Reasoning in Description Logic Ontologies for Privacy Management

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

vorgelegt an der
Technischen Universität Dresden
Fakultät Informatik

eingereicht von
Adrian Nuradiansyah, M.Sc.
geboren am 29. August 1992 in Jakarta, Indonesien

eingereicht am 20. September 2019
verteidigt am 25. November 2019

Gutachter:
Prof. Dr.-Ing. Franz Baader
Technische Universität Dresden

Prof. Piero A. Bonatti
Universita' di Napoli Federico II

Dresden, im December 2019

Acknowledgment

I would like to present my first gratitude to my wife, BERLIAN PRAWIRO, and my little son, CEDRIC ATHALLA NURADIANSYAH for the happiness and cheerfulness they always share every day.

During the work on my thesis for 3 years, I am indebted to my supervisor, FRANZ BAADER, very much for his supports, insightful comments, as well as his patience. I would also like to thank THORSTEN STRUFE for being the subject expert of my thesis, which shared many privacy scenarios that are relevant to my thesis. Furthermore, I would like to express my gratitude to PIERO BONATTI for conducting a review on my thesis.

I am also thankful to ANNI-YASMIN TURHAN that has become my first lecturer in Description Logics since I studied in Dresden 5 years ago. My sincere appreciations are also sent to RAFAEL PENALOZA that has kindly advised me for the study of *ontology repair* during my 3-months visit to Free-University of Bozen-Bolzano last year as well as to BERNARDO CUENCA GRAU and EGOR KOSTYLEV that also has generously supported me in understanding the study of *privacy-preserving ontology publishing* during my 3-months visit to the University of Oxford this year.

I highly appreciate KERSTIN ACHTRUTH and ULRIKE SCHÖBEL for their help in administrative works and their encouraging and positive words. I also share my gratitude to the 'Graduiertenkolleg' RoSI that has supported me in terms of facilities, lectures, workshops, and finance. Moreover, I also thank my colleagues in the Chair of Automata and RoSI as well as to my friends for the joy and the exciting time.

Last, but not least, I thank my families in Depok and Samarinda that always support me via video call almost every day :)

Contents

Chapter 1

Introduction

Living in an era that enables us to collect and disseminate any form of data provides us with various impacts. On the positive side, governments, companies, or individuals may employ all these data for knowledge-based decision making, which then prompt them to make own significant business or individual decisions. Notwithstanding these promising benefits, this sort of data transaction activity may cause the sacrifice of the privacy value of the data. For example, in a very famous study from Sweeney in [Swe00], it was found that 87% of the US population can be uniquely identified by gender, ZIP code, and full date of birth within an experiment using 1990 US Census summary data. In medical area, research from [Wes76] stated that the necessity of patients data dissemination may infer sensitive or personal information of the patients, such as their job status, their types of insurance, or even an information about whether the patients are permitted to drive cars or not due to the serious illnesses they suffer from. These concerns also arise in workplaces where employers' actions aimed to protect company assets or safeguard proprietary information may violate confidential data of employees if these data are used for inappropriate purposes [FR07].

Realizing all these privacy matters, one needs to understand avenues and methods of producing a sort of anonymous data that preserves privacy policies, but keep a high utility value when it is released publicly. In databases, the study of data privacy itself has been a long-standing research area. This ranges from the studies of privacy detection in databases (e.g., [MS07; DP05]), enforcement for controlled query evaluation (e.g., [BB04]), database anonymizations (e.g., [Swe02a]), to the study of attack models in databases (e.g., [FWC+10]).

That being said, all such existing works on privacy in databases mainly assume that the information in databases is *complete*. When we shift to another more expressive data representation, such as *ontologies*, then they are mainly assumed to deal with *incomplete* information. The term incompleteness in this context means that we can infer additional facts from the ontologies, which are not explicitly stated there. Furthermore, in contrast to complete databases, during query evaluations, all these inferred facts are derived by taking *all* models of the ontologies into account, which intuitively may involve complex interactions with other pieces of information within the ontologies.

It should be realized that *incomplete databases* [Lev96; MHS09] also exist in the area of databases and ontologies are strongly related with them. Moreover, researches on data privacy in the context of incomplete DBs have also been investigated in some literature, such as [BTW10; BW08]. However, ontologies itself are formulated using languages that are much more expressive than database schema languages. One of the most common languages formalizing ontology is the Web Ontology Language (OWL)[1] that has been standardized in

[1] See https://www.w3.org/TR/owl-features/

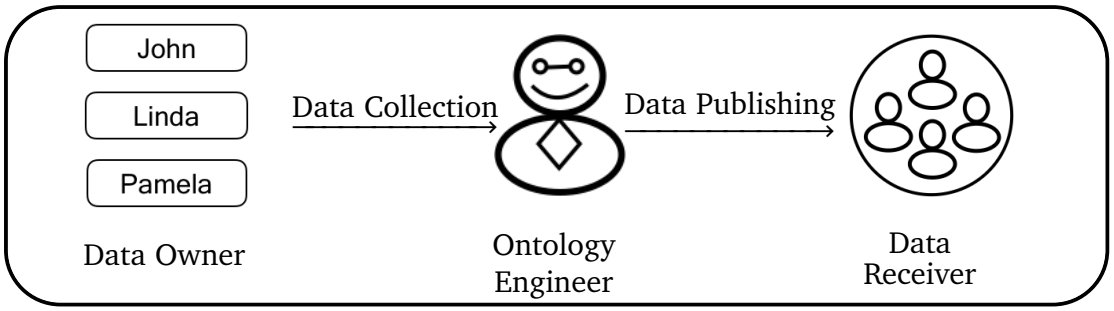


Figure 1.1: A scenario for data transactions within organizations

2004 and frequently used in many application domains. This standardization also resulted in a connection with a family of knowledge representation (KR) languages, called Description Logics (DLs) [BCM+03; BHL+17], that are more known as a logic-based semantics which, up to some different notations, are a fragment of first-order logic. This connection provides more promising features for ontologies to not only be used for giving well-defined and understandable reasons why a statement is entailed by the ontologies, but also be extensively integrated with other automated reasoning technologies, e.g., DL reasoners [2].

However, the capacity of ontologies to infer new facts is still prone to privacy violations in general. To come up with reasonable actions that can be used for privacy management in ontologies, we start with a scenario of data transactions depicted in Figure 1.1. Taking a look at the figure, a typical scenario for data transactions in some (government) organizations consists of the *data collection phase* and the *data publication phase*. In the former phase, the ontology engineer collects the data from data owners, e.g., John and Linda, and then stores them in a database that has been augmented with an ontology. Meanwhile, in the data publication phase, the ontology engineer releases the whole or some parts of the ontology to (possibly unknown) data receivers. As an example, an IT division of a hospital, acting as an ontology engineer, collects the data about medical information of patients and then share parts or the whole of the data to a data receiver, e.g., a medical center, that wants to build a cyclical pattern of some diseases based on the patient data by integrating it with other external ontologies. Nevertheless, a privacy leak is very susceptible to occur in the publication phase. For instance, it might be the case that information about patients are disclosed unauthorizedly to users or data receivers who have no access to the information. This means that the ontology engineer needs to be at least guided with the following anticipation steps:

1. Asking if the confidential information of objects is kept hidden or not w.r.t. the ontology. In particular, as mentioned by [Gra10], the identity of an (anonymous) object is a critical asset that needs to be protected in many application domains, e.g., medical or insurance.

2. Repairing the ontology in a minimal way such that the identity or other sensitive properties of individuals cannot be inferred from any user, but at the same time, the utility value of the ontology is still preserved.

[2] See http://owl.cs.manchester.ac.uk/tools/list-of-reasoners/

3. The two above steps only assume that the users' knowledge is a part of the ontology in the data publication phase. However, according to [FWC+10; Gra10], when the ontology is ready to be published on the Web or to be integrated with other external applications, the solution for ontology repairs should be featured with a provable logic-based guarantee against a linkage attack from other possible users (attackers) that may have extra knowledge from different sources. In particular, this sort of attacks may occur when the combination of the repaired ontology with background knowledge of external users (atackers) still violates a privacy policy.

In the context of privacy in OWL or DL ontologies, to the best of our knowledge, the studies of preserving identity or reckoning linkage attacks during the publication phase are still unexplored, whereas the studies of ontology repairs have been carried out by e.g., [Hor11; DQF14; LSP+08; TCG+18] with different settings and motivations. In this thesis, we only focus on ontologies written in Description Logics and, in general, we deal with the following tasks inspired by the three anticipation steps above:

1. We formulate various new reasoning problems in DLs aimed at preserving the identity of individuals. We present algorithms to solve the problems and provide complexity results for each problem.

2. Then, in case the identity or other properties of some object is not preserved, we build a general framework for *repairing ontologies* in a minimal way based on *information weakening* that can be used to get rid of unwanted consequences, e.g., to hide secret consequences.

3. Last, we study *privacy-preserving ontology publishing* where the ontology is ready to be published on the Web. Since information linkage may happen and violate privacy of individuals in this case, we specifically provide a logical mechanism against such matters.

Relating the three main tasks above with the studies on privacy in general, we realize that what we focus on this thesis is more about the confidentiality aspects of privacy and then build mechanisms to achieve those confidentiality goals. Some other privacy aspects, such as controlling personal data usage after disclosure [FAK+16] or the right to be forgotten [Ros11] are beyond the scope of this thesis.

Figure 1.2 represents a general illustration of how the tasks mentioned above will be explained in detail sequentially throughout chapters within this thesis, in particular from Chapter 3 to Chapter 6. The remaining parts of this chapter are described as follows. In Section 1.1, we give a brief introduction to Description Logics. In Section 1.2, we explain existing work on detecting whether confidential information about individuals is protected or not in some form of information systems, such as databases or ontologies. In Section 1.3, we present existing methods to repair the systems in a way that the unwanted consequences can be removed. Then, in Section 1.4, we discuss the state-of-the-art of the research on *privacy-preserving data publishing*, which motivates the study on *privacy-preserving ontology publishing* in this thesis. Last, in Section 1.5, we give an outline of the thesis as well as the list of related publications.

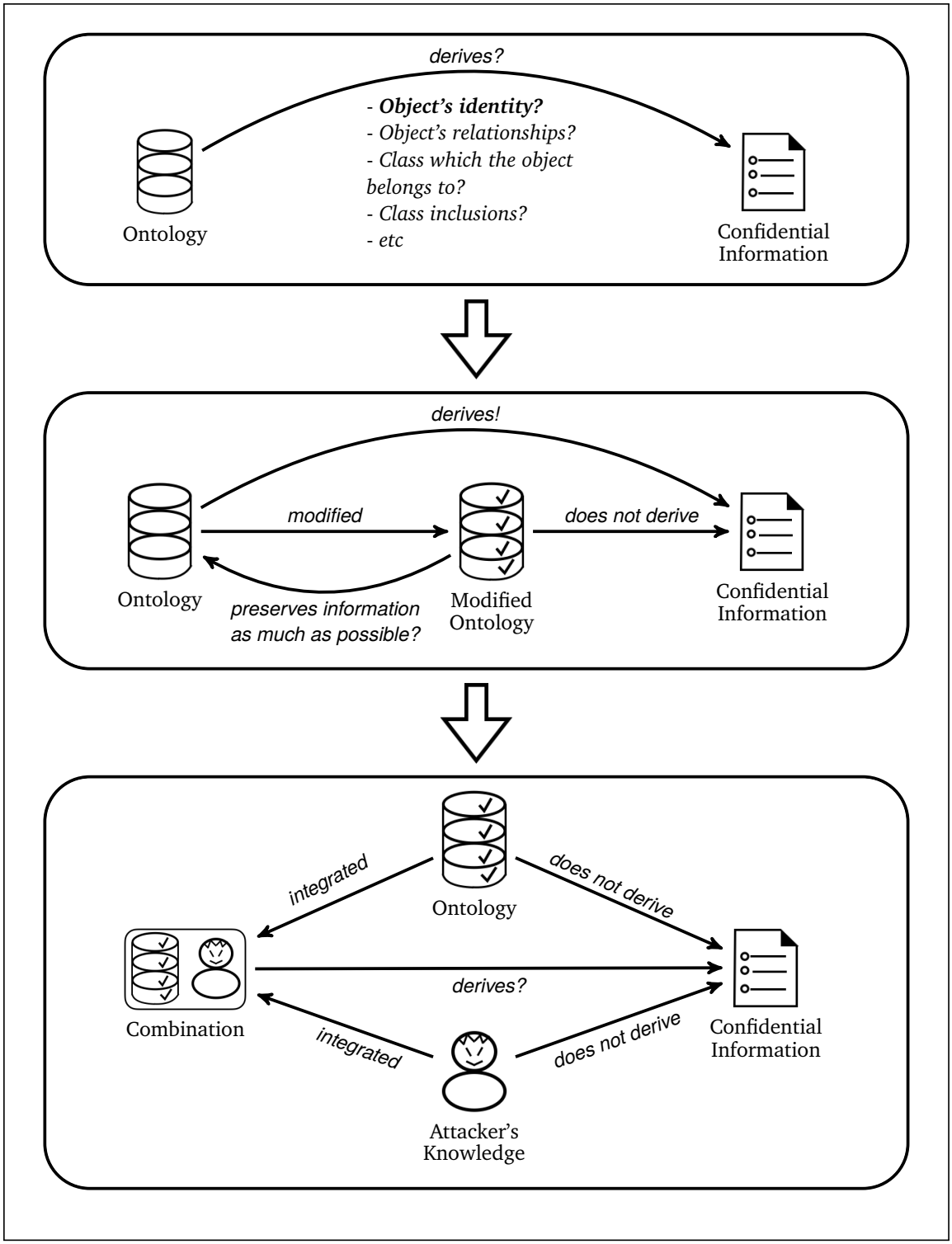


Figure 1.2: Illustrations of problem settings described from Chapter 3 to Chapter 6

1.1 Description Logics

Description logics (DLs) are a family of knowledge representation languages which can represent conceptual knowledge of an application domain in a formally structured and well-understood way [BCM+03; BHL+17]. Due to the prominent adoption of the DL-based language OWL [HPH03], which has become the standard ontology language, many important application areas use DLs as the definition of their ontologies, for instance, in the biology and medical areas [HSG15]. Prominent ontologies that have been used for many different purposes in these two areas are, such as SNOMED CT[3], GALEN[4], or GENE ONTOLOGY[5].

In terms of the syntactical features DLs have, the building blocks of these logics are *concept names* expressing the set of elements that can be viewed as unary predicates, *role names* denoting a relation between elements that can be seen as binary predicates, and *individual names* that point to single elements. From these blocks, the notion of DL *concepts* are formed using the constructors that the DL has. For example, in a workplace application domain, Employee and TechTeam are concept names expressing the set of all employees and the set of all technical teams, respectively, while worksAt is a role name denoting a binary relation between employees and the divisions where the employee is working. For instance, the DL concept

$$\mathsf{Employee} \sqcap \exists \mathsf{worksAt}.\mathsf{TechTeam}$$

defines the set of all employees working at a technical team. When we want to construct DL ontologies, it should be considered what sort of information or axioms we want to include in the ontologies. Normally, DL ontologies consist of two different types of knowledge. First, it has a terminological knowledge (TBox) consisting of axioms, called *General Concept Inclusions (GCIs)*, stating hierarchical relationships between concepts. For example, in the medical area, we may have an axiom

$$\begin{aligned}\mathsf{Patient} \sqcap \exists \mathsf{suffer}.\mathsf{BloodCancer} \quad \sqsubseteq \quad & \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \\ & \exists \mathsf{symptom}.\mathsf{Fatigue} \sqcap \exists \mathsf{symptom}.(\mathsf{Rash} \sqcap \exists \mathsf{feels}.\mathsf{Itchy}).\end{aligned}$$

In addition to TBoxes, DL ontologies also consist of assertional knowledge (ABox), where instances of concepts and the relationships between these instances are stated. For example, the assertions

$$\mathsf{Patient}(\mathsf{JOHN}), \mathsf{suffer}(\mathsf{JOHN}, \mathsf{BC}), \mathsf{BloodCancer}(\mathsf{BC})$$

say that John is a patient suffering from a blood cancer. DL ontologies provide their users with abilities to infer implicit knowledge from the stated explicit axioms. One may wonder whether there is an implicit contradiction within an ontology by asking if the ontology has a model (*satisfiability problem*). Another typical question is to ask whether a concept is more specific than another concept w.r.t. an ontology (*subsumption problem*) or whether an individual is an instance of a given concept w.r.t. an ontology (*instance problem*). For example, the combination of the terminological and assertional axioms above may derive $(\exists \mathsf{symptom}.(\mathsf{Rash} \sqcap \exists \mathsf{feels}.\mathsf{Itchy}))(\mathsf{JOHN})$ saying that John has a symptom which is a rash that feels itchy.

[3] See http://www.snomed.org/
[4] See https://bioportal.bioontology.org/ontologies/GALEN
[5] See http://geneontology.org/

A main concern occurring in many works on DLs is to develop expressive DLs that have decidable inference problems that can be solved by practical reasoning procedures. For instance, the DL $\mathcal{ALC}$ is the smallest DL closed under Boolean operator [Sch91; SS91] and the reasoning is already ExpTime-complete in this logic. Furthermore, for a DL that is more expressive than $\mathcal{ALC}$, such as $\mathcal{SROIQ}$ known as the underlying logic for the full version of the second generation of OWL, called OWL 2 [HKS06; GHM+08], the reasoning here becomes N2ExpTime-complete. This intractability issues brought motivations to DL communities to find out DLs that are tractable in practice. For instance, in the Description logic $\mathcal{EL}$, which is a sub-logic of $\mathcal{ALC}$ and only has conjunctions ($\sqcap$), existential restrictions ($\exists r.C$), and the top concept ($\top$) as its constructor, reasoning can be done in polynomial time [BBL05].

In addition to answer yes/no questions for the reasoning problems mentioned above, DL ontologies can also be viewed as collections of information used to derive answers, which are generally tuples of individuals, for queries. Commonly, many application use conjunctive queries (CQs) as a form of querying in the context of DL ontologies. A CQ is essentially a first-order formula of the form of conjunctions of atoms over unary or binary predicates with existentially quantified variables. Complexity-wise, answering conjunctive queries is already intractable for all aforementioned DLs, e.g., it is NP-complete in the Description Logic $\mathcal{EL}$ [Ros07]. However, if the complexity is measured in the size of the data only, i.e., in the size of ABox, which is called *data complexity*, then the complexity of CQ answering in $\mathcal{EL}$ becomes PTime-complete [Ros07]. Even, most members of the *DL-Lite* family, which is a family of tractable DLs mainly used in ontologies with very large ABoxes associated to relational databases, enjoy the very low data complexity AC^0 for CQ answering [CDL+13].

1.2 Detecting Privacy Breaches in Information Systems

Tracing back to very early history of philosophical discussions, we cite a concept of 'privacy' in one of the well-known Aristotle's arguments:

> *There is a distinction between two spheres of life, which are the public one associated to political activity and the private one that concerns with domestic life.*[6] (Aristotle)

Having said that, any single object in this world has their own assumptions on the notion of *privacy in their domestic life*. In information systems, the term sensitive data itself varies from one work to another and is also protected differently depending on the settings that are being considered. One of the early protections against unintended disclosures was introduced by the concept of *role-based access control* in [SCF+96] where every user of a system is labeled with a role that defines which information can be accessed by the user. In this privacy scheme, the main focus is to design a security mechanism to predefine role-permission relationship, instead of detecting whether some secret data is protected or not w.r.t. a given information system.

The work by Samarati and Sweeney in [SS98] was arguably the first one, at least in the database area, addressing the problem of releasing individual-specific data while, at the same time, safeguarding the anonymity of the individuals. In their well-defined confidentiality criteria, called *k-anonymity*, a secret is violated if there is an attempt to distinguish the

[6] https://plato.stanford.edu/entries/privacy

information for each database tuple from at least k-1 tuples whose information also appears in the released table.

A different means of shielding the secret information is also captured in the notion of *Controlled Query Evaluation* (CQE) that was studied in [BB04] for databases. The correct answer of the user queries w.r.t. the given database is judged by some censor. If the censor notices that the query answer violates one of the policies, then a *modifier* that may distort the form of the answer, either as *lying* or *refusal*, is applied. In databases, a privacy leak may occur due to a set of materialized views over the database, which is actually intended only to be an accessible layer for users to query, rather than to store the data. The various studies of checking whether a confidential query is entailed by a set of database views were investigated thoroughly in, for instance, [DP05; MS07].

Motivated by the long-standing research of privacy in databases, the concern of preserving confidential information in ontologies becomes a critical requirement in numerous applications. Deciding what can be told to a user without revealing secrets from an ontology in an access-control manner has been investigated in [BKP12], where the axioms of the ontology are labeled with access restrictions, and users can only see the (consequences of the) axioms for which they have right of access. Instead of restricting the access only on the side of the axioms, the works from [GH08; SS09; CDL+12; GMK09; GM12; GKK+15] restrict the user access on the side of the ontology consequences.

A first attempt to restrict the user access on the consequences side is by adapting the CQE paradigm over ontologies that were done by [TSH10; BS13; GKK+15]. This approach is performed by using a confidentiality-preserving layer such that any finite set of answers for the query w.r.t. this layer will not disclose any secret. Depending on the context, this layer is called *complement of a secrecy envelope* in [TSH10], *view* in [BS13], or *censor* in [GKK+15]. Moreover, in the context of ontology reuse, it might be the case that such an external ontology-based application $\mathfrak{E}$ wants to reuse an ontology $\mathfrak{O}$ whose content is not available due to privacy concerns. To avoid any leak of sensitive data during the reuse phase, [GMK09; GM12] proposed such an oracle that can be used to access the content of $\mathfrak{O}$ such that, by using a, so-called, *import-by-query* algorithm, any answer for user queries about $\mathfrak{E} \cup \mathfrak{O}$ are based on the union of $\mathfrak{E}$ and the oracle only. In another case, the practice of view-based query answering studied in [GH08; SS09; CDL+12] only allowed the user to access the ontology $\mathfrak{O}$ via a query interface that only enable them to ask permissible queries. These sort of queries, together with their answers, are represented as a set $\mathcal{V}$ of views over $\mathfrak{O}$ and the data to be protected is defined as a query q.

In [SS09], the ontologies were written in $\mathcal{ALC}$ and the queries were limited to either a *subsumption query* which asks whether a concept C is subsumed by another concept D w.r.t. $\mathfrak{O}$ or *a retrieval query* which computes all individuals that belong to a given concept w.r.t. $\mathfrak{O}$. Their privacy problem asks whether the data privacy is preserved for q w.r.t. the combination of $\mathcal{V}$ and additional background knowledge of the user. In contrast to the latter problem that only concentrates on standard semantics in DLs, the privacy problem considered in [CDL+12] provides additional semantics for view-based query answering, each one capturing additional properties for the answers of the queries. Then, applying these various semantics in their framework, they ask whether the secret facts captured as answers of q logically follow from $\mathfrak{O}$ w.r.t. $\mathcal{V}$. They apply this framework, in particular, to the *DL-Lite* family [CDL+07; ACK+09]. Differing from [SS09; CDL+12] that principally only use classical semantics in their logics, [GH08] also takes additional background knowledge of attackers captured as

a probabilistic distribution into account. They consider the notion of *perfect privacy* which guarantees that the attackers should not learn anything about the possible answer of q w.r.t. $\mathcal{V}$ whenever their knowledge is increased.

Nevertheless, as argued by [Gra10], among the confidential information about individuals, that should be commonly protected, is their *identity*. However, none of the works mentioned above explicitly considers the notion of identity of an object and provides a mechanism to detect whether the identity is hidden or not. In their works, properties of individuals, i.e., the memberships of an individual (or a tuple of individuals) are the common information they try to hide in their settings. For instance, in the medical example we had in Section 1.1, if it is secret to know that John has a symptom of the form itchy rash, then the property 'symptom' of John is not protected w.r.t. the combination of axioms written in Section 1.1. In the context of DL ontologies, however, the formal term of identity itself is also not defined yet. In Chapter 3, we define the notion of *identity* of individuals formulated in DLs and then study various reasoning problems related to identity preservation.

1.3 Repairing Information Systems

In the previous section, we described several works identifying whether a secret is deduced from either databases, ontologies, a set of views, or a combination of a protected system with an oracle. Some of the works complement this identification process with some censors that may twist the real answers of user queries or pointing the users to other subsets of real answers that do not disclose any secret. Alternatively, another possible action that can also save sensitive valuable data from unintended disclosures is by directly modifying or repairing a given information system in a way that the modified one does not entail any sensitive data. Indeed, as long as the secret is not a tautology, one may easily modify the system by emptying its information, and obviously the secret will not be deduced. However, one also needs to take care of the utility of the modified system whose data still can be learned by other parties. Thus, the modification needs to ensure that the modified ones remain maximally informative to users. Similar to previous sections, for literature study, we focus on such information systems in the structure of database and ontologies.

In database studies, two common approaches that are used to modify the data are by applying either a *perturbation* or an *anonymization* technique. A popular instance of the former technique is called *differential privacy* introduced by Dwork in [Dwo06]. Being commonly used in statistical databases for survey of population, this approach introduces noise in each of the numeric quantities of personal data in a way that the modified database will be used for further statistical analysis. However, in other cases, it is more necessary to release actual data, and not just statistical information. To this end, an anonymization approach, namely k-anonymity [SS98; Swe02b], is more suitable for this demand. To realize this approach, [Swe02a] provided a *generalization* technique which involves replacing a value with a less specific but semantically consistent value, and a *suppression* technique which decides to not release the sensitive data at all. The work on data suppression as well as complexity of optimal k-anonymity was investigated further in [MW04].

The notion of *repair* in DL ontologies initially did not come from a privacy scenario, but from a more general situation where engineers of DL-based ontologies have problems to understand why an ontology is inconsistent or why a consequence (subsumption or instance

relationships) computed by the DL reasoner actually follows from the ontology. This leads to the question on how to repair the ontology in case the consequence is not intended. This question becomes more challenging since the size of DL-based ontologies grows and the tools that support improving the quality of such ontologies are highly demanded, which imply that the need for an ontology maintenance in the sense of repairing ontologies becomes important.

Axiom pinpointing [SC03] was introduced to help developers or users of DL-based ontologies to understand the reasons why a certain consequence holds by computing so-called *justifications*, i.e., minimal subsets of the ontology that have the consequence in question. Black-box approaches for computing justifications such as [SHC+07; KPH+07; BS08] use repeated calls of existing highly-optimized DL reasoners for this purpose, but it may be necessary to call the reasoner an exponential number of times. In contrast, glass-box approaches such as [BH95; SC03; PSK05; MLB+06] compute all justifications by a single run of a modified, but usually less efficient reasoner.

Given all justifications of an unwanted consequence, one can then repair the ontology by removing one axiom from each justification. This approach gives a close connection with the well-known model-based diagnosis from [Rei87]. In [KPS+06a], this approach is extended to the very expressive DL $\mathcal{SHOIN}$ and the authors consider a ranking of axioms that can be used to select preferred repairs. An implementation of this approach was made available in the ontology editor SWOOP [KPS+06b]. However, removing complete axioms may also eliminate consequences that are actually wanted.

Approaches for repairing ontologies while keeping more consequences than the classical approach based on completely removing axioms have already been considered in the literature. On the one hand, there are approaches that first modify the given ontology, and then repair this modified ontology using the classical approach. In [Hor11], a specific syntactic structural transformation is applied to the axioms in an ontology, which replaces them by sets of logically weaker axioms. More recently, the authors of [DQF14] have generalized this idea by allowing for different specifications of the structural transformation of axioms. They also introduced a specific structural transformation that is based on specializing left-hand sides and generalizing right-hand sides of axioms in a way that ensures finiteness of the obtained set of axioms. However, it is inevitable that such an approach replacing a single axiom with several axioms might blow up the size of the ontology.

Using a different strategy, the approach in [LSP+08] adapts the tracing technique from [BH95] to identify not only the axioms that cause a consequence, but also the parts of these axioms that are actively involved in deriving the consequence. This provides them with information for how to weaken these axioms. In [TCG+18], repairs are computed via axiom weakening with the help of refinement operators that were originally introduced for the purpose of concept learning [LH10].

However, we show later in Chapter 4 that the approach in [LSP+08] does not necessarily yield a repair since, in general, computing such an optimal repair needs iterations. The authors of [TCG+18] had already realized that this iteration is needed, but they did not give an example explicitly demonstrating their iterative algorithm, and they had no termination proof of the algorithm. To this end, we also provide the termination proof in Chapter 4. This investigation supports our main purpose in Chapter 4 that actually wants to build a *gentle repair framework* which repairs the ontology based on axiom weakening with the help of the

notion of *weakening relations*. We will apply this framework to the Description Logics $\mathcal{EL}$ and $\mathcal{ALC}$.

1.4 Privacy-Preserving Data Publishing

Repairing information systems, such as ontologies, can be an effective way to get rid of unwanted consequences, e.g., the consequences we want to keep hidden in the context of privacy. However, it may only work if we assume that the users' knowledge is a part of the input ontology we want to repair. In the context of ontology publishing, it becomes sensible to deal with users that have extra knowledge which they obtain from different sources. In this case, this additional knowledge may not violate any privacy policy, but if it is combined with our input ontology, then the privacy of an individual can be revealed. This vigilant assumption somehow reminds us to a very stringent definition of privacy protection in statistical databases provided by Dalenius in [Dal77]. He enunciated the following desideratum:

> *Nothing about an individual should be learnable from the database that cannot be learned without access to the database*. (Dalenius)

This ideal aim was, however, smashed by Dwork in [Dwo06] saying that such absolute privacy protection is impossible due to the presence of background knowledge of the attacker. If we lift up Dwork's argument to any type of information system, then the background knowledge of an attacker is indeed something that needs to be carefully considered. Suppose that the attacker only knows that John lives in the same apartment room with his wife, Pamela. At the same time, the attacker is given access to an ontology disclosing the address information of Pamela. Using this access in conjunction with his background knowledge, the attacker can infer the address information of John, too. In many literatures of *Privacy-Preserving Data Publishing* (PPDP), this sort of unintended disclosure is called *linkage attack*.

In fact, some popular privacy protections in database, such as k-anonymity, was initially designated to not only publish informative individual-specific data in the anonymized table, but also to secure the data from possible attacks in the type of *record linkage*. However, as investigated thoroughly by [FWC+10], some possible linkage attacks in databases, such as *attribute linkage*, *table linkage*, or *probabilistic attack*, still can attack the privacy of an individual, even though the database has satisfied the k-anonymity criteria. This argument was also firmly supported by [MKM+07] showing a formal study of worst-case background knowledge for PPDP, and was furthermore confirmed by several confidentiality criteria in the setting of linkage attacks, such as ℓ-diversity [MKG+07] or t-closeness [LLV07], that strengthen and refine the notion of k-anonymity.

