity, considered in other definitions of DL [Horrocks et al. 2006]. Such restrictions, imposed in order to maintain decidability of tableau based reasoning algorithms, are unnecessary for our specialized algorithms and thus, we ignore them.

Without loss of generality, instead of dealing with conjunctive query (CQ) answering, we restrict our attention to the simpler task of CQ entailment of boolean conjunctive queries (BCQs). This is without loss of generality since CQ answering can be reduced to checking entailment of BCQs. A BCQ, or simply a query, is a formula of the form $\exists \vec{y} \eta(\vec{y})$ where $\eta$ is a conjunction of atoms not containing occurrences of constants, function symbols nor $\approx$.

As customary, the semantics of ontologies is given through the definition of interpretations. An interpretation $\mathcal{I}$ of some ontology $\mathcal{O}$ is a tuple $(\Delta^\mathcal{I}, \cdot^\mathcal{I})$ where $\Delta^\mathcal{I}$ is a non-empty set, referred to as the domain of $\mathcal{I}$, and $\cdot^\mathcal{I}$ is a function that maps each individual, concept and role in $\mathcal{O}$ to an element, a subset and a binary relation in $\Delta^\mathcal{I}$, respectively, such that $\top^\mathcal{I} = \Delta^\mathcal{I}$ and $\bot^\mathcal{I} = \emptyset$.

Let $\mathcal{I}$ be an interpretation of some ontology $\mathcal{O}$. We say that $\mathcal{I}$ satisfies an axiom $\alpha$, written $\mathcal{I} \models \alpha$, if the corresponding condition shown in Figure 3.1 holds. Furthermore, we say that $\mathcal{I}$ satisfies the ontology $\mathcal{O}$, written $\mathcal{I} \models \mathcal{O}$, if $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{O}$. If this is indeed the case, we say that $\mathcal{I}$ is a model of $\mathcal{O}$.

An ontology $\mathcal{O}$ is satisfiable if there exists at least some interpretation $\mathcal{I}$ which is also a model of $\mathcal{O}$. An ontology $\mathcal{O}$ entails an axiom $\alpha$, written $\mathcal{O} \models \alpha$, if, for every model $\mathcal{I}$ of $\mathcal{O}$, $\mathcal{I} \models \alpha$.

Finally, we introduce several reasoning tasks which will be considered across this document. Let $\mathcal{O}$ be some ontology.

- Ontology satisfiability: Determine whether $\mathcal{O}$ is satisfiable.
- ABox retrieval: Compute all ABox axioms that are entailed by $\mathcal{O}$.
- Classification: Compute all axioms of the form $A \sqsubseteq B$ with $A, B \in N_C$ that are entailed by $\mathcal{O}$. 

3.3. THE RELATION BETWEEN DLs AND EXISTENTIAL RULES

In subsequent chapters, we employ mappings from ontologies to programs in such a way that the latter can be used to solve reasoning tasks over the former. In this section, we briefly introduce one of such mappings which is reused across further sections in this document.

Definition 3.3.1 Given a TBox $T$, let $R(T) = \{\pi(\alpha) | \alpha \in T\}$ where $\pi$ is the function from Figure 3.2. Given an ontology $O = \langle T, A \rangle$, let $P(O) = \langle R(T), A \rangle$.

3.4 Axiomatizations of the Equality Predicate

In FOL, the equality predicate is commonly assumed to have a predefined interpretation. Nevertheless, equality can also be treated as an ordinary predicate with an explicit axiomatization. In this section, we introduce two different approaches to axiomatize this predicate.

3.4.1 The Standard Equality Axiomatization

Definition 3.4.1 Given a formula $\phi$, let $Eq(\phi)$ be the formula that results from substituting every atom of the form $t \approx u$ by $Eq(t, u)$ where $Eq$ is a fresh predicate. Given a program $P = \langle R, I \rangle$, let $Eq(P) = \langle Eq(R), I \rangle$.

Let $P$ be some program. Then, the standard equality axiomatization of $P = \langle R, I \rangle$ is the program $SEA(P) = \langle R', I \rangle$ where $R'$ is the set of rules containing the set $Eq(R)$, the rules (3.15-3.17) from Figure 3.3, and an instance of rule (3.18) for every $n$-ary predicate $p$ in $P$ for every $1 \leq i \leq n$.

As shown by the following result, the standard equality axiomatization preserves CQ entailment.
3.4. AXIOMATIZATIONS OF THE EQUALITY PREDICATE

\[
\top(x) \rightarrow \text{Eq}(x, x) \quad (3.15)
\]

\[
\text{Eq}(y, x) \rightarrow \text{Eq}(x, y) \quad (3.16)
\]

\[
\text{Eq}(x, z) \land \text{Eq}(z, y) \rightarrow \text{Eq}(x, y) \quad (3.17)
\]

\[
p(x_1, \ldots, x_i, \ldots, x_n) \land \text{Eq}(x_i, x'_i) \rightarrow p(x_1, \ldots, x'_i, \ldots, x_n) \quad (3.18)
\]

Figure 3.3: Rules Employed in the Axiomatization of Equality.

**Proposition 3.4.2** Given a program \( P \) and a query \( \gamma \), \( P \models \gamma \) if and only if \( \text{SEA}(P) \models \gamma \).

### 3.4.2 Singularization

Another method to axiomatize the meaning of equality is singularization, first described in [Marnette 2009].

**Definition 3.4.3** A singularization of a conjunct of atoms \( \beta \) is a conjunct of atoms that results from performing the following transformation to \( \beta \): For every variable \( x \) occurring in \( \beta \), (i) arbitrarily select some occurrence of \( x \), (ii) replace each other occurrences of \( x \) with different fresh variables \( x_1, \ldots, x_n \) and, (iii) for every \( i = 1, \ldots, n \), add the conjunct \( \text{Eq}(x, x_i) \) to \( \beta \), where \( \text{Eq} \) is a fresh predicate.

A singularization of a rule \( \rho \) is a rule that results from substituting the body of \( \rho \) by some of its singularizations and replacing every atom of the form \( t \approx u \) in the head with \( \text{Eq}(t, u) \). A singularization of a query \( \exists \vec{x} \beta \) is a query with a singularization of \( \beta \) as body.

A singularization of a set of rules \( \mathcal{R} \) is the set of rules containing rules (3.15-3.17) and exactly one singularization for every rule in \( \mathcal{R} \). The singularization of a program \( P \) is a program that results from replacing the set of rules in \( P \) with some of its singularizations.

Given a conjunction of atoms, query, rule, set of rules or program \( \phi \), let \( \text{Sing}(\phi) \) be the set of all possible singularizations of \( \phi \).

**Example 3.4.4** Let \( \rho = A(x) \land B(x) \land C(x) \rightarrow \exists y (R(x, y) \land B(y)) \). Such rule admits the following singularizations.

\[
A(x) \land B(x_1) \land C(x_2) \land \text{Eq}(x, x_1) \land \text{Eq}(x, x_2) \rightarrow \exists y (R(x, y) \land B(y))
\]

\[
A(x_1) \land B(x) \land C(x_2) \land \text{Eq}(x, x_1) \land \text{Eq}(x, x_2) \rightarrow \exists y (R(x, y) \land B(y))
\]

\[
A(x_1) \land B(x_2) \land C(x) \land \text{Eq}(x, x_1) \land \text{Eq}(x, x_2) \rightarrow \exists y (R(x, y) \land B(y))
\]

As shown in [Marnette 2009], any singularization of a program \( P \) can alternatively be employed in place of \( P \) to solve BCQ entailment.

**Proposition 3.4.5** A program entails a query if and only if every singularization of such program entails every singularization of such query.
3.5 The Chase Algorithm

BCQ entailment over programs and Horn ontologies can, in some cases, be addressed via application of the chase algorithm, a technique where all relevant consequences of an ontology are precomputed, allowing queries to be directly evaluated on the materialized set of facts. In this section we present two variants of the chase algorithm, which are somewhat similar to the oblivious and restricted chase from [Calì et al. 2013].

Definition 3.5.1 A fact \( \rho \) is an oblivious consequence of a TGD \( \sigma = \beta(x, y) \rightarrow \exists \exists \eta(x, z) \) on a set of facts \( F \) if and only if there is some substitution \( \sigma' = \beta(x, y)\sigma \subseteq F \) and \( \phi \in \text{sk}(\eta(x, z))\sigma \) where \( \text{sk}(\eta(x, y)) \) is the head of the (skolemized) TGD \( \text{sk}(\rho) \). A fact \( \phi \) is a restricted consequence of \( \rho \) on \( F \) if and only if there is a substitution \( \sigma \) with (1) \( \beta(x, z)\sigma \subseteq F \) and \( \phi \in \text{sk}(\eta(x, y))\sigma \), and (2) there is no substitution \( \tau \supseteq \sigma \) with \( \eta(x, y)\tau \subseteq F \).

The result of obliviously applying \( \rho \) to \( F \), written \( \rho_{O}(F) \), is the set of all oblivious consequences of \( \rho \) on \( F \). The result of obliviously applying a set of TGDs \( R \) to \( F \), written \( \mathcal{R}_{O}(F) \), is the set \( \bigcup_{\rho \in R} \rho_{O}(F) \cup F \). The result of restrictively applying \( \rho \) to \( F \) (resp., \( R \) to \( F \)), written \( \rho_{R}(T) \) (resp., \( \mathcal{R}_{R}(T) \)), is analogously defined.

Definition 3.5.2 Let \( \sim \) be some total strict order over the set of all terms such that \( t \sim u \) only if \( \text{dep}(t) \leq \text{dep}(u) \). Furthermore, we say that \( t \) is greater than \( u \) with respect to \( \sim \) to indicate \( t \sim u \).

Given a set of EGDs \( R \) and a set of facts \( F \), let \( \rightarrow_{R}^{F} \) be the minimal congruence relation over terms such that \( t \rightarrow_{R}^{F} u \) if and only if there exists some \( \beta(x) \rightarrow t_{\sim} \in R \) and some substitution \( \sigma \) with \( \beta(x)\sigma \subseteq F \) and \( \sigma(t) = t \) and \( \sigma(y) = u \). Let \( \mathcal{R}(F) \) be the set that is obtained from \( F \) by replacing all occurrences of every term \( t \) by \( u \) where \( u \) is the greatest term with respect to \( \sim \) such that \( t \rightarrow_{R}^{F} u \).

Note that we define consequences with respect to sets of rules instead of simply (single) rules as it is customary [Calì et al. 2013]. This allows us to define the chase as a deterministic procedure (modulo \( \sim \)). Also, unlike in [Calì et al. 2013], where a lexicographic order is used to direct the replacement of terms, we employ a type of order which ensures that terms are always replaced by terms of equal or lesser depth. This effectively precludes some terms with larger depth from being introduced during the computation of the chase.

Definition 3.5.3 Let \( \mathcal{P} = \langle \mathcal{R}, \mathcal{I} \rangle \) be some program. The oblivious chase sequence of \( \mathcal{P} \) is the sequence \( F_{0}, F_{1}, \ldots \) such that \( F_{1} = \mathcal{I} \) and, for all \( i \geq 1 \), \( F_{i} \) is the set of facts defined as follows.

- If \( \mathcal{R}^{\sim}(F_{i-1}) \neq F_{i-1} \), then \( F_{i} = \mathcal{R}^{\sim}(F_{i-1}) \).
- If \( F_{i-1} = \mathcal{R}^{\sim}(F_{i-1}) \) and \( F_{i-1} \neq \mathcal{R}_{O}^{\sim}(F_{i-1}) \), then \( F_{i} = \mathcal{R}_{O}^{\sim}(F_{i-1}) \).
- Otherwise, \( F_{i} = \mathcal{R}_{O}^{\sim}(F_{i-1}) \).

The restricted chase sequence of \( \mathcal{P} \) is defined analogously.

For the sake of brevity, we frequently denote the oblivious (resp., restricted) chase of a program \( \mathcal{P} \) with \( \mathcal{P}_{O}^{1}, \mathcal{P}_{O}^{2}, \ldots \) (resp., \( \mathcal{P}_{R}^{1}, \mathcal{P}_{R}^{2}, \ldots \)).
3.5. THE CHASE ALGORITHM

Let \( \mathcal{T} = \{ \text{Film} \sqsubseteq \exists \text{isProdBy}. \text{Producer}, \text{Producer} \sqsubseteq \exists \text{prod}. \text{Film}, \text{isProdBy}^- \sqsubseteq \text{prod}, \text{prod}^- \sqsubseteq \text{isProdBy} \} \)

\( \mathcal{O} = \langle \mathcal{T}, \{ \text{Film}(\mathcal{A}) \} \rangle \)

\( \mathcal{R}(\mathcal{T}) = \{ \rho = \text{Film}(x) \rightarrow \exists y[\text{isProdBy}(x, y) \land \text{Producer}(y)], \) 
\( v = \text{Producer}(x) \rightarrow \exists y[\text{prod}(x, y) \land \text{Film}(y)], \) 
\( \text{isProdBy}(y, x) \rightarrow \text{prod}(x, y), \text{prod}(y, x) \rightarrow \text{isProdBy}(x, y) \} \)

\( \mathcal{P}(\mathcal{O}) = \langle \mathcal{R}(\mathcal{T}), \{ \text{Film}(\mathcal{A}) \} \rangle \)

\( \mathcal{P}(\mathcal{O})^1_R = \{ \text{Film}(\mathcal{A}), \text{isProdBy}(\mathcal{A}, f^\rho_\mathcal{O}(\mathcal{A})), \text{Producer}(f^\rho_\mathcal{O}(\mathcal{A})) \} \)

\( \mathcal{P}(\mathcal{O})^2_R = \{ \text{prod}(f^\rho_\mathcal{O}(\mathcal{A}), \mathcal{A}) \} \cup \mathcal{P}(\mathcal{O})^1_\mathcal{O} \)

\( \mathcal{P}(\mathcal{O})^3_R = \mathcal{P}(\mathcal{O})^2_R \)

\( \mathcal{P}(\mathcal{O})^4_R = \mathcal{P}(\mathcal{O})^3_R \cup \{ \text{prod}(f^\rho_\mathcal{O}(\mathcal{A}), f^\rho_\mathcal{O}(f^\rho_\mathcal{O}(\mathcal{A}))), \text{Film}(f^\rho_\mathcal{O}(f^\rho_\mathcal{O}(\mathcal{A}))), \ldots \} \)

Figure 3.4: Ontology \( \mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle \), Program \( \mathcal{P}(\mathcal{O}) \) and the (Restricted and Oblivious) Chase of \( \mathcal{P}(\mathcal{O}) \).

**Definition 3.5.4** Let \( \mathcal{P} \) be some program and let \( \mathcal{R} \) be some set of rules. Then, the oblivious chase of \( \mathcal{P} \) is the set \( \mathcal{O}(\mathcal{P}) = \bigcup_{i \in \mathbb{N}} \mathcal{P}(\mathcal{O}) \). The restricted chase of \( \mathcal{P} \), written \( \mathcal{RC}(\mathcal{P}) \), is defined analogously.

The oblivious (resp., restricted) chase of \( \mathcal{P} \) terminates if and only if there is some \( i \) such that, for all \( j \geq i \), \( \mathcal{P}(\mathcal{O})^j = \mathcal{P}(\mathcal{O})^i \). Furthermore, the oblivious (resp., restricted) chase of a set of rules \( \mathcal{R} \) terminates if the oblivious (resp., restricted) chase of every program of the form \( \langle \mathcal{R}, \mathcal{T} \rangle \) terminates.

The (restricted or oblivious) chase of a program can be employed to solve CQ entailment [Calì et al. 2013]. I.e., a program \( \mathcal{P} \) entails a query \( \gamma \), written \( \mathcal{P} \models \gamma \), if and only if either \( \mathcal{O}(\mathcal{P}) \models \gamma \) or \( \mathcal{O}(\mathcal{P}) \models \gamma \) (resp., \( \mathcal{RC}(\mathcal{P}) \models \gamma \) or \( \mathcal{RC}(\mathcal{P}) \models \gamma \)). Thus, we may also use the chase to solve CQ entailment over ontologies: An ontology \( \mathcal{O} \) entails a query \( \gamma \) if and only if \( \mathcal{O}(\mathcal{P}(\mathcal{O})) \models \gamma \) or \( \mathcal{O}(\mathcal{P}(\mathcal{O})) \models \gamma \) (resp., \( \mathcal{RC}(\mathcal{P}(\mathcal{O})) \models \gamma \) or \( \mathcal{RC}(\mathcal{P}(\mathcal{O})) \models \gamma \)).

In the above paragraph, \( \models \) represents the standard FOL entailment relationship. As customary, given a set of facts \( F \) and a query \( \gamma = \exists \exists \beta \), \( F \models \gamma \) if and only if there is some substitution \( \sigma \) such that \( \beta \sigma \subseteq F \).

For readability purposes, we say that the oblivious (resp. restricted) chase of some ontology \( \mathcal{O} \) terminates if and only if the oblivious (resp. restricted) chase of \( \mathcal{P}(\mathcal{O}) \) terminates. The oblivious (resp. restricted) chase of some TBox \( \mathcal{T} \) terminates if and only if the oblivious (resp. restricted) chase of \( \mathcal{R}(\mathcal{T}) \) terminates.

The restricted chase has a relevant advantage over the oblivious chase: in some cases, the former might terminate whereas the latter does not.

**Example 3.5.5** Let \( \mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle \) be as in Figure 3.4. This figure depicts the oblivious and restricted chase sequence of \( \mathcal{P}(\mathcal{O}) \). In this case, \( \mathcal{RC}(\mathcal{P}(\mathcal{O})) \) terminates whilst \( \mathcal{O}(\mathcal{P}(\mathcal{O})) \) does not.
4

The \( \text{RCA}_n \) Fragment of Horn DL

In some cases, CQ answering over Horn DL can be addressed via application of the chase algorithm (see Section 3.5 for a thorough description of this reasoning algorithm). However, the chase is not guaranteed to terminate for all ontologies, and checking whether it does is not a straightforward procedure. It is thus an ongoing research endeavor to establish so-called acyclicity conditions; i.e., sufficient conditions which characterize ontologies for which the chase does indeed terminate.

The main contribution presented in this chapter is the definition of restricted chase acyclicity (\( \text{RCA}_n \)), a novel acyclicity condition applicable to Horn-\( \Sigma_1 \text{IQ} \) ontologies. If an ontology is proven to be \( \text{RCA}_n \), then \( n \)-cyclic terms do not occur during the computation of the chase of such ontology and thus, the chase is guaranteed to terminate. Note that, an acyclicity notion such as \( \text{RCA}_n \) may be regarded as the definition of a language; i.e., the \( \text{RCA}_n \) fragment of Horn DL is the set of all Horn ontologies that are \( \text{RCA}_n \).

In contrast with existing acyclicity notions [Cuenca Grau et al. 2013] which deal with termination of the oblivious chase of arbitrary sets of existential rules, we restrict our attention to the language Horn-\( \Sigma_1 \text{IQ} \) and seek to achieve termination of the restricted chase algorithm. As described in Section 3.5, this is a special variant of the chase in which the inclusion of further terms to satisfy existential restrictions is avoided if such restrictions are already satisfied. By considering such a chase algorithm we are able to devise acyclicity conditions which are empirically more general than any other of the notions previously described.

On the theoretical side, we show that \( \text{RCA}_n \) is more general than model-faithful acyclicity (MFA) provided \( n \) is linearly large size of ontology. As shown in [Cuenca Grau et al. 2013], MFA is one of the most general acyclicity conditions for ontologies described to date, as it encompasses many other existing notions such as joint acyclicity [Krötzsch and Rudolph 2011], super-weak acyclicity [Marnette 2009] or the hybrid acyclicity notions presented in [Baget et al. 2014]. Furthermore, we show that deciding \( \text{RCA}_n \) membership of an ontology is not harder than deciding MFA membership.

On the practical side, we empirically show that (i) \( \text{RCA}_n \) characterizes more real-world ontologies as acyclic than MFA. Furthermore, we demonstrate that (ii) the use of \( \text{RCA}_n \) results in a more efficient reasoning procedure. This is because acyclicity is still preserved in the case when employing renaming techniques when reasoning in the presence of equality. Thus, the use of cumbersome axiomatizations
of equality such as singularization [Marnette 2009] can be avoided. Moreover, we report on an implementation of the restricted chase algorithm based on the datalog engine RDFOx [Nenov et al. 2015] and show that (iii) it vastly outperforms state-of-the-art DL reasoners. To verify (i-iii), we complete an extensive evaluation with very encouraging results.

4.1 Model Faithful Acyclicity

In this section we briefly describe MFA [Cuenca Grau et al. 2013], one of the most general acyclicity notions for sets of existential rules. MFA guarantees the termination of the oblivious chase of a program by precluding the occurrence of cyclic terms in the chase.

When one is interested in checking the termination of the oblivious chase of a set of rules with respect to every possible instance, it is enough to check termination with respect to a special instance, the critical instance [Marnette 2009]. The critical instance is the minimal set which contains all possible atoms that can be formed using the relational symbols which occur in TGDs and the special constant ⋆. Such a strategy is used by MFA to guarantee termination of a set of rules.

Definition 4.1.1 The critical instance $I_⋆(R)$ of a set of rules $R$ is the set of all facts that can be constructed using relational symbols in $R$ and the fresh constant ⋆.