More recently, the study of PPDP and linkage attack was applied in the context of Linked Data [GK16; GK19]. In [GK16; GK19], PPDP was investigated in a setting where the information to be published is given as a relational dataset with (labeled) null values, and the privacy policy is given by a conjunctive query whose answer is not allowed to be disclosed publicly. The notion of *compliant* is introduced to make sure that the published information (dataset) does not entail the privacy policy (query) and the notion of *safety* ensuring that for any information compliant with the query, the combination of this information with our published information is again compliant. In order to make a given dataset compliant or safe, one is basically allowed to replace constants or null values by new null values. The

paper investigates the complexity of deciding compliance (Is a given modification of a dataset policy compliant?), safety (Is a given modification of a dataset safe w.r.t. a policy?), and optimality (Is a given modification of a dataset safe w.r.t. a policy and does it change the dataset in a minimal way?). The obtained complexity results depend on whether *combined* or *data* complexity is considered, and whether closed- or open-world semantics are used.

In the context of DL ontologies, the study of linkage attacks is rather unexplored. To this end, we initiate the study of *Privacy-Preserving Ontology Publishing* (PPOP) in Chapter 5 and 6. In general, the setting we design for this study is composed of published information that is given either by $\mathcal{EL}$ instance stores or $\mathcal{EL}$ ABoxes, background knowledge of attackers ranging from the form of concepts in Description Logics $\mathcal{EL}$, $\mathcal{FL}_0$, $\mathcal{FLE}$, to the form of $\mathcal{EL}$ ABoxes, and privacy policies in the form of $\mathcal{EL}$ concepts or a conjunctive query. We adopt similar decision problems defined in [GK16; GK19], which are *compliance*, *safety*, and *optimality*. Additionally, we will provide algorithms for computing optimal compliant (safe) generalizations of $\mathcal{EL}$ concepts and introduce an anonymization function for $\mathcal{EL}$ ABoxes.

1.5 Outline and Contributions of the Thesis

In the following, we will briefly give an outline of the remainder of the thesis.

In Chapter 2, we study basic notions in Description Logics, ontologies, and computational complexities. Initially, we discuss a basic Description Logic, called $\mathcal{ALC}$ in terms of its syntax and semantics. This will be followed by a discussion on DL ontologies together with reasoning problems in DLs that are commonly investigated in the literature. Then, we see a relationship between DLs and first-order logic and look at fragments of $\mathcal{ALC}$ that will be used in this thesis. One fragment of $\mathcal{ALC}$ that we concentrate on is the DL $\mathcal{EL}$ that is known for its limited, but sufficient expressiveness and its tractability for most of the classical reasoning problems in DLs. Finally, we end this chapter with the discussion on complexity of reasoning problems in DLs.

As mentioned in Section 1.2, the type of confidential information we first investigate is the one aiming at hiding the identity of individuals. Accordingly, in Chapter 3, we introduce the *identity problem* that asks whether two individuals are equal w.r.t. a given ontology. We investigate which DLs that are non-trivial to this problem and analyze its complexity, both upper bounds and lower bounds. We move to the extended problem called the *view-based identity problem* where users can only learn information about individuals of the ontology through views which they are permitted to access or store. The question is then whether the identity of anonymous individuals are hidden w.r.t. to this collection of views. Then, we look at another privacy problem that is violated if the identity of anonymous individuals belongs to a set of known individual of cardinality smaller than k. This problem is called the *k-hiding problem*. Contents, notions, and results written in Chapter 3 particularly on the identity problem and the view-based identity problem are mainly based on the following publications [BBN17a; BBN17b].

- Franz Baader, Daniel Borchmann, and Adrian Nuradiansyah. 'Preliminary Results on the Identity Problem in Description Logic Ontologies'. In *Proceedings of the 30th International Workshop on Description Logics*, Montpellier, France, 2017.

- Franz Baader, Daniel Borchmann, and Adrian Nuradiansyah. 'The Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding', In *Proceedings of Semantic Technology - 7th Joint International Conference* (JIST), Gold Coast, QLD, Australia, 2017.

In Chapter 4, we present an approach for modifying DL ontologies in order to eliminate hidden or unwanted consequences in general. We introduce a *gentle repair framework* that repairs ontologies based on weakening axioms. Then, instead of allowing arbitrarily many ways to weaken axioms, we propose the notion of a *weakening relation* restricting the way in which axioms should be weakened. We then introduce two weakening relations for $\mathcal{EL}$ axioms, especially for $\mathcal{EL}$ GCIs, that are based on generalizing the right-hand side of GCIs semantically and syntactically. Likewise, we present two weakening relations for $\mathcal{ALC}$ GCIs, where right-hand side and left-hand side of GCIs are generalized and specialized, respectively. The first one generalizes and specializes concept w.r.t. a finite signature and a fixed role-depth, while the second one performs the generalizations and specializations syntactically. All notions and results on the gentle repair framework, weakening relations and its application to $\mathcal{EL}$ axioms are mainly based on the following publications [BKN+18a; BKN+18b; BKN+18c]

- Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza. 'Making Repairs in Description Logics More Gentle'. In *Proceedings of the Sixteenth International Conference* (KR), Tempe, Arizona, US, 2018.
- Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza. 'Making Repairs in Description Logics More Gentle (Extended Abstract)'. In *Proceedings of the 31st International Workshop on Description Logics*, Tempe, Arizona, US, 2018.
- Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza. 'Repairing Description Logic Ontologies by Weakening Axioms'. *CoRR*, 2018.

Next, we discuss a study on handling ontologies in the context of privacy-preserving data publishing in Chapter 5. All sections within this chapter are set with the goal of investigating three properties, namely *compliance*, *safety*, and *optimality* in different settings. We begin with a quite restricted setting where information about individuals are contained in $\mathcal{EL}$ Instance Stores without TBoxes, which implies that information to be published about an individual is given as an $\mathcal{EL}$ concept. In addition, privacy policies are encoded as a finite set of $\mathcal{EL}$ concepts whereas the knowledge of possible attackers are written as an $\mathcal{EL}$ concept. We shift this setting to a condition where the way of encoding information about individuals and privacy policies remains the same, but knowledge of possible attackers is now written as $\mathcal{FL}_0$ and $\mathcal{FLE}$ concepts. All contents, notions, and results on privacy-preserving ontology publishing for $\mathcal{EL}$ instance stores that are written in Chapter 5 are mainly based on the following publications [BN18; BKN19; BN19].

- Franz Baader and Adrian Nuradiansyah. 'Towards Privacy-Preserving Ontology Publishing'. In *Proceedings of the 31st International Workshop in Description Logics*, Tempe, Arizona, US, 2018.
- Franz Baader, Francesco Kriegel, and Adrian Nuradiansyah. 'Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ Instance Stores'. In *Proceedings of the 16th European Conference on Logics in Artifical Intelligence* (JELIA), Rende, Italy, 2019.

- Franz Baader and Adrian Nuradiansyah. 'Mixing Description Logics in Privacy-Preserving Ontology Publishing'. In *Proceedings of Künstliche Intelligenz (KI) - the 42nd German Conference on AI*, Kassel, Germany, 2019.

In Chapter 6, we extend the problem setting formulated in Chapter 5. Here, the infomation about individuals as well as the knowledge of attackers are given by $\mathcal{EL}$ ABoxes consisting of concept and role assertions. Then, the privacy policies are given either as an $\mathcal{EL}$ concept or a conjunctive query. If one policy is violated, then we provide an anonymization approach by using a function, called an *anonymizer* that is applied to the given $\mathcal{EL}$ ABox, which either generalizes $\mathcal{EL}$ concepts or renames individuals with a new anonymous individual in ABox assertions. We present algorithms to decide whether an anonymization of the ABox fulfills compliance and safety properties and then check whether the anonymizer we perform over the ABox is also *optimal* in the sense that it keeps information from the original ABox as much as possible. The complexity of checking such properties are also analyzed within this chapter.

We end this thesis by providing conclusions and future work in Chapter 7.

Chapter 2

Preliminaries

In this chapter, we begin with the basic definition for the Description Logic (DL) $\mathcal{ALC}$ containing all Boolean operators and being the basis of many more expressive Description Logics. After discussing the syntactical representations and the semantics of $\mathcal{ALC}$, we describe Description Logic ontologies that are built over $\mathcal{ALC}$ concept descriptions and then it is continued with an introduction of reasoning problems that are relevant for this thesis and have been well investigated. Following this, we look at fragments of $\mathcal{ALC}$ that will be considered in this thesis. Most notably, we dedicate one section to discuss one fragment of $\mathcal{ALC}$, called $\mathcal{EL}$, forming the basis of Chapter 4, 5, and 6. This chapter is closed with a brief summary of the foundation foundation of computational complexity and its applications in Description Logics.

2.1 Description Logic $\mathcal{ALC}$

As mentioned in Section 1.1, Description Logics (DLs) are useful to represent the conceptual knowledge of an application domain in a well-understood way. DLs allow their user to define important notions in an application domains as *concepts* by stating sufficient and necessary conditions for individuals to belong to a concept. In general, the building block for the notion of DL concepts consists of three disjoint sets $\mathsf{N_C}, \mathsf{N_R}, \mathsf{N_I}$, which are sets of *concept names*, *role names*, and *individual names*, respectively. For more detail explanations about Description Logics, readers may refer to [BCM+03; BHL+17].

In this section, we focus on the Description Logic $\mathcal{ALC}$ since it is the most widely used in many DL reasoning services and a very basic one in the sense of containing all Boolean operators (*conjunction*, *disjunction*, and *negation*). The name of $\mathcal{ALC}$ itself stands for 'Attributive concept Language with Complement', which was first introduced in [SS91].

Definition 2.1. *Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be sets of concepts names and role names, respectively. The set of $\mathcal{ALC}$ concepts is the smallest set satisfying the following conditions:*

- $\top$ (Top), $\bot$ (Bottom)*, and every concept name $A \in \mathsf{N_C}$ are $\mathcal{ALC}$ concepts.*
- *if C, D are concepts, then $C \sqcap D$* (conjunction)*, $C \sqcup D$* (disjunction)*, and $\neg C$* (negation) *are also $\mathcal{ALC}$ concepts.*
- *if C is a concept and $r \in \mathsf{N_R}$, then $\forall r.C$* (value restriction) *and $\exists r.C$* (existential restriction) *are also $\mathcal{ALC}$ concepts.* ♢

For any DL $\mathcal{L}$, in the following we often use to write 'complex' concept of the DL $\mathcal{L}$ to distinguish this notion from concept names. If it is clear from the context, we drop the word

'complex'. The semantics of Description Logic concepts is defined in a model-theoretic way using a Tarsky-style set theoretic *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, consisting of a non-empty set $\Delta^{\mathcal{I}}$ of domain elements and the *interpretation function* $\cdot^{\mathcal{I}}$, which maps

- every $a \in \mathsf{N_I}$ to an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$,
- every $A \in \mathsf{N_C}$ to a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and
- every $r \in \mathsf{N_R}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

Definition 2.2. *Let* $\mathsf{N_C}$ *and* $\mathsf{N_R}$ *be sets of concept names and role names, respectively, and let* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *be an interpretation. This interpretation function is defined recursively for* $\mathcal{ALC}$ *concepts as follows.*

- $(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}}$,
- $(C \sqcup D)^{\mathcal{I}} := C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\neg C)^{\mathcal{I}} := \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$,
- $(\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}}.(d,e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\}$, *and*
- $(\forall r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \forall e \in \Delta^{\mathcal{I}}.(d,e) \in r^{\mathcal{I}} \rightarrow e \in C^{\mathcal{I}}\}$ ♢

The following example illustrates an $\mathcal{ALC}$ concept together with its interpretations.

Example 2.3. *We write an* $\mathcal{ALC}$ *concept* C *as follows:*

$$C = \neg\mathsf{German} \sqcap \exists \mathsf{worksAt}.(\mathsf{IT\_Dept} \sqcap \forall \mathsf{located}.(\mathsf{Germany} \sqcup \mathsf{Austria})).$$

The concept above expresses elements who are not German and work at an IT Department which is only located in either Germany or Austria. In Figure 2.4, a graphical representation for interpretations $\mathcal{I}_1$ *and* $\mathcal{I}_2$ *of* C *are illustrated. If we take a closer look at* $\mathcal{I}_1$*, all elements in* $\mathcal{I}_1$ *are in the extension of* $\top$ *and none of them are in the extension of* $\bot$*. For instance, the element* e_0 *is in the extension of* IT_Dept*, whereas the element* e_1 *is in the extension of* Austria*. Thus, these elements are labeled by* IT_Dept *and* Austria*, respectively. Now, if we extend* $\mathcal{I}_1$ *to complex concepts, we may see that* d_0 *is in the extension of the concept* C *since* d_0 *is not in the extension of* German *and connected via a role name* worksAt *to* e_0*, which is in the extension of* IT_Dept *and has a relationship 'located' to* e_1 *which is in the extension of at least one of disjuncts in* Germany ⊔ Austria*. Likewise, in the other picture, the element* d_1 *is also an element of* C *under* $\mathcal{I}_2$*, which is justified with the same argument for* d_0*, except that the* worksAt*-successor* e_2 *of* d_1*, which is an element of* IT_Dept *under* $\mathcal{I}_2$*, is also in the extension of* ∀located.(Germany⊔Austria)*. This is due to the fact that* e_2 *does not have any* located*-successors and hence* all *their* located*-fillers vacuously satisfy any imposed condition.*

One may call the union of $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ the *signature*. This notion is extended to Description Logic concepts by collecting all concept names, role names, and individual names occurring in the concepts. Besides this notion, one might be interested in counting the *size* of concepts that is obtained by counting the number of occurrences of concept names,

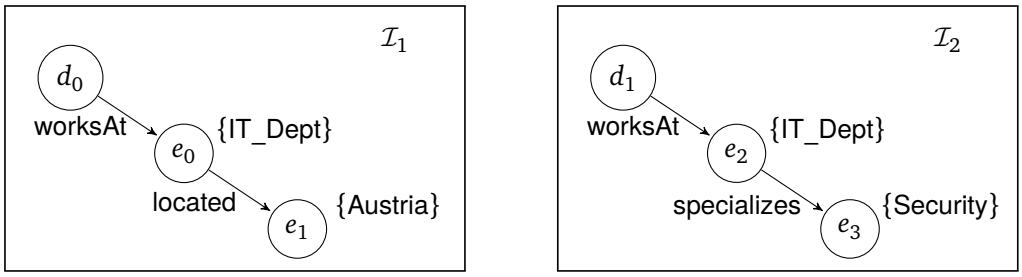


Figure 2.4: A graphical representation of interpretations of the concept C defined in example 2.3

role names, individual names, and Boolean operators in the concepts [1]. In addition to those two notions, we also define the notion of *subconcepts* that are intuitively concepts that syntactically occur in a given concept. For $\mathcal{ALC}$ concepts, the formal definitions of *signature*, *size*, and *subconcepts* are defined as follows.

Definition 2.5. *Let C be an $\mathcal{ALC}$ concept. The* signature *of C, denoted by $sig(C)$ is the set of all concept names and role names occurring in C, while the* size *and the set of subconcepts of C, denoted by $|\ C\ |$ and $sub(C)$, are defined inductively as follows:*

- *if $C = A \in \mathsf{N_C}$, then $|\ C\ | := 1$ and $sub(C) = \{A\}$;*
- *if $C = C_1 \sqcap C_2$ or $C = C_1 \sqcup C_2$, then $|\ C\ | := 1 + |\ C_1\ | + |\ C_2\ |$ and $sub(C) = \{C\} \cup sub(C_1) \cup sub(C_2)$;*
- *if $C = \neg D$ or $C = \exists r.D$ or $C = \forall r.D$, then $|\ C\ | := 1 + |\ D\ |$ and $sub(C) = \{C\} \cup sub(D)$.* ♢

For example, the concept C that is defined in Example 2.3 has the following signature $sig(C) := \{\mathsf{German}, \mathsf{Germany}, \mathsf{Austria}, \mathsf{IT\_Dept}, \mathsf{worksAt}, \mathsf{located}\}$ and the size that is counted as follows $|\ C\ | := \ 1+1+1+1+(1+1+1+(1+1+1)) = 10$.

2.1.1 Reasoning in $\mathcal{ALC}$ Ontologies

Mainly, Description Logic ontologies consist of two parts, which are terminological knowledge, called *TBoxes*, and assertional knowledge, called *ABoxes*. Intuitively, *TBoxes* provide constraints on the interpretation of concepts and define how concepts are related each other, while *ABoxes* formulate knowledge about instances of concepts and relationship between individuals. We start with the definition of $\mathcal{ALC}$ TBoxes.

Definition 2.6. *Let C, D be $\mathcal{ALC}$ concepts. We call the following axiom $C \sqsubseteq D$ a* general concept inclusion *(GCI). An $\mathcal{ALC}$* TBox *$\mathcal{T}$ is a finite set of GCIs.* ♢

If there are two GCI axioms $C \sqsubseteq D$ and $D \sqsubseteq C$ in $\mathcal{T}$, then we may abbreviate them with an *equivalence axiom* $C \equiv D$. We will sometimes use the word *axiom* to refer to a GCI

[1] There are also Description Logics that has individual names as one of their constructors that will be introduced in the next chapter

$$
\begin{aligned}
\mathcal{T}_{ex} = \quad & \{\ \mathsf{Disease} \sqsubseteq \neg \mathsf{Patient}, \\
& \mathsf{Cancer} \sqsubseteq \mathsf{Disease}, \\
& \exists \mathsf{suffer}.\mathsf{Disease} \sqsubseteq \mathsf{Person}, \\
& \mathsf{Patient} \sqsubseteq \mathsf{Person} \sqcap \exists \mathsf{seen\_by}.\mathsf{Doctor}, \\
& \mathsf{Oncologist} \sqsubseteq \mathsf{Doctor}\}
\end{aligned}
$$

Figure 2.8: The example TBox $\mathcal{T}_{ex_1}$

or an equivalence axiom. The semantics of TBoxes is also defined using the notion of interpretations.

Definition 2.7. *Let $\mathcal{I}$ be an interpretation. $\mathcal{I}$* satisfies *a GCI $C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. $\mathcal{I}$ is a* model of a TBox *$\mathcal{T}$ iff $\mathcal{I}$ satisfies all GCIs in $\mathcal{T}$.* ♢

In another sense, one may also perceive TBoxes as knowledge consisting of hierarchical relationships between concepts. Given $C \sqsubseteq D$, we say that C is more specific than D, and vice versa, D is more general than C. We may alternatively call the notion of TBox above with the name *general TBox*. This is due to the fact that there is also another type of TBox, defined as *terminologies*, whose axioms are of the form $A \equiv C$, called a *concept definition*, where $A \in \mathsf{N_C}$ and C is a complex concept. However, in this thesis we just focus on general TBoxes and sometimes we drop the word 'general' if it is already clear from the context. Next, we illustrate an interpretation $\mathcal{I}_{ex_1}$ that interprets all concept names and role names in a TBox $\mathcal{T}_{ex_1}$ presented in Figure 2.8.

$$
\begin{aligned}
\Delta^{\mathcal{I}_{ex_1}} &= \{cn_1, cn_2, ds_1, dt_1, on_1, pt_1, pt_2\}, \\
\mathsf{Patient}^{\mathcal{I}_{ex_1}} &= \{pt_1, pt_2\}, \\
\mathsf{Doctor}^{\mathcal{I}_{ex_1}} &= \{dt_1, on_1\}, \\
\mathsf{Oncologist}^{\mathcal{I}_{ex_1}} &= \{on_1\}, \\
\mathsf{Person}^{\mathcal{I}_{ex_1}} &= \{dt_1, on_1, pt_1, pt_2\}, \\
\mathsf{Cancer}^{\mathcal{I}_{ex_1}} &= \{cn_1, cn_2\}, \\
\mathsf{Disease}^{\mathcal{I}_{ex_1}} &= \{cn_1, cn_2, ds_1\}, \\
\mathsf{suffer}^{\mathcal{I}_{ex_1}} &= \{(pt_1, cn_1), (pt_1, cn_2), (pt_2, ds_1)\}, \\
\mathsf{seen\_by}^{\mathcal{I}_{ex_1}} &= \{(pt_1, dt_1), (pt_2, dt_1)\}
\end{aligned}
$$

To confirm that $\mathcal{I}_{ex_1}$ is a model of $\mathcal{T}_{ex_1}$, we check whether for each GCI $C \sqsubseteq D \in \mathcal{T}_{ex_1}$, we have $C^{\mathcal{I}_{ex_1}} \subseteq D^{\mathcal{I}_{ex_1}}$. For the GCI $\mathsf{Disease} \sqsubseteq \neg\mathsf{Patient}$, we have

$$\mathsf{Disease}^{\mathcal{I}_{ex_1}} = \{cn_1, cn_2, ds_1\} \not\subseteq \{pt_1, pt_2\} = \mathsf{Patient}^{\mathcal{I}_{ex_1}}.$$

Then, for the second GCI $\mathsf{Cancer} \sqsubseteq \mathsf{Disease}$, it holds that

$$\mathsf{Cancer}^{\mathcal{I}_{ex_1}} = \{cn_1, cn_2\} \subseteq \{cn_1, cn_2, ds_1\} = \mathsf{Disease}^{\mathcal{I}_{ex_1}}.$$

$$
\begin{aligned}
\mathcal{A}_{ex} = \{\ & \mathsf{Patient}(\mathsf{ALICE}), && \mathsf{Doctor}(\mathsf{DIANA}),\\
& \mathsf{Patient}(\mathsf{BOB}), && \mathsf{Cancer}(\mathsf{AML}),\\
& \mathsf{Patient}(\mathsf{CAROL}), && \mathsf{Cancer}(\mathsf{CLL}),\\
& \mathsf{seen\_by}(\mathsf{ALICE}, \mathsf{DIANA}), && \mathsf{suffer}(\mathsf{ALICE}, \mathsf{AML}),\\
& \mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA}), && \mathsf{suffer}(\mathsf{BOB}, \mathsf{CLL}),\\
& \mathsf{seen\_by}(\mathsf{CAROL}, \mathsf{DIANA}), && (\forall \mathsf{suffer}.\neg \mathsf{Cancer})(\mathsf{CAROL})\}
\end{aligned}
$$

Figure 2.11: The example ABox $\mathcal{A}_{ex}$

For the concept $(\exists \mathsf{suffer}.\mathsf{Disease})^{\mathcal{I}}_{ex_1}$, we know that the elements pt_1, pt_2 belong to the extension of this concept, since their suffer-successors cn_1, cn_2, and ds_1 belong to $\mathsf{Disease}^{\mathcal{I}_{ex_1}}$, and thus it is verified that

$$(\exists \mathsf{suffer}.\mathsf{Disease})^{\mathcal{I}}_{ex_1} = \{pt_1, pt_2\} \subseteq \{dt_1, on_1, pt_1, pt_2\} = \mathsf{Person}^{\mathcal{I}_{ex_1}}.$$

The fourth GCI that contains an existential restriction, too, is also true under $\mathcal{I}_{ex_1}$ such that

$$\mathsf{Patient}^{\mathcal{I}_{ex_1}} = \{pt_1, pt_2\} \subseteq \{pt_1, pt_2\} = (\mathsf{Person} \sqcap \exists \mathsf{seen\_by}.\mathsf{Doctor})^{\mathcal{I}_{ex_1}}$$

since the only elements belong to the intersection of $\mathsf{Person}^{\mathcal{I}_{ex_1}}$ and $(\exists \mathsf{seen\_by}.\mathsf{Doctor})^{\mathcal{I}_{ex_1}}$ are pt_1 and pt_2. Last but not least, it obviously holds that

$$\mathsf{Oncologist}^{\mathcal{I}_{ex_1}} = \{on_1\} \subseteq \{dt_1, on_1\} = \mathsf{Doctor}^{\mathcal{I}_{ex_1}}.$$

Next, we introduce how information about individuals and their relationships are stored in *ABoxes*

Definition 2.9. *Let C be an $\mathcal{ALC}$ concept, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. We call $C(a)$ and $r(a, b)$ a concept assertion and a role assertion, respectively. An* ABox *$\mathcal{A}$ is a finite set of concept assertions and role assertions.* ♢

Likewise, we will sometimes use the word *axiom* to refer a concept assertion or a role assertion. We also define the semantics of ABoxes with the use of interpretations.

Definition 2.10. *Let $\mathcal{I}$ be an interpretation. $\mathcal{I}$* satisfies *a concept assertion $C(a)$ and a role assertion $r(a, b)$ iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$, respectively. $\mathcal{I}$ is a* model of an ABox *$\mathcal{A}$ iff $\mathcal{I}$ satisfies all concept assertions and role assertion in $\mathcal{A}$.* ♢

As an illustration, we have presented an ABox $\mathcal{A}_{ex}$ in Figure 2.11. According to $\mathcal{A}_{ex}$, we may say that Alice, Bob, and Carol are patients, while Dan is a doctor. Then, we have *Acute Myeloid Leukemia* (AML) and *Chronic Lymphocytic Leukemia* (CLL) as names of cancer diseases from which Alice and Bob suffer, respectively. Further, it is stated that everything from which Alice suffers is not an instance of cancer. Additionally, the ABox also explicitly states that Dan examines Alice, Bob, and Carol.

Now, we construct an interpretation $\mathcal{I}'_{ex_1}$ which coincides with $\mathcal{I}_{ex_1}$ on its domain elements as well as all concept and role names occurring in $\mathcal{T}_{ex_1}$, but it interprets all individual names in $\mathcal{A}_{ex_1}$ as follows:

$$\begin{aligned} \mathsf{ALICE}^{\mathcal{I}'_{ex}} &= pt_1, & \mathsf{DIANA}^{\mathcal{I}'_{ex}} &= dt_1, \\ \mathsf{BOB}^{\mathcal{I}'_{ex}} &= pt_1, & \mathsf{AML}^{\mathcal{I}'_{ex}} &= cn_1, \\ \mathsf{CAROL}^{\mathcal{I}'_{ex}} &= pt_2, & \mathsf{CLL}^{\mathcal{I}'_{ex}} &= cn_2. \end{aligned}$$

According to this construction, it is clear to see that $\mathcal{I}'_{ex_1}$ is also a model of $\mathcal{T}_{ex_1}$. Observe that the individual names ALICE and BOB are interpreted to the same element pt_1. This is due to the fact that we do not assume *Unique Name Assumption* (UNA) in this thesis. In many literatures, UNA is enforced to their semantics such that it requires $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ in the case $a \neq b$.

Now, we verify whether $\mathcal{I}'_{ex_1}$ is also a model of $\mathcal{A}_{ex_1}$ by checking if $a^{\mathcal{I}'_{ex_1}} \in C^{\mathcal{I}'_{ex_1}}$ for all concept assertion $C(a) \in \mathcal{A}_{ex_1}$ and $(a^{\mathcal{I}'_{ex_1}}, b^{\mathcal{I}'_{ex_1}}) \in r^{\mathcal{I}'_{ex_1}}$ for all role assertions $r(a,b) \in \mathcal{A}_{ex_1}$. It clearly holds that for all concept assertions whose all the concepts are concept names, they are satisfied by $\mathcal{I}'_{ex_1}$. Then, for each role assertion in $\mathcal{A}_{ex_1}$, it is also trivial to see that $\mathcal{I}'_{ex_1}$ satisfies them. Now, we see that $\mathsf{CAROL}^{\mathcal{I}'_{ex_1}} = pt_2$ has a suffer-successor ds_1 that is not in extension of any concept name in $\mathcal{I}'_{ex_1}$. Then, since the universal restriction $\forall$suffer.$\neg$Cancer propagates the concept $\neg$Cancer to any suffer-successor of pt_2, we have $ds_1 \in (\neg\mathsf{Cancer})^{\mathcal{I}'_{ex_1}}$. Thus, $\mathcal{I}'_{ex_1}$ satisfies ($\forall$suffer.$\neg$Cancer)(CAROL) and it concludes that $\mathcal{I}'_{ex_1}$ is a model of $\mathcal{A}_{ex_1}$.

Finally, we arrived at the notion of DL ontologies that are formed by the combination of TBoxes and ABoxes.

Definition 2.12. *A* Description Logic ontology $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ *consists of a TBox* $\mathcal{T}$ *and an ABox* $\mathcal{A}$*. An interpretation* $\mathcal{I}$ *is a* model of $\mathfrak{O}$ *iff* $\mathcal{I}$ *is a model of both* $\mathcal{T}$ *and* $\mathcal{A}$*.* ♢

Now, we introduce the notions of *signature* and *size* that are extended to Description Logic ontologies, in particular to $\mathcal{ALC}$ ontologies, that are defined in the following

Definition 2.13. *Let* $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ *be an* $\mathcal{ALC}$ *ontology. The signatures of* $\mathcal{T}$ *and* $\mathcal{A}$ *are defined as follows*

$$sig(\mathcal{T}) = \bigcup_{C \sqsubseteq D \in \mathcal{T}} sig(C) \cup sig(D) \;\; and \;\; sig(\mathcal{A}) = \bigcup_{C(a) \in \mathcal{A}} sig(C) \cup \{a\} \cup \bigcup_{r(a_1,a_2) \in \mathcal{A}} \{r, a_1, a_2\}.$$

The signature *of* $\mathfrak{O}$*, written sig* $(\mathfrak{O})$*, is simply defined as* $sig(\mathfrak{O}) = sig(\mathcal{T}) \cup sig(\mathcal{A})$*. Then, the* size *of each* $\mathcal{T}$ *and* $\mathcal{A}$ *is defined as follows*

$$|\,\mathcal{T}\,| = \sum_{C \sqsubseteq D \in \mathcal{T}} |\,C\,| + |\,D\,| \quad and \quad |\,\mathcal{A}\,| = \sum_{C(a) \in \mathcal{A}} |\,C\,| + 1 + \sum_{r(a_1,a_2)} 3.$$

Finally, the size *of* $\mathfrak{O}$*, denoted by* $|\,\mathfrak{O}\,|$*, is obtained by adding* $|\,\mathcal{T}\,|$ *with* $|\,\mathcal{A}\,|$*.* ♢

Note that the number 1 used to define $|\,\mathcal{A}\,|$ above is considered in the summation to indicate one occurrence of the individual a in an assertion $C(a) \in \mathcal{A}$, while the number 3 is used to indicate three occurences of the role name r and the individual names a_1 and a_2 in an assertion $r(a_1, a_2) \in \mathcal{A}$.