Definition 4.1.2 A set of TGDs $R$ is MFA if and only if no cyclic term occurs in the oblivious chase sequence of the program $⟨R, I_⋆(R)⟩$.

The previous definition may also be applied to set of axioms in the following manner. A TBox $T$ is MFA if and only if the set $R(T)$ is MFA (the set $R(T)$ is introduced in Definition 3.3.1). Note that, a condition such as MFA can be applied to check whether a TBox $T$ is acyclic; i.e., $T$ is MFA if and only if $R(T)$ is MFA.

While the actual definition of MFA does not preclude the existence of EGDs, equality is assumed to be axiomatized, and thus it is treated as a regular predicate (EGDs are de facto TGDs). Nevertheless, the use of the standard equality approach to axiomatize equality tends to make the MFA membership check fail (see [Cuenca Grau et al. 2013] for a lengthier discussion on this topic). Thus, the use of singularization [Marnette 2009] (introduced in Section 3.4.2), a somewhat “less-harmful” axiomatization of equality, is proposed in [Cuenca Grau et al. 2013]. The use of singularization results in the definition of several variants of MFA, which are introduced in the following definition.

Definition 4.1.3 For a set of TGDs $R$, if there is some set of rules in $\text{Sing}(R)$ which is MFA, then $R$ is MFA$^\exists$. If every set in $\text{Sing}(R)$ is MFA, then $R$ is MFA$^\forall$.

Due to the high number of possible singularizations, it is frequently not feasible to check MFA$^\exists$ or MFA$^\forall$ membership. Given a set of TGDs $R$, a simpler alternative is to check whether $\bigcup_{R' \in \text{Sing}(R)} R'$ is MFA. If that is the case, then $R$ is said to be MFA$^\cup$. Note that given some Horn-$\text{SRI}\text{Q}$ TBox $T$, the set $|\bigcup_{R' \in \text{Sing}(R(T))} R'|$ is actually polynomial in $|T|$ and, as such, MFA$^\cup$ is more feasible to check. Thus, we will use MFA$^\cup$ as a baseline for the empirical evaluation of our novel acyclicity condition RCA$^\forall$, which is introduced in the next section.
4.2 Restricted Chase Acyclicity

While MFA is quite a general acyclicity condition, it has two main drawbacks:

1. It only considers the oblivious chase, which as we have seen in Example 3.5.5 might not terminate (even though the restricted chase does!), and

2. its treatment of equality via singularization is cumbersome and inefficient in practice. Not only MFA $\exists$ and MFA $\forall$ are difficult to check, but even after a set of TGDs are established to belong to some MFA subclass, one has to employ a singularized program for reasoning purposes.

In this section, we present RCA$_n$, an acyclicity notion with neither of these drawbacks: RCA$_n$ verifies termination of the restricted chase of a TBox and does not require the use of cumbersome axiomatizations of the equality predicate.

Since we are primarily interested in termination of the restricted chase of a Horn-$\mathcal{SRIQ}$ TBox, one might wonder why we do not simply check for termination of the restricted chase for such a TBox with respect to the critical instance, as it is done in the previous section with the oblivious chase. Unfortunately, this is not possible: The restricted chase of any set of existential rules always terminates with respect to the critical instance. Thus, we have to devise more sophisticated techniques to check the termination of the restricted chase. We start by introducing the notion of an overchase for a TBox.

**Definition 4.2.1** A set of facts $\mathcal{V}$ is an overchase for some TBox $\mathcal{T}$ if and only if, for every $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, $\text{RC}(\mathcal{P}(\mathcal{O})) \ast \subseteq \mathcal{V}$.

As defined in Chapter 3, the set $\text{RC}(\mathcal{P}(\mathcal{O})) \ast$ is the set that results from substituting every syntactic occurrence of constant in $\text{RC}(\mathcal{P}(\mathcal{O}))$ by $\ast$. Given some TBox $\mathcal{T}$, an overchase for $\mathcal{T}$ may be intuitively regarded as an over-approximation of the restricted chase of $\mathcal{T}$.

**Lemma 4.2.2** If there exists a finite overchase for a TBox, then the restricted chase of such TBox terminates.

**Proof 1** Let $\mathcal{V}$ be some finite overchase of $\mathcal{T}$. By the definition of an overchase, we have that, for every ontology of the form $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, $|\text{RC}(\mathcal{P}(\mathcal{O}))| \cdot |\text{N}_I(\mathcal{O})|^2 \leq |\mathcal{V}|$ where $\text{N}_I(\mathcal{O})$ is the set of all individuals in $\mathcal{O}$. Note that, the maximum arity of a fact occurring in $\text{RC}(\mathcal{P}(\mathcal{O}))$ is 2. By Lemma 4.4.2 we have that, once a fact $\phi$ is removed from the chase sequence, then $\phi$ may not reoccur. Thus, having an upper bound on the number of facts that may occur in $\text{RC}(\mathcal{O})$ for every program of the form $\mathcal{P}(\mathcal{O})$ where $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, termination of the restricted chase of $\mathcal{T}$ is guaranteed.

Thus, to determine whether the chase of a TBox $\mathcal{T}$ terminates, we introduce a procedure to compute an overchase for $\mathcal{T}$ and a means to check its termination. We proceed with some preliminary notions and notation.
4.2. RESTRICTED CHASE ACYCLICITY

Definition 4.2.3 Let $T$ be some TBox and $t$ a term. Let $\mathcal{I}(t)$ be the set of facts defined as follows: If $t$ is of the form $f^\rho_p(s)$ where $\rho = A(x) \rightarrow \exists y[R(x,y) \land B(y)]$, then $\mathcal{I}(t) = \{A(s), R(s,t), B(t)\} \cup \mathcal{I}(s)$; otherwise, $\mathcal{I}(t) = \emptyset$. Furthermore, we introduce the program $U(T, t) = (\mathcal{R}(T)^\forall \cup \mathcal{R}(T)^\exists, \mathcal{I}(t))$.

Intuitively, the restricted chase of the program $U(T, t)$ can be regarded as some kind of underapproximation of the facts that must occur in the chase of every program of the form $P(\langle T, A \rangle)$ where $t$ occurs. I.e., if $t$ occurs in the restricted chase sequence of any program $P(\langle T, A \rangle)$, then the facts in the restricted chase of $U(T, t)$ must also occur (up to renaming) in the chase sequence of such program. Furthermore, due to the special priority of application of the rules during the computation of the chase, the facts in the restricted chase of $U(T, t)$ must occur in the restricted chase sequence of every program of the form $P(\langle T, A \rangle)$ before any successors of $t$ are introduced.

Example 4.2.4 Let $O$, $\rho$ and $\psi$ be the ontology and rules from Example 3.5.5. Then, by Definition 4.2.3,

$\mathcal{I}(f^\rho_p(AI)) = \{\text{Film}(AI), \text{isProdBy}(AI, f^\rho_p(AI)), \text{Producer}(f^\rho_p(AI))\}$ and

$\mathcal{R}(U(T, f^\rho_p(AI))) = \{\text{prod}(f^\rho_p(AI), AI)\} \cup \mathcal{I}(f^\rho_p(AI))$.

All the facts in the restricted chase of $U(T, t)$ occur in the restricted chase sequence of $P(O)$ before any successors of term $f^\rho_p(AI)$ are introduced. This is because the rule $\text{isProdBy}(y,x) \rightarrow \text{prod}(x,y)$ is applied with a higher priority than the rule $\psi = \text{Producer}(x) \rightarrow \exists y[\text{prod}(x,y) \land \text{Film}(y)]$.

Given a TBox $T$ and some term of the form $f^\rho_p(t)$, we can in some cases conclude that such a term may never occur during the computation of the restricted chase of every program of the form $P(\langle T, A \rangle)$ by carefully inspecting the facts in the set $U(T, t)$.

Definition 4.2.5 Let $T$ be a TBox and $t$ a term of the form $f^\rho_p(s)$ where $\rho = A(x) \rightarrow \exists y[R(x,y) \land B(y)]$. We say that a term $t$ is restricted with respect to $T$ if and only if there is some term $u$ with $\{R([s], u), B(u)\} \subseteq \mathcal{R}(U(T, s))$ where $[s] = [v]$, if $s$ is replaced by $v$ during the computation of the restricted chase sequence; and $[s] = s$, otherwise.

We often simply say that a term is “restricted”, instead of “restricted with respect to $T$,” if the TBox $T$ is clear from the context.

Lemma 4.2.6 Let $T$ be a TBox and $t$ a restricted term. Then, for every possible $O = \langle T, A \rangle$, $t \notin \mathcal{R}(P(O))$.

Proof 2 (Sketch) Let $t$ be a term of the form $f^\rho_p(s)$ where $\rho = A(x) \rightarrow \exists y[R(x,y) \land B(y)]$. We can verify that, if $t$ occurs during the computation of the chase sequence, then every fact $\mathcal{R}(U(T, s))$ will also be included in such chase sequence before any new terms are introduced. Thus, if $t$ is indeed restricted, there must be some $u$ with $R([s], u)$ and $B(u)$ occurring in the chase sequence. Therefore, by the definition of the chase, the term $t$ may never be derived.

For a complete proof of the previous lemma, see Section 4.4.2.
4.2. RESTRICTED CHASE ACYCLICITY

\[ \forall \text{-rule if there is some TGD of the form } \rho = \beta(x, y) \rightarrow \eta(x) \in \mathcal{R}(T) \]

\[ \text{then } \mathcal{V}_T \rightarrow \rho_R(\mathcal{V}_T) \cup \mathcal{V}_T \]

\[ \exists \text{-rule if there is some TGD of the form } \rho = A(x) \rightarrow \exists y[R(x, y) \land B(y)] \in \mathcal{R}(T) \text{ and there exists some substitution } \sigma \text{ such that (i) } A(x)\sigma \subseteq \mathcal{V}_T \text{ and (ii) } f^\rho_y(x)\sigma \text{ is not restricted with respect to } T \]

\[ \text{then } \mathcal{V}_T \rightarrow \{ R(x, f^\rho_y(x)), B(f^\rho_y(x)) \}\sigma \cup \mathcal{V}_T \]

\[ \approx \text{-rule if there is some EGD } \beta(x, y) \rightarrow x \approx y \in \mathcal{R}(T) \text{ and there exists some substitution } \sigma \text{ such that } \beta(x, y)\sigma \subseteq \mathcal{V}_T \]

\[ \text{then } \mathcal{V}_T \rightarrow \{ \text{Eq}(x, y), \text{Eq}(y, x) \}\sigma \cup \mathcal{V}_T \]

\[ \text{Eq-rule if there are some terms } t, u \text{ and } u_i \text{ where } i = 1, \ldots, n \text{ and some predicate } p \text{ such that (i) } p \neq \text{Eq}, (ii) } \{ \text{Eq}(t, u), p(u_1, \ldots, u_n) \} \subseteq \mathcal{V}_T, (iii) \text{dep}(t) \leq \text{dep}(u) \text{ and (iv) } u = u_j \text{ for some } j = 1, \ldots, n \]

\[ \text{then } \mathcal{V}_T \rightarrow \{ p(u_1, \ldots, u_n) \}[u/t] \cup \mathcal{V}_T \]

Figure 4.1: Expansion rules for the construction of \( \mathcal{V}_T \).

**Example 4.2.7** Let \( T, \rho \) and \( \nu \) be the TBox and rules from Example 3.5.5. We proceed to show that the term \( f^\nu_y(f^\rho_y(\mathcal{A})) \) is restricted. First, we compute the restricted chase of \( \mathcal{U}(T, f^\nu_y(\mathcal{A})) \).

\[
\mathcal{R}(\mathcal{U}(T, f^\nu_y(\mathcal{A}))) = \{ \text{Producer}(\mathcal{A}), \text{prod}(\mathcal{A}), f^\nu_y(\mathcal{A}), \text{Film}(f^\nu_y(\mathcal{A})), \text{ProdBy}(f^\nu_y(\mathcal{A}), \mathcal{A}) \}
\]

Note that \( \{ \text{ProdBy}(f^\nu_y(\mathcal{A}), \mathcal{A}), \text{Producer}(\mathcal{A}) \} \subseteq \mathcal{R}(\mathcal{U}(T, f^\nu_y(\mathcal{A}))) \). Thus, \( f^\nu_y(f^\rho_y(A)) \) is restricted with respect to \( T \) and, by Lemma 4.2.6, it may not occur in the restricted chase of a program of the form \( \mathcal{P}(\langle T, \mathcal{A} \rangle) \). Furthermore, by Definition 4.2.5, if \( f^\nu_y(f^\rho_y(\mathcal{A})) \) is restricted, then every term of the form \( f^\nu_y(f^\rho_y(c)) \), where \( c \) is a constant, is also restricted.

With Definition 4.2.3 and Lemma 4.2.6 in place, we proceed with the definition of a procedure to construct an overchase for some given TBox \( T \).

**Definition 4.2.8** Let \( T \) be a TBox. We define \( \mathcal{V}_T \) as the set initially containing every fact in \( \mathcal{I}_s(\mathcal{R}(T)) \) which is then expanded by exhaustively applying the rules in Figure 4.1 (in non-deterministic order).

**Lemma 4.2.9** The set \( \mathcal{V}_T \) is an overchase of the TBox \( T \).

**Proof (Sketch)** The lemma can be proven via induction on chase sequence of any ontology of the form \( \mathcal{O} = \langle T, \mathcal{A} \rangle \). Note that, \( \mathcal{O}_R^0 \subseteq \mathcal{V}_T \) by the definition of \( \mathcal{V}_T \). It can be verified that, for every possible derivation of a set of facts during the computation of the chase of \( \mathcal{O} \), such facts will always be contained in \( \mathcal{V}_T \).

See Section 4.4.3 for a complete proof of the previous lemma.

**Corollary 4.2.10** The restricted chase of some TBox \( T \) terminates if \( \mathcal{V}_T \) is finite.
Example 4.2.11 Let $\mathcal{T}$ be the TBox from Example 3.5.5. Then $\mathcal{V}_T$ is as follows.

$$
\mathcal{V}_T = \{ \text{Film}(\star), \text{isProdBy}(\star), \text{Producer}(\star), \text{prod}(\star, \star),
\text{isProdBy}(\star, f\rho(\star)), \text{Producer}(f\rho(\star)), \text{prod}(\star, f\rho(\star)), \text{Producer}(f\rho(\star)) \}
$$

Note that terms $f\rho(f\rho(\star))$ and $f\rho(f\rho(\star))$ are restricted and thus, they are not included in $\mathcal{V}_T$. Since $\mathcal{V}_T$ is finite, we can conclude termination of the restricted chase of the TBox $\mathcal{T}$.

In the previous example, we were able to ascertain termination of the restricted chase of $\mathcal{T}$ after verifying that the set $\mathcal{V}_T$ is finite. A sufficient condition for finiteness of $\mathcal{V}_T$ is to only allow cyclic terms up to a certain depth in this set. We use such condition to formally define RCA$_n$.

Definition 4.2.12 A TBox $\mathcal{T}$ is RCA$_n$ if and only if there are no $n$-cyclic terms in $\mathcal{V}_T$. An ontology $\langle \mathcal{T}, \mathcal{A} \rangle$ is RCA$_n$ if and only if $\mathcal{T}$ is RCA$_n$.

Theorem 4.2.13 If a TBox $\mathcal{T}$ is RCA$_n$ then the restricted chase of $\mathcal{T}$ terminates.

We proceed with several results regarding the complexity of deciding RCA$_n$ membership and reasoning over RCA$_n$ ontologies.

Theorem 4.2.14 Deciding whether some TBox $\mathcal{T}$ is RCA$_n$ is in ExpTime.

Theorem 4.2.15 Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be some RCA$_n$ ontology and $\gamma$ a query. Then, checking whether $\mathcal{O} \models \gamma$ is ExpTime-complete.

To close the section, we present several results in which we theoretically compare the generality of RCA$_n$ to MFA$^\cup$.

Theorem 4.2.16 MFA$^\cup$ does not cover RCA$_1$.

Proof 4 The TBox $\mathcal{T}$ from Example 3.5.5 is RCA$_1$ but not MFA$^\cup$.

Theorem 4.2.17 If $\mathcal{T}$ is MFA$^\cup$ then $\mathcal{T}$ is RCA$_n$ for every $n > |\mathcal{T}|^2$ where $\mathcal{T}^2$ is the set of all existential axioms in $\mathcal{T}$.

The proof of the previous result can be found in Section 4.4.5.

4.3 Evaluation

4.3.1 An Empirical Comparison of RCA$_n$ and MFA$^\cup$

In this section we include an empirical comparison of the generality of RCA$_n$ and MFA$^\cup$. For our experiments, we use the TBoxes of the ontologies in the OWL Reasoner Evaluation workshop (ORE, https://www.w3.org/community/owled/ore-2015-workshop/) and Ontology Design Patterns (ODP, http://www.ontologydesignpatterns.org) datasets. The former is a large repository
used in the ORE competition containing a large corpus of ontologies. The latter contains a wide range of smaller ontologies that capture design patterns commonly used in ontology modeling. The ORE dataset is rather large, and thus we restrict our experiments to the 294 ontologies with the smallest number of existential axioms, while skipping the 77 ontologies with the largest number of existential axioms. The number of such axioms contained in an ontology is a useful metric to predict the “hardness” of acyclicity membership tests; i.e. running these experiments would be very time-intensive, while our results, reported below, already indicate that for such very hard TBoxes MFA and RCA will likely not differ much (while they differ significantly for ontologies with a lower count of existential axioms).

Horn-$\mathcal{SRIQ}$ TBoxes which can be expressed in $\mathcal{EL}^{++}$, DL-Lite$_R$ and DLP were not considered in our experiments. This is because all DLP TBoxes are acyclic (with respect to every applicable acyclicity notion known to us), and there already exist effective algorithms and efficient implementations that solve CQ answering over $\mathcal{EL}^{++}$ and DL-Lite$_R$ ontologies [Stefanoni et al. 2014] (albeit, if these do not include complex roles).

The results from our experiments are summarized in Figure 4.1. The evaluated TBoxes are sorted into brackets depending on the number of existential axioms they contain. For each bracket we provide the average number of axioms in the ontologies (“Avg. Size”), the number of ontologies (“Count”), and, for every condition “X” considered, the percentage of “X acyclic” ontologies.

RCA$_2$ and RCA$_3$ turned out to be indistinguishable with respect to the TBoxes considered and thus, we limit our evaluation to RCA$_n$ with $n \leq 3$. Our tests reveal that RCA$_2$ is significantly more general than MFA, particularly when it comes to TBoxes with a low count of existential axioms. However note that reasoning over ontologies with few (existential) axioms is in general not trivial: All of the ontologies considered in our materialization tests (see Figure 4.2) contain less than 20 existential axioms. For TBoxes containing from 1 to 10 existential axioms in the ORE dataset, more than half of the ontologies which are not MFA are RCA$_2$. Furthermore, the 4 ontologies in the ODP dataset which are not MFA are RCA$_2$. Interestingly, in both repositories we could not find any ontology that is MFA but not RCA$_1$. Thus, with respect to the TBoxes in our corpus, RCA$_1$ already proves to be more general than MFA$_\cup$.

In total, we looked at 312 ontologies, 62% and 75% of which are MFA$_\cup$ and RCA$_2$, respectively. To gauge the significance of this improvement, we roughly compare these numbers with the results presented in [Cuenca Grau et al. 2013]. In that paper, the authors consider a total of 336 ontologies, of which 49%, 58% and 68% are weakly acyclic [Fagin et al. 2005], jointly acyclic [Krötzsch and Rudolph 2011] and MFA$_\cup$, respectively. Even though the comparison is not over the same TBoxes, we verify that the improvement in generality of our notion is in line with previous iterations of related work.

4.3.2 A Materialization Based Reasoner

We now report on an implementation of the restricted chase as defined in Section 3.5. Moreover, we also present an implementation of the oblivious chase with singularization, i.e., the chase as it must
4.3. EVALUATION

<table>
<thead>
<tr>
<th>3-Axioms</th>
<th>Avg. Size</th>
<th>Count</th>
<th>MFA</th>
<th>RCA₁</th>
<th>RCA₂</th>
<th>RCA₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>175</td>
<td>70</td>
<td>70.0</td>
<td>87.1</td>
<td>92.9</td>
<td>92.9</td>
</tr>
<tr>
<td>6-10</td>
<td>219</td>
<td>48</td>
<td>58.3</td>
<td>83.3</td>
<td>83.3</td>
<td>83.3</td>
</tr>
<tr>
<td>11-25</td>
<td>916</td>
<td>54</td>
<td>83.3</td>
<td>85.2</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>26-100</td>
<td>521</td>
<td>42</td>
<td>54.8</td>
<td>59.5</td>
<td>61.9</td>
<td>61.9</td>
</tr>
<tr>
<td>101-500</td>
<td>1290</td>
<td>42</td>
<td>26.2</td>
<td>26.2</td>
<td>28.6</td>
<td>28.6</td>
</tr>
<tr>
<td>501-1922</td>
<td>5052</td>
<td>38</td>
<td>60.5</td>
<td>60.5</td>
<td>60.5</td>
<td>60.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3-Axioms</th>
<th>Size</th>
<th>Total</th>
<th>MFA</th>
<th>RCA₁</th>
<th>RCA₂</th>
<th>RCA₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-12</td>
<td>39</td>
<td>18</td>
<td>73.7</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4.1: Results for the ORE and ODP Repositories.

be used if we employ MFA⁻¹ (see Section 4.1). We use the datalog engine RDFOx [Nenov et al. 2015] in both implementations.

We evaluate the performance of our chase based implementations against Konclude [Steigmiller et al. 2014], a very efficient OWL DL reasoner, and PAGOda [Zhou et al. 2015], a hybrid approach to query answering over ontologies. PAGOda combines a datalog reasoner with a fully-fledged OWL 2 reasoner in order to provide scalable ’pay-as-you-go’ performance and is, to the best of our knowledge, the only other implementation that may solve CQ answering over Horn-\textsc{Sriq} ontologies with completeness guarantees, albeit only in some cases. Nevertheless, PAGOda was able to solve all the queries (that is, all of which for which it did not time-out or run out of memory) in this evaluation in a sound and complete manner.

We consider two real-world ontologies in our experiments, Reactome and Uniprot, and two standard benchmarks, LUBM and UOBM, all of which contain a large amount of ABox axioms. Axioms in these ontologies which are not expressible in Horn-\textsc{Sriq} were pruned. Furthermore, one extra axiom had to be removed from Uniprot for it to be both MFA⁻¹ and RCA₁ acyclic.

The results from our experiments are summarized in Figure 4.2. For each ontology, we consider four samples of the original ABox. The number of triples contained in each one of these is indicated at the beginning of each row, under the column “Triples Count.” As previously mentioned, we consider four different implementations: These include the two aforementioned variants of the chase (“Restricted” and “Oblivious”), PAGOda (“PAGOda”) and Konclude (“Konc.”). For both chase based implementations, we check the time it takes to compute the chase (“C”) and then the time to solve each of the four queries crafted for each ontology (“Q1-Q4”). In a similar manner, we list the time PAGOda takes to preprocess each ontology (“P”) plus the time it takes to answer the queries (“Q1-Q4”). Finally, we list the time Konclude takes to solve realization; i.e., the task of computing every fact of the form \( A(a) \) entailed by an ontology (note that Konclude cannot solve arbitrary CQ
4.3. EVALUATION

Table 4.2: Results for Reactome, Uniprot, LUBM and UOBM (sorted from top to bottom in the above table).

<table>
<thead>
<tr>
<th>Triples Count</th>
<th>Restricted</th>
<th>Oblivious</th>
<th>PAGOda</th>
<th>Konc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Q1-Q4</td>
<td>C</td>
<td>Q1-Q4</td>
<td>P</td>
</tr>
<tr>
<td>2.8M</td>
<td>10 0 0 0 0</td>
<td>45 0 0 TO 0</td>
<td>89 OM 4 1</td>
<td>75</td>
</tr>
<tr>
<td>5.1M</td>
<td>21 0 0 0 0</td>
<td>138 0 0 TO 3</td>
<td>147 OM 1 2</td>
<td>214</td>
</tr>
<tr>
<td>6.7M</td>
<td>28 0 0 0 0</td>
<td>1029 2 0 TO 0</td>
<td>203 OM 2 3</td>
<td>506</td>
</tr>
<tr>
<td>8.1M</td>
<td>36 37 0 0 0</td>
<td>TO - - - -</td>
<td>263 OM 2 2</td>
<td>1347</td>
</tr>
<tr>
<td>9.0M</td>
<td>37 0 0 0 0</td>
<td>OM - - - -</td>
<td>113 1 1 1</td>
<td>198</td>
</tr>
<tr>
<td>17.8M</td>
<td>72 0 0 0 0</td>
<td>OM - - - -</td>
<td>232 2 2 3</td>
<td>987</td>
</tr>
<tr>
<td>26.2M</td>
<td>107 0 0 0 0</td>
<td>OM - - - -</td>
<td>378 4 10 12</td>
<td>3491</td>
</tr>
<tr>
<td>33.9M</td>
<td>141 0 1 0 0</td>
<td>OM - - - -</td>
<td>521 6 21 21</td>
<td>12  TO</td>
</tr>
<tr>
<td>2.8M</td>
<td>8 0 0 0 1</td>
<td>70 0 0 0 74</td>
<td>51 OM 0 0 0</td>
<td>51</td>
</tr>
<tr>
<td>5.7M</td>
<td>16 0 0 0 2</td>
<td>158 1 1 1 154</td>
<td>99 OM 1 1 0</td>
<td>118</td>
</tr>
<tr>
<td>8.4M</td>
<td>26 0 0 0 3</td>
<td>242 1 1 2 186</td>
<td>142 OM 2 1 1</td>
<td>220</td>
</tr>
<tr>
<td>11.4M</td>
<td>37 1 0 0 5</td>
<td>341 2 2 3 311</td>
<td>197 OM 3 1 1</td>
<td>315</td>
</tr>
<tr>
<td>2.2M</td>
<td>11 0 0 0 0</td>
<td>56 0 0 0 1</td>
<td>61 28 0 TO 1</td>
<td>53</td>
</tr>
<tr>
<td>4.5M</td>
<td>27 2 0 0 0</td>
<td>133 0 0 1 2</td>
<td>121 60 0 TO 2</td>
<td>125</td>
</tr>
<tr>
<td>6.6M</td>
<td>42 3 1 1 0</td>
<td>216 1 1 2 3</td>
<td>186 TO 0 TO 5</td>
<td>292</td>
</tr>
<tr>
<td>8.9M</td>
<td>58 5 1 2 1</td>
<td>310 1 2 4 6</td>
<td>260 TO 0 TO 5</td>
<td>644</td>
</tr>
</tbody>
</table>

answering). Time-outs, indicated with “TO,” were set at 1 hour for materialization and 5 minutes for queries. We make use of the acronym “OM” to indicate that an out-of-memory error occurred. Sometimes, a time-out or an out of memory error prevents us from answering the queries: Such a situation is indicated with “-.” All experiments were performed on a MacBook Pro with 8GB of RAM and a 2.4 GHz Intel Core i5 processor.

For each ontology, we consider four different queries which are listed in Figure 4.2 (such figure can be found at the end of the chapter). For every ontology, the query Q1 is of the form $\exists x, y, z R(x, y) \land R(z, y)$ where $R$ is an existentially quantified role occurring in the TBox. It appears that PAGOda has trouble with this kind of query, whereas the chase based implementations efficiently solve it in all but one case. This is probably due to the design of the hybrid reasoner which considers under and over approximations to provide complete answers to CQ: It appears that queries as the one previously considered find a large number of matches in the upper bound which slows down the performance of this reasoner. Queries Q2, and Q3 and Q4 are acyclic and cyclic, respectively (a query is acyclic if the shape of its body is acyclic). Even though it is well-known that answering acyclic CQs can be reduced to satisfiability [Carral and Hitzler 2012], we included such a type of query in our evaluation in an attempt to verify whether solving acyclic queries is simpler than cyclic queries (this is indeed the case theoretically). Nevertheless, our experiments do not reveal any significant differences.
4.4 Proofs

This section contains the complete formal proofs for all the lemmas and theorems presented across the chapter as well as several further intermediate and preliminary results.

4.4.1 Overchase and Termination

**Lemma 4.4.1** Let $P = \langle R, I \rangle$ be some program and $t$ be a term occurring in some $P^i_R$. If there is some $P^j_R$ such that $t \notin P^i_R$ and $j \geq i$, then $t \notin P^k_R$ for every $k \geq j$.

**Proof 5** It is clear that the Lemma holds if $t$ is a constant since constants may never occur in the set of consequences of a rule or a set of rules (note that, by definition, rules may not contain occurrences of constants). Thus, let us assume that $t$ is of the form $f(y_{\rho}(t_1, \ldots, t_n)$ where $\rho = \beta(x, z) \rightarrow \exists y(x, y)$ and $x = x_1, \ldots, x_n$. Furthermore, let $l$ be the smallest natural number such that $t \in P^l_R$.

Note that, the term $t$ may only be introduced in the chase sequence if it occurs as part as the set of consequences of the rule $\rho$ with respect to some substitution $\sigma$ such that, for every $i = 1, \ldots, n$, $\sigma(x_i) = t_i$.

The lemma follows from the following claim: For every $m \geq l$ and every substitution $\sigma$ such that $\sigma(x_i) = t_i$ with $i = 1, \ldots, n$ and $\beta(x, z)\sigma \subseteq P^m_R$, there is some substitution $\tau \supseteq \sigma$ and $\eta(x, y)\tau \subseteq P^m_R$.

Such a claim can be verified via induction on the sequence $P^l_R, P^{l+1}_R, \ldots$

**Lemma 4.4.2** Let $P$ be some program and $\phi$ be a fact occurring in some $P^i_R$. If there is some $P^j_R$ such that $\phi \notin P^i_R$ and $j \geq i$, then $\phi \notin P^k_R$ for every $k \geq j$.

**Proof 6** By the definition of the chase sequence, $\phi$ must contain some term $t$ that has been removed from the chase sequence due to some replacement as defined in Definition 3.5.2. Thus, the lemma holds since, Lemma 4.4.1, term $t$ may not be reintroduced in any $P^k_R$ where $k \geq j$.

4.4.2 Restricted Terms

We introduce some preliminary function that allows us to keep track of the replacements of terms and atoms that are performed during the computation of the restricted chase.

**Definition 4.4.3** Let $P = \langle R, I \rangle$ be a program and $t$ be some term. Then we define the term $[t]_P$ as follows:

- If the term $t$ is replaced by some different term $u$ during the computation of the restricted chase sequence of $P$ up to the set $P^i_R$, then $[t]_P = [u]_P$.
- Otherwise, $[t]_P = t$

Furthermore, given some fact $\phi = p(t_1, \ldots, t_n)$ and a set of facts $F$, let $[\phi]_P = p([t_1]_P, \ldots, [t_n]_P)$ and $[F]_P = \{[\phi]_P | \phi \in F\}$.

We often write $[\phi]$ (resp., $[F]$) instead of $[\phi]_P$ (resp., $[F]_P$) if the set $P_R$ is clear from the context.
Lemma 4.4.4 Let $\mathcal{P}$ be some ontology and $\phi$ a fact occurring in the restricted chase of $\mathcal{P}$. Furthermore, let $i$ be the smallest natural number such that $\phi \in \mathcal{P}_R^i$. Then, $[\phi]^i_R \in \mathcal{P}_R^j$ for every $j \geq i$.

Proof 7 We prove the lemma via induction on the sequence $\mathcal{P}_R^i, \mathcal{P}_R^{i+1}, \ldots$. Let $\phi = p(t_1, \ldots, t_n)$. We prove the lemma via induction on the sequence $\mathcal{P}$.

(Base Case) We proceed to show that $[\phi]^1_R \in \mathcal{P}_R^1$. Note that, since $i$ is the smallest number with $\phi \in \mathcal{P}_R^i$, $[\phi]^1_R = p([t_1]^1_R, \ldots, [t_n]^1_R) = p(t_1, \ldots, t_n) = \phi$.

(Induction Step) We proceed to show that $[\phi]^k_R \in \mathcal{P}_R^k$ where $k > i$. By induction hypothesis (IH), we have that $[\phi]^{k-1}_R = p([t_1]^{k-1}_R, \ldots, [t_n]^{k-1}_R) \in \mathcal{P}_R^{k-1}$. Two possible cases arise.

1. $[\phi]^{k-1}_R \notin \mathcal{P}_R^k$. Then $\mathcal{P}_R^k = R^\infty(\mathcal{P}_R^{k-1})$ and some of the terms $[t_i]^{k-1}_R$ must have been replaced by some different terms $u_i$. Nevertheless, $[\phi]^k_R \in \mathcal{P}_R^k$ since, by the definition of $[\cdot]^k_R$, we have that $[t_i]^k_R = u_i$ for every term $[t_i]^{k-1}_R$ that was been replaced by $u_i$.

2. $[\phi]^{k-1}_R \in \mathcal{P}_R^k$. Then none of the terms $[t_i]^{k-1}_R$ have been replaced. Then the lemma holds because $[\phi]^k_R = [\phi]^{k-1}_R$.

Lemma 4.4.5 Let $\mathcal{O} = \langle T, \mathcal{A} \rangle$ be some ontology and let $t$ be some term of the form $f^p(u)$ where $\rho = A(x) \rightarrow \exists y[R(x,y) \land B(y)]$ that occurs in the restricted chase of $\mathcal{P}(\langle T, \mathcal{A} \rangle)$. Furthermore, let $i$ be the smallest number such that $t \in \mathcal{P}(\mathcal{O})^i_R$. Then, $\{A(u), R(u,t), B(t)\} \subseteq \mathcal{P}(\mathcal{O})^i_R$.

Proof 8 Since $t$ occurs in the chase of $\mathcal{P}(\mathcal{O})$, then $A(x) \rightarrow \exists y[R(x,y) \land B(y)] \in R(T)$. Also, since $i$ is the smallest number with $t \in \mathcal{P}(\mathcal{O})^i_R$, then $\mathcal{P}(\mathcal{O})^i_R = R(T)^3(\mathcal{P}(\mathcal{O})_{R}^{i-1})$. Furthermore, the term $t$ may only occur in $\mathcal{P}(\mathcal{O})_{R}^{i}$ if there is some substitution $\sigma$ with $\sigma(x) = u$ and $A(u) \in \mathcal{P}(\mathcal{O})_{R}^{i-1}$. Thus, $\{R(x, f^p(x)), B(f^p(x))\} \sigma \subseteq \mathcal{P}(\mathcal{O})_{R}^{i}$.

Lemma 4.4.6 Let $\mathcal{O} = \langle T, \mathcal{A} \rangle$ be some ontology and let $t$ be some term occurring in the restricted chase of $\mathcal{P}(\mathcal{O})$. Furthermore, let $i$ be the smallest number such that $t \in \mathcal{P}(\mathcal{O})^i_R$. Then, $[T(t)] \subseteq \mathcal{P}(\mathcal{O})^i_R$.

Proof 9 Note that, since $t$ is a term which occurs in the restricted chase of $\mathcal{P}(\mathcal{O})$, $t$ is of the form $f_n(\ldots f_2((f_1(c))))$ where all $f_i$ are unary function symbols and $c$ is a constant. We prove the lemma via induction on the following sequence $c, f_1(c), f_2(f_1(c)), \ldots, f_n(\ldots f_2((f_1(c))))$.

(Base Case) Note that, if $t$ is a constant then $[T(t)] = \emptyset$ and thus the lemma trivially holds.

(Induction Step) Let $t$ be some term of the form $f^p(u)$, where $\rho = A(x) \rightarrow \exists y[R(x,y) \land B(y)]$ (note that every term which is not a constant occurring in $\mathcal{P}(\mathcal{O})$ must be of that form). By the definition of the set $[T(t)]$, $[T(t)] = [T(u)] \cup \{A(u), R(u,t), B(t)\}$. Let $j$ be the smallest number such that $u \in \mathcal{P}(\mathcal{O})^j_R$ (note that, necessarily $j \leq i$ since $u$ is an immediate ancestor of $t$). By IH, $[T(u)] \subseteq \mathcal{P}(\mathcal{O})^j_R$. Furthermore, by Lemma 4.4.4, $[T(u)] \subseteq \mathcal{P}(\mathcal{O})^j_R$. By Lemma 4.4.5, $\{A(u), R(u,t), B(t)\} \subseteq \mathcal{P}(\mathcal{O})^j_R$. Thus, $\{A(u), R(u,t), B(t)\} \subseteq \mathcal{P}(\mathcal{O})^j_R$ since, by Definition 4.4.3, $[u] = u$ and $[t] = t$. 
Lemma 4.4.7 Let $O = \langle T, A \rangle$ be some ontology and let $t$ be some term occurring in the restricted chase of $P(O)$. Furthermore, let $i$ be the smallest number with $t \in P(O)_R^i$. Then, there exists some $j \geq i$ such that all of the following hold.

1. For every $\phi \in RC(U(T, t))$, we have that $[\phi]_{P(O)}^j \in P(O)_R^j$.
2. $P(O)_R^j$ does not contain any successors of $t$.

Proof 10 For the rest of the proof, we simply write $U$ instead of $U(T, t)$ and $U^k$, where $k \geq 0$, as a shortcut for $U(T, t)_R^k$. In the same manner, we simply write $P$ instead of $P(O)$ and $P^k$ instead of $P(O)_R^k$. Furthermore, given some $k, l \geq 0$, we write $S(U^k, P^l)$ if and only if $[U^k] \subseteq P^l$ and $P^l$ does not contain any successors of the term $t$.

The lemma follows if, for every set in the sequence $U^0, U^1, U^2, \ldots$, there exists some chase step $P^l$ with $S(U^k, P^l)$. We proceed to show that this is indeed the case via induction on such sequence.

(Base Case) We proceed to show that $S(U^0, P^m)$. First, note that, by Lemma 4.4.6, $[I(t)] \subseteq P^i$. It is clear that $P^i$ may not contain any successors of $t$, since $i$ is the smallest number with $t \in P^i$.

(Induction Step) We verify that, for every $k \geq 2$, the existence of some $m$ with $S(U^k-1, P^m)$ implies that there is some $l$ with $S(U^k, P^l)$. We assume that there is some $\phi \in U^k$ with $[\phi] \notin P^m$ as, otherwise, $S(U^k, P^m)$ and the induction step holds. Furthermore, since $S(U^k-1, P^m)$, we have that $[\phi] \notin U^k-1$. There exist two possible cases, which are considered along the following paragraphs: Either (i) $U^k = U^m(U^k-1)$ or (ii) $U^k = U^m(U^k-1)$ (note that, by definition, program $U$ does not contain any rules with existentially quantified variables).

Let (i) $U^k = U^m(U^k-1)$. In this case, we verify that $S([U^k], P^{m+1})$.

We first show $P^{m+1} = P^m(U^m)$. As previously argued, there is at least some fact $\phi$ with $[\phi] \notin U^k$, $\phi \notin U^k-1$ and $[\phi] \notin P^m$. Therefore, one of the following cases must hold:

1. $\phi$ is of the form $C(t)$, there is some rule $\rho = A(x) \land R(x, y) \land B(y) \land R(x, z) \land B(z) \rightarrow x \approx y \in U$ and some $u$ and $v$ such that $\{A(v), R(v, u), B(u), R(v, t), B(t), C(u)\} \subseteq U^k$. By IH, $\{A(v), R(v, u), B(u), R(v, t), B(t), C(u)\} \subseteq P^m$. Furthermore, $\rho \in P$, by the definition of $U$. Since $[\phi] \notin P^m$, $P^m \neq P^m(P^m)$ and thus, by Definition 4.4.4, $P^{m+1} = P^m(P^m)$.

2. $\phi$ is of the form $R(t, s)$ (resp., $R(s, t)$), $\rho = A(x) \land R(x, y) \land B(y) \land R(x, z) \land B(z) \rightarrow x \approx y \in U$ and $\{A(v), R(v, u), B(u), R(v, t), B(t), S(u, s)(\text{resp., } S(s, u))\} \subseteq U^k$ for some $u$ and $v$. Analogous to the previous case.

In either case, we have that $P^{m+1} = P^m(P^m)$. Thus, the set $P^{m+1}$ does not contain any successors of $t$, since, by IH, $P^m$ contains no such terms.

We proceed to show that, $[U^k] \subseteq P^{m+1}$; i.e., we show that, for every $\phi \in U^k$, we have that $[\phi] \notin P^{m+1}$. A fact $\phi \in U^k$ only if one of the following cases holds.

1. Let $\phi \in U^k-1$. Then, by IH, $[\phi]_{P^m} \subseteq P^m$. By Lemma 4.4.4, $[\phi]_{P^{m+1}} \subseteq P^{m+1}$.

2. Let $\phi \notin U^k$. Several possible cases arise.
4.4. PROOFS

(a) \( \phi \) is of the form \( C(t) \), there is some rule of the form \( \rho = A(x) \land R(x, y) \land B(y) \land R(x, z) \land B(z) \rightarrow x \approx y \in U \) and there are some \( u \) and \( v \) such that \( \{A(v), R(v, u), B(u), R(v, t), B(t), C(u)\} \subseteq U^{k-1} \). By IH, \( \{\{A(v), R(v, u), B(u), R(v, t), B(t), C(u)\}\} \subseteq P^m \). Then \( \{C(t)\} \in \mathcal{P}^{m+1} \).

(b) \( \phi \) is of the form \( S(t, s) \) (resp., \( S(s, t) \)), there is some rule of the form \( \rho = A(x) \land R(x, y) \land B(y) \land R(x, z) \land B(z) \rightarrow x \approx y \in U \) and there are some \( u \) and \( v \) such that \( \{A(v), R(v, u), B(u), R(v, t), B(t), S(u, s)\} \subseteq U^{k-1} \). Analogous to the previous case.

Thus, \( S[U^k, \mathcal{P}^{m+1}] \).