Further, we discuss reasoning problems in DL ontologies that have been well-investigated. They are the *consistency problem*, *satisfiability problem*, *subsumption problem*, and the *instance problem*. We start with the first problem.

Definition 2.14. *Let $\mathfrak{O}$ be an ontology. $\mathfrak{O}$ is* consistent *if it has a model. The* consistency problem *asks whether there is a model for $\mathfrak{O}$.* ♢

Now, if we construct an $\mathcal{ALC}$ ontology $\mathfrak{O}_{ex_1} = (\mathcal{T}_{ex_1}, \mathcal{A}_{ex_1})$, then $\mathfrak{O}_{ex_1}$ is consistent since it has a model $\mathcal{I}'_{ex_1}$ that satisfies $\mathcal{T}_{ex_1}$ and $\mathcal{A}_{ex_1}$. However, if we add the following axioms to $\mathfrak{O}_{ex_1}$:

$$\alpha_1 = \exists \mathsf{seen\_by}.\mathsf{Oncologist} \sqsubseteq \exists \mathsf{suffer}.\mathsf{Cancer},$$
$$\alpha_2 = \mathsf{Oncologist}(\mathsf{DIANA})$$

such that $\mathfrak{O}'_{ex_1} = (\mathcal{T}_{ex_1} \cup \{\alpha_1\}, \mathcal{A}_{ex_1} \cup \{\alpha_2\})$, then $\mathcal{I}'_{ex_1}$ is not a model of $\mathfrak{O}'_{ex_1}$. This is due to the fact that

$$\mathsf{DIANA}^{\mathcal{I}'_{ex_1}} = \{dt_1\} \notin \{on_1\} = \mathsf{Oncologist}^{\mathcal{I}'_{ex_1}}.$$

Nevertheless, the ontology $\mathfrak{O}'_{ex_1}$ itself is also inconsistent since in every model $\mathcal{J}$ of $\mathfrak{O}'_{ex_1}$, we have $\mathsf{DIANA}^{\mathcal{J}} \in \mathsf{Oncologist}^{\mathcal{J}}$ and this implies that ALICE, BOB, and CAROL are seen by an oncology, which means that all of them suffer from Cancer. This implication however contradicts the assertion $(\forall \mathsf{suffer}.\neg\mathsf{Cancer})(\mathsf{CAROL})$ and thus the inconsistency of $\mathfrak{O}'_{ex_1}$ follows.

Next, we define the *satisfiability problem* that asks whether a given concept C has a model w.r.t. a given TBox.

Definition 2.15. *Let C be a concept and $\mathcal{T}$ be an ontology. C is* satisfiable *w.r.t. $\mathcal{T}$ iff there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}} \neq \emptyset$. The* satisfiability problem *asks whether C is satisfiable w.r.t. $\mathcal{T}$*

The subsequent problem asks whether one concept is more specific than another concept w.r.t. a given TBox.

Definition 2.16. *Let $\mathcal{T}$ be a TBox, and C, D be DL concepts. C is* subsumed by *D* w.r.t. *$\mathcal{T}$ (denoted by $C \sqsubseteq_{\mathcal{T}} D$) iff for all models $\mathcal{I}$ of $\mathcal{T}$, $C^{\mathcal{I}} \subseteq D$. If $C \sqsubseteq D$, but $D \not\sqsubseteq D$, then we say that C is* strictly subsumed *by D w.r.t. $\mathcal{T}$. The* (strict) subsumption *problems asks whether C is subsumed by D w.r.t. $\mathcal{T}$.* ♢

Last, we introduce *the instance problem* involving not only TBoxes, but also the use of ABoxes, to ask whether an individual is an *instance* of a given concept w.r.t. an ontology.

Definition 2.17. *Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ be an ontology, C be a DL concept, and $a \in \mathsf{N_I}$. The individual a is an* instance of *C* w.r.t. *$\mathfrak{O}$ (denoted by $\mathfrak{O} \models C(a)$) iff for all models $\mathcal{I}$ of $\mathfrak{O}$, $a^{\mathcal{I}} \in C^{\mathcal{I}}$. The* instance problem *asks whether a is an instance of C w.r.t. $\mathfrak{O}$.* ♢

If the last two reasoning problems above do not take TBoxes as input into account, then the subsumption problem will only ask whether $C \sqsubseteq D$ and the instance problem asks whether $\mathcal{A} \models C(a)$. Next, we introduce implicit relationships between the aforementioned reasoning problems.

Theorem 2.18 ([BHL+17]). *Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ALC}$ ontology, C, D be $\mathcal{ALC}$ concepts and a be an individual name.*

(i.) $C \equiv_{\mathcal{T}} D$ *if and only if* $C \sqsubseteq_{\mathcal{T}} D$ *and* $D \sqsubseteq_{\mathcal{T}} C$.

(ii.) $C \sqsubseteq_{\mathcal{T}} D$ *if and only if* $C \sqcap \neg D$ *is not satisfiable w.r.t.* $\mathcal{T}$.

(iii.) C *is satisfiable w.r.t.* $\mathcal{T}$ *if and only if* $C \not\sqsubseteq_{\mathcal{T}} \bot$.

(iv.) C *is satisfiable w.r.t.* $\mathcal{T}$ *if and only if* $(\mathcal{T}, \{C(a)\})$ *is consistent.*

(v.) $\mathfrak{O} \models C(a)$ *if and only if* $(\mathcal{T}, \mathcal{A} \cup \{\neg C(a)\})$ *is inconsistent.*

Note that the relationships above do not only hold for $\mathcal{ALC}$ ontologies but also for all ontologies which are formulated in DLs that have conjunction and negation of complex concepts. As a consequence of the theorem above, we can see that all reasoning problems can be *reduced* to the consistency problem, i.e., we can use an algorithm for ontology consistency to decide all reasoning problems mentioned above.

As mentioned in Chapter 1, an ontology can also be viewed as a collection of information that is used to derive answers for given queries. For instance, one basic query inherited from the definition of the instance problem is called the *instance query*. This query asks which individuals in $\mathsf{N_I}$ that are instances of a given concept C w.r.t. an ontology $\mathfrak{O}$. Using first-order logic representation, we view that this kind of query only consists of one free variables (called *answer variables* in the following) with some existentially quantified variables. The following notion is a class of query called *conjunctive query*, that may consist of more than one answer variables and quantified variables as well as is constructed over the conjunction of unary and binary predicates only.

Definition 2.19. *A* conjunctive query (CQ) q *is a first-order formula with the following expression:*

$$q(\vec{v}) \leftarrow \exists \vec{w}.conj(\vec{v}, \vec{w}),$$

where $\vec{v}$ *are answer variables,* $\vec{w}$ *are existentially quantified variables, and* the body $conj(\vec{v}, \vec{w})$ *is a conjunction of query atoms, each of the form* $A(z)$ *or* $r(z, z')$ *where* z, z' *are either individuals names or variables over* $\vec{v} \cup \vec{w}$ *and* A *and* r *are unary (concept names) and binary (role names) predicates, respectively. A CQ is* Boolean *if it has no answer variables. The* size *of* q*, written* $| q |$*, is the number of all occurrences of conjunction symbols (*$\wedge$*), concept names, role names, and individual names in* q. ♢

To emphasize that q has answer variables $\vec{v}$, we particularly write $q(\vec{v})$. Next, we write how answers to CQs are defined in two steps: first on the level of interpretations and then on the level of knowledge bases.

Definition 2.20. *Let* $q(\vec{v})$ *be a conjunctive query having* n *answer variables,* $a_1, \ldots, a_n \in N_I$ *and* $\mathcal{I}$ *an interpretation. We call a tuple* $\vec{t} = a_1, \ldots, a_n$ *of individuals an* answer *to* q *on* $\mathcal{I}$ *if* $\mathcal{I} \models q(\vec{t})$*, i.e.,* $q(\vec{v})$ *evaluates to true in* $\mathcal{I}$ *under the valuation that maps* $\vec{v}$ *to the tuple* $\vec{t}$*. We use* $ans(q, \mathcal{I})$ *to denote the* set of all answers to q on $\mathcal{I}$.

Now, given an ontology $\mathfrak{O}$ *and a CQ* q*, a tuple* $\vec{t} \in (\mathsf{N_I})^n$ *of individuals is an* answer to q w.r.t. $\mathfrak{O}$*, written as* $\mathfrak{O} \models q(\vec{t})$ *iff all individual names from* $\vec{t}$ *occur in* $\mathcal{A}$ *and* $\vec{t} \in ans(q, \mathcal{I})$ *for all* $\mathcal{I}$ *of* $\mathfrak{O}$*. We write* $ans(q, \mathfrak{O}) = \bigcap_{\mathcal{I} \text{ model of } \mathfrak{O}} ans(q, \mathcal{I})$ *to denote the* set of all answers to q w.r.t. $\mathfrak{O}$. ♢

Let us consider the following ontology $\mathfrak{O}_{ex_2} = (\mathcal{T}_{ex_2}, \mathcal{A}_{ex_2})$, where

$$\begin{aligned}\mathcal{T}_{ex_2} &:= \{\,\mathsf{Patient} \sqsubseteq \exists \mathsf{seen\_by}.\mathsf{Doctor}\}\\ \mathcal{A}_{ex_2} &:= \{\,\mathsf{Doctor}(\mathsf{DIANA}), \mathsf{Patient}(\mathsf{ALICE}), \mathsf{Patient}(\mathsf{BOB}), \mathsf{Patient}(\mathsf{CAROL})\\ &\qquad \mathsf{seen\_by}(\mathsf{ALICE}, \mathsf{DIANA}), \mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA})\}.\end{aligned}$$

Then, we consider the following samples expressing conjunctive queries:

a.) Return all pairs of individual names (a, b) such that a is a doctor who sees a patient b:
$$q_1(v_1, v_2) = \mathsf{Doctor}(v_1) \wedge \mathsf{seen\_by}(v_2, v_1) \wedge \mathsf{Patient}(v_2).$$

b.) Return all individual names a such that a is a patient who is seen by a doctor:
$$q_2(v) = \exists w.(\mathsf{Doctor}(w) \wedge \mathsf{seen\_by}(v, w) \wedge \mathsf{Patient}(v).$$

The answers for the first query q_1 are $\{(\mathsf{ALICE}, \mathsf{DIANA}), (\mathsf{BOB}, \mathsf{DIANA})\}$. Note that the individual CAROL is not included in any answer since it may be that there is a model of $\mathfrak{O}_{ex_2}$, where CAROL is seen by another doctor who is not DIANA. For the other query q_2, we have the answers for it, which are $\{\mathsf{ALICE}, \mathsf{BOB}, \mathsf{CAROL}\}$. Now, the individual CAROL is included since CAROL is a patient and the GCI in $\mathcal{T}_{ex_2}$ enforces CAROL to be seen by a doctor.

2.1.2 Relationship with First-Order Logic

It has been mentioned briefly in Chapter 1 that Description Logics can be seen as decidable fragments of first-order logic. Concept names can be interpreted as unary predicates, role names can be represented as binary predicates, while individual names can be viewed as constants. Let us consider the following example

$$\begin{aligned}\exists \mathsf{leads}.\top &\sqsubseteq \mathsf{Boss},\\ \mathsf{Employee} &\equiv \mathsf{Person} \sqcap \exists \mathsf{worksAt}.\top,\\ \mathsf{Employee}(\mathsf{JIM}).&\end{aligned}$$

The axioms above can be translated to the following first-order logic formulas

$$\begin{aligned}&\forall x.(\exists y.\mathsf{leads}(x, y) \Rightarrow \mathsf{Boss}(x)),\\ &\forall x.(\mathsf{Employee}(x) \Leftrightarrow \mathsf{Person}(x) \wedge \exists y.(\mathsf{worksAt}(x, y))), \text{ and}\\ &\mathsf{Employee}(\mathsf{JIM}).\end{aligned}$$

To translate formally Description Logic concepts and axioms, we need at least two translation functions π_v and π_w, where v and w are free variables. For this translation, we again focus on the DL $\mathcal{ALC}$ that are inductively defined as follows.

$$\begin{aligned}\pi_v(A) &= A(v)\\ \pi_v(C \sqcap D) &= \pi_v(C) \wedge \pi_v(D)\\ \pi_v(C \sqcup D) &= \pi_v(C) \vee \pi_v(D)\\ \pi_v(\exists r.C) &= \exists w.(r(v, w)) \wedge \pi_w(C),\\ \pi_v(\forall r.C) &= \forall w.r(v, w) \Rightarrow \pi_w(C),\\ \pi_v(\neg C) &= \neg \pi_v(C)\end{aligned}$$

The definition for $\pi_w(C)$ is analogously defined. Then, we extend this translation to $\mathcal{ALC}$ TBoxes $\mathcal{T}$ and $\mathcal{ALC}$ ABoxes $\mathcal{A}$. For this translation, we define $\psi[v \mapsto a]$ as the first-order formula obtained from ψ by replacing all variables v with the constant a.

$$\begin{aligned} \pi(\mathcal{T}) &= \forall v. \bigwedge_{C \sqsubseteq D \in \mathcal{T}} (\pi_v(C) \Rightarrow \pi_v(D)), \\ \pi(\mathcal{A}) &= \bigwedge_{C(a) \in \mathcal{A}} \pi_v(C)[v \mapsto a] \wedge \bigwedge_{r(a,b) \in \mathcal{A}} r(a,b). \end{aligned}$$

The following theorem is cited from [BHL+17], which describes the equivalence relationship between $\mathcal{ALC}$ ontologies and its corresponding first-order logic representation.

Theorem 2.21 ([BHL+17]). *Let $(\mathcal{T}, \mathcal{A})$ be an $\mathcal{ALC}$ ontology. It holds that $(\mathcal{T}, \mathcal{A})$ is satisfiable if and only if $\pi(\mathcal{T}) \wedge \pi(\mathcal{A})$ is satisfiable.*

2.1.3 Fragments of $\mathcal{ALC}$

Now we define three fragments of $\mathcal{ALC}$ that are relevant in this thesis, namely the Description Logics $\mathcal{FLE}$, $\mathcal{FL}_0$, and $\mathcal{EL}$. The similarity between these logics is the absence of the bottom concept ($\bot$), negation ($\neg C$) and disjunction ($C \sqcup D$), which consequently implies that their concept descriptions are always satisfiable w.r.t. their ontologies, and in addition, the logics always have a *universal model* of both concepts and ontologies. We will not describe the latter property for each DL in this thesis, but this kind of model indirectly inspires us to have a nice recursive and structural characterization for reasoning problems, such as subsumption and instance problems, in these logics that will be described in more detail in the upcoming sections. Note, since they are fragments of $\mathcal{ALC}$, all the notions, such as signature, size, subconcept, models, and ontologies of $\mathcal{FL}_0$, $\mathcal{FLE}$, and $\mathcal{EL}$ follow from the definitions of all those notions for $\mathcal{ALC}$.

Let $\mathsf{N_C}, \mathsf{N_R}$ be sets of concept names and role names. We introduce formally $\mathcal{FLE}$ concepts C, D that are built through the following grammar

$$C, D := \top \mid A \mid C \sqcap D \mid \exists r.C \mid \forall r.D,$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$, i.e., the DL $\mathcal{FLE}$ has the concept constructors $\top$ (top concept), $\sqcap$ (conjunction), $\exists r.C$ (existential restriction), and $\forall r.C$ (value restriction). We call an $\mathcal{FLE}$ concept an atom if it is a concept name, a value restriction, or an existential restriction. We define the *role depth* of an $\mathcal{FLE}$ concept C as the maximum nesting of the value restriction and existential restriction in C.

In fact, reasoning in this DL, such as deciding subsumption, has been shown to be NP-hard, such that $\mathcal{FLE}$ is claimed as the first description language not including disjunctive constructor that were proved intractable [DLN+92]. Even, the application of normalization rules for $\mathcal{FLE}$ concepts may result in normalized concepts which are necessarily exponential in the size of the original concepts [DLN+92; BKM99]. This normalization is necessary even when only considering subsumption without $\mathcal{FLE}$ GCIs [BKM99]. To gain tractability, a small fragment of $\mathcal{FLE}$, called $\mathcal{FL}_0$, is considered. This logic restricts the expressiveness of $\mathcal{FLE}$ by removing existential restrictions or formally $\mathcal{FL}_0$ concepts C, D are constructed recursively as follows

$$C, D := \top \mid A \mid C \sqcap D \mid \forall r.C.$$

We call an $\mathcal{FL}_0$ concept an *atom* if it is a concept name or a value restriction. The *role depth* of an $\mathcal{FL}_0$ concept C, written $\mathtt{rd}(C)$, is the maximum nesting of the value restrictions in C. The tractability result for this DL was first mentioned in [LB87]. Reasoning problems, such as subsumption, in $\mathcal{FL}_0$ is in polynomial time if we consider this problem without TBoxes. However, it came as a surprise since the presence of GCIs increases the complexity to ExpTime-complete [BBL05], respectively. In the next section, we introduce the Description Logic $\mathcal{EL}$ that has more benefits in terms of its tractability in classical reasoning problems than the two latter DLs.

2.2 Description Logic $\mathcal{EL}$

The intractability issue in DLs presented above arises to an idea of looking at another small fragment of $\mathcal{ALC}$ that also has limited expressiveness but becomes tractable even for reasoning problems that require GCIs. Now, we introduce the DL $\mathcal{EL}$, for which reasoning is tractable [Bra04; BBL05]. Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be sets of concept and role names, respectively. Then $\mathcal{EL}$ *concepts* over these names are constructed through the grammar rule

$$C, D ::= \top \mid A \mid C \sqcap D \mid \exists r.C,$$

where $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$, i.e., the DL $\mathcal{EL}$ has the concept constructors $\top$ (top concept), $\sqcap$ (conjunction), and $\exists r.C$ (existential restriction).

We name an $\mathcal{EL}$ concept an *atom* if it is a concept name or an existential restriction. Given an $\mathcal{EL}$ concept C, we denote the set of atoms occurring in its top-level conjunction with $\mathtt{con}(C)$. For example, if $C = A \sqcap \exists r.(B \sqcap \exists s.A)$, then $\mathtt{con}(C) = \{A, \exists r.(B \sqcap \exists s.A)\}$. We *reduce* an $\mathcal{EL}$ concept C by exhaustively replacing subconcepts of the form $E \sqcap F$ with $E \sqsubseteq F$ by E (modulo associativity and commutativity of $\sqcap$). As cited from [Küs01], two concepts C and D are equivalent iff their reduced forms are equal up to associativity and commutativity of $\sqcap$.

Important reasoning tasks in $\mathcal{EL}$ are deciding subsumption between $\mathcal{EL}$ concepts and checking whether an individual is an instance of an $\mathcal{EL}$ concept w.r.t. an ABox. For the former task, we can apply the following recursive characterization of the subsumption relation that has been proved in [BM10].

Lemma 2.22. *Let C, D be two $\mathcal{EL}$ concepts. It holds that $C \sqsubseteq D$ if and only if*

- *for all $A \in \mathsf{N_C} \cap \mathtt{con}(D)$, there is $A \in \mathtt{con}(C)$ and*
- *for all $\exists r.D' \in \mathtt{con}(D)$, there is $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqsubseteq D'$.*

Then, to decide whether an individual is an instance of an $\mathcal{EL}$ concept w.r.t an ABox, we use the following recursive characterization that is a direct consequence of Lemma 27 in [LW10].

Lemma 2.23. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox, D an $\mathcal{EL}$ concept, and $a \in \mathsf{N_I}$. It holds that $\mathcal{A} \models D(a)$ iff*

1. *for all $A \in \mathtt{con}(D)$, there is $C(a) \in \mathcal{A}$ such that $A \in \mathtt{con}(C)$ and*
2. *for all $\exists r.D' \in \mathtt{con}(D)$,*
   a.) *there is $C(a) \in \mathcal{A}$ and $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqsubseteq D'$ or*
   b.) *there is $r(a, b) \in \mathcal{A}$ such that $\mathcal{A} \models D'(b)$.*

2.3 The Complexity of Reasoning Problems in DLs

In this section, we will study briefly the subject of computational complexity that is used to measure how complex the decision problem is, in particular how hard it is to compute a solution for the problem. In fact, every problem belongs to their own *complexity class* which is determined by a (non)-deterministic turing machine and the resource bound for that machine model. Typically, one uses time and space as the resources. Another aim to study the complexity theory is to know how each complexity class is interrelated and to determine to which complexity class this given problem belongs. In previous sections, we have mentioned names of complexity classes to which reasoning problems belong, such as PTIME, NP, or EXPTIME. In the following, we list complexity classes, that are relevant for this thesis, in an increasing order according to set inclusion.

- PTIME: problems that can be solved in polynomial time.
- NP: problems that can be solved in non-deterministic polynomial time.
- PSPACE: problems that can be solved in polynomial space.
- EXPTIME: problems that can be solved in exponential time.
- NEXPTIME: problems that can be solved in non-deterministic exponential time.

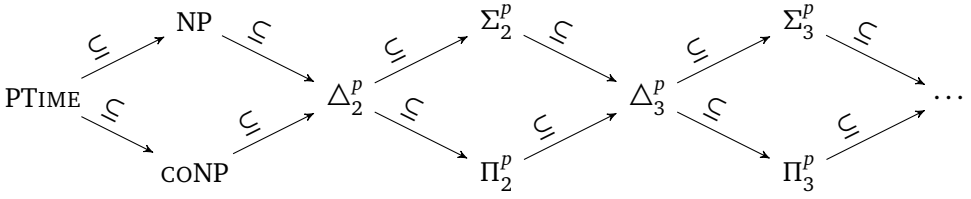
We also define CONP and CONEXPTIME as the classes of problems whose complements are in NP and NEXPTIME, respectively.

Given a complexity class $\mathfrak{C}$, a problem L is $\mathfrak{C}$-hard if there is a polynomial time-reduction from L' to L for all $L' \in \mathfrak{C}$. The most common way to prove hardness for a complexity class $\mathfrak{C}$ is to find an appropriate problem L that is already known to be $\mathfrak{C}$-hard and then show a polynomial time reduction from L to the problem at hand. We say that a problem is $\mathfrak{C}$-complete if it is in $\mathfrak{C}$ and also $\mathfrak{C}$-hard. When we prove that a problem L is *hard* for a complexity class $\mathfrak{C}$, we often call this a *lower bound* because it says that L is *at least as hard as* the other problems in $\mathfrak{C}$. Also, proving that L is contained in $\mathfrak{C}$ will be called an *upper bound* because it explains that solving L is *at least as easy as* $\mathfrak{C}$-hard problems.

For complexity classes $\mathfrak{C}, \mathfrak{C}'$, we denote by $\mathfrak{C}^{\mathfrak{C}'}$ the class of decision problems that can be solved by a Turing machine running in $\mathfrak{C}$ and using an oracle for decision problems in $\mathfrak{C}'$. The *polynomial hierarchy* is then defined to order relationships between complexity classes that lie in between PTIME and PSPACE. For $k > 0$, it is inductively defined as follows:

$$\Sigma_0^p := \Pi_0^p := \text{PTIME} \qquad \Sigma_k^p := \text{NP}^{\Sigma_{k-1}^p} \qquad \Pi_k^p := \text{CONP}^{\Sigma_k^p} \qquad \triangle_k^p := \text{PTIME}^{\Sigma_{k-1}^p}.$$

It is also known that $\triangle_1^p = \text{PTIME}$, $\Sigma_1^p = \text{NP}$, and $\Pi_1^p = \text{CONP}$. Thus, it is also possible to derive the following inclusions:



We also define the complexity class DP as the class of problems that are the intersection of NP problems and coNP problems and this class is also contained in both Σ_2^p and Π_2^p. Beyond ExpTime, there are also complexity classes consisting of all problems that can be solved in n-iterated exponentials for $n > 1$, i.e., $2^{\underbrace{2^{2^{\cdots 2^{k^d}}}}_{\text{n times}}}$.

For instance, 2ExpTime contains all problems that are solved in double exponential time. We introduce the Elementary complexity class as the union of the following classes

$$\text{Elementary} = \text{ExpTime} \cup 2\text{ExpTime} \cup 3\text{ExpTime} \cup \ldots.$$

The complement of Elementary, which is NonElementary, obviously consists of all problems that are not a member of the class Elementary. For more detailed explanations about every complexity class, readers may refer to [Pap07].

In the following, we will concentrate on basic reasoning problems in the DLs $\mathcal{EL}$ and $\mathcal{ALC}$ and then we see which complexity classes the considered reasoning problems belong to. In this thesis, we just focus on one type of complexity measures, called *combined complexity*, that is measured in the size of the whole input. Another complexity measure, mentioned in Chapter 1, that is called *data complexity*, is a different way to measure the complexity of problems based on the size of the data only, e.g., size of the ABox in the context of DL ontologies. We will not consider the data complexity in this thesis, but further details about this kind of complexity measure in DLs can be found in [BO15]. We begin this discussion with the Description Logic $\mathcal{EL}$ for which algorithms for solving subsumption and instance checking without TBoxes have been defined in Section 2.2.

It is obvious to see that the characterization shown in Lemma 2.22 can be done in polynomial time. This is because in the base case there are only quadratically many steps to check whether for all $A \in \mathsf{N_C} \cap \mathtt{con}(D)$, there is $A \in \mathtt{con}(C)$. Then for each $\exists r.D' \in \mathtt{con}(D)$, we look for $\exists r.C' \in \mathtt{con}(C)$, where $C' \sqsubseteq D'$ for which, by induction on the role depth of C and D, checking subsumption can be done in polynomial time.

To solve the instance problem in $\mathcal{EL}$, as shown in Lemma 2.23, we do quadratically many steps to check whether for all $A \in \mathtt{con}(D)$, there is $C(a)$ such that $A \in \mathtt{con}(C)$. Additionally, for each existential restriction $\exists r.D' \in \mathtt{con}(D)$, we check whether there is $C(a) \in \mathcal{A}$ and $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqsubseteq D'$, where this subsumption checking can be done in polynomial time, or alternatively, we may check whether there is $r(a,b) \in \mathcal{A}$ such that $\mathcal{A} \models D'(b)$ that can also be done in polynomial time by induction on the role depth of D. Thus, it implies that the instance problem in $\mathcal{EL}$ in in PTime. Interestingly, if we lift up this problem by adding GCIs, then the complexity remains the same [BBL05].

Lemma 2.24. *The subsumption and the instance problems in $\mathcal{EL}$ are* PTime*-complete with and without TBoxes.*

Another essential reasoning task in $\mathcal{EL}$ is answering conjunctive queries. We define the *conjunctive query entailment* problem that computes all answers to a CQ q w.r.t. a given ontology $\mathfrak{O}$. In [Ros07], it has been shown that this reasoning taks is NP-complete with and without general TBoxes. The problem belongs to NP since the query rewriting algorithm [Ros07] that is used as a pre-procedure before computing query entailment relies on a method to unify terms in q, which runs in NP by guessing one variable substitution which is

applied to q. The NP-hardness follows from NP-hardness of simple database evaluation of a CQ [AHV95].

Lemma 2.25. *The CQ entailment problem with and without general TBoxes in $\mathcal{EL}$ is* NP*-complete.*

We have seen that for other small fragments of $\mathcal{ALC}$, such as $\mathcal{FL}_0$ and $\mathcal{FLE}$, most of the basic reasoning problems are no longer tractable in these logic. These facts consequently transfer some impacts to the complexity of reasoning problems in $\mathcal{ALC}$. For $\mathcal{ALC}$, we just focus on the consistency problem, since all reasoning problems mentioned above can be reduced to it. In [BHL+17], it is proved that the consistency problem in $\mathcal{ALC}$ w.r.t. no GCIs is PSPACE-complete. The upper bound is obtained by reducing this problem to the satisfiability problem and then construct a tree model of concept using a tableau algorithm that if we use a specific strategy to explore the model, then this tableau algorithm only needs polynomial space. Moreover, the lower bound is obtained from the reduction of a well-known PSPACE-hard problem, which is the *winning strategy problem* in an *finite boolean game* [SC79]. Extending this problem by adding GCIs implies that the complexity of this problem becomes EXPTIME-complete as mentioned in [BHL+17]. The EXPTIME-hardness for this problem is obtained from the reduction of the *winning strategy problem* in an *infinite boolean game* [SC79].

Lemma 2.26. *Let $L \in$ {consistency, satisfiability, subsumption, instance} be a reasoning problem in $\mathcal{ALC}$. The complexity of L problem in $\mathcal{ALC}$ without TBoxes is* PSPACE*-complete, while it becomes* EXPTIME*-complete with respect to general TBoxes.*

Chapter 3

The Identity Problem and Its Variants in Description Logic Ontologies

After looking at basic introductions on Description Logics, ontologies, and complexity theory, we shift our attention to the first task stated in the beginning of Chapter 1 that we want to deal with mechanisms on detecting if there is a privacy breach occurring in a given ontology. As mentioned in Section 1.2, most of the works, e.g., [GH08; SS09; CDL+12] concentrated on approaches trying to hide the *properties* of individuals, i.e., the membership of an individual (or a tuple of individuals) in the answers to certain queries. In this chapter, we address a specific class of secrets, so-called *identity*, and thus we will provide the corresponding reasoning tasks related to identity-preserving problems.

In order to illustrate the privacy scenario motivating this problem, assume that you are asked to perform a survey regarding the satisfaction of employees with the management of a company. Since the boss of the company is known not to respond well to criticism, the employees insist that you perform the survey such that the identity of persons voicing criticism cannot be deduced by the boss. Thus, you let the employees use a pseudonym when answering the survey. However, the survey does ask some personal data from the participants, and you are concerned that the boss can use the provided answers, in combination with the employee database and general knowledge about how things work in the company, to deduce that a certain pseudonym corresponds to a specific employee. For example, assume that in the survey the anonymous individual x states that she is female and has expertise in logic and privacy. The boss knows that all employees with expertise logic belong to the formal verification task force and all employees with expertise privacy belong to the security task force. In addition, the employee database contains the information that the members of the first task force are John, Linda, Paul, Pattie and of the second Jim, John, Linda, Pamela. Since Linda is the only female employee belonging to both task forces, the boss can deduce that Linda hides behind the pseudonym x. The question is now whether you can use an automated system to check whether such a breach of privacy can occur in your survey.

We want to show that ontology reasoners can in principle be used for this purpose. We assume that both the information provided in the survey and the employee database are represented in a DL ABox $\mathcal{A}$, where the employees from the database are represented as known individuals in $\mathcal{A}$ and the pseudonyms used in the survey are represented as anonymous individuals in $\mathcal{A}$. Background information (such as disjointness of the concepts Male and Female, or the connection between expertise and task forces) are represented in a DL TBox $\mathcal{T}$. In order to detect a breach of privacy, we need to check whether the ontology $\mathfrak{O}$ consisting of $\mathcal{T}$ and $\mathcal{A}$ implies an identity between some anonymous individual x and a known individual a. We call the underlying reasoning task the *identity problem* for $\mathfrak{O}$, x, and a.