Let (ii) \( U^k = U^v(U^{k-1}) \). We proceed to show that \( \mathcal{P}^m \neq \mathcal{P}^v(\mathcal{P}^m) \). As in case (i), there is some fact \( \phi \) with \( \phi \in U^k, \phi \notin U^{k-1} \) \( [\phi] \notin \mathcal{P}^m \). Thus, one of the following cases must hold:

1. \( \phi \) is of the form \( B(t) \) and either of the following holds:
   (a) \( A_1(x) \land \ldots \land A_n(x) \rightarrow B(x) \in U \) and \( \{A_1(t), \ldots, A_n(t)\} \subseteq U^{k-1} \). By IH, \( \{\{A_1(t), \ldots, A_n(t)\}\} \subseteq \mathcal{P}^m \). Moreover, \( A_1(x) \land \ldots \land A_n(x) \rightarrow B(x) \in \mathcal{P} \) by the definition of \( U \). Since \( [\phi] = [B(t)] \notin \mathcal{P}^m \), we have that \( \mathcal{P}^m \neq \mathcal{P}^v(\mathcal{P}^m) \).
   (b) \( A(x) \land R(x, y) \rightarrow B(y) \in U \) and \( \{A(u), R(u, t)\} \subseteq U^{k-1} \) for some term \( u \). Analogous to the previous case.

2. \( \phi \) is of the form \( R(t, u) \)
   (a) \( S(x, y) \rightarrow R(x, y) \in U \) and \( \{S(t, u) \in U^{k-1} \). Analogous to the previous case.
   (b) \( S(y, z) \rightarrow R(x, y) \in U \) and \( \{S(u, t) \in U^{k-1} \). Analogous to the previous case.
   (c) \( S(x, z) \land V(z, y) \rightarrow R(x, y) \in U \) and \( \{S(t, v), V(v, u) \in U^{k-1} \) for some term \( v \). Analogous to the previous case.

In every case we have that \( \mathcal{P}^m \neq \mathcal{P}^v(\mathcal{P}^m) \).

Therefore, by Definition 3.5.5, we have that either of the following holds: (x) \( \mathcal{P}^{m+1} = \mathcal{P}^{\infty}(\mathcal{P}^m) \) or (xx) \( \mathcal{P}^{m+1} = \mathcal{P}^v(\mathcal{P}^m) \).

Let (x) \( \mathcal{P}^{m+1} = \mathcal{P}^{\infty}(\mathcal{P}^m) \). We first show that \( S[U^{k-1}, \mathcal{P}^{m+1}] \). By induction hypothesis we have that, for every \( \phi \in U^{k-1} \), \( [\phi]_P^m \in \mathcal{P}^m \). Note that, by Lemma 4.4.4, for every \( \phi \in \mathcal{P}^m \), \( [\phi]_P^{m+1} \in \mathcal{P}^{m+1} \). Thus, for every \( \phi \in U^{k-1} \), \( [\phi]_P^{m+1} \in \mathcal{P}^{m+1} \). Furthermore, note that, since \( \mathcal{P}^{m+1} = \mathcal{P}^{\infty}(\mathcal{P}^m) \), we can conclude that \( \mathcal{P}^{m+1} \) contains no successors of \( t \) (note that by induction hypothesis, \( \mathcal{P}^m \) contains no such terms). Hence, \( S[U^{k-1}, \mathcal{P}^{m+1}] \).

Let us assume that there is some \( \phi \in U^k \) such that \( [\phi] \notin \mathcal{P}^{m+1} \) as otherwise we would have that \( S[U^k, \mathcal{P}^{m+1}] \) and the induction step would hold. In this case, we may verify \( \mathcal{P}^{m+1} \neq \mathcal{P}^v(\mathcal{P}^{m+1}) \) making an analogous argument as the one at the beginning of case (i). Thus, the same situation as before arises: We either have that \( \mathcal{P}^{m+2} = \mathcal{R}^{\infty}(\mathcal{P}^{m+1}) \) or \( \mathcal{P}^{m+2} = \mathcal{P}^v(\mathcal{P}^{m+1}) \). If \( \mathcal{P}^{m+2} = \mathcal{R}^{\infty}(\mathcal{P}^{m+1}) \), we can make an analogous argument to the one at the beginning of case (x) to conclude that \( \mathcal{P}^{m+1} \).

Yet again, if \( S(U^k, \mathcal{P}^{m+2}) \) does not hold, the very same situation could arise: We either have that \( \mathcal{P}^{m+3} = \mathcal{R}^{\infty}(\mathcal{P}^{m+2}) \) or \( \mathcal{P}^{m+3} = \mathcal{P}^v(\mathcal{P}^{m+2}) \).
4.4. PROOFS

Eventually, there must be some $\mathcal{P}^n$, where $n = m + c$, such that

- $\mathcal{P}^{n-1} = \mathcal{R}^\approx(\mathcal{P}^{n-2}), \mathcal{P}^{n-2} = \mathcal{R}^\approx(\mathcal{P}^{n-3}), \ldots, \mathcal{P}^{m+1} = \mathcal{P}^\approx(\mathcal{P}^m)$, and
- either $\mathcal{P}^n = \mathcal{R}^\approx(\mathcal{P}^{n-1})$ or $S(\mathcal{U}^k, \mathcal{P}^n)$.

Note that, for every possible program $\mathcal{P}$ and any $o \geq 0$, if $\mathcal{P}^o = \mathcal{P}^\approx(\mathcal{P}^{o-1})$, then set $\mathcal{P}^o$ contains at least one term less than $\mathcal{P}^\approx(\mathcal{P}^{o-1})$. We proceed to show that, if $\mathcal{P}^n = \mathcal{R}^\approx(\mathcal{P}^{n-1})$, then $S(\mathcal{U}^k, \mathcal{P}^n)$.

It is clear that $\mathcal{P}^n$ does not contain any successors of $s$ since $\mathcal{P}^n = \mathcal{R}^\approx(\mathcal{P}^{n-1}), \mathcal{P}^{n-2} = \mathcal{R}^\approx(\mathcal{P}^{n-3}), \ldots, \mathcal{P}^{m+1} = \mathcal{P}^\approx(\mathcal{P}^m)$, and the set $\mathcal{P}^m$ does not contain any successors of $t$. Furthermore, note that $S(\mathcal{U}^k, \mathcal{P}^{n-1})$. Thus, it only remains to show that, for every $\phi \in \mathcal{U}^k$, $[\phi] \in \mathcal{P}^n$.

A fact $\phi \in \mathcal{U}^k$ only if one of the following holds:

1. $\phi \in \mathcal{U}^{k-1}$. Then, $[\phi] \in \mathcal{P}^{n-1}$ since $S(\mathcal{U}^{k-1}, \mathcal{P}^{n-1})$, and thus $[\phi] \in \mathcal{P}^n$.

2. $\phi \notin \mathcal{U}^{k-1}$ and $\phi$ is of the form $B(t)$. Then, since $\phi \in \mathcal{U}^k$ and $\phi \notin \mathcal{U}^{k-1}$, we have that one of the following must hold:

   (a) $A_1(x) \land \ldots \land A_n(x) \rightarrow B(x) \in \mathcal{U}$ and $\{A_1(t), \ldots, A_n(t)\} \subseteq \mathcal{U}^{k-1}$. Since $S(\mathcal{U}^{k-1}, \mathcal{P}^{n-1})$, $\{A_1(t), \ldots, A_n(t)\} \subseteq \mathcal{P}^{n-1}$. Moreover, $A_1(x) \land \ldots \land A_n(x) \rightarrow B(x) \in \mathcal{P}$ by the definition of $\mathcal{U}$.

   (b) $A(x) \land R(x, y) \rightarrow B(y) \in \mathcal{U}$ and $\{A(u), R(u, t)\} \subseteq \mathcal{U}^{k-1}$ for some term $u$. Analogous to the previous case.

3. $\phi \notin \mathcal{U}^{k-1}$ and $\phi$ is of the form $R(t, u)$

   (a) $S(x, y) \rightarrow R(x, y) \in \mathcal{U}$ and $S(t, u) \in \mathcal{U}^{k-1}$. Analogous to the previous case.

   (b) $S(y, z) \rightarrow R(x, y) \in \mathcal{U}$ and $S(u, t) \in \mathcal{U}^{k-1}$. Analogous to the previous case.

   (c) $S(x, z) \land V(z, y) \rightarrow R(x, y) \in \mathcal{U}$ and $\{S(t, v), V(v, u)\} \subseteq \mathcal{U}^{k-1}$ for some term $v$. Analogous to the previous case.

Note that, in either case, $[\phi] \in \mathcal{P}^n$ and thus, $S(\mathcal{U}^k, \mathcal{P}^n)$.

\textit{(xx)} Let $\mathcal{P}^{m+1} = \mathcal{P}^\approx(\mathcal{P}^m)$. Then, we may show that $S(\mathcal{U}^k, \mathcal{P}^{m+1})$ in an analogous way as we show that $S(\mathcal{U}^k, \mathcal{P}^n)$ in the second part of case (x).

\textbf{Lemma 4.2.6} Let $\mathcal{T}$ be a TBox and $t$ a restricted term. Then, for every possible $\mathcal{O} = (\mathcal{T}, \mathcal{A})$, $t \notin R(\mathcal{O})$.

\textbf{Proof} 11 By Definition 4.2.3, we have that $t$ is of the form $f_r^\approx(u)$ where $\rho = A(x) \rightarrow \exists y[R(x, y) \land B(y)]$. Note that, $t \notin R(\mathcal{O})$ only if $u \in R(\mathcal{O})$. Let $i$ be the smallest number such that $u \in \mathcal{P}_R^i$. Then, by Lemma 4.4.7, there is some $j$ such that $\mathcal{U}(\mathcal{T}, u) \subseteq \mathcal{P}^j R$ and $\mathcal{P}^j_R$ does not contain any successors of $u$. Furthermore, by the definition of a restricted term, we have that there is some $s$ with $\{R([u], s), B(s)\} \subseteq \mathcal{P}^j R$. By Lemma 4.4.4 for every $\mathcal{P}^k_R$ with $k \geq j$, we have that $\{R(u, s), B(s)\} \subseteq \mathcal{P}^k R$ and thus, by the definition of the restricted consequences of a TGD, term $t$ may not be introduced in any $\mathcal{P}_R^k$. 


4.4. PROOFS

4.4.3 \( \mathcal{V}_T \) is an Overchase of \( \mathcal{T} \)

**Lemma 4.4.8** Let \( \mathcal{T} \) be some TBox and let \( t_1, \ldots, t_n \) be a sequence of terms such that, for every \( i = 1, \ldots, n \), \( \text{dep}(t_i) \leq \text{dep}(t_1) \) and \( \{\text{Eq}(t_1, t_2), \ldots, \text{Eq}(t_{n-1}, t_n)\} \subseteq \mathcal{V}_T \). Then, for every \( i = 2, \ldots, n \), \( \text{Eq}(t_1, t_i) \in \mathcal{V}_T \).

**Proof 12** We prove the lemma via induction on the sequence \( t_1, \ldots, t_n \). It is clear that the base case holds, since \( \text{Eq}(t_1, t_1) \in \mathcal{V}_T \).

We proceed to show the induction step; i.e., we verify that \( \text{Eq}(t_1, t_i-1) \in \mathcal{P} \) implies \( \text{Eq}(t_1, t_i) \in \mathcal{P} \) for every \( i = 3, \ldots, n \). By definition, we have that \( \text{Eq}(t_1, t_i) \in \mathcal{V}_T \). Thus, there must be some \( A(x) \wedge R(x, y) \wedge B(y) \wedge R(x, z) \wedge B(z) \rightarrow y \approx z \in \mathcal{R}(\mathcal{T}) \) and some \( u \) such that \( \{A(u), R(u, t_{i-1}), B(t_{i-1}), R(u, t_i), B(t_i)\}\) \( \subseteq \mathcal{V}_T \) (note that, a fact of the form \( \text{Eq}(s, v) \) may only occur in \( \mathcal{V}_T \) via application of the \( \approx \)-rule and every EGD in \( \mathcal{R}(\mathcal{T}) \) is of the form \( C(x) \wedge S(x, y) \wedge D(y) \wedge S(x, z) \wedge D(z) \rightarrow y \approx z \)). By IH, we have that \( \text{Eq}(t_1, t_{i-1}) \) and \( \text{dep}(t_1) \leq \text{dep}(t_{i-1}) \). Hence, \( \{R(u, t_1), B(t_1)\}\) \( \subseteq \mathcal{V}_T \) by application of the \( \exists \)-rule. Thus, \( \text{Eq}(t_1, t_1) \in \mathcal{V}_T \) by application of the \( \approx \)-rule.

**Lemma 4.2.9** Set \( \mathcal{V}_T \) is an overchase of the TBox \( \mathcal{T} \).

**Proof 13** To show the lemma, we verify that, for every ontology \( \mathcal{O} = (\mathcal{T}, \mathcal{A}) \) and every \( i \geq 0 \), the following hold.

1. If \( B(t) \in \mathcal{P}(\mathcal{O})^1_R \) where \( B \in \mathcal{N}_C \), then \( B(t) \in \mathcal{V}_T \).
2. If \( R(t, u) \in \mathcal{P}(\mathcal{O})^1_R \) where \( R \in \mathcal{N}_R \), then \( R(t, u) \in \mathcal{V}_T \).

We proceed to show (a) and (b) via induction on the chase sequence of \( \mathcal{P}(\mathcal{O}) \). Our argument is structured as follows.

- **Base Case:** Claims (a) and (b) hold for \( \mathcal{P}(\mathcal{O})^1_R \).

- **Induction Step:** For every \( i > 1 \), claims (a) and (b) hold for \( \mathcal{P}(\mathcal{O})^i_R \) provided such claims hold for \( \mathcal{P}(\mathcal{O})^{i-1}_R \).

  *(Base Case)* The base case holds since \( \mathcal{P}(\mathcal{O})^0_R = \mathcal{A} \) and \( \mathcal{L}_*(\mathcal{T}) \subseteq \mathcal{V}_T \). Note that, by assumption, there are neither concept nor role names in \( \mathcal{A} \) which do not occur in \( \mathcal{T} \) (see Section ??).

  *(Induction Step)* Let \( \mathcal{P}(\mathcal{O}) = (\mathcal{R}, \mathcal{T}) \). By the definition of the chase sequence, three possible cases arise: For every \( i \geq 1 \), \( \mathcal{P}(\mathcal{O})^i_R \) is either (i) \( \mathcal{R}^1_R(\mathcal{P}(\mathcal{O})^{i-1}_R) \), (ii) \( \mathcal{R}^1_R(\mathcal{P}(\mathcal{O})^{i-1}_R) \), or (iii) \( \mathcal{R}^\infty_R(\mathcal{P}(\mathcal{O})^{i-1}_R) \).

  We proceed to show that, in either case, the induction hypothesis (IH) holds.

  Let (i) \( \mathcal{P}(\mathcal{O})^i_R = \mathcal{R}^1_R(\mathcal{P}(\mathcal{O})^{i-1}_R) \). We do a case by cases analysis of all the possible facts that may occur in \( \mathcal{P}(\mathcal{O})^i_R \) and verify that the induction step holds. When checking if the induction step holds for some given fact \( \phi \in \mathcal{P}(\mathcal{O})^{i-1}_R \) we assume that \( \phi \notin \mathcal{P}(\mathcal{O})^{i-1}_R \) as otherwise the induction step trivially holds by induction hypothesis (IH).

  - \( B(t) \in \mathcal{P}(\mathcal{O})^i_R \) only if \( \rho = A(x) \rightarrow \exists y[R(x, y) \wedge B(y)] \in \mathcal{R}(\mathcal{T}) \) and \( t = f^\rho_y(u) \). By Lemma 4.2.6, \( t \) is not restricted since \( t \) occurs in the restricted chase of \( \mathcal{P}(\mathcal{O}) \). Then, by Definition 4.2.5, \( t \) is
Thus, the induction step holds if $P(O)_R^i = R_T^i(P(O)_R^{i-1})$.

Let (iii) $P(O)_R^i = R_T^i(P(O)_R^{i-1})$. Again, we do a case by cases analysis of all the possible facts that may occur in $P(O)_O$ and verify that claims (a) and (b) hold.

- $B(t) \in P(O)_R^i$ only if there exists a sequence of terms $t_1, \ldots, t_n$ such that $t_1 = t$; $B(t_n) \in P(O)_R^{i-1}$; for every $j = 1, \ldots, n$, $\text{dep}(t_j) \leq \text{dep}(t_j)$; and, for every $j = 1, \ldots, n - 1$, there is some EGD $A_j(x, y, z) \land C_j(y, z) \land R_j(x, y, z) \land C_j(y, z) \in R(T)$ and some $u_j$ with $\{A_j(u_j), R_j(u_j, t_j), C_j(t_j), R_j(u_j, t_{j+1}), C_j(t_{j+1})\} \subseteq P(O)_R^i$. By IH, for every $j = 1, \ldots, n - 1$, $\{A_j(u_j), R_j(u_j, t_j), C_j(t_j), R_j(u_j, t_{j+1}), C_j(t_{j+1})\} \subseteq V_T$ by application of the Eq-rule. By Lemma 4.4.8, $\text{Eq}(t_1, t_n) \in V_T$ (note that the lemma is indeed applicable since, for every $j = 1, \ldots, n$, $\text{dep}(t_j) \leq \text{dep}(t_{j+1})$). Thus, $B(t)_* \in V_T$ by application of the $\forall$-rule.

- $R(t, u) \in P(O)_R^i$ (resp., $R(u, t) \in P(O)_R^i$) only if there exists a sequence of terms $t_1, \ldots, t_n$ such that $t_1 = t$; $R(t_n, u) \in P(O)_R^{i-1}$ (resp., $R(u_n, t) \in P(O)_R^{i-1}$); for every $j = 1, \ldots, n$, $\text{dep}(t_j) \leq \text{dep}(t_j)$; and, for every $j = 1, \ldots, n - 1$, there is some EGD $A_j(x, y, z) \land C_j(y, z) \land R_j(x, y, z) \land C_j(y, z) \in R(T)$ and some $u_j$ with $\{A_j(u_j), R_j(u_j, t_j), C_j(t_j), R_j(u_j, t_{j+1}), C_j(t_{j+1})\} \subseteq P(O)_R^i$. Analogous to the previous case.

Thus, the induction step holds if $P(O)_O = R_T^i(P(O)_R^{i-1})$. 

Thus, the induction step holds if $P(O)_O = R_T^i(P(O)_R^{i-1})$. 

Thus, the induction step holds if $P(O)_O = R_T^i(P(O)_R^{i-1})$. 

Thus, the induction step holds if $P(O)_O = R_T^i(P(O)_R^{i-1})$.
4.4.4 Complexity Results

**Lemma 4.4.9** Let $\mathcal{T}$ be some TBox and let $t$ be some term which may occur during the computation of the chase of a program $\mathcal{P}((\mathcal{T}, \mathcal{A}))$. If $t$ is not $n$-cyclic, then deciding whether $t$ is restricted is in \textit{PTime} with respect to $\mathcal{T}$.

**Proof 14** We can decide whether $t$ is restricted by computing the restricted chase of the program $\mathcal{U}(\mathcal{T}, t)$. If $t$ is non-$n$-cyclic, then $\text{dep}(t) \leq n|\mathcal{T}^3|$: Note that (i) $t$ is of the form $f_1(\ldots f_n(*))$, the number of function symbols in $\mathcal{V}_T$ is at most $|\mathcal{T}^3|$, and a term $f_1(\ldots f_n(*))$ is $n$-cyclic if and only if it contains $n + 1$ occurrences of the same function symbol. Thus, by definition, $|\mathcal{I}(t)|$ is linear with respect to $|\mathcal{T}^3|$ and so is the number of terms in it.

It is clear that the restricted chase of $\mathcal{U}(\mathcal{T}, t)$ can be computed in polynomial time. Note that: (i) $\mathcal{U}(\mathcal{T}, t)$ does not contain any TGDs featuring existentially quantified variables, (ii) every rule in $\mathcal{U}(\mathcal{T}, t)$ contains at most 3 different variables, (iii) every predicate in $\mathcal{U}(\mathcal{T}, t)$ has a maximum arity of 2 and (iv) once a functional term is removed from the restricted chase sequence due to some replacement, it is never reintroduced again (see Lemma 4.4.1).

**Lemma 4.2.14** Deciding whether some TBox $\mathcal{T}$ is $\mathcal{RCA}_n$ is in \textit{ExpTime}.

**Proof 15** By the definition of $\mathcal{RCA}_n$: $\mathcal{T}$ is $\mathcal{RCA}_n$ if and only if there is some $n$-cyclic term in $\mathcal{V}_T$. Thus, it suffices to compute the set $\mathcal{V}_T$ up to the occurrence of the first $n$-cyclic term. Hence, for the rest of the argument, we assume that $\mathcal{V}_T$ may not include $n$-cyclic terms.

Every term $t$ in $\mathcal{V}_T$ has a depth of at most $n|\mathcal{T}^3|$: Note that (i) $t$ is of the form $f_1(\ldots f_n(*))$, the number of function symbols in $\mathcal{V}_T$ is at most $|\mathcal{T}^3|$, and a term $f_1(\ldots f_n(*))$ is $n$-cyclic if and only if it contains $n + 1$ occurrences of the same function symbol. Thus, there exist at most $\mathcal{T} = \sum_{i=0}^{n|\mathcal{T}^3|} |\mathcal{T}^3| i$ non-$n$-cyclic terms that may occur in $\mathcal{V}_T$.