In Section 3.1 we formally introduce the identity problem and show that, for a large class of DLs, this problem is trivial in the sense that no identities between distinct individuals can be deduced from consistent ontologies formulated in these DLs. Not surprisingly, this class consists of the DLs that are fragments of first-order logic without equality. In Subsection 3.1.1, we introduce DLs for which the identity problem is non-trivial, i.e., the DL $\mathcal{ALCO}$ [Sch94] and $\mathcal{ELO}$ ([BBL05], [KKS12]), where nominals allow us to derive identities; *DL-Lite*$_A$ ([ACK+09], [CDL+07]), where functional roles allows us to derive identities; $\mathcal{ALCQ}$ [HB91], where number restrictions allow us to derive identities; and $\mathcal{CFD}_{nc}$ [TW13], where functional dependencies allow us to derive identities. In Subsection 3.1.2 we show that the identity problem can be reduced in polynomial time to the instance problem, and that for the DLs mentioned above this actually yields an optimal procedure w.r.t. worst-case complexity.

Section 3.2 considers the identity problem in the context of rôle-based access control [SCF+96] to ontologies. Basically, we assume that a user rôle $\hat{r}$ is associated with access to a subset $\mathfrak{O}_{\hat{r}}$ of the ontology.[1] While having rôle $\hat{r}$, the user can access $\mathfrak{O}_{\hat{r}}$ through queries, and can then store the result in a view $V_{\hat{r}}$. In a setting where rôles can dynamically change, the user may have collected (and stored) a sequence of views for different rôles. The question is then whether it is possible to derive the identity of an anonymous individual with a known one using these views. We will show that answering this question can eventually be reduced to the identity problem. In Section 3.3, we move to a scenario in which one may define that the identity of an anonymous individual is protected if it does not belong to any subset of known individuals with cardinality smaller than k. We call the reasoning problem underlying this scenario the k-hiding problem. In Subsection 3.3.1 and 3.3.2, we will show the upper bounds and the lower bounds of the k-hiding problem, which consequently show that this problem is not harder than the identity problem in most of DLs that can derive equalities between individuals.

3.1 The Identity Problem

Now we define the *identity problem* that asks whether two individuals are equal w.r.t. a given ontology. Since anything (also identities) follows from an inconsistent ontology, we consider this problem only for the case where the ontology is consistent.

Definition 3.1. *Let $a, b \in N_I$ be distinct individual names and $\mathfrak{O}$ a consistent ontology. Then a is* equal *to b w.r.t. $\mathfrak{O}$ (denoted by $\mathfrak{O} \models a \doteq b$) iff $a^{\mathcal{I}} = b^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathfrak{O}$. The* identity problem *for $\mathcal{O}, a, b$ asks whether $\mathfrak{O} \models a \doteq b$.* ♢

Not all DLs are able to derive equality between individuals. We call those that can *DLs with equality power*.

3.1.1 Description Logics with Equality Power

Definition 3.2. *A Description Logic $\mathcal{L}$ is a* Description Logic without equality power *if there is no consistent ontology $\mathcal{O}$ formulated in $\mathcal{L}$ and two distinct individual names $a, b \in \mathsf{N_I}$ such that $\mathcal{O} \models a \doteq b$. Otherwise we say that $\mathcal{L}$* has equality power. ♢

[1] To distinguish user rôles from DL roles, we write them with "ô" and also denote specific such rôles with letters with a hat.

As has been described in Subsection 2.1.2, and more detail in [BCM+03], that many DLs can be translated into first-order predicate logic (FOL). For the translation of some DLs, FOL without equality is sufficient whereas for others equality is needed.

Theorem 3.3. *If the DL $\mathcal{L}$ can be translated into* FOL without equality*, then it is a DL* without equality power.

Proof. Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ be a consistent ontology of $\mathcal{L}$ and $a, b \in \mathsf{N_I}$ be two distinct individual names. We must show that $\mathfrak{O} \not\models a \doteq b$. According to our assumption on $\mathcal{L}$, there is an FOL formula ϕ not containing the equality symbol that is equivalent to $\mathfrak{O}$. Consequently, it is sufficient to show that $\phi \not\models a = b$ according to the semantics of FOL, where the equality symbol '=' is interpreted as equality. Since $\mathfrak{O}$ is consistent, the formula ϕ is satisfiable.

Using well-known approaches and results regarding FOL [Gal15], we can transform ϕ into a formula ϕ' in Skolem form containing additional function symbols such that (i) ϕ is satisfiable iff ϕ' is satisfiable, and (ii) any model of ϕ' is a model of ϕ. Thus, ϕ' is satisfiable and since it is in Skolem form it has a Herbrand model $\mathcal{I}_H$. Since ϕ' does not contain equality, distinct terms (and thus in particular distinct constants) are interpreted by distinct elements in $\mathcal{I}_H$. Finally, we know that $\mathcal{I}_H$ is also a model of ϕ, which shows that there is a model of ϕ in which a and b are not interpreted by the same domain element. This proves $\phi \not\models a = b$.□

As a consequence of this theorem, we conclude that the basic DL $\mathcal{ALC}$ and its fragments, but also more expressive DLs such as $\mathcal{SRI}$ (see the Appendix in [BHL+17]), do not have equality power, and thus the identity problem is trivial for these DLs.

Now, we introduce four DLs that are able to derive equalities between individuals, and for which the identity problem is non-trivial. They are $\mathcal{ALCO}$, $\mathcal{ALCQ}$, *DL-Lite*$_A$, and $\mathcal{CFD}_{nc}$.

The first two DLs extend $\mathcal{ALC}$ by *nominals* and by qualified number restrictions. Nominals can be used to generate singleton concepts from individual names: if $a \in \mathsf{N_I}$, then $\{a\}$ is a concept description of $\mathcal{ALCO}$, whose semantics is defined as $\{a\}^{\mathcal{I}} := \{a^{\mathcal{I}}\}$. Qualified number restrictions are of the form $\geq n\, r.C$ and $\leq n\, r.C$, with associated semantics

- $(\geq n\, r.C)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{there are at least } n \text{ elements } e \in \Delta^{\mathcal{I}} \text{ with } (d,e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\}$;
- $(\leq n\, r.C)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{there are at most } n \text{ elements } e \in \Delta^{\mathcal{I}} \text{ with } (d,e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\}$.

The third DL is *DL-Lite*$_A$ [ACK+09; CDL+07] which derives equalities from the presence of *functionality axioms*. First of all, we introduce the syntax of basic concept descriptions in *DL-Lite*$_A$ as follows:

- every concept name $A \in \mathsf{N_C}$,
- $\top$ (the top concept),
- $\exists r$ (unqualified existential restriction), and
- $\exists r^-$ (unqualified existential restriction on inverse role).

Note that in *DL-Lite*$_A$, $\exists r$ and $\exists r^-$ are just abbreviations for $\exists r.\top$ and $\exists r^-.\top$, respectively. Semantically, $\exists r^-$ is defined as: $(\exists r^-)^{\mathcal{I}} := \{e \in \Delta^{\mathcal{I}} \mid \exists d \in \Delta^{\mathcal{I}}.(d,e) \in r^{\mathcal{I}}\}$. Then, a *DL-Lite*$_A$ TBox is a finite set of

- positive concept inclusions $B_1 \sqsubseteq B_2$,
- negative concept inclusions $B_1 \sqsubseteq \neg B_2$,
- positive role inclusions $r_1 \sqsubseteq r_2$,
- negative role inclusions $r_1 \sqsubseteq \neg r_2$, and
- global functionality axioms $\texttt{funct}\ r$,

where B_1 and B_2 are *DL-Lite*$_A$ basic concepts and r_1, r_2 range over role names and their inverses. The aforementioned concepts or axioms whose semantics has not been described in Chapter 2 are defined as follows. First, the negation of role has the following semantics $(\neg r)$ $\Delta^{\mathcal{I}} \setminus r^{\mathcal{I}}$. Then, a role inclusion is semantically defined as $r_1^{\mathcal{I}} \subseteq r_2^{\mathcal{I}}$. Last, an interpretation $\mathcal{I}$ satisfies $(\texttt{funct}\ r)$ iff for all $(b, b_1) \in r^{\mathcal{I}}$ and $(b, b_2) \in r^{\mathcal{I}}$, we have $b_1^{\mathcal{I}} = b_2^{\mathcal{I}}$. Note that functionality axioms are incorporated with conditions that the functional roles r cannot be specialized, i.e., it does not occur on the right hand side of role inclusions.

The fourth DL, called $\mathcal{CFD}_{nc}$ [TW13], derives its equality power from so-called functional dependencies. Instead of roles, this logic uses attributes, which are interpreted as total functions. We use the symbol N_A to denote the set of all attributes, replacing the set N_R. Concept descriptions C, D of $\mathcal{CFD}_{nc}$ are defined using the following syntax rules:

$$C, D ::= A \mid \neg A \mid C \sqcap D \mid \forall \texttt{Pf}.C \mid A : \texttt{Pf}_1, \ldots, \texttt{Pf}_k \rightarrow \texttt{Pf},$$

where $A \in N_C$, $k \geq 1$, and the *path functions* $\texttt{Pf}, \texttt{Pf}_i$ are words in N_A^* with the convention that the empty word is denoted by *id*. A concept description of the form $A : \texttt{Pf}_1, \ldots, \texttt{Pf}_k \rightarrow \texttt{Pf}$ is called a *path functional dependency* (PFD). In $\mathcal{CFD}_{nc}$ there is an additional restriction on PFDs to ensure that reasoning in this logic is polynomial: for any PFD of the form above there is an $i, 1 \leq i \leq k$ such that

1. $\texttt{Pf}$ is a prefix of $\texttt{Pf}_i$, or
2. $\texttt{Pf} = \texttt{Pf}' f$ for $f \in N_A$ and $\texttt{Pf}'$ is a prefix of $\texttt{Pf}_i$.

Note that PFDs whose right-hand side $\texttt{Pf}$ has length ≤ 1 trivially satisfy this restriction.

The interpretation of attributes as total functions is extended to path functions by using composition of functions and interpreting *id* as the identity function. The semantics of atomic negation $(\neg A)$ and conjunction $(C \sqcap D)$ is defined in the same way as in $\mathcal{ALC}$. For the constructors involving path functions, it is defined as follows:

$$(\forall \texttt{Pf}.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \texttt{Pf}^{\mathcal{I}}(d) \in C^{\mathcal{I}}\}, \text{ and}$$

$$(A : \texttt{Pf}_1, \ldots, \texttt{Pf}_k \rightarrow \texttt{Pf})^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \forall e \in A^{\mathcal{I}}. \left(\bigwedge_{1 \leq i \leq k} \texttt{Pf}_i^{\mathcal{I}}(d) = \texttt{Pf}_i^{\mathcal{I}}(e) \right) \Rightarrow \texttt{Pf}^{\mathcal{I}}(d) = \texttt{Pf}^{\mathcal{I}}(e)\}.$$

A TBox $\mathcal{T}$ in $\mathcal{CFD}_{nc}$ consists of a finite set of *inclusion dependencies* $A \sqsubseteq C$, where C is a complex $\mathcal{CFD}_{nc}$ concept, and an ABox $\mathcal{A}$ consists of a finite set of concept assertions $A(a)$ and path function assertions $\texttt{Pf}_1(a) = \texttt{Pf}_2(b)$, where $A \in N_C$, C is a $\mathcal{CFD}_{nc}$ concept description, $a, b \in N_I$, and $\texttt{Pf}_i \in N_A^*$.

Let $\mathcal{L} \in \{\mathcal{ALCO}, \mathcal{ALCQ}, \textit{DL-Lite}_A, \mathcal{CFD}_{nc}\}$. We define $\mathcal{L}$-ontologies analogously to $\mathcal{ALC}$-ontologies. However, different to $\mathcal{ALC}$, the presence of new symbols in these logics modifies

formal definitions of some important notions in the previous chapter. For instance, the size of an $\mathcal{L}$-ontology $\mathfrak{O}$ is defined as the number of all occurrences of constructors $(\sqcap, \sqcup, \neg)$, concept names, role names, individual names, and attribute names in $\mathfrak{O}$.

Theorem 3.4. *The DLs $\mathcal{ALCO}$, $\mathcal{ALCQ}$, DL-Lite$_A$, and $\mathcal{CFD}_{nc}$ have equality power.*

This theorem is an immediate consequence of the following four examples, which each shows for the respective DL that it can derive equality between different individuals.

Example 3.5. *Here we formulate the example, written in the beginning of Chapter 3, in the DL $\mathcal{ALCO}$. Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ where*

$$\begin{aligned}\mathcal{T} \;:=\; & \{\exists \mathsf{expert}.\{\mathsf{LOGIC}\} \sqsubseteq \mathsf{VerTF},\ \exists \mathsf{expert}.\{\mathsf{PRIVACY}\} \sqsubseteq \mathsf{SecTF},\\ & \mathsf{VerTF} \sqsubseteq \{\mathsf{JOHN}, \mathsf{LINDA}, \mathsf{PAUL}, \mathsf{PATTIE}\},\\ & \mathsf{SecTF} \sqsubseteq \{\mathsf{JIM}, \mathsf{JOHN}, \mathsf{LINDA}, \mathsf{PAMELA}\},\ \mathsf{Female} \sqsubseteq \neg \mathsf{Male}\},\\ \mathcal{A} \;:=\; & \{\mathsf{Female}(x),\ \mathsf{expert}(x, \mathsf{LOGIC}),\ \mathsf{expert}(x, \mathsf{PRIVACY}),\\ & \mathsf{Female}(\mathsf{LINDA}), \mathsf{Female}(\mathsf{PATTIE}), \mathsf{Female}(\mathsf{PAMELA}),\\ & \mathsf{Male}(\mathsf{JOHN}), \mathsf{Male}(\mathsf{JIM}), \mathsf{Male}(\mathsf{PAUL})\}.\end{aligned}$$

It is easy to see that $\mathfrak{O} \models x \doteq \mathsf{LINDA}$ since x's expertise implies that she belongs to both the verification and the security task force, but the only female employee belonging to both is Linda. ♢

For the sake of brevity, we use abstract examples to show that $\mathcal{ALCQ}$, DL-Lite$_A$, and $\mathcal{CFD}_{nc}$ have equality power. It would, however, be easy to provide intuitive examples also for these three DLs.

Example 3.6. *Consider the $\mathcal{ALCQ}$ ontology $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ where*

$$\begin{aligned}&\mathcal{T} := \{A \sqsubseteq \leq 1\, r.B\} \text{ and}\\ &\mathcal{A} := \{A(a), r(a,b), r(a,x), B(b), B(x)\}.\end{aligned}$$

Obviously, we have $\mathfrak{O} \models x \doteq b$. ♢

Example 3.7. *Consider the DL-Lite$_A$ ontology $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$, where*

$$\begin{aligned}&\mathcal{T} := \{(\mathit{funct}\ r)\} \text{ and}\\ &\mathcal{A} := \{r(a, b_1), r(a, b_2)\}\end{aligned}$$

Clearly, we have $\mathfrak{O} \models b_1 \doteq b_2$. ♢

Example 3.8. *Consider the $\mathcal{CFD}_{nc}$ ontology $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ where*

$$\begin{aligned}&\mathcal{T} := \{A \sqsubseteq A : f \rightarrow id\} \text{ and}\\ &\mathcal{A} := \{A(a), f(a) = b, A(x), f(x) = b\}.\end{aligned}$$

Since both x and a belong to A and have the same value b for the path function f, the path functional dependency in $\mathcal{T}$ implies that they must be equal, i.e., we have $\mathfrak{O} \models x \doteq a$. ♢

We leave it to the reader to come up with translations of nominals, qualified number restrictions, functional roles and path functional dependencies into FOL with equality.

3.1.2 The Complexity of the Identity Problem

In this section, we first show that the identity problem can be polynomially reduced to the *instance problem* for all DLs with equality power. Note that the instance problem is one of the basic inference problems for DLs, and thus instance checking facilities are available in most DL reasoners. Given an ontology $\mathfrak{O}$, a concept description C, and an individual name a, we say that a is an *instance* of C w.r.t. $\mathfrak{O}$ (written $\mathfrak{O} \models C(a)$) if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathfrak{O}$.

Lemma 3.9. *Let $\mathcal{L}$ be a DL with equality power, $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ an $\mathcal{L}$ ontology and a, b two distinct individual names. If B is a concept name not occurring in $\mathfrak{O}$, then we have*

$$\mathfrak{O} \models a \doteq b \;\; \textit{iff} \;\; (\mathcal{T}, \mathcal{A} \cup \{B(a)\}) \models B(b).$$

Proof. The direction from left to right is trivial. We show the other direction by contraposition. Thus, assume that $\mathfrak{O} \not\models a \doteq b$. Let $\mathcal{I}$ be a model of $\mathfrak{O}$ such that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$. Let $\mathcal{I}'$ be the interpretation that coincides with $\mathcal{I}$ on all role names, individual names, and concept names different from B. For B we define $B^{\mathcal{I}'} := \{a^{\mathcal{I}}\}$. Since B does not occur in $\mathfrak{O}$, the interpretation $\mathcal{I}'$ is still a model of $\mathcal{T}$ and $\mathcal{A}$, and it satisfies $B(a)$ by our definition of $B^{\mathcal{I}'}$. However, it does not satisfy $B(b)$ since $b^{\mathcal{I}'} = b^{\mathcal{I}} \neq a^{\mathcal{I}}$ does not belong to $B^{\mathcal{I}'}$. □

This lemma shows that the identity problem is at most as complex as the instance problem for all DLs with equality power that allow instance assertions for concept names in the ABox. Since the instance problem is polynomial for $\mathcal{CFD}_{nc}$ [TW13], this implies that also the identity problem is polynomial for this DL. In [TW13] it is mentioned that PTIME-hardness of the consistency problem for $\mathcal{CFD}_{nc}$ ontologies is an easy consequence of PTIME-hardness of satisfiability of propositional Horn formulas [CN10]. We now show that the same is true also for the identity problem.

Theorem 3.10. *The identity problem is* PTIME*-complete for $\mathcal{CFD}_{nc}$ ontologies.*

Proof. We already know that the problem is in PTIME. To show PTIME-hardness, we reduce Horn-SAT to the identity problem. Recall that a Horn-formula ϕ is a finite set of clauses of the form

(a) $p_1 \wedge \ldots \wedge p_n \rightarrow p_0$ where $n > 0$ and $p_0, \ldots, p_n$ are propositional variables;

(b) $\rightarrow p_0$, which states that the propositional variable p_0 must be true;

(c) $p_1 \wedge \ldots \wedge p_n \rightarrow$ for $n > 0$ propositional variables $p_1, \ldots, p_n$, which states that $p_1, \ldots, p_n$ cannot be true at the same time.

Given ϕ, we construct a $\mathcal{CFD}_{nc}$-ontology $\mathfrak{O}_\phi = (\mathcal{T}_\phi, \mathcal{A}_\phi)$ as follows. For every propositional variable p occurring in ϕ we introduce a functional role f_p as well as individuals c_p, d_p. In addition, we introduce the functional role $f_\perp$ and the individuals $c_\perp, d_\perp$ and a, b. Intuitively, we encode truth of the propositional variable p as equality of the individuals c_p and d_p, and inconsistency as equality of $c_\perp$ and $d_\perp$. Clauses of the form (a) and (c) are encoded

using path functional dependencies in the TBox and clauses of the form (b) as path function assertions in the ABox. To be more precise, we define:

$$\begin{array}{rcl}
\mathcal{T}_\phi & := & \{A \sqsubseteq A : f_{p_1}, \ldots f_{p_n} \to f_{p_0} \mid p_1 \wedge \ldots \wedge p_n \to p_0 \in \phi\} \cup \\
& & \{A \sqsubseteq A : f_{p_1}, \ldots f_{p_n} \to f_\bot \mid p_1 \wedge \ldots \wedge p_n \to \in \phi\} \text{ and} \\
\mathcal{A}_\phi & := & \{A(a), A(b)\} \cup \\
& & \{f_p(a) = c_p, f_p(b) = d_p \mid \ p \in \mathtt{var}(\phi)\} \cup \\
& & \{f_\bot(a) = c_\bot, f_\bot(b) = d_\bot\} \cup \\
& & \{c_{p_0} = d_{p_0} \mid \to p_0 \in \phi\}.
\end{array}$$

where A is a concept name. The ontology constructed this way satisfies the syntactic restrictions on $\mathcal{CFD}_{nc}$ ontologies. Moreover, it can be constructed in logarithmic space since it can simply be read off the representation of ϕ.

By definition, the assertions $c_{p_0} = d_{p_0}$ enforce equality of these individuals iff ϕ contains a clause $\to p_0$ of the form (b). The path functional dependencies in $\mathcal{T}_\phi$ can then be used to derive further equalities according to the clauses of the form (a) and (c) in ϕ. It is thus easy to see that equality of the individuals c_p and d_p can be derived from $\mathfrak{O}$ iff ϕ implies that the propositional variable p must be set to true. Consequently, deriving an equality of the individuals $c_\bot$ and $d_\bot$ indicates that a clause $p_1 \wedge \ldots \wedge p_n \to$ of the form (c) in ϕ is violated. In fact, deriving $c_\bot = d_\bot$ is only possible if there is such a clause of the form (c) in ϕ and the equalities $c_{p_i} = d_{p_i}$ $(i = 1, \ldots, n)$ have already been derived.

Using this intuition, it is then easy to prove the following claim:

$$\phi \text{ is unsatisfiable iff } \mathfrak{O} \models c_\bot \doteq d_\bot,$$

which states correctness of our reduction, and thus establishes P-hardness of the identity problem in $\mathcal{CFD}_{nc}$. □

Next, we investigate the complexity of the identity problem in other lightweight DLs, such as *DL-Lite*$_A$ and $\mathcal{ELO}$. The latter extends $\mathcal{EL}$ with nominals and is clearly a fragment of $\mathcal{ALCO}$. It also holds that, by Theorem 3.3 and Theorem 3.4, $\mathcal{ELO}$ has equality power.

By Lemma 3.9, using the complexity result from the instance problem, the upper bound complexity for the identity problem in both $\mathcal{ELO}$ and *DL-Lite*$_A$ is PTime [BBL05; ACK+09]. To match PTime lower bounds for the identity problem in these two logics, the following two lemmas will show how to get that lower bounds.

Theorem 3.11. *The identity problem for $\mathcal{ELO}$ is P-complete.*

Proof. For $\mathcal{ELO}$, the hardness result for the identity problem is also from the horn-satisfiability problem that is known to be PTIME-complete. Given a Horn-Sat formula ϕ containing horn clauses of the following forms:

(a) $p_1 \wedge \ldots \wedge p_n \to p_0$ where $n > 0$ and $p_0, \ldots, p_n$ are propositional variables;

(b) $\to p_0$, which states that the propositional variable p must be true;

(c) $p_1 \wedge \ldots \wedge p_n \to$ for $n > 0$ propositional variables $p_1, \ldots, p_n$, which states that $p_1, \ldots, p_n$ cannot be true at the same time.

Given ϕ, we construct an $\mathcal{ELO}$-ontology $\mathfrak{O}_\phi = \{\mathcal{T}_\phi, \mathcal{A}_\phi\}$. For each propositional variable p in ϕ, we introduce a concept name A_p. In addition, we introduce individuals a and $b_\perp$, where the latter represents $\perp$. For each clause (a) and (c), we construct a GCI in TBox, whereas for each clause (b), we build a concept name assertion in the ABox. The following is the precise definition of $\mathfrak{O}_\phi = (\mathcal{T}_\phi, \mathcal{A}_\phi)$.

$$\begin{array}{rcl} \mathcal{T}_\phi & := & \{A_{p_1} \sqcap \ldots \sqcap A_{p_n} \sqsubseteq A_{p_0} \mid p_1 \wedge \ldots \wedge p_n \rightarrow p_0 \in \phi\} \cup \\ & & \{A_{p_1} \sqcap \ldots \sqcap A_{p_n} \sqsubseteq \{b_\perp\} \mid p_1 \wedge \ldots \wedge p_n \rightarrow \in \phi\} \text{ and} \\ \mathcal{A}_\phi & := & \{A_{p_0}(a) \mid\rightarrow p_0 \in \phi\}, \end{array}$$

where $A_p, \ldots, A_{p_n}$ are concept names. By definition, one can derive that a is an instance of a concept name A_p w.r.t. $\mathfrak{O}_\phi$ iff 1.) ϕ contains a clause $\rightarrow p$ or 2.) for a clause $p_1 \wedge \ldots \wedge p_n \rightarrow p_0 \in \phi$, we have all p_i that are true imply p_0 to be true iff $A_{p_i}(a)$ $(i = 1, \ldots, n)$ implies $A_{p_0}(a)$. As a consequence, one may derive $\{a\} \sqsubseteq \{b_\perp\}$ iff a clause $p_1 \wedge \ldots \wedge p_n \rightarrow$ of the form (c) is violated. It means that $\{a\} \sqsubseteq \{b_\perp\}$ is derived iff the clause $p_1 \wedge \ldots \wedge p_n \rightarrow$ is found in ϕ and $A_{p_i}(a)$ $(i = 1, \ldots, n)$ have been derived.

Using this intuition, it is then easy to show the following claim:

$$\phi \text{ is unsatisfiable iff } \mathfrak{O}_\phi \models \{a\} \sqsubseteq \{b_\perp\}$$

It is easy to see that subsumption between nominals are equivalent with equality between two individuals. Consequently, we have P-hardness for the identity problem in $\mathcal{ELO}$. □

Next, we show a reduction from the entailment problem of the Horn-CNF formulas [BGG97] to the identity problem in *DL-Lite*$_A$. To show this reduction, we adopt the proof from ([ACK+09], Theorem 8.7).

Theorem 3.12. *The identity problem in DL-Lite*$_A$ *is* PTIME*-complete.*

Proof. To show the PTIME-hardness for this problem, first we take a Horn-CNF formula ϕ defined formally as follows:

$$\phi = \bigwedge_{k=1}^{n} (a_{k,1} \wedge a_{k,2} \rightarrow a_{k,3}) \wedge \bigwedge_{\ell=1}^{p} a_{\ell,o},$$

where each $a_{k,j}$ and each $a_{\ell,0}$ is one of the propositional variables $a_1, \ldots, a_m$ and all $a_{k,1}, a_{k,2}, a_{k,3}$ are distinct for each k, $1 \leq k \leq n$. Next, we construct an ontology $\mathfrak{O}_\phi = (\mathcal{T}_\phi, \mathcal{A}_\phi)$ by using additional propositional variables t, a_i^k, for $1 \leq k \leq n$, $1 \leq i \leq m$, and f_k, g_k for $1 \leq k \leq n$, and role names r_P, r_Q, r_S, and r_T. Note that all these variables are considered as individuals in the ABox $\mathcal{A}_\phi$ defined as follows:

$$\begin{array}{rcl} \mathcal{A}_\phi & := & \{r_S(a_i^j, a_i^{j+1}) \mid 1 \leq i \leq m \wedge 1 \leq j \leq n-1\} \cup \\ & & \{r_S(a_i^n, a_i^1) \mid 1 \leq i \leq m \wedge 1 \leq j \leq n-1\} \cup \\ & & \{r_P(a_{k,1}^k, f_k), r_P(a_{k,2}^k, f_k) \mid 1 \leq k \leq n\} \cup \\ & & \{r_Q(g_k, a_{k,3}^k), r_Q(f_k, ak, 1^k)\} \cup \\ & & \{r_T(t, a_{\ell,0}^1) \mid 1 \leq \ell \leq p\} \end{array}$$

and the TBox $\mathcal{T}_\phi$ is defined as follows:

$$\mathcal{T}_\phi \quad := \quad \{(\texttt{funct } r_P), (\texttt{funct } r_Q), (\texttt{funct } r_S), (\texttt{funct } r_T)\}$$

By the construction above, it is shown [ACK+09] that given a propositional variable a_j,

$$\phi \models a_j \text{ if and only if } \mathfrak{O}_\phi \models r_T(t, a_j^1).$$

Now, let u be a fresh propositional variable such that $\mathfrak{O}'_\phi = (\mathcal{T}_\phi, \mathcal{A}_\phi \cup \{r_T(t,u\})$. It is easy to see that $\mathfrak{O}_\phi \models r_T(t, a_j^1)$ if and only if $\mathfrak{O}'_\phi \models a_j^1 \doteq u$ and thus we have

$$\phi \models a_j \text{ if and only if } \mathfrak{O}'_\phi \models a_j^1 \doteq u. \tag{3.1}$$

Finally this produces PTIME-hardness for the identity problem in *DL-Lite*$_A$. □

For $\mathcal{ALCO}$ and $\mathcal{ALCQ}$, the instance problem is EXPTIME-complete [Sch94; Tob01]. Thus, we obtain exponential-time upper bounds for the identity problem in these DLs. To show that these upper bounds are optimal, we basically prove that there are polynomial-time reductions of the instance problem in $\mathcal{ALC}$ to the identity problem in these logics. In fact, the instance problem is already EXPTIME-hard for the common sub-logic $\mathcal{ALC}$ of $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ [Sch91].

Before introducing these reductions and proving that they are correct, we have to deal with a subtlety that shows up in these proofs. Note that, in $\mathcal{ALC}$, we can assume without loss of generality that any instance relationship that does not follow from an ontology can be refuted by a model of cardinality greater than 1.

Lemma 3.13. *Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ALC}$ ontology, C an $\mathcal{ALC}$ concept description, and a an individual name. If $\mathfrak{O} \not\models C(a)$, then there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $a^\mathcal{I} \notin C^\mathcal{I}$ and $|\Delta^\mathcal{I}| \geq 2$.*

Proof. This follows from the fact that models of $\mathcal{ALC}$ ontologies are closed under disjoint union (see [BHL+17], Theorem 3.8). In fact, if $\mathfrak{O} \not\models C(a)$, then there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $a^\mathcal{I} \notin C^\mathcal{I}$. However, this model could have cardinality 1. If we take the disjoint union $\mathcal{J} = \mathcal{I}_1 \uplus \mathcal{I}_2$ of $\mathcal{I}$ with itself, then the cardinality of $\Delta^\mathcal{J}$ is twice the cardinality of $\Delta^\mathcal{I}$, and thus at least 2. Theorem 3.8 in [BHL+17] says that $\mathcal{J}$ is a model of $\mathcal{T}$. Regarding the ABox, we assume that all individual names occurring in $\mathcal{A}$ are interpreted in $\mathcal{J}$ by their interpretation in the renaming $\mathcal{I}_1$ of $\mathcal{I}$. Using Lemma 3.7 in [BHL+17], it is easy to see that this ensures that $\mathcal{J}$ is also a model of $\mathcal{A}$. □

Note that this lemma does not hold for $\mathcal{ALCO}$ ontologies. For example, $\mathfrak{O} = (\{\top \sqsubseteq \{a\}\}, \emptyset)$ has only models of size 1, and $\mathfrak{O} \not\models A(a)$. This is the reason why we use the DL $\mathcal{ALC}$ rather than the more expressive logics $\mathcal{ALCO}$ or $\mathcal{ALCQ}$ in our reductions.