Every axiom in $\mathcal{T}$ may contain at most 3 different concept or role names and there are at most $|N_C(\mathcal{T}) + N_R(\mathcal{T})|$ different predicates in $\mathcal{V}_T$ where $N_C(\mathcal{T})$ and $N_R(\mathcal{T})$ are the sets of concept names and role names in $\mathcal{T}$. Thus, the number of predicates in $\mathcal{V}_T$ is at most $P = 3|\mathcal{T}|$. Thus, the maximum number of facts that may occur in $\mathcal{V}_T$ is $F = PT^2$ since the arity of every predicate is at most two. Hence, since all of the rules from Figure 4.1 result in the addition of at least some fact to $\mathcal{V}_T$, it takes at most $\mathcal{T}$ many applications of such rules to compute $\mathcal{V}_T$. It can be verified that $\mathcal{T}$ is exponentially large with respect to $n|\mathcal{T}|$.

Finally, note that all of the expansion rules can be applied in polynomial time. This is also the case for the $\exists$-rule: Given some term non-cyclic $t$, we can decide whether $t$ is restricted in polynomial time (see Lemma 4.4.9).

**Lemma 4.4.10** Let $\mathcal{O} = (\mathcal{T}, \mathcal{A})$ be some $\mathcal{RCA}_n$ ontology and $\gamma$ a query. Then, checking whether $\mathcal{O} \models \gamma$ is in \textit{ExpTime}.

**Proof 16** As shown in Section 3.5, it suffices to compute the restricted chase of $\mathcal{P}(\mathcal{O})$ to decide whether $\mathcal{O} \models \gamma$. We proceed to show that $\text{RC}(\mathcal{P}(\mathcal{O}))$ can be computed in \textit{ExpTime}. 


4.4. PROOFS

As shown in the proof of Lemma 4.2.14, if $T$ is $\text{RCA}_n$ then $\mathcal{V}_T$ contains at most $F$ facts, a number which is exponential with respect to $n|T^\exists|$ (where $T^\exists$ is the set of all existential axioms in $T$). By the definition of an overchase we have that $\phi \in \text{RC}(\mathcal{P}(O))$ only if $\phi_* \in \mathcal{V}_T$. Thus, the maximum number of facts in $\text{RC}(\mathcal{P}(\langle T, A \rangle))$ is $|T^\exists|N_I(O)^2$ where $N_I(O)$ is the number of individuals in the ontology $O$.

Note that every predicate in $\mathcal{P}(\langle T, A \rangle)$ has a maximum arity of 2.

Furthermore, even if some facts are removed from the restricted chase sequence during its computation, this is not problematic: See Lemma 4.4.2 to verify that, once a fact is removed from the restricted chase sequence, it may not be reintroduced again.

Lemma 4.4.11 Let $O = \langle T, A \rangle$ be some $\text{RCA}_n$ ontology and $\gamma$ a query. Then, checking whether $O|\leq \gamma$ is ExpTime-hard.

Proof 17 It is shown in [Cuenca Grau et al. 2013] that weak acyclicity is less general than universal MFA when it comes to Horn-SRI TBoxes. It can be easily verified that, when it comes to Horn-SRI TBoxes, universal MFA and $\text{RCA}_1$ coincide in terms of generality. I.e., a Horn-SRI TBox is universally MFA if and only if it is $\text{RCA}_1$. Thus, the Lemma follows from Lemma 59 from [Cuenca Grau et al. 2013] which states the following: “Let $T$ be a weak acyclic Horn-SRI TBox, let $I$ be an instance and let $\phi$ be a fact. Then, checking whether $\langle R(T), I \rangle|\leq \phi$ is ExpTime-hard.”

Lemma 4.2.15 Let $O = \langle T, A \rangle$ be some $\text{RCA}_n$ ontology and $\gamma$ a query. Then, checking whether $O|\leq \gamma$ is ExpTime-complete with respect to $|T|$ and $n$.

Proof 18 Follows from Lemmas 4.4.10 and 4.4.11.

4.4.5 $\text{RCA}_n$ vs MFA$^\cup$

In an attempt to make this document more self-contained, we present a summarized definition of MFA$^\cup$ that can be directly applied to Horn-SRI TBoxes.

Definition 4.4.12 Let $\text{Eq} = \{ \top(x) \rightarrow \text{Eq}(x,x), \text{Eq}(y,x) \rightarrow \text{Eq}(x,y), \text{Eq}(x,z) \land \text{Eq}(z,y) \rightarrow \text{Eq}(x,y) \}$ and let $\Xi$ be the function mapping axioms to rules which is identical to $\pi$ from Figure 3.2 in all but the following cases:

Given some TBox $T$, let $\text{MFA}^\cup(T) = \langle \pi(T) \cup \text{Eq}, I_\star(T) \rangle$. A TBox $T$ is $\text{MFA}^\cup$ if and only if there are no cyclic terms in the oblivious chase of $\text{MFA}^\cup(T)$.

It can be easily verified that our definition coincides with the on from [Cuenca Grau et al. 2013]. We proceed a theoretical comparison of $\text{MFA}^\cup$ and $\text{RCA}_n$.

Lemma 4.2.17 If $T$ is $\text{MFA}^\cup$ then $T$ is $\text{RCA}_n$ for every $n > |T^\exists|$.

Proof 19 Let $T$ be some $\text{MFA}^\cup$ TBox. We proceed to show that $T$ is $\text{RCA}_n$ for every $n > |T^\exists|$.

By the definition of $\text{MFA}^\cup$, we have that the set $\text{OC}(\text{MFA}^\cup(T))$ does not contain any cyclic terms. Thus, for every term $\text{OC}(\text{MFA}^\cup(T))$, $\text{dep}(t) \leq |T^\exists|$ since, by definition, a term $s$ with $\text{dep}(s) > |T^\exists|$ is necessarily cyclic.
We proceed to show that this upper bound on the depth of terms also holds for \( V_T \). Note that, if such is the case, then \( T \) is necessarily RCA\(_n\) for every \( n > |T^2| \). This is because \( T \) is RCA\(_n\) if and only if some \( n \)-cyclic term occurs in \( V_T \), and, for every \( n \)-cyclic term \( u \), \( \text{dep}(u) > |T^3| \). Thus, the lemma follows if we verify that, for every term \( t \in V_T \), \( \text{dep}(t) \leq |T^3| \).

We show that \( |T^3| \) is an upper bound on the depth on the terms in \( V_T \) by constructing a set \( W_T \) with the following properties:

1. If there is some term \( t \in W_T \), then there is some term \( u \in MFA^\uparrow(T) \) with \( \text{dep}(t) \leq \text{dep}(u) \).
2. A term \( t \) occurs in \( V_T \) only if it occurs in \( W_T \).

We proceed with the construction of \( W_T \). We first present a function, which will be used in the definition of such set. Let \( \Psi \) be the function mapping TBox axioms into rules defined as follows:

\[
\rho = A(x) \rightarrow \exists y R(x, y) \land B(y).
\]

Let \( W_T^1, W_T^2, W_T^3, \ldots \) be the sequence such that \( W_T^1 = \{ A(\ast, \ast) \mid A \in N_C(T) \} \cup \{ R(\ast, \ast, \ast, \ast) \mid R \in N_R(T) \} \) and, for every \( i \geq 2 \), \( W_T^i \) is the set that results by applying some rule in Figure 4.3 to \( W_T^{i-1} \). Furthermore, let \( W_T = \bigcup_{i \in \mathbb{N}} W_T^i \).

We proceed to show some properties of \( W_T \), which will be used in subsequent parts of this argument. For all terms \( t', u, u' \), and predicates \( A \) and \( R \), the following hold.

1. If \( A(t, t') \), \( R(t, t', u, u') \) or \( R(u, u', t, t') \) in \( W_T \), then \( \text{dep}(t) \leq \text{dep}(t') \).
2. If \( A(t, t') \in W_T \), then \( A(t', t') \in W_T \).
3. If \( R(t, t', u, u') \in W_T \), then \( R(t', t', u, u') \in W_T \).
4. If \( R(t, t', u, u') \in W_T \), then \( R(t, t', u', u') \in W_T \).

These claims can be proven via induction via induction on the sequence \( W_T^1, W_T^2, W_T^3, \ldots \). It is clear that the base case holds since \( W_T^1 = \{ A(\ast, \ast) \mid A \in N_C(T) \} \cup \{ R(\ast, \ast, \ast, \ast) \mid R \in N_R(T) \} \). The induction step can be shown with a simple case by case analysis on all the possible facts that may be derived during the computation of the sequence \( W_T^1, W_T^2, W_T^3, \ldots \).

4.4.5.1 Claim (a):

We proceed to show the following claim: If there is some term \( t \in W_T \), then there is some term \( u \in MFA^\uparrow(T) \) with \( \text{dep}(t) \leq \text{dep}(u) \).

To do so, we first introduce a function \( \pi \) which maps the terms in \( W_T \) into terms in \( MFA^\uparrow(T) \).

Let \( \pi \) be some function such that \( \pi(\ast) = \ast \) and, for every term of the form \( t = f^\rho_p(u) \) where \( \rho = A(x) \rightarrow \exists y R(x, y) \land B(y) \), \( \pi(t) = f^\rho_p(\pi(u')) \) where \( u' \) is some term such that \( A(u, u') \in W_T \). By claim (ii), term \( t \) may not occur in \( W_T \) unless there is some fact of the form \( A(u, u') \in W_T \) and thus, \( \pi \) is well-defined. Furthermore, since \( \text{dep}(u) \leq \text{dep}(u') \) for every \( A(u, u') \in W_T \) by claim (i), we have that \( \text{dep}(t) \leq \text{dep}(\pi(t)) \) for every term \( t \).

We proceed to show the following claims via induction.

1. If \( t \) occurs in \( W_T \), then \( \pi(t) \) occurs in \( OC(MFA^\uparrow(T)) \).
4.4. PROOFS

2. \( A(t, t') \in W_T \) implies \( \{ \text{Eq}(\pi(t), \pi(t')) \}, A(\pi(t')) \} \subseteq OC(MFA^T(T)) \).

3. \( R(t, t', u, u') \in W_T \) implies \( \{ \text{Eq}(\pi(t), \pi(t')), R(\pi(t'), \pi(u')) \}, \text{Eq}(\pi(u'), \pi(u)) \} \subseteq OC(MFA^T(T)) \).

4. \( \text{Eq}(t, u) \in W_T \) implies \( \text{Eq}(\pi(t), \pi(u)) \in OC(MFA^T(T)) \).

Note that, claim (a) directly follows from (1) and the fact that \( \text{dep}(t) \leq \text{dep}(\pi(t)) \) for every term \( T \) in \( W_T \). The remainder of the claims (2-4) are just introduced to properly structure our induction argument.

The base case of the induction trivially holds since \( I_0(T) \subseteq OC(MFA^T(T)) \) and \( W_T^0 = \{ A(\ast, \ast) \mid A \in N_C(T) \} \cup \{ R(\ast, \ast, \ast, \ast) \mid R \in N_R(T) \} \). We proceed to show the induction step; i.e., we show that (1-4) hold for every fact \( \phi \in W_T^i \) provided they hold for every fact in \( W_T^{i-1} \). When showing that the claims hold for some term or fact \( \phi \in W_T^i \) we assume that \( \phi \notin W_T^{i-1} \) as otherwise the induction step trivially holds by induction hypothesis.

Let \( t \) be some term occurring in \( W_T^i \). Then, \( t \) is of the form \( t = f^n_\rho(u) \) for some term \( u \) where \( \rho = A(x) \rightarrow \exists y[R(x, y) \land B(y)] \), \( A(x, x') \rightarrow R(x, x', f^n_\rho(x), f^n_\rho(x)) \land B(f^n_\rho(x), f^n_\rho(x)) \in \Psi(T) \) and \( A(u, u') \in W_T^{i-1} \). By the definition of \( \pi \), we have that \( t = f^n_\rho(u) = f^n_{\rho'}(s') \) where \( s' \) is some term such that \( A(u, s') \in W_T^{i-1} \). By IH, \( A(s') \in OC(MFA^T(T)) \). It is clear that claim (1) holds since \( \rho \in MFA^T(T) \).

Let \( \phi \) be a fact of the form \( A(t, t') \in W_T^i \). Then, one of the following cases holds:

- \( \bigwedge_{i=1}^n A(x, x'_i) \rightarrow \bigwedge_{i=1}^n (B(x, x'_i) \land B(x'_i, x'_i)) \in \Psi(T) \) and \( \{ A_i(t, t'_i) \mid i = 1, \ldots, n \} \). Then, \( \bigwedge_{i=1}^n A_i(x_i) \land \bigwedge_{i=1}^{n-1} \text{Eq}(x_i, x_{i+1}) \rightarrow \bigwedge_{i=1}^n B(x_i) \in MFA^T(T) \). By IH, \( \{ \text{Eq}(\pi(t), \pi(t'_i)), A_i(\pi(t'_i)) \mid i = 1, \ldots, n \} \).

- \( A(x, x') \land R(x, x'', y, y') \rightarrow B(y, y') \in \Psi(T) \) and \( \{ A(u, u'), R(u, u', t, t') \} \subseteq W_T^{i-1} \). Then \( A(x) \land \text{Eq}(x, x') \land R(x, y) \rightarrow B(y) \in MFA^T(T) \). By IH, \( \{ \text{Eq}(\pi(u), \pi(u')), A(\pi(u')), \text{Eq}(\pi(u), \pi(u'')) \}, R(\pi(u''), \pi(t')) \} \subseteq OC(MFA^T(T)) \).

- \( A(x, x') \rightarrow R(x, x', f^n_\rho(x), f^n_\rho(x)) \land B(f^n_\rho(x), f^n_\rho(x)) \in \Psi(T) \) where \( \rho = A(x) \rightarrow \exists y[R(x, y) \land B(y)] \), \( A(u, u') \in W_T^{i-1} \), \( t = t' = f^n_\rho(u) \). By the definition of \( \pi \), we have that \( \pi(t) = f^n_{\rho'}(s') \) where \( s' \) is some term such that \( A(u, s') \in W_T^{i-1} \). By IH, \( A(\pi(s')) \in OC(MFA^T(T)) \). Also, note that \( A(x) \rightarrow \exists y[R(x, y) \land B(y)] \in MFA^T(T) \).

- \( \{ \text{Eq}(t, s), A(s, t') \} \subseteq W_T^{i-1} \) and \( \text{dep}(t) \leq \text{dep}(s) \). By IH, \( \{ \text{Eq}(\pi(t), \pi(s)), \text{Eq}(\pi(s), \pi(t')), A(\pi(t')) \} \subseteq OC(MFA^T(T)) \).

It is clear that in either case, claim (2) holds. Note that \( \top(x) \rightarrow \text{Eq}(x, x), \text{Eq}(x, y) \rightarrow \text{Eq}(x, y), \text{Eq}(x, z) \land \text{Eq}(y, z) \rightarrow \text{Eq}(x, y) \) \subseteq MFA^T(T) \).

Let \( \phi \in W_i \) be a fact of the form \( S(t, t', u, u') \). Then, \( S(t, t', u, u') \in W_T^i \) only if one of the following cases hold:

- \( A(x, x') \rightarrow S(x, x', f^n_\rho(x), f^n_\rho(x)) \land B(f^n_\rho(x), f^n_\rho(x)) \in \Psi(T) \) and \( A(t, t') \in W_T^{i-1} \). Note that, in this case, \( f^n_\rho(t) = u = u' \). By the definition of \( \pi \), we have that \( \pi(f^n_\rho(t)) = \pi(u) = f^n_{\rho}(s') \)
where \( s' \) is some term such that \( A(t, s') \in W_{T}^{m-1} \). By IH, \( A(\pi(s')) \in OC(MFA^{\pi}(T)) \). Also, \( A(x) \rightarrow \exists y [R(x, y) \land B(y)] \in MFA^{\pi}(T) \).

- \( R(x, x', y, y') \rightarrow S(x, x', y, y') \in \Psi(T) \) and \( R(t, t', u, u') \in W_{T}^{m-1} \). Then, \( R(x, y) \rightarrow S(x, y) \in MFA^{\pi}(T) \). By IH, \( \{ Eq(\pi(t), \pi(t')), R(\pi(t'), \pi(u')) \} \subseteq OC(MFA^{\pi}(T)) \).

- \( R(y, y', x, x') \rightarrow S(x, x', y, y') \in \Psi(T) \) and \( R(u, u', t, t') \in W_{T}^{m-1} \). Then, \( R(y, x) \rightarrow S(x, y) \in MFA^{\pi}(T) \). By IH, \( \{ Eq(\pi(u), \pi(u')), R(\pi(u'), \pi(t')) \} \subseteq OC(MFA^{\pi}(T)) \).

- \( R(y, y', z, s, t) \in V(z, z', y, y') \rightarrow S(x, x', y, y') \in \Psi(T) \) and \( \{ R(t, t', s, s', S, s'', u, u') \} \subseteq W_{T}^{m-1} \). Then, \( R(y, x) \land Eq(z, z') \land R(z', y) \rightarrow S(x, y) \in MFA^{\pi}(T) \). By IH, \( \{ Eq(\pi(t), \pi(t')), R(\pi(t'), \pi(t')), \} \subseteq OC(MFA^{\pi}(T)) \).

It is clear that in either case, claim (3) holds. Again, note that note that \( \{ \top(x) \rightarrow Eq(x, x), Eq(y, x) \rightarrow Eq(x, y), Eq(x, z) \land Eq(z, y) \rightarrow Eq(x, y) \} \subseteq MFA^{\pi}(T) \).

Let \( \phi \) be a fact of the form \( Eq(t, u) \). Then, \( Eq(t, u) \in W_{T}^{m} \) only if one of the following cases hold.

- \( A(x, x') \land R(x, x', y, y') \land B(y, y') \land R(x, x''', y, z, z') \land B(z, z'') \rightarrow Eq(y, z) \in W_{T}^{m} \) and \( \{ A(s, s'), R(s, s', t, t'), B(t, t''), R(s, s'', u, u'), B(u, u'') \} \subseteq W_{T}^{m-1} \). By IH, \( \{ Eq(\pi(s), \pi(s')), A(\pi(s')) \} \subseteq W_{T}^{m} \).

- \( Eq(u, t) \in W_{T}^{m-1} \). By IH, \( Eq(\pi(u), \pi(t)) \in OC(MFA^{\pi}(T)) \).

- \( \{ Eq(t, s), Eq(s, u) \} \subseteq W_{T}^{m-1} \). By IH, \( \{ Eq(t, s), Eq(s, u) \} \subseteq OC(MFA^{\pi}(T)) \).

It is clear that in either case, claim (3) holds. Again, note that note that \( \{ \top(x) \rightarrow Eq(x, x), Eq(y, x) \rightarrow Eq(x, y), Eq(x, z) \land Eq(z, y) \rightarrow Eq(x, y) \} \subseteq MFA^{\pi}(T) \).

### 4.4.5.2 Claim (b): A term \( t \) occurs in \( V_{T} \) only if it occurs in \( W_{T} \).

Claim (a) follows from the following statements.

1. For every term \( t \), if \( t \in V_{T} \), then \( t \in W_{T} \).

2. For every term \( t \) and every \( \Lambda \in N_{C} \), if \( A(\Lambda) \in V_{T} \), then \( A(t, \Lambda) \in W_{T} \).

3. For every terms pair of terms \( t \) and \( u \), and every role name \( R \), if \( R(t, u) \in V_{T} \), then \( R(t, t, u, u) \in W_{T} \).

4. For all terms \( t \) and \( u \), if \( Eq(t, u) \in V_{T} \), then \( Eq(t, u) \in W_{T} \).

Claim (b) follows from (1): The rest of the claims (2-4) are in placed to properly structure the induction argument.