Lemma 3.14. *Let $\mathcal{L} \in \{\mathcal{ALCO}, \mathcal{ALCQ}\}$, $\mathfrak{O}$ be an $\mathcal{ALC}$ ontology, C an $\mathcal{ALC}$ concept description, and a an individual name. Then we can construct in polynomial time an $\mathcal{L}$ ontology $\mathfrak{O}'$ and individuals a', b' such that*

$$\mathfrak{O} \models C(a) \text{ iff } \mathfrak{O}' \models a' \doteq b'.$$

Proof. Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$. We consider the two DLs separately.

1.) $\mathcal{L} = \mathcal{ALCO}$:
We define $\mathfrak{O}' := (\mathcal{T} \cup \{C \sqsubseteq \forall r.\{b'\}\}, \mathcal{A} \cup \{r(a,a'), r(a,b')\})$, where a', b' are distinct individual names and r is a role name such that a', b', r do not occur in $\mathfrak{O}$. The direction from left to right is again trivial. The other direction is shown by contraposition. Let $\mathcal{I}$ be a model of $\mathfrak{O}$ such that $a^{\mathcal{I}} \notin C^{\mathcal{I}}$. By Lemma 3.13, we can assume without loss of generality that the domain of $\mathcal{I}$ contains at least two distinct elements $d_1 \neq d_2$. We construct an interpretation $\mathcal{I}'$ that coincides with $\mathcal{I}$ on all concept, role, and individual names occurring in $\mathfrak{O}$, and thus is also a model of $\mathfrak{O}$. In addition, $\mathcal{I}'$ interprets r as $r^{\mathcal{I}'} := \{(a^{\mathcal{I}}, d_1), (a^{\mathcal{I}}, d_2)\}$ and the new individual names as $a'^{\mathcal{I}'} := d_1$ and $b'^{\mathcal{I}'} := d_2$. By construction, $\mathcal{I}'$ satisfies the assertional part of $\mathfrak{O}'$. To see that it also satisfies the GCI $C \sqsubseteq \forall r.\{b'\}$, note that $a^{\mathcal{I}} = a^{\mathcal{I}'}$ is the only element of $\mathcal{I}'$ that has successors w.r.t. the role r. Since it does not belong to $C^{\mathcal{I}} = C^{\mathcal{I}'}$, the elements of $C^{\mathcal{I}'}$ trivially satisfy the value restriction $\forall r.\{b'\}$. Thus, $\mathcal{I}'$ is a model of $\mathfrak{O}'$ in which the individuals a', b' are interpreted by different elements, which shows $\mathfrak{O}' \not\models a' \doteq b'$.

2.) $\mathcal{L} = \mathcal{ALCQ}$:
We define $\mathfrak{O}' := (\mathcal{T} \cup \{C \sqsubseteq\leq 1 r \top\}, \mathcal{A} \cup \{r(a,a'), r(a,b')\})$, where a', b' are distinct new individuals and r is a new role name not occurring in $\mathfrak{O}$. The direction from left to right is again trivial. To show the other direction, assume that $\mathcal{I}$ is a model of $\mathfrak{O}$ such that $a^{\mathcal{I}} \notin C^{\mathcal{I}}$. Again, we assume without loss of generality that the domain of $\mathcal{I}$ contains at least two distinct elements $d_1 \neq d_2$. We construct an interpretation $\mathcal{I}'$ in the same way as in case 1. above. Also, the argument why $\mathcal{I}'$ is a model of $\mathfrak{O}'$ in which a', b' are interpreted by different elements is identical to the one above.

As an easy consequence of Lemma 3.9 and Lemma 3.14 we obtain the exact complexity of the identity problem in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$. In fact, Lemma 3.9 yields ExpTime upper bounds. To show that Lemma 3.14 indeed yields ExpTime lower bounds, we need to take into account the fact that we have defined the identity problem with only consistent ontologies as possible input. In fact, since the consistency problem can be reduced to the instance problem in $\mathcal{ALC}$, it could potentially be the case that the reason for the ExpTime-hardness of the instance problem comes from the hardness of consistency only. However, we will show now that this is not the case, i.e., we show that ExpTime-hardness of the instance problem in $\mathcal{ALC}$ also holds if we consider the instance problem only for consistent $\mathcal{ALC}$ ontologies $\mathfrak{O}$.

Lemma 3.15. *The instance problem w.r.t. consistent $\mathcal{ALC}$ ontologies is* ExpTime*-hard.*

Proof. We show this by a reduction of the (un)satisfiability problem for $\mathcal{ALC}$-concepts w.r.t. TBoxes, which is also known to be ExpTime-complete ([BHL+17], Theorem 5.13). Recall that C is satisfiable w.r.t. $\mathcal{T}$ iff there is a model $\mathcal{I}$ of $\mathcal{T}$ satisfying $C^{\mathcal{I}} \neq \emptyset$.

Thus, let C be an $\mathcal{ALC}$ concept description and $\mathcal{T}$ an $\mathcal{ALC}$ TBox. We can assume without loss of generality that $\mathcal{T}$ consists of a single GCI $\top \sqsubseteq D$ for an $\mathcal{ALC}$ concept description D (see [BHL+17], page 117). Note that $\mathcal{T}$ may actually be inconsistent.

Given C and D, we now construct a consistent $\mathcal{ALC}$ ontology $\mathfrak{O}_{C,D} = (\mathcal{T}_{C,D}, \emptyset)$ as follows:

$$\begin{aligned} \mathcal{T}_{C,D} \quad := \quad & \{B \sqsubseteq \exists r.(C \sqcap A), A \sqsubseteq D\} \cup \\ & \{A \sqsubseteq \forall s.A \mid s \text{ occurs in } C, D\}, \end{aligned}$$

where A, B are concept names not occurring in C, D and r is a role name not occurring in C, D. It is easy to see that $\mathfrak{O}_{C,D}$ is consistent. In fact, any interpretation $\mathcal{I}$ with $A^{\mathcal{I}} = B^{\mathcal{I}} = \emptyset$ is obviously a model of $\mathcal{T}_{C,D}$. Thus, to prove the lemma it is sufficient to show that the following holds (for an arbitrary individual name a):

$$C \text{ is satisfiable w.r.t. } \{\top \sqsubseteq D\} \text{ iff } \mathfrak{O}_{C,D} \not\models \neg B(a).$$

First, assume that $\mathfrak{O}_{C,D} \not\models \neg B(a)$. This means that there is a model $\mathcal{I}$ of $\mathfrak{O}_{C,D}$ that interprets B as a non-empty set. Then the first GCI ensures that there is an element d_0 of A that also belongs to C. In addition, all the elements connected via roles occurring in C, D with d_0 also belong to A, and thus to D because of the second GCI. Consequently, if we restrict $\mathcal{I}$ to these elements, we obtain a model of $\top \sqsubseteq D$ in which d_0 belongs to C. This shows that C is satisfiable w.r.t. $\{\top \sqsubseteq D\}$.

Conversely, assume that $\mathcal{I}$ is a model of $\{\top \sqsubseteq D\}$ with $d_0 \in C^{\mathcal{I}}$. Then $\mathcal{I}$ can easily be extended to a model of $\mathcal{T}_{C,D}$ in which a belongs to B by (i) introducing an additional element d belonging to B, (ii) interpreting a as d, (iii) interpreting r as $\{(d, d_0)\}$, and (iv) putting d_0 as well as all the elements reachable from it into A. □

In addition, if $\mathfrak{O}$ is a consistent $\mathcal{ALC}$ ontology, then so are the ontologies $\mathfrak{O}$' constructed from it in the proof of Lemma 3.14. Thus, Lemma 3.14 together with Lemma 3.15 yields the matching ExpTime lower bounds for the identity problem in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$.

Theorem 3.16. *The identity problem is* ExpTime*-complete for $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ ontologies.*

One also wonder whether the complexity of the instance problem can be transferred to the identity problem also for DLs where the instance problem has a higher complexity than ExpTime. For example, the DL $\mathcal{ALCOIQ}$ which extends both $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ and additionally allows the use of inverse roles, has a NExpTime-complete satisfiability problem [Tob00], even w.r.t. the empty TBox. This implies that the instance problem w.r.t. consistent $\mathcal{ALCOIQ}$ ontologies is coNExpTime-complete. In fact, the $\mathcal{ALCOIQ}$ concept description C is unsatisfiable iff $(\emptyset, \emptyset) \models \neg C(a)$ (for a new individual name a), which shows coNExpTime-hardness also w.r.t. consistent ontologies. The complexity upper bound follows from the NExpTime upper bound of satisfiability in C^2, i.e., two-variable fragment of first-order logic with counting quantifiers [Pra05].

Since $\mathcal{ALCOIQ}$ contains $\mathcal{ALCO}$, it has equality power and can force models to have cardinality 1. Lemma 3.9 implies that the identity problem in $\mathcal{ALCOIQ}$ is in coNExpTime. Regarding hardness, the reductions employed in the proof of Lemma 3.14 can in principle both be used since the constructors employed in them are available in $\mathcal{ALCOIQ}$. However, Lemma 3.14 uses an $\mathcal{ALC}$ ontology $\mathfrak{O}$ in the reduction, which yields only an ExpTime lower bound. Simply using an $\mathcal{ALCOIQ}$ ontology instead does not work since the proof depends on the fact that $\mathfrak{O}$ has models refuting the instance relation of cardinality at least 2. However, by looking at the NExpTime-hardness proof for satisfiability in $\mathcal{ALCOIQ}$ in [Tob00], it is easy to see that the following modified instance problem is also coNExpTime-hard for consistent $\mathcal{ALCOIQ}$ ontologies: is a an instance of C in all models of $\mathfrak{O}$ of cardinality ≥ 2? Now, let us call this problem the *instance problem w.r.t.* 2*-consistency*. Thus, one can without loss of generality restrict the attention to models of cardinality ≥ 2 when reducing the instance problem for $\mathcal{ALCOIQ}$ to the identity problem for this logic.

Theorem 3.17. *The identity problem is* coNExpTime*-complete for $\mathcal{ALCOIQ}$ ontologies.*

For the DLs with equality power considered so far in this chapter, the identity problem has the same complexity as the instance problem. A natural question to ask is whether this is always the case. A simple example shows that the answer to this question is negative. In fact, let $\mathcal{ALC}^{=}$ be the DL $\mathcal{ALC}$, with the only difference that $\mathcal{ALC}^{=}$ ABoxes may contain equality assertions $a \doteq b$ between individual names. It is easy to see that the identity problem in this DL is non-trivial, but it can be solved in polynomial time. In fact, to check whether a consistent $\mathcal{ALC}^{=}$ ontology implies an equality $a \doteq b$, we only need to construct the reflexive, transitive, and symmetric closure of the explicitly stated equalities. However, since $\mathcal{ALC}$ is a sub-logic of $\mathcal{ALC}^{=}$, the instance problem in this DL is ExpTime-hard (and it is easy to show that it is also in ExpTime).

3.2 The View-Based Identity Problem

In this section, we will adapt the approach of [SS06; SS09] for view-based information hiding such that it can formalize the rôle-based access control scenario sketched in the beginning of this chapter. We assume that ontologies are written using some DL $\mathcal{L}$ with equality power.

To define what kind of information is to be hidden, we divide the set $\mathsf{N_I}$ of individual names into the disjoint sets $\mathsf{N_{AI}}$ and $\mathsf{N_{KI}}$ consisting of anonymous and known individuals, respectively. As before, we do not make the unique name assumption for these individuals.

Now, given an anonymous individual $x \in \mathsf{N_{AI}}$ and an ontology $\mathfrak{O}$, we define the *identity* of x w.r.t. $\mathfrak{O}$ as

$$idn(x, \mathfrak{O}) := \{b \in \mathsf{N_{KI}} \mid \mathfrak{O} \models x \doteq b\}.$$

Note that $b, b' \in idn(x, \mathfrak{O})$ implies that $\mathfrak{O} \models b' \doteq b$. Thus, if the cardinality of $idn(x, \mathfrak{O})$ is greater 1, this does not mean that x is equal to one of these individuals, but rather that it is equal to all of them (and thus that all of them are equal). We say that x is *hidden* if $idn(x, \mathfrak{O}) = \emptyset$.

In the rôle-based access control scenario we assume that there is a "large" *input ontology* $\mathfrak{O}_I$ that is always consistent, but users can only see a part of it depending on which rôle they currently have. More formally, we assume that there is a finite set of *user rôles* $\mathfrak{R}$, and that playing the rôle $\hat{r} \in \mathfrak{R}$ gives access to a subset $\mathfrak{O}_{\hat{r}} \subseteq \mathfrak{O}_I$ of the input ontology. Here "access" does not mean that a user with rôle $\hat{r}$ can download the ontology $\mathfrak{O}_{\hat{r}}$. Instead, the user can ask queries to $\mathfrak{O}_{\hat{r}}$, both in the form of a subsumption query $C \sqsubseteq D$ for concept descriptions C, D and a conjunctive query (CQ) q. Note that the former query just simply asks whether for all models $\mathcal{I}$ of $\mathfrak{O}_{\hat{r}}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. We next formally define the answer to the queries w.r.t. each rôle.

Definition 3.18 (Answer to Queries). *Let $\mathfrak{O}_I$ be the input ontology, $\mathfrak{O}_{\hat{r}} \subseteq \mathfrak{O}_I$ the ontology accessible by users with rôle $\hat{r} \in \mathfrak{R}$, and q be a query. The* answer *to q w.r.t. $\hat{r}$, denoted by $ans(q, \hat{r})$, is defined as follows:*

- *If $q = C \sqsubseteq D$ or a CQ without answer variables, then $ans(q, \hat{r}) := \{\mathit{true}\}$ if $\mathfrak{O} \models q$ or $ans(q, \hat{r}) := \emptyset$ if $\mathfrak{O} \not\models q$*
- *if q is a CQ with $n > 0$ answer variables, then $ans(q, \hat{r}) := \{\vec{t} \in (N_I)^n \mid \mathfrak{O}_{\hat{r}} \models q(\vec{t})\}$.* $\diamondsuit$

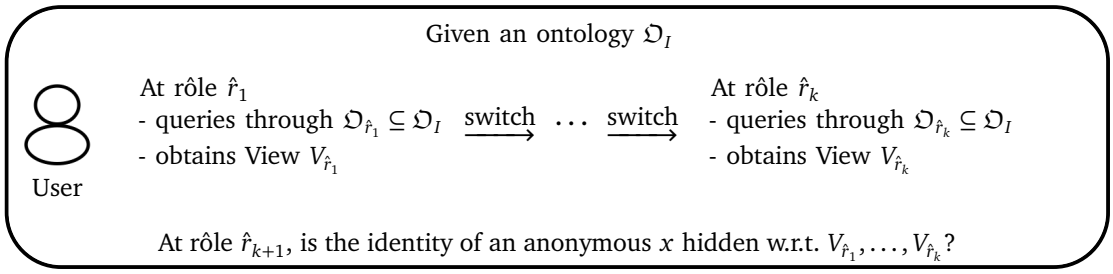


Figure 3.20: An illustration for the identity problem in a rôle-based setting

Since $\mathfrak{O}_{\hat{r}} \subseteq \mathfrak{O}_I$, *positive answers* to queries, i.e., $ans(C \sqsubseteq D, \hat{r}) = \{\texttt{true}\}$ or $\vec{t} \in ans(q, \hat{r})$ imply that this subsumption or conjunctive queries also hold in $\mathfrak{O}_I$. In contrast, negative answers do not tell us anything about what holds in $\mathfrak{O}_I$ since the inclusion may be strict. Answers to queries w.r.t. rôle $\hat{r}$ is stored in a view.

Definition 3.19 (View). *A* view for $\hat{r} \in \mathfrak{R}$ *(written* $\hat{r} \models V$*) is a finite set of pairs* $\langle q_i, ans(q_i, \hat{r})\rangle$*, where each* q_i *is a query. The* size *of the view* V *is defined as*

$$\sum_{(q_i, ans(q_i,\hat{r}))\in V} |\, q_i \,| + (k \cdot |ans(q_i, \hat{r})|), \text{ where } k \text{ is the arity of } q_i$$

In a setting where user rôles can dynamically change, a user may successively play rôles $\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_k$, in each rôle $\hat{r}_i$ generating (and storing) a view $V_{\hat{r}_i}$ for $\hat{r}_i$ by asking queries. The question is now whether these views can be used to find out the identity of a given anonymous individual $x \in \mathsf{N_{AI}}$. An illustration for this setting is depicted in Figure 3.20.

Assume that the user wants to know whether there is $b \in N_{KI}$ such that $b \in idn(x, \mathfrak{O}_I)$. However, the user cannot access $\mathfrak{O}_I$ as a whole, all she knows is that the positive answers to the queries in the views $V_{\hat{r}_i}$ are justified by subsets of $\mathfrak{O}_I$. Consequently, instead of one (unknown) ontology $\mathfrak{O}_I$, the user needs to consider all possible ontologies, i.e., all ontologies that are compatible with the positive answers in the views.

Definition 3.21 (Possible Ontology). *The ontology* $\mathfrak{P}$ *is a* possible ontology *for the sequence of views* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ *if* $\mathfrak{P}$ *is consistent and compatible with all positive answers in these views, where* $\mathfrak{P}$ *is* compatible *with*

- $\langle C \sqsubseteq D, \{true\}\rangle \in V_{\hat{r}_i}$ *if* $\mathfrak{P} \models C \sqsubseteq D$,
- $\langle q, ans(q, \hat{r}_i)\rangle \in V_{\hat{r}_i}$ *if for all* $\vec{t} \in ans(q, \hat{r}_i)$*, we have* $\mathfrak{P} \models q(\vec{t})$. ♢

We denote the set of all possible ontologies *for* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ *with* $\texttt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$*. The* certain identity *of* x *w.r.t.* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ *is defined as*

$$cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) := \bigcap_{\mathfrak{P}\in \texttt{Poss}(V_{\hat{r}_1},\ldots,V_{\hat{r}_k})} idn(x, \mathfrak{P}).$$

Definition 3.22 (View-Based Identity Problem). *Given* $x \in \mathsf{N_{AI}}$ *and views* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$*, the identity of* x *is* hidden *w.r.t.* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ *if*

$$cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) = \emptyset.$$

The view-based identity problem *asks whether the identity of x is hidden or not w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$.* ◇

Since $\mathfrak{O}_I \in \texttt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$, we know that $b \in \mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ implies that $b \in \mathit{idn}(x, \mathfrak{O}_I)$. Thus, if $\mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) \neq \emptyset$, the identity of x in $\mathfrak{O}_I$ is no longer hidden. Conversely, if $\mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) = \emptyset$, then for all $b \in N_{KI}$, there is a possible ontology $\mathfrak{P} \in \texttt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ such that $\mathfrak{P} \not\models x \doteq b$. Since, according to the information available to the user, $\mathfrak{O}_I$ could be this $\mathfrak{P}$, she cannot conclude for any $b \in N_{KI}$ that $\mathfrak{O}_I \models x \doteq b$. This shows that $\mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) = \emptyset$ indeed corresponds to the fact that the views $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ do not disclose the identity of x.

Since the set $\texttt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ consists of infinitely many ontologies, $\mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ does not directly yield an approach for computing this set. We will now show that we can reduce this computation to the identity problem for the *canonical ontology* of $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$. Basically, this ontology consists of axioms obtained from the positive answers in the views. However, for some views $V_{\hat{r}}$, there are still some pairs $\langle q, \mathit{ans}(q, \hat{r})\rangle \in V_{\hat{r}_i}$, where q is a CQ and contains existentially quantified variables. To remove these variables, first we restrict our attention to all pairs $\langle q, \mathit{ans}(q, \hat{r})\rangle$ in all views $V_{\hat{r}_i}$, where q is a conjunctive query. Then, we construct a first-order representation for this collection of pairs. Formally, we construct a first-order sentence obtained from $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ as follows:

$$\alpha_{FO}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) \;=\; \bigwedge_{V_{\hat{r}_i}} \wedge \bigwedge_{\substack{\langle q_j, \mathit{ans}(q_j, \hat{r}_i)\rangle \in V_{\hat{r}_i}, \\ q_j \text{ is a CQ}}} \wedge \bigwedge_{\vec{t}_k \in \mathit{ans}(q_j, \hat{r}_i)} \exists \vec{w}.\mathit{conj}(\vec{t}_k, \vec{w})$$

Since $\alpha_{FO}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ still contains existentially quantified variables w from $\vec{w}$, we can remove them by simply replacing w with fresh constants a_w not occurring in $\alpha_{FO}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$. These replacements yields the ground first-order representation $\alpha_G(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ of $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$. Now, we start constructing the *canonical ontology* $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ from a given $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$.

Definition 3.23 (Canonical Ontology). *The* canonical ontology $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ *of* $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ *is defined as* $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) := (\mathcal{T}, \mathcal{A})$ *where*

$$\begin{aligned} \mathcal{T} &:= \{C \sqsubseteq D \mid \langle C \sqsubseteq D, \{true\}\rangle \in V_{\hat{r}_i} \text{ for some } i, 1 \leq i \leq k\} \\ \mathcal{A} &:= \{A(a) \mid A(a) \text{ is an atom in } \mathit{conj}(\alpha_G(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))\} \cup \\ &\quad\; \{r(a_1, a_2) \mid r(a_1, a_2) \text{ is an atom in } \mathit{conj}(\alpha_G(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))\} \end{aligned}$$

Note that the construction of $\alpha_{FO}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ above only considers linearly many queries q_j in $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ and then for every answers of q_j, we transform the form of q_j by replacing each answer variable x with t from $\vec{t}_k$. Additionally, $\alpha_G(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ is obtained from $\alpha_{FO}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ by replacing linearly many existentially quantified variables occurring in $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ with fresh constants. Since the construction of $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ is performed by taking GCIs whose answer is positive in each $V_{\hat{r}_i}$ and taking each atom from $\mathit{conj}(\alpha_G(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$, we know that this construction is also done in linear time and its size is linearly bounded in the sum of the sizes of the views $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$.

Now, for simplicity to prove the next theorem, we treat the fresh constants a_w in $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ as anonymous individuals. The following theorem says that the sets $\mathit{cert\_idn}(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ and $\mathit{idn}(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$ of the identities of an anonymous individual x are the same.

Theorem 3.24. *Given views $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$ and an anonymous individual $x \in N_{AI}$, we have*

$$cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}) = idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})).$$

Proof. Let us assume that there is $b \in \mathsf{N_{KI}}$ such that $b \in cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$. This implies that for all $\mathfrak{P} \in \mathtt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$, we have $\mathfrak{P} \models x \doteq b$. By contradiction, we assume that b is not in $idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$. This implies that there is a model $\mathcal{I}$ of $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ such that $\mathcal{I} \not\models x \doteq b$. However, $\mathcal{I}$ is also a model of a possible ontology $\mathfrak{P} \in \mathtt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ by assigning to each existential variable w the object $a_w^{\mathcal{I}}$. Hence, $b \notin cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$, which is a contradiction to the fact that b is in $cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$.

For the converse direction, we assume that $b \in idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$. By contradiction, we assume that $b \notin cert\_idn(x, V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$. This implies that there is $\mathfrak{P} \in \mathtt{Poss}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ such that $\mathfrak{P} \not\models x \doteq b$ and thus there is a model $\mathcal{I}$ of $\mathfrak{P}$ such that $\mathcal{I} \not\models x \doteq b$. Then, we can extend $\mathcal{I}$ to an interpretation $\mathcal{I}'$ by interpreting each fresh anonymous individual a_w as the value assigned to the existentially quantified variable w that makes $\mathcal{I}$ satisfying $\mathfrak{P}$. Thus, $\mathcal{I}'$ is also a model of $\mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k})$ such that $\mathcal{I}' \not\models x \doteq b$, which is a contradiction to $b \in idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$ □

This theorem shows that, to check whether x is *hidden* w.r.t. $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$, it is sufficient to compute $idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$. If the employed ontology language $\mathcal{L}$ allows for unrestricted GCIs, concept assertions, and role assertions, the set $idn(x, \mathcal{C}(V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}))$ can clearly be computed using an algorithm that solves the identity problem for $\mathcal{L}$ ontologies a polynomial number of times.

Note that this applies to the DLs $\mathcal{ELO}$, $\mathcal{ALCO}$, $\mathcal{ALCQ}$, and $\mathcal{ALCOIQ}$ considered in the previous sections, but not to *DL-Lite*$_A$ and $\mathcal{CFD}_{nc}$ since their GCIs and concept assertions need to satisfy certain restrictions. One may ask whether it makes sense to have views where the answers for the queries are obtained from ontologies written in the DLs $\mathcal{ELO}$, $\mathcal{ALCO}$, $\mathcal{ALCQ}$, and $\mathcal{ALCOIQ}$. The answer is yes since the subsumption problem and computing query answering in these four DLs are decidable and computable, respectively [OS12].

The upper bounds for $\mathcal{ELO}$, $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ are obvious. For $\mathcal{ALCOIQ}$, one considers all the (polynomially many) known individuals $a_1, \ldots, a_p$. Using a NExpTime procedure for the complement of the identity problem, one then checks whether x is not identical to a_1. The non-successful paths of this non-deterministic computation stop with failure whereas the successful ones continue with the same test for a_2, etc. It is easy to see that this yields the desired NExpTime procedure. In fact, any path of this procedure has only exponential length, and a successful path indicates that inequality with x holds for all known individuals.

Corollary 3.25. *For $\mathcal{L} = \mathcal{ELO}$, we can check in polynomial time whether an anonymous individual x is hidden w.r.t. views $V_{\hat{r}_1}, \ldots, V_{\hat{r}_k}$. For $\mathcal{L} \in \{\mathcal{ALCO}, \mathcal{ALCQ}\}$ this problem can be checked in exponential time. Meanwhile, for $\mathcal{L} = \mathcal{ALCOIQ}$, this problem can be solved in NExpTime.*

To show that the upper bounds above are optimal, we investigate the lower bounds for the view-based identity problem for each DL. We start with a reduction from the (un)satisfiability problem of Horn-SAT formulas to the view-based identity problem in $\mathcal{ELO}$.

Theorem 3.26. *The view-based identity problem in $\mathcal{ELO}$ is* PTIME*-complete*

Proof. As written in Theorem 3.11, we take a Horn-SAT formula ϕ containing horn clauses of forms (a), (b), and (c) and then construct an $\mathcal{ELO}$-ontology $\mathfrak{O}_\phi = \{\mathcal{T}_\phi, \mathcal{A}_\phi\}$. For this construction, we emphasize that $b_\perp$ is the known individual and rename a with x, where x is an anonymous individual. From this construction, we can deduce that

$$\phi \text{ is unsatisfiable iff } \mathfrak{O}_\phi \models \{x\} \sqsubseteq \{b_\perp\} \tag{3.2}$$

Now, let V_ϕ be a view that is constructed from $\mathfrak{O}_\phi$ as follows:

$$\begin{aligned} V_\phi \quad := \quad & \{\langle C \sqsubseteq D, \texttt{true}\rangle \mid C \sqsubseteq D \in \mathfrak{O}_\phi\} \cup \\ & \{\langle A(v), x\rangle \mid A(x) \in \mathfrak{O}_\phi \wedge v \text{ is an answer variable }\} \end{aligned}$$

It can be readily seen that $\mathfrak{O}_\phi$ is the canonical ontology of V_ϕ. Since $\{b_\perp\}$ is the only known individual in $\mathfrak{O}_\phi$, x is not hidden in V_ϕ iff $\mathfrak{O}_\phi \models x \doteq b_\perp$. Consequently, by Equation 3.2, we show that

$$\phi \text{ is unsatisfiable iff the identity of } x \text{ is not hidden w.r.t.} V_\phi$$

This shows us that the view-based identity problem is PTIME-hard. □

Next, we show that the EXPTIME complexity of the view-based identity problem in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ is also optimal. This is shown by doing a reduction from the instance problem in consistent $\mathcal{ALC}$ ontologies.

Theorem 3.27. *The view-based identity problem in $\mathcal{L} \in \{\mathcal{ALCO}, \mathcal{ALCQ}\}$ is* EXPTIME*-complete.*

Proof. We consider the proof from Lemma 3.14, where we take the $\mathcal{ALC}$ ontology $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$, the $\mathcal{ALC}$ concept C and the individual a. Then, we construct the same $\mathcal{L}$ ontology $\mathfrak{O}' = (\mathcal{T}', \mathcal{A}')$ as in the proof of Lemma 3.14, but we rename the individual a' with an anonymous individual $x \in N_{AI}$ and now we treat the individual b' as a known individual. Next, we construct a new ontology $\mathfrak{O}'' = (\mathcal{T}'', \mathcal{A}'')$ that is obtained from $\mathfrak{O}'$ by replacing every concept assertion $D(a) \in \mathcal{A}'$ with $A_D(a)$, where A_D is a new concept name not occurring in $\mathfrak{O}'$, and adding $A_D \sqsubseteq D$ to $\mathcal{T}$'. From the proof in Lemma 3.14 and due to the fact that $\mathfrak{O}'$ and $\mathfrak{O}''$ are equisatisfiable, we know that

$$\mathfrak{O} \models C(a) \text{ iff } \mathfrak{O}'' \models x \doteq b'. \tag{3.3}$$

Now, we construct a view V'' obtained from $\mathfrak{O}''$ as follows:

$$\begin{aligned} V'' \quad := \quad & \{\langle C \sqsubseteq D, \texttt{true}\rangle \mid C \sqsubseteq D \in \mathfrak{O}''\} \cup \\ & \{\langle A(v), a\rangle \mid A(a) \in \mathfrak{O}'' \wedge v \text{ is an answer variable }\} \cup \\ & \{\langle r(u_1, u_2), (a_1, a_2)\rangle \mid r(a_1, a_2) \in \mathfrak{O}'' \ \wedge \ u_1, u_2 \text{ are answer variables}\} \end{aligned}$$

It is obvious to see that $\mathfrak{O}''$ is the canonical ontology of V''. Due to Theorem 3.24 and Equation 3.3, the following statement also holds if $\mathfrak{O}$ is formulated in $\mathcal{ALCO}$ or $\mathcal{ALCQ}$.

$$\mathfrak{O} \models C(a) \text{ iff the identity of } x \text{ is not hidden w.r.t. } V''. \qquad \square$$

Last, we show that the NEXPTIME complexity of $\mathcal{ALCOIQ}$ is also tight. Similar to the arguments above Theorem 3.17, we reduce the instance problem in $\mathcal{ALCOIQ}$ ontologies w.r.t.

2-consistency to the view-based identity problem in $\mathcal{ALCOIQ}$ by only considering models of cardinality ≥ 2. This reduction can also use the same proof construction described in Theorem 3.27.

Theorem 3.28. *The view-based identity problem in $\mathcal{ALCOIQ}$ is* NEXPTIME*-complete*

3.3 The k-Hiding Problem

So far, in the identity problem, we investigated whether an anonymous individual x belongs to a singleton set of known individual. If it does not belong to any singleton set consisting of one known individual, then the identity of x is hidden. However, in some cases, an attacker does not really want to known the exact identity of some anonymous objects. Instead, he wants to deduce whether x belongs to a set of known individuals such that the set has a cardinality smaller than k. From this deduction, he may infer that the identity of x is one of the known individuals that belong to that set. In this situation, it is sufficient to see that the identity of x is 'k-hidden' if x does not belong to any (k-1)-subsets of known individuals.

Now, we provide the definition for the *k-hiding problem* that depends on the definition of a subprobem called *$\mathfrak{K}$-membership*, which asks whether an anonymous individuals belongs to a given set of known individuals in all models of an ontology.

Definition 3.29. *Let $\mathfrak{O}$ be a consistent ontology, $x \in N_{AI}$, and $\mathfrak{K} \subseteq N_{KI}$, where $\mathfrak{K} = \{a_1, \ldots, a_{k-1}\}$. The individual x is* in $\mathfrak{K}$-membership w.r.t. $\mathfrak{O}$ *iff $x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\}$ for all models $\mathcal{I}$ of $\mathfrak{O}$. Then, x is* not k-hidden w.r.t. $\mathfrak{O}$ *iff there is $\mathfrak{K} \subseteq N_{KI}$, where $\mathfrak{K} = \{a_1, \ldots, a_{k-1}\}$, such that x is in $\mathfrak{K}$-membership w.r.t. $\mathfrak{O}$.*

The definition above implies that the *k-hiding* (resp. *$\mathfrak{K}$-membership*) problem asks whether x is k-hidden (resp. in $\mathfrak{K}$-membership) w.r.t. $\mathfrak{O}$. Since the number k is included to the input, we need to be careful when considering the size of the k-hiding problem. Here we assume that the number k is written in unary encoding and thus the size of the k-hiding problem is $|\mathfrak{O}| + |\mathsf{N_{KI}}| + k$. We can also infer that if x is not k-hidden w.r.t. $\mathfrak{O}$, then for all ℓ, where $\ell > k$, the individual x is not ℓ-hidden either. To provide a concrete illustration on how users or attackers can infer that x is (not) k-hidden w.r.t. a given ontology, we provide the following example inspired from Example 3.5.

Example 3.30. *Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ where*

$$
\begin{aligned}
\mathcal{T} \;:=\; & \{\exists \mathsf{expert}.\{\mathsf{CODING}\} \sqsubseteq \mathsf{TechTeam},\ \mathsf{TechTeam} \equiv \mathsf{VerTF} \sqcup \mathsf{SecTF},\\
& \mathsf{VerTF} \sqsubseteq \{\mathsf{JOHN}, \mathsf{LINDA}, \mathsf{PAUL}\},\\
& \mathsf{SecTF} \sqsubseteq \{\mathsf{JIM}, \mathsf{PATTIE}, \mathsf{PAMELA}\},\ \mathsf{Female} \sqsubseteq \neg \mathsf{Male}\},\\
\mathcal{A} \;:=\; & \{\mathsf{Female}(x),\ \mathsf{expert}(x, \mathsf{CODING}),\\
& \mathsf{Female}(\mathsf{LINDA}), \mathsf{Female}(\mathsf{PATTIE}), \mathsf{Female}(\mathsf{PAMELA}),\\
& \mathsf{Male}(\mathsf{JOHN}), \mathsf{Male}(\mathsf{JIM}), \mathsf{Male}(\mathsf{PAUL})\}.
\end{aligned}
$$

Let $x \in \mathsf{N_{AI}}$. First, we show that $\mathfrak{O} \not\models x \doteq a$ for all $a \in \mathsf{N_{KI}}$. Since the concept of female is disjoint with the concept of male, x is not equal to any male individual. It remains to check whether x is equal to one of female individuals. However, due to the disjunction rule, x's expertise enforces x to become a member of either the verification team or the security team. Let $\mathcal{I}$ and $\mathcal{I}'$

be models of $\mathfrak{O}$ such that $x^{\mathcal{I}} \in \mathsf{VerTF}^{\mathcal{I}}$, but $x^{\mathcal{I}'} \notin \mathsf{VerTF}^{\mathcal{I}'}$ and $x^{\mathcal{I}'} \in \mathsf{SecTF}^{\mathcal{I}'}$, but $x^{\mathcal{I}} \notin \mathsf{SecTF}^{\mathcal{I}}$. Since both verification and security teams are disjoint, it implies that $x^{\mathcal{I}} = b^{\mathcal{I}}$, but $x^{\mathcal{I}'} \neq b^{\mathcal{I}'}$ for $b \in \{\mathsf{LINDA}, \mathsf{PATTIE}, \mathsf{PAMELA}\}$. This implies that x is 2-hidden w.r.t. $\mathfrak{O}$.

Next, we show that the ontology above does not entail any set consisting of two known individuals either, i.e., $x^{\mathcal{I}} \notin \{a^{\mathcal{I}}, b^{\mathcal{I}}\}$ for all models $\mathcal{I}$ of $\mathfrak{O}$ and all $a, b \in \mathsf{N_{KI}}$. If both a and b are males, or the genders of a and b are different, or a and b are females belonging to the same team, then it is easy to see that there is a model $\mathcal{I}'$ of $\mathfrak{O}$ such that $x^{\mathcal{I}'} \notin \{a^{\mathcal{I}'}, b^{\mathcal{I}'}\}$. It remains to show whether in all models $\mathcal{I}$ of $\mathfrak{O}$, $x^{\mathcal{I}} \in \{\mathsf{LINDA}^{\mathcal{I}}, \mathsf{PATTIE}^{\mathcal{I}}\}$ or $x^{\mathcal{I}} \in \{\mathsf{LINDA}^{\mathcal{I}}, \mathsf{PAMELA}^{\mathcal{I}}\}$. However, for the case that $x^{\mathcal{I}} \in \{\mathsf{LINDA}^{\mathcal{I}}, \mathsf{PATTIE}^{\mathcal{I}}\}$, due to the disjunction rule, there are models $\mathcal{I}_1, \mathcal{I}_2$ of $\mathfrak{O}$ such that

- *$x^{\mathcal{I}_1} \in \mathsf{VerTF}^{\mathcal{I}_1}$, but $x^{\mathcal{I}_1} \notin \mathsf{SecTF}^{\mathcal{I}_1}$ and*
- *$x^{\mathcal{I}_1} = \mathsf{LINDA}^{\mathcal{I}_1}$ and, additionally,*
- *$x^{\mathcal{I}_2} \in \mathsf{SecTF}^{\mathcal{I}_2}$, but $x^{\mathcal{I}_2} \notin \mathsf{VerTF}^{\mathcal{I}_2}$ and*
- *$x^{\mathcal{I}_2} = \mathsf{PAMELA}^{\mathcal{I}_2}$ and $x^{\mathcal{I}_2} \neq \mathsf{PAMELA}^{\mathcal{I}_2}$*

This implies that $x^{\mathcal{I}_1} \in \{\mathsf{LINDA}^{\mathcal{I}_1}, \mathsf{PATTIE}^{\mathcal{I}_1}\}$, but $x^{\mathcal{I}_2} \notin \{\mathsf{LINDA}^{\mathcal{I}_2}, \mathsf{PATTIE}^{\mathcal{I}_2}\}$. The same argument also holds to check the case that $x^{\mathcal{I}} \in \{\mathsf{LINDA}^{\mathcal{I}}, \mathsf{PAMELA}^{\mathcal{I}}\}$. This implies that x is 3-hidden w.r.t. $\mathfrak{O}$.

Nevertheless, if we extend k to 4, then there is a subset $\mathfrak{K}$ of known individuals, where $\mathfrak{K} = \{\mathsf{LINDA}, \mathsf{PATTIE}, \mathsf{PAMELA}\}$ such that $x^{\mathcal{I}} \in \{\mathsf{LINDA}^{\mathcal{I}}, \mathsf{PATTIE}^{\mathcal{I}}, \mathsf{PAMELA}^{\mathcal{I}}\}$ in all models $\mathcal{I}$ of $\mathfrak{O}$. This can be justified by the fact that in every model of $\mathfrak{O}$, due to the x's expertise, x belongs to either the verification team or security team. Then, in every possible team to which x belongs, there is a female individual to which x is equal, which directly implies that x is not 4-hidden w.r.t. $\mathfrak{O}$. ♢

Since the definition of the k-hiding problem above relies on the $\mathfrak{K}$-membership probem, we first investigate the complexities for the $\mathfrak{K}$-membership problem in some DLs with equality power and then show the complexities for the k-hiding problem. We start with the upper bound.

3.3.1 Upper Bounds

We first show that the $\mathfrak{K}$-membership problem actually can also be reduced to the instance problem for all DLs with equality power.

Lemma 3.31. *Let $\mathcal{L}$ be a DL with equality power, $\mathfrak{O}$ be a consistent ontology formulated in $\mathcal{L}$, $x \in \mathsf{N_{AI}}$, and $a_i \in \mathsf{N_{KI}}$ for all $1 \leq i \leq k-1$. It holds that for all models $\mathcal{I}$ of $\mathfrak{O}$,*

$$x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\} \text{ iff } \mathfrak{O}' \models A(x),$$

where $\mathfrak{O}' := (\mathcal{T}, \mathcal{A} \cup \{A(a_i) \mid 1 \leq i \leq k-1\})$ and A is a new concept name not occurring in $\mathfrak{O}$.

Proof. It is easy to show the direction from left to right. Now we prove from the other direction via contraposition. Assume that there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $x^{\mathcal{I}} \notin \{a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\}$. Now, we extend $\mathcal{I}$ to $\mathcal{I}'$ such that $\mathcal{I}'$ coincides with $\mathcal{I}$ on all concept names, role names, and individual names occurring in $\mathfrak{O}$. Additionally, $A^{\mathcal{I}'} = \{a_1^{\mathcal{I}'}, \ldots, a_{k-1}^{\mathcal{I}'}\}$ and thus $\mathcal{I}'$ satisfies the

terminological and assertional parts of $\mathfrak{O}'$. It implies that $\mathcal{I}'$ is a model of $\mathfrak{O}'$. Due to the fact that A does not occur in $\mathfrak{O}$ and $a_1, \ldots, a_{k-1}$ are the only instances of A w.r.t. $\mathcal{I}'$ as well as $x^{\mathcal{I}'} \neq a_i^{\mathcal{I}'}$ for all $1 \leq i \leq k-1$, it holds that $x^{\mathcal{I}'} \notin A^{\mathcal{I}'}$. Thus, $\mathfrak{O}' \not\models A(x)$. □

For Description Logics with equality power that contain nominal, it is also easy to reduce the $\mathfrak{K}$-membership problem to the subsumption problem.

Lemma 3.32. *Let $\mathcal{L}$ be a DL with equality power containing nominals, $\mathfrak{O}$ be a consistent ontology formulated in $\mathcal{L}$, $x \in \mathsf{N_{AI}}$, and $\{a_1, \ldots, a_k\} = \mathfrak{K} \subseteq \mathsf{N_{KI}}$. It holds for all models $\mathcal{I}$ of $\mathfrak{O}$ that*

$$x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, \ldots, a_k^{\mathcal{I}}\} \text{ iff } \mathfrak{O} \models \{x\} \sqsubseteq \{a_1, \ldots, a_k\}.$$

As a consequence, to check whether an anonymous x is k-hidden, we simply enumerate all (k-1)-subsets $\mathfrak{K}$ of known individuals and then for each subset $\mathfrak{K}$, we check whether x is in $\mathfrak{K}$-membership based on Lemma 3.31 or Lemma 3.32 depending whether the underlying logic contains nominal or not. For some DLs with equality power, such as $\mathcal{ALCO}$ or $\mathcal{ALCQ}$, the instance problem or the subsumption problem for both is in ExpTime. Since there are exponentially many subsets of known individuals to be enumerated and for each subset call an ExpTime algorithm to check whether x belongs to the subset w.r.t. $\mathfrak{O}$, it means that solving the k-hiding problem is also in ExpTime for both logics. For a more expressive DL $\mathcal{ALCOIQ}$, where the subsumption and the instance problems in this logic can be solved in coNExpTime, we need to perform an NExpTime procedure in each subset of known individuals for checking whether x is not contained in the subset w.r.t. $\mathfrak{O}$. Since the complexity class NExpTime subsumes ExpTime, the whole procedure for checking whether x is k-hidden w.r.t. $\mathfrak{O}$ takes non-deterministic exponential time.

Theorem 3.33. *Let $\mathcal{L}_1 \in \{\mathcal{ALCO}, \mathcal{ALCQ}\}$ and $\mathcal{L}_2 \in \{\mathcal{ALCOIQ}\}$. The $\mathfrak{K}$-membership problem in $\mathcal{L}_1$ and $\mathcal{L}_2$ can be solved in* ExpTime *and* coNExpTime*, respectively, while the k-hiding problem in $\mathcal{L}_1$ and $\mathcal{L}_2$ can be solved in* ExpTime *and* NExpTime*, respectively.*

However, the two reductions in Lemma 3.31 and 3.32 still look costly for some tractable DLs with equality power, such as $\mathcal{ELO}$, *DL-Lite*$_A$, or $\mathcal{CFD}_{nc}$, because in order to check if x is k-hidden we have to generate all $(k-1)$-subsets of $\mathsf{N_{KI}}$ and then run a decision procedure to solve the instance problem of which the complexity for these logics are mostly in polynomial time.

To deal with this issue, one needs to investigate whether the ontology is formulated in a *convex* DL or not. In general, a DL $\mathcal{L}$ is *convex* if it satisfies the following property: Given an $\mathcal{L}$-ontology $\mathfrak{O}$, an anonymous $x \in \mathsf{N_{AI}}$, and $a_1, \ldots, a_{k-1} \in \mathsf{N_{KI}}$, it holds that for all models $\mathcal{I}$ of $\mathfrak{O}$, $x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\}$ iff for all models $\mathcal{J}$ of $\mathfrak{O}$, $x^{\mathcal{J}} = a_i^{\mathcal{J}}$ for some $1 \leq i \leq k-1$.

We show that for each ontology $\mathfrak{O}$ formulated in convex DLs that have equality power, the k-hiding ($\mathfrak{K}$-membership) problem w.r.t. $\mathfrak{O}$ can be reduced to the identity problem w.r.t. $\mathfrak{O}$ and the converse direction also holds. For the convex logics, these reductions show that k-hiding is also the same with 2-hiding and it implies that the k-hiding problem is not an interesting generalization of the identity problem since they share the same complexity classes for the considered convex logics.

Obviously, this property does not hold for $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ since we may find the following counterexamples.

Example 3.34. *We define an ontology $\mathfrak{O}_1 = (\mathcal{T}_1, \emptyset)$ formulated in $\mathcal{ALCO}$ as follows:*

$$\mathcal{T}_1 = \{\{x\} \sqsubseteq \{a_1\} \sqcup \{a_2\}\}.$$

It is easy to see that $x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, a_2^{\mathcal{I}}\}$ for all models $\mathcal{I}$ of $\mathfrak{O}_1$, but it does not hold that $x^{\mathcal{I}} \in \{a_i^{\mathcal{I}}\}$ in all models $\mathcal{I}$ of $\mathfrak{O}_1$ for any $i \in \{1,2\}$. Now, we define an ontology $\mathfrak{O}_2 = (\mathcal{T}_2, \mathcal{A}_2)$ formulated in $\mathcal{ALCQ}$ as follows:

$$\begin{aligned}\mathcal{T}_2 &:= \{A_1 \sqcap A_2 \sqsubseteq \bot\}\\ \mathcal{A}_2 &:= \{A_1(a_1), A_2(a_2), \leq 2r.\top(a)\} \cup\\ &\quad\ \{r(a,x), r(a,a_1), r(a,a_2)\}\end{aligned}$$

In all models of $\mathfrak{O}_2$, the individual a is enforced by $\leq 2r.\top(a)$ to only have at most two r-successors. Since a_1 and a_2 are also enforced to be not equal w.r.t. $\mathfrak{O}_2$ due to the GCI in $\mathcal{T}_2$, in every model $\mathcal{J}$ of $\mathfrak{O}_2$ we can only have either $x^{\mathcal{J}} = a_1^{\mathcal{J}}$ or $x^{\mathcal{J}} = a_2^{\mathcal{J}}$. This implies that $x^{\mathcal{J}} \in \{a_1^{\mathcal{J}}, a_2^{\mathcal{J}}\}$, but similar to the previous counterexample, it does not hold that $x^{\mathcal{J}} \in \{a_i^{\mathcal{J}}\}$ in all models $\mathcal{J}$ of $\mathfrak{O}_2$ for any $i \in \{1,2\}$. ♢

One advantage to ensure that our ontology $\mathfrak{O}$ is formulated in a convex logic is that we can solve the k-hiding problem as the same as solving the identity problem. One just need to take the set N_{KI} of all known individuals and then check whether a given anonymous x is equal to one of the elements from N_{KI} w.r.t. $\mathfrak{O}$. If there is at least one known individual that is equal to x w.r.t. $\mathfrak{O}$, then we know that x is not k-hidden for any $k \geq 1$. This procedure is better than the previous one which requires us to first enumerate all $(k-1)$-subsets $\mathfrak{K}$ of N_{KI}.

Lemma 3.35. *Let $\mathfrak{O}$ be a consistent ontology formulated in a convex logic $\mathcal{L}$, $x \in \mathsf{N}_{\mathsf{AI}}$, and N_{KI} be the set of all known individuals. The individual x is k-hiding w.r.t. $\mathfrak{O}$ iff x is not in $\mathfrak{K}$-membership w.r.t. $\mathfrak{O}$, where $\mathfrak{K} = \mathsf{N}_{\mathsf{KI}}$.*

Next, we show that $\mathcal{ELO}$, *DL-Lite*$_A$, and $\mathcal{CFD}_{nc}$ are convex.

Lemma 3.36. *Let $\mathfrak{O}$ be a consistent ontology formulated in $\mathcal{L} \in \{\mathcal{ELO}, \textit{DL-Lite}_A, \mathcal{CFD}_{nc}\}$. If $x \in \mathsf{N}_{\mathsf{AI}}$, $a_1, \ldots, a_{k-1} \in \mathsf{N}_{\mathsf{KI}}$, then for all models $\mathcal{I}$ of $\mathfrak{O}$, $x^{\mathcal{I}} \in \{a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\}$ iff for all models $\mathcal{J}$ of $\mathfrak{O}$, $x^{\mathcal{J}} = a_i^{\mathcal{J}}$ for some $1 \leq i \leq k-1$.*

Proof. We show this by distinguishing the following three cases, where for each case, the 'only if' direction is trivial, and thus it remains to show the 'if' direction.

1. Let $\mathcal{L} = \mathcal{ELO}$.
   We assume that $\mathfrak{O} \models \{x\} \sqsubseteq \{a_1, \ldots, a_{k-1}\}$ since $\mathcal{ELO}$ has nominals by Lemma 3.32. As argued in [KKS12], given $\mathfrak{O}$, reasoning in $\mathcal{ELO}$ should consider the set $\mathcal{M}_{\mathfrak{O}}$ of axioms closed under the inference rules for $\mathcal{ELO}$, which may contain *conditional subsumption axiom* written as $G : C \sqsubseteq D$ with the semantics: if $G^{\mathcal{I}} \neq \emptyset$, then $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, where C, D, G are concepts occurring in $\mathfrak{O}$. Additionally, as stated in [KKS12], to deduce subsumption in $\mathcal{ELO}$, we can construct the canonical model $\mathcal{I}_c$ of $\mathfrak{O}$ such that for the concept G and all concepts D occurring in $\mathfrak{O}$, the following holds:

$$\mathcal{I}_c \models G \sqsubseteq D \text{ implies } G : G \sqsubseteq D \in \mathcal{M}_{\mathfrak{O}}, \tag{3.4}$$

where $\mathcal{M}_\mathfrak{O}$ is the completion set closed under the completion rule. Without loss of generality, we assume that $\{x\}$ and $\{a_i\}$, for all $1 \leq i \leq k-1$, are in $\mathfrak{O}$. Now we assume that $\mathfrak{O} \models \{x\} \sqsubseteq \{a_1\} \sqcup \ldots \sqcup \{a_{k-1}\}$. Since $\mathcal{I}_c$ is a model of $\mathfrak{O}$, we have $\{x^{\mathcal{I}_c}\} \in \{a_1^{\mathcal{I}_c}, \ldots, a_{k-1}^{\mathcal{I}_c}\}$. Consequently, there exists $1 \leq i \leq k-1$ such that $x^{\mathcal{I}_c} = a_i^{\mathcal{I}_c}$ or $\mathcal{I}_c \models \{x\} \sqsubseteq \{a_i\}$ and thus $\{x\} : \{x\} \sqsubseteq \{a_i\} \in \mathcal{M}_\mathfrak{O}$ follows by equation (3.4) . Due to the soundness of the completion rule stated in (3.4), if $\{x\} : \{x\} \sqsubseteq \{a_i\} \in \mathcal{M}_\mathfrak{O}$, then we obtain $\mathfrak{O} \models \{x\} \sqsubseteq \{a_i\}$.

2. Let $\mathcal{L} = \textit{DL-Lite}_A$
   We assume that $x^\mathcal{I} \in \{a_1^\mathcal{I}, \ldots, a_k^\mathcal{I}\}$ for all models $\mathcal{I}$ of $\mathfrak{O}$. In [ACK+09], it is said that every *DL-Lite* family has canonical model $\mathcal{I}_c$ for the ontologies, if they are consistent. Since $\mathcal{I}_c$ is also a model, $x_c^\mathcal{I} \in \{a_1^{\mathcal{I}_c}, \ldots, a_{k-1}^{\mathcal{I}_c}\}$. It implies that there exists i, for $1 \leq i \leq k-1$, such that $x^{\mathcal{I}_c} = a^{\mathcal{I}_c}$. In [CDL+07], it is also said that for every model $\mathcal{I}$ of $\mathfrak{O}$, there is a homomorphism from $\mathcal{I}_c$ to $\mathcal{I}$ that maps the objects in the extension of concept, role, and individual names in $\mathcal{I}$ to objects in the extension of concept, role, and individual names in $\mathcal{I}$. In other words, $x^{\mathcal{I}_c} = a_i^{\mathcal{I}_c}$ implies $x^\mathcal{I} = a_i^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathfrak{O}$.

3. Let $\mathcal{L} = \mathcal{CFD}_{nc}$
   We assume that $x^\mathcal{I} \in \{a_1^\mathcal{I}, \ldots, a_{k-1}^\mathcal{I}\}$ for all models $\mathcal{I}$ of $\mathfrak{O}$. In [TW13], it is said that in $\mathcal{CFD}_{nc}$ ontologies $\mathfrak{O}$, we can build a non-deterministic finite (NFA) automaton $\mathfrak{A}$ whose states Q are concept names, their negations, and individuals occurring in $\mathfrak{O}$. Additionally, it also has transition relations $\delta_\mathfrak{A}(\mathfrak{O})$, which is a set of transitions from one state q_1 to another state q_2 via path functional relation f, denoted by $q_1 \xrightarrow{f} q_2$, where $f \in N_F^* \cup \{\epsilon\}$ for $q_1, q_2 \in Q$. We may also write $q \xrightarrow{\mathtt{Pf}} q'$, where $q, q' \in Q$ and $\mathtt{Pf} \in N_F^*$. This kind of automaton $\mathfrak{A}$ is very useful for various reasoning problems related in $\mathcal{CFD}_{nc}$. For instance

$$a \xrightarrow{\epsilon} b, b \xrightarrow{\epsilon} a \in \delta_\mathfrak{A}(\mathfrak{O}) \text{ iff } \mathfrak{O} \models a \doteq b \tag{3.5}$$

   It turns out that $\mathfrak{A}$ also accepts such a canonical model $\mathcal{I}_c$, if $\mathfrak{O}$ is consistent, whose domain $\Delta^{\mathcal{I}_c}$ consists of equivalent classes $[a]$ defined as:

$$[a] := \{b \in N_I \mid a \xrightarrow{\epsilon} b, b \xrightarrow{\epsilon} a \in \delta_\mathfrak{A}(\mathfrak{O})\} \text{ for all } a \in \mathsf{N_I}.$$

   Since $\mathcal{I}_c$ is a model of $\mathfrak{O}$, it implies that $x^{\mathcal{I}_c} \in \{a_1^{\mathcal{I}_c}, \ldots, a_{k-1}^{\mathcal{I}_c}\}$ and thus there exists a_i such that $x^{\mathcal{I}_c} = a_i^{\mathcal{I}_c}$. Immediately, we have $x, a_i \in [a_i]$ and hence $x \xrightarrow{\epsilon} a_i, a_i \xrightarrow{\epsilon} x \in \delta_\mathfrak{A}(\mathfrak{O})$. Due to (3.5), we have $\mathfrak{O} \models x \doteq a_i$. □

From the proof above, we can see that the convexity of those three logics exists if and only if they have a canonical model. The following theorem is a consequence of Lemma 3.36, Theorem 3.10, and Theorem 3.11.

Theorem 3.37. *The $\mathfrak{K}$-membership problem and the k-hiding problem for $\mathcal{ELO}$, DL-Lite$_A$, and $\mathcal{CFD}_{nc}$ are in PTime , respectively.*

3.3.2 Lower Bounds

Now, we investigate the lower bounds for the k-hiding and the $\mathfrak{K}$-membership problems. We start with the tractable ones, which are $\mathcal{ELO}$, *DL-Lite*$_A$, and $\mathcal{CFD}_{nc}$, where the hardness result for the $\mathfrak{K}$-membership problem can be trivially obtained from the identity problem.

Theorem 3.38. *For DLs $\mathcal{ELO}$, DL-Lite$_A$, and $\mathcal{CFD}_{nc}$, the $\mathfrak{K}$-membership problem is in P-hard.*

Proof. We take the identity problem and reduce it to the $\mathfrak{K}$-membership problem. It is sufficient to show that given a consistent ontology $\mathfrak{O}$ formulated in $\mathcal{L}$, $x \in \mathsf{N_{AI}}$, $a \in \mathsf{N_{KI}}$, and $a_i \in \mathsf{N_{KI}}$ as the known individuals not occurring in $\mathfrak{O}$, for all $1 \leq i \leq k-1$, we have

$$\mathfrak{O} \models x \doteq a \text{ iff } x^{\mathcal{I}} \in \{a^{\mathcal{I}}, a_1^{\mathcal{I}}, \ldots, a_{k-1}^{\mathcal{I}}\} \tag{3.6}$$

for all models $\mathcal{I}$ of $\mathfrak{O}$', where $\mathfrak{O}' := \mathfrak{O} \cup \{\top(a_i) \mid 1 \leq i \leq k-1\}$. It can be see that $\mathfrak{O}'$ basically is equivalent to $\mathfrak{O}$, where adding $\top(a_i)$ to $\mathfrak{O}$ just means a tautology. Consequently, Equation 3.6 trivially holds. □

For more expressive DLs such as $\mathcal{ALCO}$ and $\mathcal{ALCQ}$, we also obtain the lower bound for the $\mathfrak{K}$-membership problem through a reduction from the instance problem in consistent $\mathcal{ALC}$ ontologies.

Lemma 3.39. *Let $\mathfrak{O} = (\mathfrak{T}, \mathfrak{A})$ be an ontology formulated in $\mathcal{ALC}$, C an $\mathcal{ALC}$ concept, and $a \in \mathsf{N_I}$. It holds that*

$$\mathfrak{O} \models C(a) \textit{ iff } x^{\mathcal{I}'} \in \{b_1^{\mathcal{I}'}, \ldots, b_{k-1}^{\mathcal{I}}\}$$

for all models $\mathcal{I}'$ of $\mathfrak{O}'$, where $\mathfrak{O}' = (\mathcal{T}', \mathcal{A}')$ is defined as

$$\begin{aligned} \mathcal{T}' &:= \mathcal{T} \cup \{C \sqsubseteq \forall r.\{b_1, \ldots, b_{k-1}\} \cup \\ &\qquad \{\{b_i\} \sqcap \{b_j\} \sqsubseteq \bot \mid 1 \leq i, j \leq k-1 \wedge i \neq j\} \\ \mathcal{A}' &:= \mathcal{A} \cup \{r(a, x)\} \end{aligned}$$

and x and b_i, for all $1 \leq i \leq k-1$, are new anonymous and known individuals, respectively, and $r \in N_R$ is a new role name.