Let \( V_{T}^{1}, V_{T}^{2}, \ldots \) be a sequence constructed as follows: \( V_{T}^{1} = \{ \top(T) \} \) and, for every \( i \geq 2 \), \( V_{T}^{i} \) is obtained by applying some (randomly chosen) rule in Figure 4.1 to \( V_{T}^{i-1} \). Claims (1-4) can be proven...
via induction on the sequence $V_1, V_2, \ldots$ It is clear that the base case holds since $V_1 = I_\star(R(T))$ and set $W_T$ contains every fact in $\{A(\star, \star) \mid A \in N_C(T)\} \cup \{R(\star, \star, \star, \star) \mid R \in N_R(T)\}$. The induction step can be easily verified by doing a case by case analysis of all the possible facts that may be derived when computing some set $V_i$ with $i \geq 2$. 
Q1(w, y) : \{physicalEntity(w, z), physicalEntity(y, z)\}
Q2(x, z) : \{Complex(z), Pathway(x), memberPhysicalEntity(z, w), memberPhysicalEntity(z, w),
            participant(y, z), pathwayComponent(x, y)\}
Q3(x, z) : \{SequenceSite(x), featureLocation(x, w), featureLocation(x, y), sequenceIntervalBegin(w, z),
            sequenceIntervalBegin(y, z), SequenceSite(z)\}
Q4(x, z) : \{Pathway(x), Protein(x), participant(w, z), participant(y, z), pathwayComponent(x, w),
            pathwayComponent(x, y)\}

Q1(x, y) : \{cellularComponent(x, z), cellularComponent(y, z)\}
Q2(x) : \{translatedFrom(w, x), NucleotideResource(x), locatedOn(x, y), database(x, z), Database(z)\}
Q3(x) : \{translatedFrom(w, y), Resource(y), translatedFrom(w, x), Resource(x), database(y, z),
            Database(z), database(x, z)\}
Q4(x) : \{CellularComponent(z), locatedIn(x, w), cellularComponent(w, z), locatedIn(x, y),
            cellularComponent(y, z)\}

Q1(x, z) : \{worksFor(x, y), worksFor(z, y), publicationAuthor(x, z)\}
Q2(x) : \{advisor(x, y), Faculty(y), teacherOf(y, z), Course(z), memberOf(y, w), Department(w)\}
Q3(x, y, z) : \{teacherOf(y, z), Course(z), advisor(x, y), Faculty(y), takesCourse(x, z), Student(x)\}
Q4(x) : \{publicationAuthor(x, z), publicationAuthor(x, y), advisor(z, y), Professor(y),
            memberOf(z, w), memberOf(y, w)\}

Q1(x, y) : \{takesCourse(z, x), takesCourse(z, y)\}
Q2(x) : \{teachingAssistantOf(y, x), TeachingAssistant(x), publicationAuthor(z, x),
            Book(z), takesCourse(w, y), worksFor(x, v), ResearchGroup(v)\}
Q3(x, y) : \{isFriendOf(x, y), UndergraduateStudent(x), GraduateStudent(y), like(x, z)\}
Q4(x, y) : \{hasDoctoralDegreeFrom(x, z), hasDoctoralDegreeFrom(y, z), hasMasterDegreeFrom(x, w),
            hasMasterDegreeFrom(y, w), worksFor(x, v), worksFor(y, v)\}

Figure 4.2: Queries for Reactome, Uniprot, LUBM and UOBM.
4.4. PROOFS

<table>
<thead>
<tr>
<th>( \forall )-rule</th>
<th>if there is some TGD ( \rho \in \Psi(T) ) then ( W_T \rightarrow \rho \circ (W_T \cup W_T) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Eq_1 )-rule</td>
<td>if there is some ( A(t',t) \in W_T ) and some ( s ) with ( \text{Eq}(t,s) \in W_T ) and ( \text{dep}(t) \leq \text{dep}(s) ) then ( W_T \rightarrow { A(s,t') } \cup W_T )</td>
</tr>
<tr>
<td>( Eq_2 )-rule</td>
<td>if there is some ( R(t',t,u,u') \in W_T ) and some ( s ) with ( \text{Eq}(t,s) \in W_T ) and ( \text{dep}(t) \leq \text{dep}(s) ) (resp., ( \text{Eq}(u,s) \in W_T ) and ( \text{dep}(u) \leq \text{dep}(s) )) then ( W_T \rightarrow { R(s,t',u,u') } \cup W_T ) (resp., ( { R(t,t',s,u') } \cup W_T ))</td>
</tr>
<tr>
<td>( Eq_3 )-rule</td>
<td>if there are some ( t, u ) and ( s ) with ( \text{Eq}(u,t) \in W_T ), ( { \text{Eq}(t,s), \text{Eq}(s,u) } \subseteq W_T ) or ( t = u ) then ( W_T \rightarrow { \text{Eq}(t,u) } \cup W_T )</td>
</tr>
</tbody>
</table>

Figure 4.3: Expansion rules for the construction of \( W_T \).
5

The RSA Fragment of Horn DL

In recent years there has been a growing interest in so-called lightweight DL languages, which are based on logics with favorable computational properties. The most prominent examples of lightweight ontology languages are the $\mathcal{EL}^{++}$, DL-Lite$_R$ and DLP (formally introduced in Section 3.2). These languages are often referred to as profile languages, because they serve as the formal basis for the profile fragments EL, QL and RL of the Web Ontology Language (OWL) [Hitzler et al. 2009]. Standard reasoning tasks, such as classification and fact entailment, are feasible in polynomial time for all profiles, and many highly scalable profile-specific reasoners have been developed [Baader et al. 2006; Bishop et al. 2011; Calvanese et al. 2011; Kazakov et al. 2014; Motik et al. 2014; Rodriguez-Muro and Calvanese 2012]. In contrast to the lightweight logics underpinning the profiles, the logics required to capture expressive Horn ontologies are intractable: standard reasoning is ExpTime-complete for Horn-SHOF and and 2-ExpTime-complete for the more expressive Horn-SROIQ [Kazakov 2008].

In an attempt to push the expressivity of lightweight DLs whilst still preserving its beneficious complexity properties, we present the role safety acyclic (RSA) fragment of Horn DL. This fragment, for which standard reasoning tasks can be solved in polynomial time, encompasses the profile languages in terms of expressiveness; i.e., every ontology defined over a profile language is also an RSA ontology. As in the previous chapter, we also present a reasoning algorithm to solve standard reasoning tasks over this fragment and moreover, we empirically verify that many real-world ontologies, which cannot be defined within the profile languages, are RSA.

5.1 The Notion of Role Safety

As mentioned in the previous chapter, reasoning over Horn-SROIQ, the most expressive of the Horn DL fragments, is intractable. In particular, satisfiability is ExpTime-hard already for Horn-ALCI (the fragment of Horn-SROIQ without nominals, cardinality restrictions or complex roles).

A closer look at existing complexity results reveals that the main source of intractability is the phenomenon typically known as and-branching: due to the interaction between existential quantifiers over a role $R$ (i.e., axioms of the form $A \sqsubseteq \exists R.B$) and universal quantifiers over $R$ (encoded by axioms of type $A \sqsubseteq \forall R.B$), an ontology may only be satisfied in models of exponential size which cannot be summarized. The same effect can be achieved via the interaction between existential quantifiers and
5.1. THE NOTION OF ROLE SAFETY

Lazy ⊑ Student

Student ⊑ ∃Attends.Course

∃Attends.MorningCourse ⊑ Diligent

Lazy ∩ Diligent ⊑ ⊥

Course ⊑ ∃IsAttendedBy.Student

Attends− ⊑ IsAttendedBy

IsAttendedBy− ⊑ Attends

Lazy(David)

Figure 5.1: Example ontology \( O \)

cardinality restrictions (axioms of type \( A ⊑ \leq 1 R.B \)): reasoning in the extension of the EL profile with counting is also known to be ExpTime-hard [Baader et al. 2005a].

And-branching can be tamed by precluding the harmful interactions between existential and universal quantifiers, on the one hand and existential quantifiers and cardinality restrictions, on the other hand. If we disallow existential quantifiers altogether (i.e., axioms of the form \( A ⊑ ∃R.B \)), then we obtain the DLP profile, and ontologies become a subset in terms of expressivity to Datalog programs with equality. Similarly, if we disallow the use of inverse roles and cardinality restrictions, thus precluding both universal quantification over roles and counting, then we obtain the EL profile.

The main idea behind our notion of role safety is to identify a subset of the roles in an ontology over which these potentially harmful interactions between language constructs do not occur. On the one hand, if a role does not occur existentially quantified in axioms of type \( A ⊑ R.B \), then its “behavior” is similar to that of a role in an DLP ontology, and hence it is safe. On the other hand, if a role occurs existentially quantified, but no axioms involving inverse roles or counting apply to any of its super-roles, then the role behaves like a role in an EL ontology, and hence it is also safe.

Definition 5.1.1 A role \( R \) is safe with respect to some Horn-SHOIQ ontology \( O \) if at least one of the following conditions holds.

1. There are no axioms of the form \( A ⊑ ∃R.B ∈ O \).
2. For every axiom of the form \( A ⊑ ≤ 1 S.B ∈ O \), \( R ∩ O S \) and \( R − ∩ O S \); and for every axiom of the form \( ∃S.A ⊑ B ∈ O \) with \( A ≠ ⊤ \), \( R − ∩ O S \).

See Section 3.2 for the definition of the relation \( ⊑ O \). Also, if the ontology \( O \) is clear from the context, we simply say that a role is “safe” instead of “safe with respect to \( O \).”

Example 5.1.2 Consider the example ontology from Figure 5.1, which is not captured by any of the logics underpinning the OWL profile languages. Note that the role Attends is safe: although it occurs existentially quantified in axiom [5.3], its inverse IsAttendedBy does not occur in an axiom of the form
∃R.A ⊑ B, and the ontology does not contain at most restrictions. In contrast, the role IsAttendedBy is unsafe since it occurs existentially quantified in (5.3) and its inverse role Attends in (5.3).

Definition 5.1.1 intuitively explains why Horn-SHOTIQ ontologies captured by any of the logics underpinning the OWL profile languages contain only safe roles:

- In $\mathcal{E}L^{++}$, roles can be existentially quantified, but there are no inverse roles or cardinality restrictions. Thus, the second condition in Definition 5.1.1 trivially holds for every role.

- In DLP, roles do not occur axioms of the form $A ⊑ ∃R.B$. Thus, the first condition of Definition 5.1.1 trivially holds.

- In DL-Lite$_R$, there are no cardinality restrictions and nor axioms of the form $∃R.A ⊑ B$. Thus, the second condition in Definition 5.1.1 trivially holds.

5.2 Role Safety Acyclicity

In this section, we propose a novel role safety acyclicity (RSA) condition that is applicable to Horn-SHOTIQ ontologies and that does not completely preclude unsafe roles. Instead, our condition restricts the way in which unsafe roles are used so that they cannot lead to the interactions between language constructs that are at the root of $\mathcal{E}XP$TIME-hardness proofs; in particular, and-branching.

To check whether an ontology $\mathcal{O}$ is RSA we first generate a directed graph $G(\mathcal{O})$ by means of a Datalog program $P_G(\mathcal{O})$. The edges in $G(\mathcal{O})$ are generated from the extension of a fresh “edge” predicate $E$ in the materialization of $P_G(\mathcal{O})$. Intuitively, the relevant facts over $E$ in the materialization stem from the presence in $\mathcal{O}$ of existential restrictions over unsafe roles. Once the directed graph $G(\mathcal{O})$ has been generated, we check that it is a directed acyclic graph (DAG) and that it does not contain “diamond-shaped” subgraphs; the former requirement will ensure termination of our reasoning algorithm in Section 5.3, while the latter is critical to ensure tractability. Furthermore, we define a weaker version of RSA (WRSA) where $G(\mathcal{O})$ is only required to be a DAG. Although this relaxed notion does not ensure tractability of reasoning, it does guarantee termination of our reasoning algorithm, and hence is still of relevance in practice.

**Definition 5.2.1** Let $\mathcal{O}$ be an ontology and let $\pi$ be the mapping defined in Figure 5.2. Let $PE$ and $E$ be fresh binary predicates, and let $U$ be a fresh unary predicate. Furthermore, for each pair of concepts $A, B$ and each role $R$, let $v_{A,R,B}$ be a fresh constant. Finally, let $Ξ$ be the function mapping each axiom $\alpha$ in $\mathcal{O}$ to a datalog rule as given next, and let $Ξ(\mathcal{O}) = \{Ξ(\alpha) | \alpha \in \mathcal{O}\}$:

$$Ξ(\alpha) = \begin{cases} A(x) \rightarrow R(x, v_{A,R,B}^A) \land B(v_{A,R,B}^A) \land PE(x, v_{A,R,B}^A) & \text{if } \alpha = A ⊑ ∃R.B \\ \pi(\alpha) & \text{Otherwise.} \end{cases}$$

Then, let $P_G(\mathcal{O})$ be the standard equality axiomatization of the following program

$$Ξ(\mathcal{O}) \cup \{U(x) \land PE(x, y) \land U(y) \rightarrow E(x, y)\} \cup \{U(v_{A,R,B}^A) | R \text{ is unsafe}\}$$
5.2. ROLE SAFETY ACYCLICITY

Furthermore, we say that

\[\text{Lazy}(x) \rightarrow \text{Student}(x)\]

\[\text{Student}(x) \rightarrow \text{Attends}(x, v_{At,Co}^{St}) \land \text{Course}(v_{At,Co}^{St}) \land \text{PE}(x, v_{At,Co}^{St})\]

\[\text{Attends}(x, y) \land \text{MorningCourse}(y) \rightarrow \text{Diligent}(y)\]

\[\text{Lazy}(x) \land \text{Diligent}(x) \rightarrow \bot(x)\]

\[\text{Course}(x) \rightarrow \text{AttendedBy}(x, v_{IA,St}^{Co}) \land \text{Student}(v_{IA,St}^{Co}) \land \text{PE}(x, v_{IA,St}^{Co})\]

\[\text{Attends}(y, x) \rightarrow \text{AttendedBy}(x, y)\]

\[\text{AttendedBy}(x, y) \rightarrow \text{Attends}(y, x)\]

\[U(x) \land \text{PE}(x, y) \land U(y) \rightarrow \text{E}(x, y)\]

\[\text{Lazy}(\text{David})\]

\[U(v_{IA,St}^{Co})\]

Figure 5.2: Checking acyclicity of our example ontology \(O\).

Let \(G(O)\) be the smallest directed graph having an edge \((c, d)\) for each fact \(E(c, d)\) with \(E(c, d) \in OC(\mathcal{P}_G(O))\). Then, we say that \(O\) is Role Safety Acyclic (RSA) if \(G(O)\) is an oriented forest. Furthermore, we say that \(O\) is weakly RSA (WRSA) if \(G(O)\) is a DAG.

An oriented forest is a disjoint union of oriented trees; that is, a DAG whose underlying undirected graph is a forest.

The core of the program \(\mathcal{P}_G(O)\) is obtained from \(O\) by translating its axioms into first-order logic in the usual way with the single exception of existentially quantified axioms \(\alpha\), which are translated into Datalog by transforming the (unique) existential variable in \(\pi(\alpha)\) into a constant. The fresh predicate \(\text{PE}\) is used to track all facts over roles \(R\) generated by the application of the rules, regardless of whether the relevant role \(R\) is safe or not. In this way, \(\text{PE}\) records “possible edges” in the graph. The safety distinction is realised by the unary predicate \(U\), which is populated with all fresh constants introduced by the Skolemisation of existential restrictions over the unsafe roles. Finally, the rule \(U(x) \land \text{PE}(x, y) \land U(y) \rightarrow E(x, y)\) ensures that only possible edges between Skolem constants in the extension of \(U\) eventually become edges in the graph.

Example 5.2.2 Figure 5.2 depicts the rules in the program \(\mathcal{P}_G(O)\) for our example ontology \(O\). Consider the application of the chase on \(\mathcal{P}_G(O)\), which applies to the initial facts \(\mathcal{P}_G(O)^0 = \{\text{Lazy}(\text{David}), U(v_{IA,St}^{Co})\}\). The chase terminates after the following iterations:

\[
\mathcal{P}_G(O)^1 = \mathcal{P}_G(O)^0 \cup \{\text{Student}(\text{David})\}
\]

\[
\mathcal{P}_G(O)^2 = \mathcal{P}_G(O)^1 \cup \{\text{Attends}(\text{David}, v_{IA,Co}^{St}), \text{Course}(v_{IA,Co}^{St}), \text{PE}(\text{David}, v_{IA,Co}^{St})\}
\]

\[
\mathcal{P}_G(O)^3 = \mathcal{P}_G(O)^2 \cup \{\text{AttendedBy}(v_{IA,Co}^{St}, v_{IA,St}^{Co}), \text{Student}(v_{IA,St}^{Co}), \text{PE}(v_{IA,Co}^{St}, v_{IA,St}^{Co})\}
\]

\[
\mathcal{P}_G(O)^4 = \mathcal{P}_G(O)^3 \cup \{\text{Attends}(v_{IA,St}^{Co}, v_{IA,St}^{Co}), \text{PE}(v_{IA,Co}^{St}, v_{IA,St}^{Co})\}
\]

No more atoms are derived in subsequent steps and hence \(OC(\mathcal{P}_G(O)) = \mathcal{P}_G(O)^4\). Note that the graph
induced by the auxiliary PE predicate is cyclic; in contrast, the extension of E is empty and G(O) has no edges. Clearly, O is thus RSA.

The following theorem establishes that checking RSA and WRSA is tractable. Intuitively, the program P(O) is linear in the size of O and each of its rules contains at most three variables regardless of O; as a result, the materialization (and hence also the resulting graph) is polynomially bounded.

**Theorem 5.2.3** Checking whether an ontology O is RSA (resp. WRSA) is feasible in polynomial time in the size of O.

**Proof 20** The program P(O) is linear in the size of O. Furthermore, each rule in P(O) contains at most three variables. Thus, the set OC(P(O)) is bounded in size by O(|O|^3). Finally, checking whether a directed graph is an oriented tree (resp. acyclic) is feasible in polynomial time by means of standard graph traversal algorithms.

### 5.3 Reasoning Over Acyclic Ontologies

In this section, we show that standard reasoning tasks are tractable for RSA ontologies. To this purpose, we propose a translation from a Horn-SHɔIQ ontology O into a set N(O) of existential rules, which may contain existentially quantified variables in the head. Axioms in O are translated directly into existential rules as specified in Figure 3.2.

**Definition 5.3.1** Let O be an ontology and let π be the mapping defined in Figure 3.2. Furthermore, for each pair of concepts A, B and each safe role R in O, let v_{A,R,B} be a fresh constant. Let Λ be the function mapping each axiom α in O to a Datalog rule as given next:

\[
Λ(α) = \begin{cases} 
A(x) → R(x,v_{A,R,B}) ∧ B(v_{A,R,B}) & \text{if } α = A ⊑ ∃R.B \text{ with } R \text{ safe} \\
π(α) & \text{Otherwise.}
\end{cases}
\]

Finally, we define the program N(O) as the set \{Λ(α) | α ∈ O\}.

**Example 5.3.2** Figure 5.3 depicts the rules of the program N(O) for our running example O.

We next show that this translation preserves satisfiability, subsumption, and instance retrieval reasoning outcomes, regardless of whether the ontology O is acyclic or not. Thus, we can reason over N(O) instead of O without sacrificing correctness.

**Lemma 5.3.3** The following properties hold for each ontology O, concept names A, B and named individuals a and b, and c is a fresh constant not occurring O:

1. O is satisfiable if and only if N(O) is satisfiable if and only if OC(N(O)) contains no fact over ⊥.

2. O ⊨ A(a) if and only if N(O) ⊨ A(a) if and only if A(a) ∈ OC(N(O)).
5.3. REASONING OVER ACYCLIC ONTOLOGIES

Lazy\((x) \rightarrow \text{Student}(x)\)

\(\text{Student}(x) \rightarrow \text{Attends}(x, v_{\text{St,Co}}) \land \text{Course}(v_{\text{St,Co}})\)

\(\text{Attends}(x, y) \land \text{MorningCourse}(y) \rightarrow \text{Diligent}(y)\)

Lazy\((x) \land \text{Diligent}(x) \rightarrow \bot(x)\)

\(\text{Course}(x) \rightarrow \exists y(\text{AttendendBy}(x, y) \land \text{Student}(y))\)

\(\text{Attends}(y, x) \rightarrow \text{AttendedBy}(x, y)\)

\(\text{AttendedBy}(x, y) \rightarrow \text{Attends}(y, x)\)

Lazy\((\text{David})\)

Figure 5.3: Running Example: Reasoning.

3. \(\mathcal{O} \models A \subseteq B\) if and only if \(\mathcal{N}(\mathcal{O}) \cup \{A(c)\} \models B(c)\) if and only if \(B(c) \in OC(\mathcal{N}(\mathcal{O}) \cup \{A(c)\})\).

So far, we have established that we can dispense with the input ontology \(\mathcal{O}\) and reason over the program \(\mathcal{N}(\mathcal{O})\) instead. The chase of \(\mathcal{N}(\mathcal{O})\), however, may still be infinite. We next show, if \(\mathcal{O}\) is RSA, then there exist a polynomial upper bound on the size of the chase of \(\mathcal{N}(\mathcal{O})\). Intuitively, every functional term occurring in an atom of the chase of \(\mathcal{N}(\mathcal{O})\) corresponds to a single path in \(G(\mathcal{O})\), and the size of the graph is polynomial in \(\mathcal{O}\). In an oriented forest there is at most one path between any two nodes, which bounds polynomially the number of possible functional terms. In contrast, the latter condition does not hold for DAGs, where only a bound in the length of paths can be guaranteed.

Lemma 5.3.4 Let \(\mathcal{O}\) be an RSA ontology. Then, the chase of \(\mathcal{N}(\mathcal{O})\) terminates and \(OC(\mathcal{N}(\mathcal{O}))\) is of polynomial size. Furthermore, if \(\mathcal{O}\) is WRSA, then the chase of \(\mathcal{N}(\mathcal{O})\) terminates and \(OC(\mathcal{N}(\mathcal{O}))\) is of exponential size.

Lemmas 5.3.3 and 5.3.4 suggest a reasoning algorithm for acyclic ontologies \(\mathcal{O}\). First, compute the program \(\mathcal{N}(\mathcal{O})\) as in Definition 5.3.1. Then, run the chase for \(\mathcal{N}(\mathcal{O})\) and read out the reasoning outcomes from the computed Herbrand model. If \(G(\mathcal{O})\) is an oriented forest (i.e., \(\mathcal{O}\) is RSA) we can implement our algorithm efficiently, which yields the following result as a corollary of the previous theorems.

Theorem 5.3.5 Satisfiability and unary fact entailment is feasible in polynomial time for the class of RSA ontologies.

In contrast to RSA, our algorithm runs in exponential time for WRSA ontologies. We next show that, indeed, reasoning with WRSA ontologies is intractable under standard complexity-theoretic assumptions.

Theorem 5.3.6 Unary fact entailment is \(\text{PSPACE}\)-hard for WRSA ontologies.

Proof 21 In [Cuenca Grau et al. 2013] Lemma 63 validity checking for QBF formulas which is a standard \(\text{PSPACE}\)-complete problem is reduced to fact entailment w.r.t. weakly-acyclic Horn-SHI.
ontologies. While weak-acyclicity and WRSA are two distinct conditions, the particular reduction provided as proof of that lemma produces a set of existential rules which can be translated into a WSRA Horn-\textit{SHI} ontology. As such, the reduction shows as well that fact entailment w.r.t. WSRA Horn-\textit{SHOIQ} ontologies is \textit{PSPACE}-hard.

5.4 Stronger Notions of Acyclicity

Note that Theorem 5.3.5 does not make any claims about the tractability of concept subsumption for RSA ontologies. To check whether $\mathcal{O} \models A \sqsubseteq B$ we need to extend $\mathcal{N}(\mathcal{O})$ with an assertion $A(c)$ over a fresh individual $c$, run the Skolem chase, and check whether $B(c)$ is derived (see Lemma 5.3.3). However, as illustrated by the following example, RSA membership is not robust under addition of ABox assertions.

\textbf{Example 5.4.1} Let $\mathcal{O}$ consist of a fact $B(c)$ and the following axioms:

\begin{align*}
A & \sqsubseteq B \\
B & \sqsubseteq C \\
A & \sqsubseteq \exists R.A
\end{align*}

Ontology $\mathcal{O}$ is RSA because the rule corresponding to the “dangerous” axiom $A \sqsubseteq \exists R.A$ involving the unsafe role $R$ does not fire during materialization; as a result, the graph generated by $\mathcal{P}_G(\mathcal{O})$ is empty. Indeed, the chase terminates on $\mathcal{N}(\mathcal{O})$ and determines satisfiability as well as all the facts entailed by $\mathcal{O}$. In contrast, if we add the fact $A(c)$ to $\mathcal{N}(\mathcal{O})$ to determine the subsumers of $A$, the chase will no longer terminate because the ontology $\mathcal{O}$ extended with $A(c)$ is now cyclic.

To ensure tractability of subsumption and classification, we therefore propose the following stronger notion of acyclicity.

\textbf{Definition 5.4.2} Let $\mathcal{O}$ be some ontology. For each concept name $A$ in $\mathcal{O}$, let $c_A$ be a fresh constant and let $A_{\text{cl}} = \{ A(c_A) \mid A \text{ is a concept name occurring in } \mathcal{O} \}$. We say that $\mathcal{O}$ is RSA for classification if $\mathcal{O}$ extended with $A_{\text{cl}}$ is RSA.

Tractability of subsumption immediately follows from our results in Section 5.3.

\textbf{Proposition 5.4.3} Checking whether $\mathcal{O} \models A \sqsubseteq B$ is feasible in polynomial time for ontologies $\mathcal{O}$ that are RSA for classification.

Although this notion is well-suited for TBox reasoning, data-intensive applications where the ABox changes frequently require a further strengthening.

\textbf{Definition 5.4.4} An ontology $\mathcal{O}$ is universally RSA if $\mathcal{O} \cup \mathcal{A}'$ is RSA for every ABox $\mathcal{A}'$.

Checking whether $\mathcal{O} = \mathcal{R} \cup \mathcal{T} \cup \mathcal{A}$ is universally RSA can be reduced to checking whether the ontology $\mathcal{O}$ extended with the critical instance $\mathcal{I}_b(\mathcal{A})$ which is defined as the set of facts containing all facts of the form $A(\star)$ and $R(\star, \star)$ for every concept name $A$ and role $R$ occurring in $\mathcal{A}$.

\textbf{Proposition 5.4.5} An ontology $\mathcal{O}$ is universally RSA if and only if $\mathcal{O} \cup \mathcal{I}_b(\mathcal{A})$ is RSA.
Proof 22 Assume \( O \) is universally RSA. Then, it is RSA also for \( O \cup I_*(A) \). It remains to be shown that if \( O \cup I_*(A) \) is RSA, \( O \cup A \) is RSA for every ABox \( A \). Let \( O^A \) be the extension of \( O \) with an arbitrary such ABox \( A \) and let \( O^* = O \cup I_*(O) \). Also let \( N^*_I \) and \( N^*_A \) be the set of named individuals occurring in \( O^* \) and \( O^A \), respectively. Then, we define a mapping \( \mu : \text{terms}(P_G(O^A)) \to \text{terms}(P_G(O^*)) \) as follows:

\[
\mu(x) = \begin{cases} 
x, & \text{if } x \in \text{terms}(P_G(O^*)), 
* , & \text{otherwise}
\end{cases}
\]

It can be shown by induction on the level of atoms in \( OC(P_G(O^A)) \) that:

- for every \( A(t) \in OC(P_G(O^A)) \): \( A(\mu(t)) \in OC(P_G(O^*)) \), and
- for every \( R(t,u) \in OC(P_G(O^A)) \): \( R(\mu(t),\mu(u)) \in OC(P_G(O^*)) \).
- for every \( x \approx y \in OC(P_G(O^A)) \): \( \mu(x) \approx \mu(y) \in OC(P_G(O^*)) \).

Thus, the graph \( G(O^A) \) is a subgraph of \( G(O^*) \).

Example 5.4.6 The critical ABox for our example ontology \( O \) consists of all facts \( A(*) \) and \( R(*,*) \) for \( A \) a concept name and \( R \) a role name from \( O \). It can be checked that \( O \) is universally RSA, and hence also RSA for classification.

Universal RSA is, however, a rather strict condition, especially in the presence of equality. The following example illustrates that, e.g., every ontology with a functional role used in an existential restriction is not universally RSA.

Example 5.4.7 Consider \( O \) consisting of axioms \( A \subseteq \exists R.B \) and \( \top \subseteq \leq 1 R.\top \). The critical ABox contains facts \( A(*) \), \( B(*) \), and \( R(*,*) \). The corresponding Datalog program entails a fact \( R(*,v^A_{R,B}) \) due to axiom \( A \subseteq \exists R.B \). Due to the functionality of \( R \), the individuals \(* \) and \( v^A_{R,B} \) become equal, and hence we have \( A(v^A_{R,B}) \) and eventually also \( R(v^A_{R,B},v^A_{R,B}) \). Since \( R \) is unsafe, the graph contains a cyclic edge \( E(v^A_{R,B},v^A_{R,B}) \). Indeed, the chase of both \( O \) and \( N(O) \) is infinite.

It is well-known that the Skolem chase often does not terminate in the presence of equality [Cuenca Grau et al. 2013; Marnette 2009]. The standard approach to circumvent this issue is to exploit the so-called singulization technique [Marnette 2009] described in Section 3.4.2. After application of the singulization transformation, the ontology is thus equality-free. Singulization preserves reasoning outcomes in a well-understood way, and it is effective in addressing non-termination problems.

5.5 Related Work

In recent years the computational properties of Horn DL have been extensively investigated. The logical underpinnings for the EL and QL profiles of OWL 2 are provided by, respectively, the Horn logics \( \mathcal{EL}^{++} \) [Baader et al. 2005a] and DL-Lite\( _R \) [Calvanese et al. 2007], while the RL profile is based
on Datalog and its intersection with DL [Grosof et al. 2003]. Hustadt et al. proposed the expressive logic Horn-SRIQ, and established its complexity [Hustadt et al. 2005]. Krötzsch et al. studied the complexity of a wide range of Horn DLs with complexities in-between the tractable logics underpinning the profiles and Horn-SROIQ [Krötzsch et al. 2013; Krötzsch et al. 2007]. Finally, the exact complexity of Horn-SHOIQ and Horn-SROIQ was determined by Ortiz et al. [Ortiz et al. 2010].

Our techniques in Section 5.3 extend the so-called combined approach to reasoning in EL [Kontchakov et al. 2011; Stefanoni et al. 2013], where ontologies are transformed into Datalog programs by means of Skolemisation of all existentially quantified variables into constants. Skolemisation into constants was also exploited by Zhou et al. [Zhou et al. 2015] to compute upper bounds to query answers.

Finally, in the literature we can find a wide range of acyclicity conditions that are sufficient to ensure chase termination. Weak acyclicity [Fagin et al. 2005] was one of the first such notions, and was subsequently extended to joint acyclicity [Krötzsch and Rudolph 2011], acyclicity of a graph of rule dependencies [Baget et al. 2014], and super-weak acyclicity [Marnette 2009], amongst others. The notion of acyclicity closest to ours is model summarizing acyclicity (MSA) [Cuenca Grau et al. 2013], where acyclicity can also be determined by the materialization of a Datalog program. Unlike existing acyclicity notions, ours was designed to ensure tractability of reasoning rather than chase termination. In particular, the Skolem chase of our example RSA ontology $O$ is infinite and hence $O$ cannot be captured by any acyclicity condition designed for chase termination. Instead, our notion ensures termination of the Skolem chase over a particular transformed Horn program $N(O)$, which we can use for reasoning over $O$. Another important difference is that, in contrast to the chase of $O$, the chase of the transformed program $N(O)$ is not a universal model of $O$, and hence it does not preserve answers to general CQs (but only for satisfiability and fact entailment). Finally, although existing acyclicity conditions guarantee termination of the chase, none of them ensures polynomiality of the computed Herbrand model. Indeed, checking fact entailment over Horn-SHI ontologies that are weakly acyclic [Fagin et al. 2005] (the most basic acyclicity notion for chase termination) is PSPACE-hard [Cuenca Grau et al. 2013].

5.6 Proof of Concept

We have implemented RSA and WRSA checkers using RDFox [Motik et al. 2014] as a Datalog reasoner. For testing, we used the ontologies in the Oxford Repository and the Design Patterns repository. The former is a large repository currently containing 761 real-world ontologies; the latter contains a wide range of smaller ontologies that capture design patterns commonly used in ontology modeling (these ontologies are particularly interesting as they highlight common interactions between language constructs). Experiments were performed on a laptop with 16 GB RAM and an Intel Core 2.9 GHz processor running Java v.1.7.0_21, with a timeout of 30 min.

Our results are summarised in Table 5.1. For each repository, we first selected those ontologies that are Horn-SHOIQ and are not captured by any of the OWL 2 profiles. We found 126 such ontologies in the Oxford Repository and 23 in the Design Patterns repository. We then tested our
5.7. PROOFS

Table 5.1: Acyclicity evaluation results for ontologies outside the OWL 2 profiles.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford</td>
<td>Satisfiability</td>
<td>126</td>
<td>37</td>
<td>37+43</td>
<td>37+44</td>
<td>46</td>
<td>39</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Ontology</td>
<td>Classification</td>
<td>126</td>
<td>37</td>
<td>37+35</td>
<td>37+35</td>
<td>52</td>
<td>49</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Repository</td>
<td>Universality</td>
<td>126</td>
<td>37</td>
<td>37+2</td>
<td>37+31</td>
<td>87</td>
<td>57</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ontology</td>
<td>Satisfiability</td>
<td>23</td>
<td>14</td>
<td>14+9</td>
<td>14+9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Design</td>
<td>Classification</td>
<td>23</td>
<td>14</td>
<td>14+8</td>
<td>14+8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Patterns</td>
<td>Universality</td>
<td>23</td>
<td>14</td>
<td>14+4</td>
<td>14+8+4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

acyclicity conditions for satisfiability (Def. 5.2.1), classification (Def. 5.4.2) and universality (Def. 5.4.4) on all these ontologies (for classification and universality, we disregarded the ABox part of the ontologies). We performed tests both with and without employing singularization. Interestingly, in both repositories we could not find any ontology that is WRSA but not RSA, and hence the two notions coincided for all our tests.

As we can observe, 37 ontologies in the Oxford Repository contained only safe roles, and hence are RSA. Without singularization, we found 43 additional ontologies with unsafe roles that are RSA. 35 of which were also RSA for classification and only 2 universally acyclic. When using singularization the number of additional RSA ontologies increased significantly, and we obtained 29 additional universally RSA ontologies, but unfortunately our tests timed-out for several ontologies. This can be explained by the fact that the use of singularization leads to more complicated Datalog rules for which RDFox is not optimised.

In the case of the Design Patterns repository, all ontologies are RSA. We only found one ontology that was not universally RSA when using singularization. Ontologies in this repository are smaller, and we encountered no time-outs.

Note that, in this chapter, we do not include an evaluation of the reasoning algorithm. This is because, being so similar to the procedure presented in the previous section, we foresee this evaluation not being very informative. Thus, to gain an insight of the performance of our algorithm presented in this section, see the results presented in [?].

Furthermore, the algorithm presented in this section may only be used to solve standard reasoning tasks over RSA ontologies. To solve the problem of CQA over this fragment see Feier et al. 2015.

5.7 Proofs

Lemma 5.3.3 The following properties hold for each ontology \( \mathcal{O} \), concept names \( A, B \) and named individuals \( a, b \), where \( c \) is a fresh constant.

1. \( \mathcal{O} \) is satisfiable if and only if \( \mathcal{N}(\mathcal{O}) \) is satisfiable if and only if \( OC(\mathcal{N}(\mathcal{O})) \) contains no fact over \( \bot \).
2. \( \mathcal{O} \models A(a) \) if and only if \( \mathcal{N}(\mathcal{O}) \models A(a) \) if and only if \( A(a) \in OC(\mathcal{N}(\mathcal{O})) \).

3. \( \mathcal{O} \models A \subseteq B \) if and only if \( \mathcal{N}(\mathcal{O}) \cup \{A(c)\} \models B(c) \) if and only if \( B(c) \in OC(\mathcal{N}(\mathcal{O}) \cup \{A(c)\}) \).

**Proof 23** For each claim of the form \( A \iff B \iff C \) in the theorem it is enough to show that \( A \iff C \) as \( B \iff C \) follows from the properties of the chase (see Section 3.5). We also reformulate all ‘\( A \)’ statements regarding satisfiability of and entailments w.r.t. \( \mathcal{O} \) in terms of properties of the chase of \( \mathcal{O} \). For the third claim in the theorem, we note that for a Horn ontology \( \mathcal{O} \), it is well-known that
\[
\mathcal{O} \models A \subseteq B \iff \mathcal{O} \cup \{A(c)\} \models B(c),
\]
where \( c \) is a fresh constant. It remains to be shown that:

a) \( OC(\mathcal{P}(\mathcal{O})) \) contains no fact over \( \bot \) iff \( OC(\mathcal{N}(\mathcal{O})) \) contains no fact over \( \bot \); and

b) \( A(a) \in OC(\mathcal{P}(\mathcal{O})) \) iff \( A(a) \in OC(\mathcal{N}(\mathcal{O})) \).

To prove the ‘if’ part of these claims (soundness) we map each term occurring in \( OC(\mathcal{N}(\mathcal{O})) \) to a set of terms occurring in \( OC(\mathcal{P}(\mathcal{O})) \) and show inductively that certain properties hold between atoms/terms in \( OC(\mathcal{N}(\mathcal{O})) \) and atoms over mapped terms/mapped terms in \( OC(\mathcal{P}(\mathcal{O})) \).

We first introduce some notions which will make the formulation of the IH more straightforward.

For a Horn-SHOIQ ontology \( \mathcal{O} \), its skolemization \( sk(\mathcal{O}) \) is the program obtained from \( \mathcal{P}(\mathcal{O}) \) by standard Skolemisation of existentially quantified variables into functional terms. For a Horn program \( \mathcal{P} \), its grounding, ground(\( \mathcal{P} \)), is the program obtained by replacing each variable occurring in \( \mathcal{P} \) with each term that can be formed using constants and functional symbols occurring in \( \mathcal{P} \). The derivation level of a ground atom \( a \in OC(\mathcal{P}) \), level(\( a, OC(\mathcal{P}) \)), is a natural number \( k \) s.t.: \( a \in \mathcal{P}(\mathcal{O}) \cup \mathcal{H} \) and \( a \notin \mathcal{P}(\mathcal{O})^{k-1} \), where \( \mathcal{S} \) is the set of facts occurring in \( \mathcal{P} \) and \( \mathcal{H} \) is the set of rules occurring in \( \mathcal{P} \). The derivation level of a ground term \( t \in OC(\mathcal{P}) \), level(\( t, OC(\mathcal{P}) \)), is a natural number \( k \) s.t.: \( t \) occurs in some atom \( a \) with level(\( a, OC(\mathcal{P}) \)) = \( k \), but \( t \) does not occur in any atom \( a \) with level(\( a, OC(\mathcal{P}) \)) < \( k \).

For a set of ground atoms \( \mathcal{S} \), terms(\( \mathcal{S} \)) is the set of all terms occurring in some atom in \( \mathcal{S} \).

**Definition 5.7.1** Let \( \mathcal{O} \) be a Horn-SHOIQ ontology, and let \( \mathcal{R} \) be a role name occurring in \( \mathcal{O} \). We say that \( \mathcal{R} \) is a forward-sound role iff for every axiom of type \( A \subseteq \exists S.B \) in \( \mathcal{O} \), with \( S \) being a safe role: \( \mathcal{O} \upmodels^* \mathcal{R} \). Conversely, \( \mathcal{R} \) is a backward-sound role iff for every axiom of type \( A \subseteq \exists S.B \) in \( \mathcal{O} \), with \( S \) being a safe role: \( \mathcal{O} \upmodels \mathcal{R}^{-} \).

**Lemma 5.7.2** Let \( \mathcal{O} \) be a Horn-SHOIQ ontology and let \( \mu : \text{terms}(OC(\mathcal{N}(\mathcal{O}))) \rightarrow 2^{\text{terms}(OC(\mathcal{P}(\mathcal{O})))} \) be the following function:
\[
\mu(x) = \begin{cases} 
\{x\}, & \text{if } x \in N_t \\
\{f_{RD}^c(t) \mid t \in \mu(y)\}, & \text{if } x = f_{RD}^c(t) \\
\{f_{RD}^c(y) \mid f_{RD}^c(y) \in \text{terms}(OC(\mathcal{P}(\mathcal{O})))\}, & \text{if } x = v_{RD}^c 
\end{cases}
\]

Then, all of the the following hold.

i) \( \text{For every } x \in \text{terms}(OC(\mathcal{N}(\mathcal{O}))) : \mu(x) \neq \emptyset \).

ii) \( A(x) \in OC(\mathcal{N}(\mathcal{O})) \) implies \( A(t) \in OC(\mathcal{P}(\mathcal{O})) \), for every \( t \in \mu(x) \) and unary predicate \( A \in N_C \).


### 5.7. PROOFS

#### iii) \( R(x, y) \in OC(N(O)) \), where \( R \) is a backward-sound role implies: for every \( t_1 \in \mu(x) \), there exists a \( t_2 \in \mu(y) \) s.t. \( R(t_1, t_2) \in OC(P(O)) \).

#### iv) \( R(x, y) \in OC(N(O)) \), where \( R \) is forward-sound role implies: for every \( t_2 \in \mu(y) \), there exists a \( t_1 \in \mu(x) \) s.t. \( R(t_1, t_2) \in OC(P(O)) \).

#### v) \( R(x, y) \in OC(N(O)) \), where \( R \) is a simple role implies: for every \( t_1 \in \mu(x) \), there exists a \( t_2 \in \mu(y) \) s.t. \( R(t_1, t_2) \in OC(P(O)) \), and for every \( t_2 \in \mu(y) \), there exists a \( t_1 \in \mu(x) \) s.t. \( R(t_1, t_2) \in OC(P(O)) \).

#### vi) \( x \approx y \in OC(N(O)) \) implies: for every \( t_1 \in \mu(x) \), there exists a \( t_2 \in \mu(y) \) s.t. \( t_1 \approx t_2 \in OC(P(O)) \), and for every \( t_2 \in \mu(y) \), there exists a \( t_1 \in \mu(x) \) s.t. \( t_1 \approx t_2 \in OC(P(O)) \).

**Proof 24** By induction on the derivation level of atoms and terms in \( OC(N(O)) \).

IH: the hypothesis holds for every \( ABax \) assertion, named individual \( a \in N_1 \) and facts of type \( x \approx x \in OC(N(O)) \).

IH: the hypothesis holds for every atom/term \( a \) with \( \text{level}(a, OC(N(O))) < k \). We show that it holds also for every atom/term \( a \) with \( \text{level}(a, OC(N(O))) = k \):

#### i) \( a \in \text{terms}(OC(N(O))) \) (other than some \( i \in N_1 \)). Then \( a \) is either of the form:

1. \( \nu^{C}_{RD} \): then, it has been introduced in \( OC(N(O)) \) via a rule of the form \( C(x) \rightarrow R(x, \nu^{C}_{RD}) \land D(\nu^{C}_{RD}) \) and \( sk(O) \) contains a counterpart rule \( C(x) \rightarrow R(x, f^{C}_{RD}(x)) \land D(f^{C}_{RD}(x)) \). Then: \( \text{level}(x, OC(N(O))) < \text{level}(a, OC(N(O))) \). From the IH: \( \mu(x) \neq \emptyset \) and for every \( y \in \mu(x) \): \( C(y) \in OC(P(O)) \). Thus, there exists a \( u \) s.t. \( C(u) \in OC(P(O)) \) and from (1) it follows that: \( D(f^{C}_{RD}(u)) \in OC(P(O)) \). Then \( f^{C}_{RD}(u) \in \mu(f^{C}_{RD}) \), and thus \( \mu(\nu^{C}_{RD}) \neq \emptyset \).