Proof. It is obvious to see the direction from left to right. For the other direction, we prove it via contraposition. We assume that $\mathfrak{O} \not\models C(a)$. As written in Lemma 3.13, there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $\mathcal{I} \not\models C(a)$ and $|\Delta^{\mathcal{I}}| \geq k$ for $k > 1$. Suppose that there are $d_0, d_1, \ldots d_{k-1} \in \Delta^{\mathcal{I}}$ with $d_i \neq d_j$, for all $0 \leq i \leq k-1$. Now we extend $\mathcal{I}$ to $\mathcal{I}'$ that coincides on all individual names, role names, and concept names occurring $\mathfrak{O}$. Additionally, $x^{\mathcal{I}'} := d_0, r^{\mathcal{I}'} := \{(a^{\mathcal{I}'}, x^{\mathcal{I}'})\}$, and $b_i^{\mathcal{I}'} := d_i$, for all $1 \leq i \leq k-1$. By construction, $\mathcal{I}$' clearly satisfies the assertional part of $\mathfrak{O}$'. Then, since each b_i and b_j are interpreted differently and b_i is disjoint with b_j, it implies that $\mathcal{I}'$ satisfies the GCI $\{b_i\} \sqcap \{b_j\} \sqsubseteq \bot$. Next, since a is the only individual that has an r-successor and $a^{\mathcal{I}'} \in C^{\mathcal{I}'}$, all elements of $C^{\mathcal{I}'}$ trivially satisfy $\forall r.\{b_1, \ldots, b_{k-1}\}$. It implies that $\mathcal{I}$' is a model of $\mathfrak{O}$', but interprets $x^{\mathcal{I}'} \notin \{b_1^{\mathcal{I}'}, \ldots, b_{k-1}^{\mathcal{I}'}\}$. □

Lemma 3.40. *Let $\mathfrak{O} = (\mathcal{T}, \mathcal{A})$ be a consistent ontology formulated in $\mathcal{ALC}$, C an $\mathcal{ALC}$ concept, and $a \in \mathsf{N_I}$. It holds that*

$$\mathfrak{O} \models C(a) \textit{ iff } x^{\mathcal{I}'} \in \{b_1^{\mathcal{I}'}, \ldots, b_{k-1}^{\mathcal{I}'}\}$$

for all models $\mathcal{I}$' of $\mathfrak{O}$', where $\mathfrak{O}' := (\mathcal{T}', \mathcal{A}')$ is defined as

$$\begin{aligned}\mathcal{T}' &:= \mathcal{T} \cup \{C \sqsubseteq \leq (k-1) r.\top\} \cup \\ &\qquad \{A_i \sqcap A_j \sqsubseteq \bot \mid 1 \leq i, j \leq k-1 \wedge i \neq j\} \\ \mathcal{A}' &:= \mathcal{A} \cup \{A_i(b_i) \mid 1 \leq i \leq k-1\} \cup \{r(a, x)\} \cup \\ &\qquad \{r(a, b_i) \mid 1 \leq i \leq k-1\}\end{aligned}$$

and $x \in \mathsf{N_{AI}}$, $b_i \in \mathsf{N_{KI}}$ are new anonymous and known individuals, respectively, $A_i \in \mathsf{N_C}$ are also new concept names, and last we have $r \in \mathsf{N_R}$ as a new role name.

Proof. Clearly, the direction from left to right holds. Now, it is sufficient to show the direction from right to left via contraposition. Assume that there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $a^{\mathcal{I}} \notin C^{\mathcal{I}}$. By Lemma 3.13, it follows that $|\Delta^{\mathcal{I}}| \geq k$ for $k > 1$. Suppose that there are $d_0, d_1, \ldots d_{k-1} \in \Delta^{\mathcal{I}}$ with $d_i \neq d_j$ for all $0 \leq i \leq k-1$. Then, we extend $\mathcal{I}$ to $\mathcal{I}'$ such that $\mathcal{I}'$ coincides with $\mathcal{I}$ on all role names, individual names, and concept names occurring in $\mathfrak{O}$. Additionally, $x^{\mathcal{I}'} := d_0$, $b_i^{\mathcal{I}'} := d_i$ for all $i \in \{1, \ldots, k-1\}$, $r^{\mathcal{I}'} := \{(a^{\mathcal{I}'}, x^{\mathcal{I}'})\} \cup \{(a^{\mathcal{I}'}, b_i^{\mathcal{I}'}) \mid 1 \leq i \leq k-1\}$, and $A_i^{\mathcal{I}'} := \{b_i^{\mathcal{I}'}\}$ for all $i \in \{1, \ldots, k-1\}$. It implies that $\mathcal{I}'$ satisfies the assertional part of $\mathfrak{O}$'. Further, since each A_i and A_j is disjoint and the only elements in $\mathcal{I}'$ belonging to A_i and A_j are d_i and d_j, respectively, that are defined differently, it implies that $\mathcal{I}'$ satisfies the second form of GCI in $\mathcal{T}'$. Then, since $a^{\mathcal{I}'} \notin C^{\mathcal{I}'}$ and a is the only individual that has an r-successor, all elements of $C^{\mathcal{I}'}$ satisfiy $\leq (k-1) r.\top$. It implies that $\mathcal{I}'$ satisfies $\mathcal{T}'$ and thus $\mathcal{I}'$ is a model of $\mathfrak{O}$' but interprets $x^{\mathcal{I}'} \notin \{b_1^{\mathcal{I}'}, \ldots, b_{k-1}^{\mathcal{I}'}\}$. □

Combining Lemma 3.31, 3.39, and 3.40, we have the following theorem.

Theorem 3.41. *The $\mathfrak{K}$-membership problem for $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ is* ExpTime*-complete.*

Next, we show that we can also reduce the instance problem w.r.t. consistent $\mathcal{ALC}$ ontologies to the k-hiding problem for DLs $\mathcal{ALCO}$ and $\mathcal{ALCQ}$.

Theorem 3.42. *The k-hiding problem in $\mathcal{ALCO}$ and $\mathcal{ALCQ}$ is* ExpTime*-complete.*

Proof. By Theorem 3.33, we know that the k-hiding problem is in ExpTime for $\mathcal{ALCO}$ and $\mathcal{ALCQ}$. For the hardness, we take a consistent $\mathcal{ALC}$ ontology $\mathfrak{O}$, an $\mathcal{ALC}$ concept C, and an individual a. Then, we show the following claim

$$\mathfrak{O} \models C(a) \text{ iff } x \text{ is not } k\text{-hidden w.r.t. } \mathfrak{O}',$$

where $x \in N_{AI}$ and $\mathfrak{O}'$ is defined as in Lemma 3.39 or Lemma 3.40 such that $\mathfrak{O}'$ is formulated in $\mathcal{ALCO}$ or $\mathcal{ALCQ}$, respectively. Now, let $\mathfrak{O}$ be formulated in $\mathcal{ALCO}$. If $\mathfrak{O} \models C(a)$, then by Lemma 3.39 and Lemma 3.40, we know that x is not k-hidden w.r.t. $\mathfrak{O}'$.

Conversely, we prove it by adopting the proofs from Lemma 3.39 and Lemma 3.40. If $\mathfrak{O} \not\models C(a)$, then there is a model $\mathcal{I}$ of $\mathfrak{O}$ such that $\mathcal{I} \not\models C(a)$ and $|\Delta^{\mathcal{I}}| \geq 2$ by Lemma 3.13. Suppose there are n distinct individuals a_i occurring in $\mathfrak{O}$ for $n \geq 1$ and $|\Delta^{\mathcal{I}}|$ has at least $n + (k-1) + 1$ distinct elements $e_1, \ldots, e_n, d_0, d_1, \ldots, d_{k-1}$ for $k > 1$. Without loss of generality, we may assume that each a_i is interpreted to one e_j in $\mathcal{I}$ for all $1 \leq i, j \leq n$. Then, we construct $\mathcal{I}'$ obtained from $\mathcal{I}$ as introduced in the proof of Lemma 3.39 or Lemma 3.40,

where $x^{\mathcal{I}'} = d_0$, $b_\ell^{\mathcal{I}'} = d_\ell$ for all ℓ, where $1 \leq \ell \leq k-1$. As argued in Lemma 3.39 and Lemma 3.40, it holds that $\mathcal{I}'$ is a model of $\mathfrak{O}'$. From this construction, it can be shown that for any given $\mathfrak{K} \subseteq N_{KI}$, where $| \mathfrak{K} | = k-1$, we can always build a model $\mathcal{I}'$ from a model $\mathcal{I}$ of $\mathfrak{O}$ in which $a^{\mathcal{I}} \notin C^{\mathcal{I}}$, which interprets x differently with all individuals in $\mathfrak{K}$. Particularly, this implies that x is not in $\mathfrak{K}$-membership w.r.t. $\mathfrak{O}'$ for all $\mathfrak{K} \subseteq N_{KI}$, where $|\mathfrak{K}| = k-1$ and thus x is not in k-hiding w.r.t. $\mathfrak{O}'$. □

For the DL $\mathcal{ALCOIQ}$, the lower bound for the $\mathfrak{K}$-membership problem and k-hiding problem can also be obtained by reduction from the instance problem in $\mathcal{ALCOIQ}$ ontologies w.r.t. 2-consistency by again restricting attentions only to models of cardinality ≥ 2. The argument for the proof construction of this reduction is similar to the ones in Lemma 3.39, Lemma 3.40, and Theorem 3.42.

Theorem 3.43. *The $\mathfrak{K}$-membership problem in $\mathcal{ALCOIQ}$ is* coNEXPTIME*-complete and thus the k-hiding problem in $\mathcal{ALCOIQ}$ is* NEXPTIME*-complete.*

Chapter 4

Repairing Description Logic Ontologies

The previous chapter has supplied various reasoning problems associated to identity-preserving problems, provided mechanisms to detect whether the identity of an object is deduced w.r.t. the considered setting, and produced the complexity results of these problems varying from one DL to another. Moreover, this sort of investigation augments more mechanisms to detect privacy breaches in DL ontologies besides other known procedures which check if protected consequences (e.g., subsumption relationships, instance relationships, or membership of (tuple of) individuals) are revealed from the given ontologies or not. However, we have not considered what to do when this is the case that this confidential information can be revealed. One possibility would be to modify the input ontology containing this confidential information such that it yields a new ontology from which the hidden information cannot be deduced. In this chapter, we are about to discuss the notion of *ontology repair*, where the motivation behind this notion is not only restricted to privacy problems, but also to more general problems that principally want to get rid of unwanted consequences.

In the classical approach of repairing ontologies [KPS+06a; Sch05a; Sch05b], one compute all minimal subsets of the ontology, called *justifications*, that are responsible to derive an unwanted consequences and then, given all of them, the ontology is repaired by removing one axiom from each justification. However, removing complete axioms may also eliminate consequences that are actually wanted. For example, assume that our ontology contains the following general concept inclusions:

$$\exists \mathsf{owns}.(\mathsf{GermanCar} \sqcap \mathsf{Diesel}) \sqsubseteq \exists \mathsf{gets}.\mathsf{Compensations},$$
$$\mathsf{GermanTaxiDriver} \sqsubseteq \exists \mathsf{owns}.(\mathsf{GermanCar} \sqcap \mathsf{Diesel}).$$

These two axioms are a justification for the incorrect consequence that every German taxi driver gets compensation. Suppose that we are not allowed by the ontology administrator to remove the first axiom. Thus, removing the second axiom will get rid of the incorrect conseqeunce. However, removing this completely would also remove the correct consequence that every German taxi driver owns a German car. Thus one may weaken the second axiom to $\mathsf{GermanTaxiDriver} \sqsubseteq \exists \mathsf{owns}.\mathsf{GermanCar}$, which together with the first axiom will also remove the unwanted consequence. However, this weakening is still not 'gentle' since the consequence stating that every German taxi driver owns diesel is also lost. Thus, it would be more appropriate if the second axiom is replaced with a weaker axiom $\mathsf{GermanTaxiDriver} \sqsubseteq \exists \mathsf{owns}.\mathsf{GermanCar} \sqcap \exists \mathsf{owns}.\mathsf{Diesel}$, which is clearly seen that this weaker axiom together with the first axiom also eliminate the unwanted consequence. This is the basic idea underlying our *gentle repair approach*. In general, in this approach we weaken

one axiom from each justification such that the modified justifications no longer have the consequence.

In Section 4.1, we formally introduce the notion of repair and show that optimal repair need not exist in general. Then, in Section 4.2, we introduce a general framework for repairing ontologies based on axiom weakening. This framework is independent of the concrete method employed for weakening axioms and of the concrete ontology language used to write axiom. It only assumes that ontologies are finite sets of axioms, that there is a monotonic consequence operator defining which axiom follows from which, and that weaker axioms have less consequences. Our first important result is that, in general, the gentle repair approach needs to be iterated, i.e., applying it once does not necessarily remove the consequence. This problem has actually been overlooked in [LSP+08], which means that their approach does not always yield a repair. Our second result is that at most exponentially many iterations are always sufficient to reach a repair. The authors of [TCG+18] had already realized that iteration is needed, but they did not give an example explicitly demonstrating this, and they had no termination proof.

Instead of allowing for arbitrary ways of weakening axioms, in Section 4.3, we then introduce the notion of a *weakening relation*, which restricts the way in which axioms can be weakened. Subsequently, we define conditions on such weakening relations that equip the gentle repair approach with better algorithmic properties if they are satisfied. Further, in Section 4.4 we address the task of defining specific weakening relations for the DL $\mathcal{EL}$. After showing that two quite large such relations do not behave well, we introduce two restricted relations, which are based on generalizing the right-hand sides of axioms semantically or syntactically. Both of them satisfy most of our conditions, but from a complexity point of view the syntactic variant behaves considerably better. Likewise, in Section 4.5, we introduce two weakening relations for $\mathcal{ALC}$, where the first relation generalizes and specializes concepts C based on a finite set of signature and a fixed role-depth, while the second relation does generalizations and specializations syntactically.

4.1 Repairing Ontologies

For the purpose of this section, we leave it open what sort of axioms and ontologies are allowed in general. We only assume that there is a monotonic consequence relation $\mathfrak{O} \models \alpha$ between ontologies (i.e., finite sets of axioms) and axioms, and that $\mathsf{Con}(\mathfrak{O})$ consists of all consequences of $\mathfrak{O}$.

Assume in the following that the ontology $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ is the disjoint union of a *static* ontology $\mathfrak{O}_{st}$ and a *refutable* ontology $\mathfrak{O}_{rt}$. When repairing the ontology, only the refutable part may be changed. For example, the static part of the ontology could be a carefully hand-crafted TBox whereas the refutable part is an ABox that is automatically generated from (possibly erroneous) data. It may also make sense to classify parts of a TBox as refutable, for example if the TBox is obtained as a combination of ontologies from different sources, some of which may be less trustworthy than others. In a privacy application, it may be the case that parts of the ontology are publicly known whereas other parts are hidden. In this setting, in order to hide critical information, it only makes sense to change the hidden part of the ontology.

Definition 4.1. *Let $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ be an ontology consisting of a static and a refutable part, and α an axiom such that $\mathfrak{O} \models \alpha$ and $\mathfrak{O}_{st} \not\models \alpha$. The ontology $\mathfrak{O}'$ is a* repair of $\mathfrak{O}$ w.r.t. α *if*

$$\mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}') \subseteq \mathsf{Con}(\mathfrak{O}) \setminus \{\alpha\}.$$

The repair $\mathfrak{O}'$ is an optimal repair *of $\mathfrak{O}$ w.r.t. α if there is no repair $\mathfrak{O}''$ of $\mathfrak{O}$ w.r.t. α with $\mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}') \subset \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}'')$. The repair $\mathfrak{O}'$ is a* classical repair *of $\mathfrak{O}$ w.r.t. α if $\mathfrak{O}' \subset \mathfrak{O}_{rt}$, and it is an* optimal classical repair *of $\mathfrak{O}$ w.r.t. α if there is no classical repair $\mathfrak{O}''$ of $\mathfrak{O}$ w.r.t. α such that $\mathfrak{O}' \subset \mathfrak{O}''$.* ♢

The condition $\mathfrak{O}_{st} \not\models \alpha$ ensures that $\mathfrak{O}$ does have a repair w.r.t. α since obviously the empty ontology $\emptyset$ is such a repair. In general, *optimal* repairs need not exist.

Proposition 4.2. *There is an $\mathcal{EL}$ ontology $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ and an $\mathcal{EL}$ axiom α such that $\mathfrak{O}$ does not have an optimal repair w.r.t. α.* ♢

Proof. We set $\alpha := A(a)$, $\mathfrak{O}_{st} := \mathcal{T}$, and $\mathfrak{O}_{rt} := \mathcal{A}$ where

$$\mathcal{T} := \{A \sqsubseteq \exists r.A, \exists r.A \sqsubseteq A\} \quad \text{and} \quad \mathcal{A} := \{A(a)\}.$$

To show that there is no optimal repair of $\mathfrak{O}$ w.r.t. α, we consider an arbitrary repair $\mathfrak{O}'$ and show that it cannot be optimal. Thus, let $\mathfrak{O}'$ be such that

$$\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}') \subseteq \mathsf{Con}(\mathfrak{O}) \setminus \{A(a)\}.$$

Without loss of generality we assume that $\mathfrak{O}'$ contains assertions only. In fact, if $\mathfrak{O}'$ contains a GCI that does not follow from $\mathcal{T}$, then $\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}') \not\subseteq \mathsf{Con}(\mathfrak{O})$. This is an easy consequence of the fact that, in $\mathcal{EL}$, a GCI follows from a TBox together with an ABox iff it follows from the TBox alone. It is also easy to see that $\mathfrak{O}'$ cannot contain role assertions since no such assertions are entailed by $\mathfrak{O}$. In addition, concept assertions following from $\mathcal{T} \cup \mathfrak{O}'$ must have a specific form.

Claim: If the assertion $C(a)$ is in $\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$, then C does not contain A.

Proof of claim. By induction on the role depth n of C.

Base case: If $n = 0$ and A is contained in C, then A is a conjunct of C and thus $C(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$ implies $A(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$, which is a contradiction.

Step case: If $n > 0$ and A occurs at role depth n in C, then $C(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$ implies that there are roles $r_1, \ldots, r_n$ such that $(\exists r_1. \cdots \exists r_n.A)(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$. Since $\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}') \subseteq \mathsf{Con}(\mathfrak{O})$, this can only be the case if $r_1 = \ldots = r_n = r$ since $\mathfrak{O}$ clearly has models in which all roles different from r are empty. Since $\mathcal{T}$ contains the GCI $\exists r.A \sqsubseteq A$ and $r_n = r$, $(\exists r_1. \cdots \exists r_n.A)(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$ implies $(\exists r_1. \cdots \exists r_{n-1}.A)(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$. Induction now yields that this is not possible, which completes the proof of the claim.

Furthermore, as argued in the proof of the claim, any assertion belonging to $\mathsf{Con}(\mathfrak{O})$ cannot contain roles other than r. The same is true for concept names different from A. Consequently, all assertions $C(a) \in \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$ are such that C is built using r and $\top$ only. Any such concept C is equivalent to a concept of the form $(\exists r.)^n \top$.

Since $\mathfrak{O}'$ is finite, there is a maximal n_0 such that $((\exists r.)^{n_0} \top)(a) \in \mathfrak{O}'$, but $((\exists r.)^n \top)(a) \notin \mathfrak{O}'$ for all $n > n_0$. Since $(\exists r.)^n \top \sqsubseteq (\exists r.)^m \top$ if $m \leq n$, we can assume without loss of generality

that $\mathfrak{O}' = \{((\exists r.)^{n_0}\top)(a)\}$. We claim that $((\exists r.)^n\top)(a) \notin \mathsf{Con}(\mathcal{T} \cup \mathfrak{O}')$ if $n > n_0$. To this purpose, we construct a model $\mathcal{I}$ of $\mathcal{T} \cup \mathfrak{O}'$ such that $a^{\mathcal{I}} \notin ((\exists r.)^n\top)^{\mathcal{I}}$. This model is defined as follows:

$$\begin{array}{lcl} \Delta^{\mathcal{I}} & = & \{d_0, d_1, \ldots, d_{n_0}\}, \\ r^{\mathcal{I}} & = & \{(d_{i-1}, d_i) \mid 1 \leq i \leq n_0\}, \\ A^{\mathcal{I}} & = & \emptyset, \\ a^{\mathcal{I}} & = & d_0. \end{array}$$

Clearly, $\mathcal{I}$ is a model of $\mathfrak{O}'$, and it does not satisfy $((\exists r.)^n\top)(a)$ if $n > n_0$. In addition, it is a model of $\mathcal{T}$ since $A^{\mathcal{I}} = (\exists r.A)^{\mathcal{I}} = \emptyset$.

Consequently, if we choose n such that $n > n_0$ and define $\mathfrak{O}'' := \{((\exists r.)^n\top)(a)\}$, then $\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}') \subset Con(\mathcal{T} \cup \mathfrak{O}'')$. In addition, $\mathsf{Con}(\mathcal{T} \cup \mathfrak{O}'') \subseteq \mathsf{Con}(\mathfrak{O}) \setminus \{A(a)\}$, i.e., $\mathfrak{O}''$ is a repair. This shows that $\mathfrak{O}'$ is not optimal. Since we have chosen $\mathfrak{O}'$ to be an arbitrary repair, this shows that there cannot be an optimal repair. □

In contrast, optimal *classical* repairs always exist. One approach for computing such a repair uses justifications and hitting sets [Rei87].

Definition 4.3 (Justifications and Hitting Sets). *Let $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ be an ontology and α an axiom such that $\mathfrak{O} \models \alpha$ and $\mathfrak{O}_{st} \not\models \alpha$. A* justification *for α in $\mathfrak{O}$ is a minimal subset J of $\mathfrak{O}_{rt}$ such that $\mathfrak{O}_{st} \cup J \models \alpha$. Given justifications $J_1, \ldots, J_k$ for α in $\mathfrak{O}$, a* hitting set *of these justifications is a set H of axioms such that $H \cap J_i \neq \emptyset$ for $i = 1, \ldots, k$. This hitting set is* minimal *if there is no other hitting set strictly contained in it.* ♢

Note that the condition $\mathfrak{O}_{st} \not\models \alpha$ implies that justifications are non-empty. Consequently, hitting sets and thus minimal hitting sets always exist.

The *algorithm for computing an optimal classical repair* of $\mathfrak{O}$ w.r.t. α proceeds in two steps: (i) compute all justifications $J_1, \ldots, J_k$ for α in $\mathfrak{O}$; and then (ii) compute a minimal hitting set H of $J_1, \ldots, J_k$ and remove the elements of H from $\mathfrak{O}_{rt}$, i.e., output $\mathfrak{O}' = \mathfrak{O}_{rt} \setminus H$.

It is not hard to see that, independently of the choice of the hitting set, this algorithm produces an optimal classical repair. Conversely, all optimal classical repairs can be generated this way by going through all hitting sets.

4.2 Gentle Repairs

Instead of removing axioms completely, as in the case of a classical repair, a gentle repair replaces them by weaker axioms.

Definition 4.4. *Let β, γ be two axioms. We say that γ is* weaker than *β if* $\mathsf{Con}(\{\gamma\}) \subset \mathsf{Con}(\{\beta\})$. ♢

Alternatively, we could have introduced *weaker w.r.t the strict part of the ontology*, by requiring $\mathsf{Con}(\mathfrak{O}_{st} \cup \{\gamma\}) \subset \mathsf{Con}(\mathfrak{O}_{st} \cup \{\beta\})$.[1] In this paper, we will not consider this alternative definition, although most of the results in this section would also hold w.r.t. it (e.g., Theorem 4.6). The difference between the two definitions is, however, relevant in the next section, where we consider concrete approaches for how to weaken axioms. In the case where the whole ontology is refutable, there is of course no difference between the two definitions.

[1] Defining weaker w.r.t the whole ontology $\mathfrak{O}$ does not make sense since this ontology is possibly erroneous.

Obviously, the *weaker-than* relation from Definition 4.4 is transitive, i.e., if α is weaker than β and β is weaker than γ, then α is also weaker than γ. In addition, a tautology is always weaker than a non-tautology. Replacing an axiom by a tautology is obviously the same as removing this axiom. We assume in the following that there exist tautological axioms, which is obviously true for Description Logics such as $\mathcal{EL}$.

Gentle repair algorithm: we still compute all justifications $J_1, \ldots, J_k$ for α in $\mathfrak{O}$ and a minimal hitting set H of $J_1, \ldots, J_k$. But instead of removing the elements of H from $\mathfrak{O}_{rt}$, we replace them by weaker axioms. To be more precise, if $\beta \in H$ and $J_{i_1}, \ldots, J_{i_\ell}$ are all the justifications containing β, then replace β by a weaker axiom γ such that

$$\mathfrak{O}_{st} \cup (J_{i_j} \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha \text{ for } j = 1, \ldots, \ell. \tag{4.1}$$

Note that such a weaker axiom γ always exists. In fact, we can choose a tautology as the axiom γ. If γ is a tautology, then replacing β by γ is the same as removing β. Thus, we have $\mathfrak{O}_{st} \cup (J_{i_j} \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$ due to the minimality of J_{i_j}. In addition, minimality of J_{i_j} also implies that β is not a tautology since otherwise $\mathfrak{O}_{st} \cup (J_{i_j} \setminus \{\beta\})$ would also have the consequence α. In general, different choices of γ yield different runs of the algorithm.

In principle, the algorithm could always use a tautology γ, but then this run would produce a classical repair. To obtain more gentle repairs, the algorithm needs to use a strategy that chooses stronger axioms (i.e., axioms γ that are less weak than tautologies) if possible. In contrast to what is claimed in the literature (e.g. [LSP+08]), this approach does not necessarily yield a repair.

Lemma 4.5. *Let $\mathfrak{O}'$ be the ontology obtained from $\mathfrak{O}_{rt}$ by replacing all the elements of the hitting set by weaker ones such that the condition (4.1) is satisfied. Then $\mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}') \subseteq \mathsf{Con}(\mathfrak{O})$, but in general we may still have $\alpha \in \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}')$.*

Proof. The definition of "weaker than" (see Definition 4.4) implies that $\mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}') \subseteq \mathsf{Con}(\mathfrak{O})$. We now give an example where this approach nevertheless does not produce a repair. Let $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ where $\mathfrak{O}_{st} = \emptyset$ and $\mathfrak{O}_{rt} = \mathcal{T} \cup \mathcal{A}$ with $\mathcal{T} = \{B \sqsubseteq A\}$ and $\mathcal{A} = \{(A \sqcap B)(a)\}$, and α be the consequence $A(a)$. Then α has a single justification $J = \{(A \sqcap B)(a)\}$, and thus $H = \{\beta = (A \sqcap B)(a)\}$ is the only hitting set. The assertion $\gamma = B(a)$ is weaker than β and it satisfies $(J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$. However, if we define $\mathfrak{O}' = (\mathfrak{O} \setminus \{\beta\}) \cup \{\gamma\}$, then $\mathfrak{O}' \models \alpha$ still holds. □

A similar example that uses only GCIs is the following, where now we consider a refutable ontology $\mathfrak{O} = \mathfrak{O}_{rt} = \{C \sqsubseteq A \sqcap B, B \sqsubseteq A\}$ and we assume that α is the consequence $C \sqsubseteq A$. Then α has a single justification $J = \{C \sqsubseteq A \sqcap B\}$ and thus $H = \{\beta = C \sqsubseteq A \sqcap B\}$ is the only hitting set. The GCI $\gamma = C \sqsubseteq B$ is a weaker than β and it satisfies $(J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$. However, if we define $\mathfrak{O}' = (\mathfrak{O} \setminus \{\beta\}) \cup \{\gamma\}$, then $\mathfrak{O}' \models \alpha$.

These examples show that applying the gentle repair approach only once may not lead to a repair. For this reason, we need to *iterate this approach*, i.e., if the resulting ontology $\mathfrak{O}_{st} \cup \mathfrak{O}'$ still has α as a consequence, we again compute all justifications and a hitting set for them, and then replace the elements of the hitting set with weaker axioms as described above. This is iterated until a repair is reached. We can show that this iteration indeed always terminates after finitely many steps with a repair.

Theorem 4.6. *Let $\mathfrak{O}^{(0)} = \mathfrak{O}_{st}^{(0)} \cup \mathfrak{O}_{rt}^{(0)}$ be a finite ontology and α an axiom such that $\mathfrak{O}^{(0)} \models \alpha$ and $\mathfrak{O}_{st}^{(0)} \not\models \alpha$. Applied to $\mathfrak{O}^{(0)}$ and α, the iterative algorithm described above stops after a finite number of iterations that is at most exponential in the cardinality of $\mathfrak{O}_{rt}^{(0)}$, and yields as output an ontology that is a repair of $\mathfrak{O}_{st}^{(0)}$ w.r.t. the consequence α.*

Proof. Assume that $\mathfrak{O}_{rt}^{(0)}$ contains n axioms, and that there is an infinite run R of the algorithm on input $\mathfrak{O}^{(0)}$ and α. Take a bijection ℓ_0 between $\mathfrak{O}_{rt}^{(0)}$ and $\{1,\dots,n\}$ that assigns unique labels to axioms. Whenever we weaken an axiom during a step of the run, the new weaker axiom inherits the label of the original axiom. Thus, we have bijections $\ell_i : \mathfrak{O}_{rt}^{(i)} \to \{1,\dots,n\}$ for all ontologies $\mathfrak{O}_{rt}^{(i)}$ considered during the run R of the algorithm. For $i \geq 0$ we define

$$S_i := \{K \subseteq \{1,\dots,n\} \mid \mathfrak{O}_{st} \cup \{\beta \in \mathfrak{O}_{rt}^{(i)} \mid \ell_i(\beta) \in K\} \models \alpha\},$$

i.e., S_i contains all sets of indices such that the corresponding subset of $\mathfrak{O}_{rt}^{(i)}$ together with $\mathfrak{O}_{st}$ has the consequence α.

We claim that $S_{i+1} \subset S_i$. Note that $S_{i+1} \subseteq S_i$ is an immediate consequence of the fact that $\ell_i(\gamma) = j = \ell_{i+1}(\gamma')$ implies that $\gamma = \gamma'$ or γ' is weaker than γ. Thus, it remains to show that the inclusion is strict. This follows from the following observations. Since the algorithm does not terminate with the ontology $\mathfrak{O}_{rt}^{(i)}$, we still have $\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(i)} \models \alpha$, and thus there is at least one justification $\emptyset \subset J \subseteq \mathfrak{O}_{rt}^{(i)}$. Consequently, the hitting set H used in this step of the algorithm contains an element β of $\mathfrak{O}_{rt}^{(i)}$. When going from $\mathfrak{O}_{rt}^{(i)}$ to $\mathfrak{O}_{rt}^{(i+1)}$, β is replaced by a weaker axiom β' such that $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\beta'\} \not\models \alpha$. But then the set $\{\ell(\gamma) \mid \gamma \in J\}$ belongs to S_i, but not to S_{i+1}.

Since S_0 contains only exponentially many sets, the strict inclusion $S_{i+1} \subset S_i$ can happen only exponentially often, which contradicts our assumption that there is an infinite run R of the algorithm on input $\mathfrak{O}^{(0)}$ and α. This shows termination after exponentially many steps. However, if the algorithm terminates with output $\mathfrak{O}_{rt}^{(i)}$, then $\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(i)} \not\models \alpha$. In fact, otherwise, there would be a possibility to weaken $\mathfrak{O}_{rt}^{(i)}$ into $\mathfrak{O}_{rt}^{(i+1)}$ since it would always be possible to replace the elements of a hitting set by tautologies, i.e., perform a classical repair. □

When computing a classical repair, considering all justifications and then removing a minimal hitting set of these justifications guarantees that one immediately obtains a repair. We have seen in the proof of Lemma 4.5 that with our gentle repair approach this need not be the case. Nevertheless, we were able to show that, after a finite number of iterations of the approach, we obtain a repair. The proof of termination actually shows that for this it is sufficient to weaken only one axiom of one justification such that the resulting set is no longer a justification. This motivates the following modification of our approach.