2. or of the form \( f^{C}_{RD}(y) \). From the IH: \( \mu(y) \neq \emptyset \) and thus, \( \mu(f^{C}_{RD}(y)) \neq \emptyset \).

#### ii) \( a \) is of the form \( A(x) \). Then, \( N(O) \) must contain a rule with head \( a \) whose body is satisfied in \( OC(N(O)) \):

1. \( C_1(x) \land \ldots \land C_n(x) \rightarrow D(x) \) (where \( a = D(x) \)): from the IH, for every \( t \in \mu(x) \): \( C_1(t), \ldots, C_n(t) \in OC(P(O)) \). Then, by applying the counterpart rule in \( sk(O) \) we obtain that for every \( t \in \mu(x) \): \( D(t) \in OC(P(O)) \).

2. \( C(x) \rightarrow R(x, f^{C}_{RD}(x)) \land D(f^{C}_{RD}(x)) \) (where \( a = D(f^{C}_{RD}(x)) \)): from the IH, for every \( t \in \mu(x) \): \( C(t) \in OC(P(O)) \). Then, for every \( t \in \mu(x) \): \( D(f^{C}_{RD}(t)) \in OC(P(O)) \), or for every \( t \in \mu(f^{C}_{RD}) \): \( D(t) \in OC(P(O)) \).

3. \( C(x) \rightarrow R(x, \nu^{C}_{RD}) \land D(\nu^{C}_{RD}) \) (where \( a = D(\nu^{C}_{RD}) \)). Then, there exists a GCI of type: \( C \subseteq \exists R.D \in O \) and \( sk(O) \) contains a rule of type \( C(x) \rightarrow D(f^{C}_{RD}(x)) \). Note that this is the only rule which introduces functional terms of type \( f^{C}_{RD}(\ldots) \). Thus, for every such term \( t = f^{C}_{RD}(y) \) occurring in \( \text{terms}(OC(P(O))) \) it holds that \( D(t) \in OC(P(O)) \). But \( \mu(\nu^{C}_{RD}) \) is exactly the set of all such terms.
5.7. PROOFS

4. \( R(x, y) \land C(y) \rightarrow D(x) \) (where \( a = D(x) \)). Then, \( R \) must be a backward-sound role (from the definition of safe roles). From the IH: for every \( t \in \mu(x) \), \( C(t) \in OC(\mathcal{P}(\mathcal{O})) \) and there exists a \( t' \in \mu(y) \) s.t. \( R(t, t') \in OC(\mathcal{P}(\mathcal{O})) \). Then, by applying the counterpart rule in \( sk(\mathcal{O}) \) for every \( t \in \mu(x) \) we obtain \( D(t) \in OC(\mathcal{P}(\mathcal{O})) \).

5. \( C(x) \land x \approx y \rightarrow C(y) \). From the IH: for every \( t \in \mu(x) \), \( C(t) \in OC(\mathcal{P}(\mathcal{O})) \) and for every \( t_2 \in \mu(y) \) there exists a \( t_1 \in \mu(x) \) s.t. \( t_1 \approx t_2 \). Then, \( C(t_1) \in OC(\mathcal{P}(\mathcal{O})) \) for every such \( t_1 \), and by applying the counterpart rule in \( sk(\pi(\mathcal{O})) \) we obtain \( C(t_2) \in OC(\mathcal{P}(\mathcal{O})) \), for every \( t_2 \in \mu(y) \).

iii) \( a = R(x, y) \), where \( R \) is a backward-sound role. Then, there must be a ground rule with head \( R(x, y) \) whose body is satisfied in ground(\( \mathcal{N}(\mathcal{O}) \)):

1. \( U(x, y) \rightarrow R(x, y) \). As \( R \) is a backward-sound role, \( U \) is a backward-sound role as well. From the IH: for every \( t_1 \in \mu(x) \), there exists \( t_2 \in \mu(y) \) s.t. \( U(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and thus, by applying the counterpart rule in \( sk(\pi(\mathcal{O})) \): for every \( t_1 \in \mu(x) \), there exists \( t_2 \in \mu(y) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \).

2. \( U(y, x) \rightarrow R(x, y) \). Then, \( U^{-} \subseteq R \), and \( U \) must be a forward-sound role (otherwise \( R \) would not be a backward-sound role). Then from the IH: for every \( t_2 \in \mu(x) \), there exists \( t_1 \in \mu(y) \) s.t. \( U(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and thus by applying the counterpart rule in \( sk(\pi(\mathcal{O})) \): for every \( t_2 \in \mu(x) \), there exists \( t_1 \in \mu(y) \) s.t. \( R(t_2, t_1) \in OC(\mathcal{P}(\mathcal{O})) \).

3. \( C(x) \rightarrow R(x, f_RD(x)) \land D(f_RD(x)) \). Similar to case ii) 2) above: for every \( t \in \mu(x) \):
   \[ R(t, f(t)) \in OC(\mathcal{P}(\mathcal{O})) \]

4. \( C(x) \rightarrow R(x, v_{RD}^{C}) \land D(v_{RD}^{C}) \). Then, there exists a GCI of type: \( C \subseteq \exists RD \) in \( \mathcal{O} \) and \( sk(\pi(\mathcal{O})) \) contains a rule of type \( C(x) \rightarrow D(f_{RD}(x)) \) (\( \dagger \)). From the IH: for every \( t \in \mu(x) \):
   \( C(t) \in OC(\mathcal{P}(\mathcal{O})) \). Then, by applying (\( \dagger \)) we obtain \( R(t, f_{RD}(t)) \in OC(\mathcal{P}(\mathcal{O})) \), for every \( t \in \mu(x) \).

5. \( R(x, s) \land R(s, y) \rightarrow R(x, y) \). From the IH it follows that: for every \( t_1 \in \mu(x) \), there exists \( t_2 \in \mu(s) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \) and for every \( t_2 \in \mu(s) \), there exists \( t_3 \in \mu(y) \) s.t. \( R(t_2, t_3) \in OC(\mathcal{P}(\mathcal{O})) \). By applying the counterpart rule in \( sk(\pi(\mathcal{O})) \), we obtain that for every \( z \in \mu(x) \) there exists \( u \in \mu(y) \) s.t. \( R(z, u) \in OC(\mathcal{P}(\mathcal{O})) \).

6. \( R(x, y) \land x \approx z \rightarrow R(z, y) \). From the IH: for every \( t_1 \in \mu(x) \), there exists \( t_2 \in \mu(y) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and for every \( t_3 \in \mu(z) \), there exists \( t_1 \in \mu(x) \) s.t. \( t_1 \approx t_3 \). Then, by applying the counterpart rule in \( sk(\pi(\mathcal{O})) \), we obtain that for every \( t_3 \in \mu(z) \) there exists \( t_2 \in \mu(y) \) s.t. \( R(t_3, t_2) \in OC(\mathcal{P}(\mathcal{O})) \).

7. \( R(x, y) \land y \approx z \rightarrow R(x, y) \). From the IH: for every \( t_1 \in \mu(x) \), there exists \( t_2 \in \mu(y) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and for every \( t_2 \in \mu(y) \), there exists \( t_3 \in \mu(z) \) s.t. \( t_2 \approx t_3 \). Then, by applying the counterpart rule in \( sk(\pi(\mathcal{O})) \), we obtain that for every \( t_1 \in \mu(x) \) there exists \( t_3 \in \mu(y) \) s.t. \( R(t_1, t_3) \in OC(\mathcal{P}(\mathcal{O})) \).
iv) \( a = R(x, y) \), where \( R \) is a forward-sound role. Then, there must be a ground rule with head \( R(x, y) \) whose body is satisfied in \( \text{ground}(\mathcal{N}(\mathcal{O})) \):

1. \( U(x, y) \rightarrow R(x, y) \). Similar to case iii) 1) above.

2. \( U(y, x) \rightarrow R(x, y). \) Then, \( U^- \subseteq_R R \) and thus, \( U \) is a backward-sound role. From the IH:
   for every \( t_1 \in \mu(y) \), there exists \( t_2 \in \mu(x) \) s.t. \( U(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and thus by applying the counterpart rule, for every \( t_1 \in \mu(y) \), there exists \( t_2 \in \mu(x) \) s.t. \( R(t_2, t_1) \in OC(\mathcal{P}(\mathcal{O})) \).

3. \( C(x) \rightarrow R(x, f^C_{RD}(x)) \land D(f^C_{RD}(x)). \) Similar to case iii) 2) above.

4. \( C(x) \rightarrow R(x, v^C_{RD}) \land D(v^C_{RD}): \) then \( R \) must be safe. Contradiction with \( R \) being a forward-sound role.

5. \( R(x, s) \land R(s, y) \rightarrow R(x, y). \) From the IH it follows that: for every \( t_3 \in \mu(y) \), there exists \( t_2 \in \mu(s) \) s.t. \( R(t_2, t_3) \in OC(\mathcal{P}(\mathcal{O})) \) and for every \( t_2 \in \mu(s) \), there exists \( t_1 \in \mu(x) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \). By applying the counterpart rule, we obtain that for every \( t_3 \in \mu(y) \) there exists \( t_1 \in \mu(x) \) s.t. \( R(t_1, t_3) \in OC(\mathcal{P}(\mathcal{O})) \).

6. \( R(x, y) \land x \approx z \rightarrow R(z, y). \) From the IH: for every \( t_2 \in \mu(y) \), there exists \( t_1 \in \mu(x) \) s.t. \( R(t_1, t_2) \in OC(\mathcal{P}(\mathcal{O})) \), and for every \( t_1 \in \mu(x) \), there exists \( t_3 \in \mu(z) \) s.t. \( t_1 \approx t_3 \). Then, by applying the counterpart rule in \( \text{sk}(\pi(\mathcal{O})) \), we obtain that for every \( t_2 \in \mu(y) \) there exists \( t_3 \in \mu(z) \) s.t. \( R(t_3, t_2) \in OC(\mathcal{P}(\mathcal{O})) \).

7. Similar to case iii) 7) above.

v) \( a = R(x, y) \), with \( R \) being a simple role. Then, there must be a ground rule with head \( R(x, y) \) whose body is satisfied in \( \text{ground}(\mathcal{P}(\mathcal{O}) \uparrow \omega) \):

1. \( U(x, y) \rightarrow R(x, y): \) \( U \) is a simple role as well, follows directly from the IH.

2. \( U(y, x) \rightarrow R(x, y): \) \( U^- \) is a simple role as well, follows from the symmetry of the IH.

3. \( C(x) \rightarrow D(v^C_{RD}) \land R(x, v^C_{RD}): \) then \( R \) must be safe and there exists a GCI of type: \( C \subseteq \exists R.D \) in \( \mathcal{O} \) and \( \text{sk}(\mathcal{O}) \) contains a rule of type \( C(x) \rightarrow D(f^C_{RD}(x)) \land R(x, f^C_{RD}(x)) \). Note that this is the only rule which introduces functional terms of type \( f^C_{RD}(\ldots) \). Thus, for every such term \( t = f^C_{RD}(y) \) occurring in \( \text{terms}(OC(\mathcal{P}(\mathcal{O}))) \) it holds that \( R(y, f^C_{RD}(y)) \in OC(\mathcal{P}(\mathcal{O})) \). But \( \mu(v^C_{RD}) \) is exactly the set of all such terms. Also, from the IH for every \( t \in \mu(x) \): \( C(t) \in OC(\mathcal{P}(\mathcal{O})) \). Then, for every \( t \in \mu(x) \): \( R(t, f^C_{RD}(t)) \in OC(\mathcal{P}(\mathcal{O})) \).

4. \( C(x) \rightarrow D(f^C_{RD}(y) \land R(x, f^C_{RD}(y)): \) from the IH, for every \( t \in \mu(x) \): \( C(t) \in OC(\mathcal{P}(\mathcal{O})) \).

Then, for every \( t \in \mu(x) \): \( R(t, f^C_{RD}(t)) \in OC(\mathcal{P}(\mathcal{O})) \), or for every \( t \in \mu(f^C_{RD}(x)) \): \( D(t) \in OC(\mathcal{P}(\mathcal{O})) \).

5. \( R(x, y) \land x \approx z \rightarrow R(z, y). \) Similar to cases iii) 6) (in one direction) and iv) 6) (in the other direction) above.

6. \( R(x, y) \land y \approx z \rightarrow R(x, z). \) Similar to cases iii) 7) (in one direction) and iv) 7) (in the other direction) above.
vi) a is an equality atom: \( a = x \approx y \). Then, there must be a ground rule whose body is satisfied in
\( \text{ground}(P(O) \uparrow \omega) \):

1. \( C(s) \land R(s,x) \land D(x) \land R(s,y) \land D(y) \rightarrow x \approx y \): Then, \( R \) is a simple role and from the IH:
   - for every \( t_1 \in \mu(x) \), \( D(t_1) \in OC(P(O)) \) and there exists \( t_2 \in \mu(s) \) s.t. \( R(t_2,t_1) \in OC(P(O)) \) and for every \( t_2 \in \mu(s) \) there exists \( t_3 \in \mu(y) \) s.t. \( R(t_2,t_3) \in OC(P(O)) \).
   
   Also, for every \( t_2 \in \mu(s) \), \( C(t_2) \in OC(P(O)) \), and for every \( t_3 \in \mu(y) \), \( D(t_3) \in OC(P(O)) \).
   
   Thus, by applying the counterpart equality rule in \( sk(\pi(O)) \), we obtain
   that for every \( t_1 \in \mu(x) \), there exists \( t_3 \in \mu(y) \) s.t. \( t_1 \approx t_3 \);
   
   • similarly as above one can show that for every \( t_3 \in \mu(y) \), there exists \( t_1 \in \mu(x) \) s.t. \( t_1 \approx t_3 \);

2. \( C(x) \rightarrow x \approx a \). From the IH: for every \( t \in \mu(x) \), \( C(t) \in OC(P(O)) \) and thus for every
   \( t \in \mu(x) \): \( t \approx a \in OC(P(O)) \). As \( \mu(x) \neq \emptyset \) (also from the IH), it follows that for every
   \( t_2 \in \mu(a) = \{a\} \), there exists \( t_1 \in \mu(x) \) s.t. \( t_1 \approx t_2 \).

3. \( x \approx y \rightarrow y \approx x \): follows from the symmetry of the IH.

4. \( x \approx y \land y \approx z \rightarrow x \approx z \): follows from the IH, similar to case iv) 5), but bidirectional.

Claims a) and b) follow directly from Lemma 5.7.2 point ii).

Lemma 5.7.3 Let \( O \) be a Horn-\( SHOIQ \) ontology. If \( f_{RD}^C(f_{SB}^A(t)) \in \text{terms}(OC(N(O))) \), then
\( E(v_{RD}^C,v_{SB}^A) \in OC(P_G(O)) \).

Proof 25 Let \( \mathcal{V} = \{v_{RD}^C \mid v_{RD}^C \in \text{terms}(OC(N(O)))\} \) and \( \mu \) be the function defined over the set of
\( \text{terms}(OC(N(O))) \) as follows:

\[
\mu(x) = \begin{cases} 
  x, & \text{if } x \in N_j \cup \mathcal{V} \\
  v_{RD}^C, & \text{if } x = f_{RD}^C(y) 
\end{cases}
\]

Then, it can be shown by straightforward induction that: \( C(x) \in OC(N(O)) \) implies \( C(x) \in OC(P_G(O)) \) (all rules in \( N(O) \) are also in \( P_G(O) \) except for rules of type \( C(x) \rightarrow R(x,f_{RD}^C(x)) \land D(f_{RD}^C(x)) \) which are replaced with rules of type \( C(x) \rightarrow R(x,v_{RD}^C) \land D(v_{RD}^C) \land \text{PE}(x,v_{RD}^C) \)).

Assume \( f_{RD}^C(f_{SB}^A(t)) \in \text{terms}(OC(N(O))) \). Then, \( \text{ground}(N(O)) \) must contain the following two rules:

• \( A(t) \rightarrow S(t,f_{SB}^A(t)) \land B(f_{SB}^A(t)), \) and

• \( C(f_{SB}^A(t)) \rightarrow R(f_{SB}^A(t),f_{RD}^C(f_{SB}^A(t))) \land D(f_{RD}^C(f_{SB}^A(t))), \)

and it must also be the case that: \( A(t) \in OC(N(O)) \) and \( C(f_{SB}^A(t)) \in OC(N(O)) \). Then, \( A(\mu(t)) \) \( \in \)
\( OC(P_G(O)) \), \( C(\mu(f_{SB}^A(t))) \) \( \in OC(P_G(O)) \), and \( \text{ground}(P_G(O)) \) contains the following rules:

• \( A(\mu(t)) \rightarrow S(\mu(t),v_{RD}^C) \land B(v_{RD}^C) \land \text{PE}(\mu(t),v_{RD}^C), \)

• \( C(v_{RD}^C) \rightarrow R(v_{RD}^C,v_{SB}^A) \land D(v_{SB}^A) \land \text{PE}(v_{RD}^C,v_{SB}^A), \)
5.7. PROOFS

- \( U(v_{RD}^C) \land PE(v_{RD}^C, v_{SB}^A) \land U(v_{SB}^A) \rightarrow E(v_{RD}^C, v_{SB}^A) \), and
- facts: \( U(v_{RD}^C) \) and \( U(v_{SB}^A) \).

From the above it follows that: \( P_G(O) \models E(v_{RD}^C, v_{SB}^A) \), and thus: \( E(v_{RD}^C, v_{SB}^A) \in OC(P_G(O)) \).

Lemma 5.3.5 Let \( O \) be an RSA ontology with signature \( \Sigma \). Then, the Skolem chase of \( N(O) \) terminates with a Herbrand model of polynomial size. Furthermore, if \( O \) is WRSA, then the Skolem chase of \( N(O) \) terminates with a Herbrand model of size at most exponential.

Proof Let \( t \in \text{terms}(OC(N(O))) \). Then, \( t \) is of the form \( g_n(\ldots(g_1(u))\ldots) \), where each \( g_i \) is of the form \( f_{R_i,D_i}^C \) and \( u \in N_I \) or \( u \) is of the form \( v_{RD}^C \). From Lemma 5.7.3 it follows that \( E(v_i, v_{i+1}) \in OC(P_G(O)) \), where \( v_i = f_{R_i,D_i}^C \), for every \( 1 \leq i < n \).

If \( d \) is a polytree, for every two nodes \( v_1 \) and \( v_n \) there is at most one path in \( G(O) \): \( (v_1,\ldots,v_n) \) which connects them. Thus for given \( g_n, g_1, \) and \( u \), \( OC(N(O)) \) contains at most one term \( t \) as above.

As both the number of function symbols and of terms of form \( v_{RD}^C \) in \( N(O) \) is polynomial in the size of the \( O \) and the number of unary and binary atoms which occur in \( N(O) \) is also polynomial, it follows that the size of \( OC(N(O)) \) if also polynomial in the size of \( O \).

If \( G(O) \) is acyclic, every path of the form \( (v_1,\ldots,v_n) \) in \( G(O) \) must not contain the same node twice. Then, the number of terms \( t \) of form \( g_n(\ldots(g_1(u))\ldots) \) is bounded by \( c^k n \), where \( c \) is the number of named individuals and terms of form \( v_{RD}^C \) occurring in \( N(O) \) and \( k \) is the number of function symbols occurring in \( N(O) \). Thus, the total number of terms occurring in \( N(O) \) is finite and bounded by \( c \Sigma_{0 \leq i \leq k} k^i \), which is exponential in the size of \( O \). Consequently, the size of \( OC(N(O)) \) is also bounded by an exponential in the size of \( O \).
Conclusions and Future Work

We have proposed the new classes of $\text{RCA}_n$ and $\text{RSA}$ ontologies, presented in the previous two chapters. Our experiments suggest that a significant proportion of out-of-profile ontologies are $\text{RCA}_n$ and $\text{RSA}$; as a result, we can exploit efficient algorithms, also presented in the previous sections, to efficiently solve reasoning tasks over ontologies within these fragments.

As for future work, we intend to follow different directions with both approaches presented in the previous chapters.

- First of, we intend to lift $\text{RCA}_n$ so it can be applied to the more general fragment of disjunctive existential rules; i.e., rules that, apart from existentially quantified variables in the head, also allow for the use of disjunction. This extension would allow the application of $\text{RCA}_n$ to (possibly non-Horn) ontologies, for which there are no implementations that can solve CQ entailment.

- The reasoning techniques presented in the previous chapter may be extended to the whole fragment of Horn-$\text{ALCIOQ}$. Despite the fact that tractability of reasoning would be lost, we foresee that this would enable the development of very efficient reasoning algorithms for this fragment. Furthermore, the notions from [Feier et al. 2015], an implementation to solve CQ entailment over such fragment could also be produced.
7

References
References


Stefanoni, G., Motik, B., and Horrocks, I. 2013. Introducing nominals to the combined query answering approaches for $\mathcal{EL}$. In AAAI.