Modified gentle repair algorithm: compute one justification J for α in $\mathfrak{O}$ and choose an axiom $\beta \in J$. Replace β by a weaker axiom γ such that

$$\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha. \tag{4.2}$$

Clearly, one needs to iterate this approach, but it is easy to see that the termination argument used in the proof of Proposition 4.6 also applies here.

Corollary 4.7. *Let $\mathfrak{O}^{(0)} = \mathfrak{O}^{(0)}_{st} \cup \mathfrak{O}^{(0)}_{rt}$ be a finite ontology and α an axiom such that $\mathfrak{O}^{(0)} \models \alpha$ and $\mathfrak{O}^{(0)}_{st} \not\models \alpha$. Applied to $\mathfrak{O}^{(0)}$ and α, the modified iterative algorithm stops after a finite number of iterations that is at most exponential in the cardinality of $\mathfrak{O}^{(0)}_{rt}$, and yields as output an ontology $\widehat{\mathfrak{O}}_s$ that is a repair of $\mathfrak{O}^{(0)}_{st}$ w.r.t. α.*

An important advantage of this modified approach is that the complexity of a single iteration step may decrease considerably. For example, for the DL $\mathcal{EL}$, a single justification can be computed in polynomial time, while computing all justifications may take exponential time [BPS07]. In addition, to compute a minimal hitting set one needs to solve an NP-complete problem [GJ90] whereas choosing one axiom from a single justification is easy. However, as usual, there is no free lunch: we can show that the modified gentle repair algorithm may indeed need exponentially many iteration steps.[2]

Proposition 4.8. *There is a sequence of $\mathcal{EL}$ ontologies $\mathfrak{O}^{(n)} = \mathfrak{O}^{(n)}_{st} \cup \mathfrak{O}^{(n)}_{rt}$ with $\mathfrak{O}^{(n)}_{st} = \emptyset$ and an $\mathcal{EL}$ axiom α such that the modified gentle repair algorithm applied to $\mathfrak{O}^{(n)}$ and α has a run with exponentially many iterations in the size of $\mathfrak{O}^{(n)}$.* ♢

Proof. For $n \geq 1$, consider the set of concept names $I^{(n)} = \{P_i, Q_i \mid 1 \leq i \leq n\}$, and define $\mathfrak{O}^{(n)} := \mathfrak{O}^{(n)}_{rt} := \mathcal{T}^{(n)}_1 \cup \mathcal{T}^{(n)}_2$, where

$$\begin{aligned}
\mathcal{T}^{(n)}_1 := {} & \{A \sqsubseteq \exists r.\textstyle\bigsqcap I^{(n)},\ \exists r.(P_n \sqcap Q_n) \sqsubseteq B\} \cup \\
& \{P_i \sqcap Q_i \sqsubseteq P_{i+1},\ P_i \sqcap Q_i \sqsubseteq Q_{i+1} \mid 1 \leq i < n\}, \\
\mathcal{T}^{(n)}_2 := {} & \{\exists r.(X \sqcap Y) \sqsubseteq D_{XY},\ D_{XY} \sqcap X \sqsubseteq Y \mid \\
& X \in \{P_i, Q_i\}, Y \in \{P_{i+1}, Q_{i+1}\}, 1 \leq i < n\} \cup \\
& \{\exists r.P_1 \sqsubseteq P_1,\ \exists r.Q_1 \sqsubseteq Q_1,\ P_n \sqsubseteq B,\ Q_n \sqsubseteq B\}.
\end{aligned}$$

It is easy to see that the size of $\mathfrak{O}^{(n)}$ is polynomial in n and that $\mathfrak{O}^{(n)} \models A \sqsubseteq B$. Suppose that we want to get rid of this consequence using the modified gentle repair approach. First, we can find the justification

$$\{A \sqsubseteq \exists r.\textstyle\bigsqcap I^{(n)},\ \exists r.(P_n \sqcap Q_n) \sqsubseteq B\}.$$

We repair it by weakening the first axiom to

$$\gamma := A \sqsubseteq \exists r.\textstyle\bigsqcap (I^{(n)} \setminus \{P_n\}) \ \sqcap\ \exists r.\bigsqcap (I^{(n)} \setminus \{Q_n\}).$$

At this point, we can find a justification that uses γ and $P_{n-1} \sqcap Q_{n-1} \sqsubseteq P_n$. We further weaken γ to

$$\begin{aligned}
A \sqsubseteq \ & \exists r.\textstyle\bigsqcap (I^{(n)} \setminus \{P_n, P_{n-1}\}) \sqcap \\
& \exists r.\textstyle\bigsqcap (I^{(n)} \setminus \{P_n, Q_{n-1}\}) \sqcap \exists r.\bigsqcap (I^{(n)} \setminus \{Q_n\}).
\end{aligned}$$

Repeating this approach, after $2n$ weakenings we have only changed the first axiom, weakening it to the axiom

$$A \sqsubseteq \bigsqcap_{X_i \in \{P_i, Q_i\}, 1 \leq i \leq n} \exists r.(X_1 \sqcap \cdots \sqcap X_n), \tag{4.3}$$

[2] It is not clear yet whether this is also the case for the unmodified gentle repair algorithm.

whose right-hand side is a conjunction with 2^n conjuncts, each of them representing a possible choice of P_i or Q_i at every location $i, 1 \leq i \leq n$.

So far, we have just considered axioms from $\mathcal{T}_1^{(n)}$. Taking also axioms from $\mathcal{T}_2^{(n)}$ into account, we obtain for every conjunct $\exists r.(X_1 \sqcap \cdots \sqcap X_n)$ in axiom (4.3) a justification for $A \sqsubseteq B$ that consists of (4.3) and the axioms

$$\{\exists r.X_1 \sqsubseteq X_1,\ X_n \sqsubseteq B\} \cup$$
$$\{\exists r.(X_i \sqcap X_{i+1}) \sqsubseteq D_{X_iX_{i+1}},\ D_{X_iX_{i+1}} \sqcap X_i \sqsubseteq X_{i+1} \mid 1 \leq i < n\}.$$

This justification can be removed by weakening (4.3) further by deleting one concept name appearing in the conjunct. The justifications for other conjuncts are not influenced by this modification. Thus, we can repeat this for each of the exponentially many conjuncts, which shows that overall we have exponentially many iterations of the modified gentle repair algorithm in this run. □

4.3 Weakening Relations

In order to obtain better bounds on the number of iterations of our algorithms, we restrict the way in which axioms can be weakened. Before introducing concrete approaches for how to do this for $\mathcal{EL}$ axioms in the next section, we investigate such restricted weakening relations in a more abstract setting.

Definition 4.9. *Given a pre-order $\succ$ (i.e., an irreflexive and transitive binary relation) on axioms, we say that it*

- *is a* weakening relation *if $\beta \succ \gamma$ implies that $\mathsf{Con}(\{\gamma\}) \subset \mathsf{Con}(\{\beta\})$;*
- *is* bounded (linear, polynomial) *if, for every axiom α, the length of the longest chain $\succ -$ generated from β is linearly (polynomially) bounded by the size of β.*
- *is* complete *if, for any axiom β that is not a tautology, there is a tautology γ such that $\beta \succ \gamma$.* ♢

If we use a linear (polynomial) and complete weakening relation, then termination with a repair is guaranteed after a linear (polynomial) number of iterations.

Proposition 4.10. *Let $\succ$ be a linear (polynomial) and complete weakening relation. If in the above (modified) gentle repair algorithm we have $\beta \succ \gamma$ whenever β is replaced by γ, then the algorithm stops after a linear (polynomial) number of iterations and yields as output an ontology that is a repair of $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$ w.r.t. the consequence α.* ♢

Proof. For every axiom β in $\mathfrak{O}_{rt}$ we consider the length of the longest $\succ$-chain issuing from it, and then sum up these numbers over all axioms in $\mathfrak{O}_{rt}$. The resulting number is linearly (polynomially) bounded by the size of the ontology (assuming that this size is given as sum of the sizes of all its axioms). Let us call this number the chain-size of the ontology. Obviously, if β is replaced by β' with $\beta \succ \beta'$, then the length of the longest $\succ$-chain issuing from β' is smaller than the length of the longest $\succ$-chain issuing from β. Consequently, if $\mathfrak{O}_{rt}^{(i+1)}$ is obtained from $\mathfrak{O}_{rt}^{(i)}$ in the i-th iteration of the algorithm, then the chain-size of $\mathfrak{O}_{rt}^{(i)}$

is strictly larger than the chain-size of $\mathfrak{O}_{rt}^{(i+1)}$. This implies that there can be only linearly (polynomially) many iterations.

Consider a terminating run of the algorithm that has produced the sequence of ontologies $\mathfrak{O}_{rt} = \mathfrak{O}_{rt}^{(0)}, \mathfrak{O}_{rt}^{(1)}, \ldots, \mathfrak{O}_{rt}^{(n)}$. Then we have

$$\mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}) \supseteq \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(1)}) \supseteq \ldots \supseteq \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(n)})$$

since $\succ$ is a weakening relation. If the algorithm has terminated due to the fact that $\alpha \notin \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(n)})$, then $\mathfrak{O}_{rt}^{(n)}$ is a repair of $\mathfrak{O}$ w.r.t. α. Otherwise, the only reason for termination could be that, although $\alpha \in \mathsf{Con}(\mathfrak{O}_{st} \cup \mathfrak{O}_{rt}^{(n)})$, the algorithm cannot generate a new ontology $\mathfrak{O}_{rt}^{(n+1)}$. In the unmodified gentle repair approach this means that there is an axiom β in the hitting set H such that there is no axiom γ with $\beta \succ \gamma$ such that (4.1) is satisfied. However, using a tautology as the axiom γ actually allows us to satisfy the condition (4.1). Thus, completeness of $\succ$ implies that this reason for termination without success cannot occur. An analogous argument can be used for the modified gentle repair approach.□

When describing our (modified) gentle repair algorithm, we have said that the chosen axiom β needs to be replaced by a weaker axiom γ such that (4.1) or (4.2) holds. But we have not said how such an axiom γ can be found. This of course depends on which ontology language and which weakening relation is used. In the abstract setting of this section, we assume that an "oracle" provides us with a weaker axiom.

Definition 4.11. *Let $\succ$ be a weakening relation. An* oracle for $\succ$ *is a computable function W that, given an axiom β that is not $\succ$-minimal, provides us with an axiom $W(\beta)$ such that $\beta \succ W(\beta)$. For $\succ$-minimal axioms β we assume that $W(\beta) = \beta$.* ♢

If the weakening relation is complete and *well-founded* (i.e., there are no infinite descending $\succ$-chains $\beta_1 \succ \beta_2 \succ \beta_2 \succ \cdots$), we can effectively find an axiom γ such that (4.1) or (4.2) holds. We show this formally only for (4.2), but condition (4.1) can be treated similarly.

Lemma 4.12. *Assume that J is a justification for the consequence α, and $\beta \in J$. If $\succ$ is a well-founded and complete weakening relation and W is an oracle for $\succ$, then there is an $n \geq 1$ such that (4.2) holds for $\gamma = W^n(\beta)$. If $\succ$ is additionally linear (polynomial), then n is linear (polynomial) in the size of β.*

Proof. Well-foundedness implies that the $\succ$-chain $\beta \succ W(\beta) \succ W(W(\beta)) \succ \ldots$ is finite, and thus there is an n such that $W^{n+1}(\beta) = W^n(\beta)$, i.e., $W^n(\beta)$ is $\succ$-minimal. Since $\succ$ is complete, this implies that $W^n(\beta)$ is a tautology. Minimality of the justification J then yields $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{W^n(\beta)\} \not\models \alpha$. Linearity (polynomiality) of $\succ$ ensures that the length of the $\succ$-chain $\beta \succ W(\beta) \succ W(W(\beta)) \succ \ldots$ is linearly (polynomially) bounded by the size of β.□

Thus, to find an axiom γ satisfying (4.1) or (4.2), we iteratively apply W to β until an axiom satisfying the required property is found. The proof of Lemma 4.12 shows that at the latest this is the case when a tautology is reached, but of course the property may already be satisfied before that by a non-tautological axiom $W^i(\beta)$.

In order to weaken axioms as gently as possible, W should realize small weakening steps. The smallest such step is one where there is no step in between.

Definition 4.13. *Let $\succ$ be a pre-order. The* one-step relation[3] *induced by $\succ$ is defined as*

$$\succ_1 := \{(\beta, \gamma) \in \succ \mid \text{ there is no } \delta \text{ such that } \beta \succ \delta \succ \gamma\}.$$

We say that $\succ_1$ covers *$\succ$ if its transitive closure is again $\succ$, i.e., $\succ_1^+ = \succ$. In this case we also say that $\succ$ is* one-step generated. ♢

If $\succ$ is one-step generated, then every weaker element can be reached by a finite sequence of one-step weakenings, i.e., if $\beta \succ \gamma$, then there are finitely many elements $\delta_0, \ldots, \delta_n$ ($n \geq 1$) such that $\beta = \delta_0 \succ_1 \delta_1 \succ_1 \ldots \succ_1 \delta_n = \gamma$. This leads us to the following characterization of pre-orders that are *not* one-step generated.

Lemma 4.14. *The pre-order $\succ$ is* not *one-step generated iff there exist two comparable elements $\beta \succ \gamma$ such that every finite chain $\beta = \delta_0 \succ \delta_1 \succ \ldots \succ \delta_n = \gamma$ can be* refined *in the sense that there is an $i, 0 \leq i < n$, and an element δ such that $\delta_i \succ \delta \succ \delta_{i+1}$.*

If $\beta \succ \gamma$ are such that any finite chain between them can be refined, then obviously there cannot be an upper bound on the length of the chains issuing from β. Thus, Lemma 4.14 implies the following result.

Proposition 4.15. *If $\succ$ is bounded, then it is one-step generated.* ♢

The following example shows that well-founded pre-orders need not be one-step generated.

Example 4.16. *Consider the pre-order $\succ$ on the set*

$$P := \{\beta\} \cup \{\delta_i \mid i \geq 0\},$$

where $\beta \succ \delta_i$ for all $i \geq 0$, and $\delta_i \succ \delta_j$ iff $i > j$. It is easy to see that $\succ$ is well-founded and that $\succ_1 = \{(\delta_{i+1}, \delta_i) \mid i \geq 0\}$. Consequently, ${\succ_1}^+$ contains none of the tuples (β, δ_i) for $i \geq 0$, which shows that $\succ_1$ does not cover $\succ$. In particular, any finite chain between β and δ_i can be refined. Interestingly, if we add elements γ_i ($i \geq 0$) with $\beta \succ \gamma_i \succ \delta_i$ to this pre-order, then it becomes one-step generated. ♢

One-step generated weakening relations allow us to find maximally strong weakenings satisfying (4.1) or (4.2). Again, we consider only condition (4.2), but all definitions and results can be adapted to deal with (4.1) as well.

Definition 4.17. *Let J be a justification for the consequence α, and $\beta \in J$. We say that γ is a* maximally strong weakening *of β in J if $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$, but $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\delta\} \models \alpha$ for all δ with $\beta \succ \delta \succ \gamma$.* ♢

In general, maximally strong weakenings need not exist. As an example, assume that the pre-order introduced in Example 4.16 (without the added axioms γ_i) is a weakening relation on axioms, and assume that $J = \{\beta\}$ and that none of the axioms δ_i have the consequence. Obviously, in this situation there is no maximally strong weakening of α in J.

Next, we introduce conditions under which maximally strong weakenings always exist, and can also be computed. We say that the one-step generated weakening relation $\succ$ is *effectively finitely branching* if for every axiom β the set $\{\gamma \mid \beta \succ_1 \gamma\}$ is finite and can effectively be computed.

[3] This is sometimes also called the transitive reduction of $\succ$.

Proposition 4.18. *Let $\succ$ be a well-founded, one-step generated, and effectively finitely branching weakening relation and assume that the consequence relation $\models$ is decidable. Then all maximally strong weakenings of an axiom in a justification can effectively be computed.* ♢

Proof. Let J be a justification for the consequence α, and $\beta \in J$. Since $\succ$ is well-founded, one-step generated, and finitely branching, König's Lemma implies that there are only finitely many γ such that $\beta \succ \gamma$, and all these γ can be reached by following $\succ_1$. Thus, by a breadth-first search, we can compute the set of all γ such that there is a path $\beta \succ_1 \delta_1 \succ_1 \ldots \succ_1 \delta_n \succ_1 \gamma$ with $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$, but $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\delta_i\} \models \alpha$ for all $i, 1 \leq i \leq n$. If this set still contains elements that are comparable w.r.t. $\succ$ (i.e., there is a $\succ_1$-path between them), then we remove the weaker elements. It is easy to see that the remaining set consists of all maximally strong weakenings of β in J. □

Note that the additional removal of weaker elements in the above proof is really necessary. In fact, assume that $\beta \succ_1 \delta_1 \succ_1 \gamma$ and $\beta \succ_1 \delta_2 \succ_1 \gamma$, and that $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\gamma\} \not\models \alpha$, $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\delta_1\} \models \alpha$, but $\mathfrak{O}_{st} \cup (J \setminus \{\beta\}) \cup \{\delta_2\} \not\models \alpha$. Then both δ_2 and γ belong to the set computed in the breadth-first search, but only δ_2 is a maximally strong weakening (see Example 4.27, where it is shown that this situation can really occur when repairing $\mathcal{EL}$ ontologies). In particular, this also means that iterated application of a one-step oracle, i.e., an oracle W satisfying $\beta \succ_1 W(\beta)$, does not necessarily yield a maximally strong weakening.

4.4 Weakening Relations for $\mathcal{EL}$ Axioms

In this section, we restrict the attention to ontologies written in $\mathcal{EL}$, but some of our approaches and results could also be transferred to other DLs. We start with observing that weakening relations for $\mathcal{EL}$ axioms need not be one-step generated.

Proposition 4.19. *If we define $\beta \succ^g \gamma$ if $\mathsf{Con}(\gamma) \subset \mathsf{Con}(\beta)$, then $\succ^g$ is a weakening relation on $\mathcal{EL}$ axioms that is not one-step generated.*

Proof. It is obvious that $\succ^g$ is a weakening relation.[4] To see that it is not one-step generated, consider a GCI β that is not a tautology and an arbitrary tautology γ. Then we have $\beta \succ \gamma$. Let $\beta = \delta_0 \succ^g \delta_1 \succ^g \ldots \succ^g \delta_n = \gamma$ be a finite chain leading from β to γ. Then δ_{n-1} must be a GCI that is not a tautology. Assume that $\delta_{n-1} = C \sqsubseteq D$. Then $\delta := \exists r.C \sqsubseteq \exists r.D$ satisfies $\delta_{n-1} \succ^g \delta \succ^g \gamma$. By Lemma 4.14, this shows that $\succ$ is not one-step generated. □

Our main idea for obtaining more well-behaved weakening relations is to weaken a GCI $C \sqsubseteq D$ by generalizing the right-hand side D and/or by specializing the left-hand side C. Similarly, a concept assertion $D(a)$ can be weakened by generalizing D. For role assertions we can use as weakening an arbitrary tautological axiom, but will no longer consider them explicitly in the following.

Proposition 4.20. *If we define*

$$\begin{array}{rll} C \sqsubseteq D \succ^s C' \sqsubseteq D' & \text{if} & C' \sqsubseteq C,\ D \sqsubseteq D' \text{ and } \{C' \sqsubseteq D'\} \not\models C \sqsubseteq D, \\ D(a) \succ^s D'(a) & \text{if} & D \sqsubset D', \end{array}$$

[4] In fact, it is the greatest one w.r.t. set inclusion.

then $\succ^s$ is a complete weakening relation. ♢

Proof. To prove that $\succ^s$ is a weakening relation we must show that $\beta \succ^s \gamma$ consequently implies $\mathsf{Con}(\{\gamma\}) \subset \mathsf{Con}(\{\beta\})$. If $C' \sqsubseteq C$ and $D \sqsubseteq D'$ hold, then it follows that

$$\mathsf{Con}(\{C' \sqsubseteq D'\}) \subseteq \mathsf{Con}(\{C \sqsubseteq D\}) \text{ and } \mathsf{Con}(\{a : D'\}) \subseteq \mathsf{Con}(\{a : D\}).$$

The second inclusion is strict iff $D \sqsubset D'$. For the first inclusion to be strict, $C' \sqsubset C$ or $D \sqsubset D'$ is a necessary condition, but it is not sufficient. This is why we explicitly require $\{C' \sqsubseteq D'\} \not\models C \sqsubseteq D$, which yields strictness of the inclusion. Completeness is trivial due to the availability of all tautologies of the form $C \sqsubseteq \top$ and $\top(a)$. □

To see why, e.g., $D \sqsubset D'$ does not imply $\mathsf{Con}(\{C \sqsubseteq D'\}) \subset \mathsf{Con}(\{C \sqsubseteq D\})$, notice that $A \sqcap \exists r.A \sqsubset \exists r.A$, but $\mathsf{Con}(\{A \sqsubseteq \exists r.A\}) = \mathsf{Con}(\{A \sqsubseteq A \sqcap \exists r.A\})$. Unfortunately, the weakening relation $\succ^s$ introduced in Proposition 4.20 is *not well-founded* since left-hand sides can be specialized infinitely. For example, we have

$$\top \sqsubseteq A \succ^s \exists r.\top \sqsubseteq A \succ^s \exists r.\exists r.\top \sqsubseteq A \succ^s \cdots.$$

To avoid this problem, we now restrict the attention to sub-relations of $\succ^s$ that only generalize the right-hand sides of GCIs. We will not consider concept assertions, but they can be treated similarly.

4.4.1 Generalizing the Right-Hand Sides of GCIs

We define

$$C \sqsubseteq D \succ^{sub} C' \sqsubseteq D' \text{ if } C' = C \text{ and } C \sqsubseteq D \succ^s C' \sqsubseteq D'.$$

Theorem 4.21. *The relation $\succ^{sub}$ on $\mathcal{EL}$ axiom is a well-founded, complete, and one-step generated weakening relation, but it is not polynomial.*

Proof. Proposition 4.20 implies that $\succ^{sub}$ is a weakening relation and completeness follows from the fact that $C \sqsubseteq D \succ^{sub} C \sqsubseteq \top$ whenever $C \sqsubseteq D$ is not a tautology. In $\mathcal{EL}$, the inverse subsumption relation is well-founded, i.e., there cannot be an infinite sequence $C_0 \sqsubset C_1 \sqsubset C_2 \sqsubset \ldots$ of $\mathcal{EL}$ concepts. Looking at the proof of this result given in [BM10], one sees that it actually shows that $\sqsubset$ is bounded. Obviously, this implies that $\succ^{sub}$ is bounded as well, and thus one-step generated by Proposition 4.15.

It remains to show that $\succ^{sub}$ is not polynomial. Let $n \geq 1$ and $N_n := \{A_1, \ldots, A_{2n}\}$ be a set of $2n$ distinct concept names. Then we have

$$\exists r.\bigsqcap N_n \sqsubset \bigsqcap_{X \subseteq N_n \wedge |X| = n} \exists r.\bigsqcap X.$$

Note that the size of $\exists r.\bigsqcap N_n$ is linear in n, but that the conjunction on the right-hand side of this strict subsumption consists of exponentially many concepts $\exists r.\bigsqcap X$ that are incomparable w.r.t. subsumption. Consequently, by removing one conjunct at a time, we can generate an ascending chain w.r.t. $\sqsubset$ of $\mathcal{EL}$ concepts whose length is exponential in n. Using these concepts as right-hand sides of GCIs with left-hand side B for a concept name $B \notin N_n$, we obtain an exponentially long descending chain w.r.t. $\succ^{sub}$. □

To be able to apply Proposition 4.18, it remains to show that $\succ^{sub}$ is effectively finitely branching. For this purpose, we first investigate the one-step relation $\sqsubset_1$ induced by $\sqsubset$. Given an $\mathcal{EL}$ concept C, we want to characterize the set of its *upper neighbors*

$$Upper(C) := \{D \mid C \sqsubset_1 D\},$$

and show that it can be computed in polynomial time. To show this, we first assume that the given $\mathcal{EL}$ concepts needs to be reduced that can be done in polynomial time.

Definition 4.22. *Given a reduced $\mathcal{EL}$ concept C, we define the set $U(C)$ by induction on the role depths of C. More precisely, $U(C)$ consists of the concepts D that can be obtained from C as follows:*

- *Remove a concept name A from the top-level conjunction of C.*
- *Remove an existential restriction $\exists r.E$ from the top-level conjunction of C, and replace it by the conjunction of all existential restrictions $\exists r.F$ for $F \in U(E)$.* ♢

For example, if $C = A \sqcap \exists r.(B_1 \sqcap B_2 \sqcap B_3)$, then $U(C)$ consists of the two concepts, which are $\exists r.(B_1 \sqcap B_2 \sqcap B_3)$ and $A \sqcap \exists r.(B_1 \sqcap B_2) \sqcap \exists r.(B_1 \sqcap B_3) \sqcap \exists r.(B_2 \sqcap B_3)$. While the former is obvious, the latter is obtained since each $B_1 \sqcap B_2, B_1 \sqcap B_3$, and $B_2 \sqcap B_3$ is an element of $U(C)$. The following characterization for $Upper(C)$ is shown in [Kri18].

Proposition 4.23. *Let C be a reduced $\mathcal{EL}$ concept. Then we have $Upper(C) = U(C)$ up to equivalence. In particular, this implies that the cardinality of $Upper(C)$ is polynomial in the size of C and that this set can be computed in polynomial time in the size of C.* ♢

Following the proposition above, a given $\mathcal{EL}$ concept has only polynomially many upper neighbors, each of which is of polynomial size. As an easy consequence we obtain the following lemma:

Lemma 4.24. *The one-step relation $\sqsubset_1$ induced by $\sqsubset$ on $\mathcal{EL}$ concepts is decidable in polynomial time.*

Unfortunately, this result does not transfer immediately from concept subsumption to axiom weakening. In fact, as we have seen before, strict subsumption need not produce a weaker axiom (see the remark below Proposition 4.20). Thus, to find all GCIs $C \sqsubseteq D'$ with $C \sqsubseteq D \succ_1^{sub} C \sqsubseteq D'$, it is not sufficient to consider only concepts D' with $D \sqsubset_1 D'$. In case $C \sqsubseteq D'$ is equivalent to $C \sqsubseteq D$, we need to consider upper neighbors of D', etc.

Proposition 4.25. *The one-step relation $\succ_1^{sub}$ induced by $\succ^{sub}$ is effectively finitely branching.* ♢

Proof. Since $\sqsubset$ is one-step generated, finitely branching, and well-founded, for a given concept D, there are only finitely many concepts D' such that $D \sqsubset D'$. Thus, a breadth first search along $\sqsubset_1$ can be used to compute all concepts D' such that there is a path $D \sqsubset_1 D_1 \sqsubset_1 \dots D_n \sqsubset_1 D'$ where $C \sqsubseteq D$ is equivalent to $C \sqsubseteq D_i$ for $i = 1, \dots, n$, and $C \sqsubseteq D \succ^{sub} C \sqsubseteq D'$. Since $\sqsubset$ is one-step generated, it is easy to see that all axioms γ with $C \sqsubseteq D \succ_1^{sub} \gamma$ can be obtained this way. However, the computed set of axioms may contain elements that are not one-step successors of $C \sqsubseteq D$. Thus, in a final step, we remove all axioms that are weaker than some axiom in the set. □

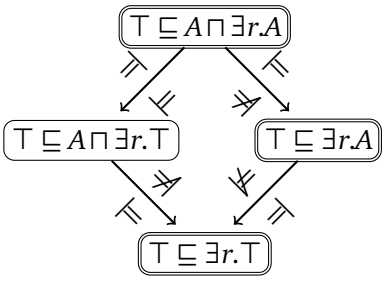


Figure 4.26: One-step weakening

Example 4.27. *To see that the final step of removing axioms in the proof of Proposition 4.25 is needed, consider the axiom $\beta = \top \sqsubseteq A \sqcap \exists r.A$. The right-hand side $A \sqcap \exists r.A$ has two upper neighbors, namely $\exists r.A$ and $A \sqcap \exists r.\top$. The first yields the axiom $\top \sqsubseteq \exists r.A$, which satisfies $\top \sqsubseteq A \sqcap \exists r.A \succ_1^{sub} \top \sqsubseteq \exists r.A$. The second yields the axiom $\top \sqsubseteq A \sqcap \exists r.\top$, which is equivalent to β. Thus, the only upper neighbor $\top \sqsubseteq \exists r.\top$ is considered, but this concept yields an axiom that is actually weaker than $\top \sqsubseteq \exists r.A$, and thus needs to be removed.*

A similar, but simpler example can be used to show that the additional removal of weaker elements in the proof of Proposition 4.18 is needed. Let α be the consequence $\top \sqsubseteq A$, $J = \{\beta\}$ for $\beta := \top \sqsubseteq A \sqcap B$, $\delta_1 := \top \sqsubseteq A$, $\delta_2 := \top \sqsubseteq B$, and $\gamma := \top \sqsubseteq \top$. Then we have exactly the situation described below the proof of Proposition 4.18, with $\succ^{sub}$ as the employed weakening relation. ♢

Corollary 4.28. *All maximally strong weakenings w.r.t. $\succ^{sub}$ of an axiom in a justification can effectively be computed.*

Proof. By Proposition 4.18, this is an immediate consequence of the fact that $\succ^{sub}$ is well-founded, one-step generated, and effectively finitely branching. □

The algorithm for computing maximally strong weakenings described in the proof of Proposition 4.18 has non-elementary complexity for $\succ^{sub}$. In fact, the bound for the depth of the tree that must be searched grows by one exponential for every increase in the role-depth of the concept on the right-hand side. It is not clear how to obtain an algorithm with a better complexity. Example 4.38 below yields an exponential lower-bound, which still leaves a huge gap. We can also show that even deciding whether a given axiom is a maximally strong weakening w.r.t. $\succ^{sub}$ is coNP-hard.

Before we can prove this hardness result, we must introduce the coNP-complete problem that will be used in our proof by reduction. A *monotone Boolean formula* φ is built from propositional variables using the connectives conjunction ($\wedge$) and disjunction ($\vee$) only. If V is the set of propositional variables occurring in φ, then propositional valuations can be seen as subsets W of V. Since φ is monotone, the valuation V clearly satisfies φ, and the valuation $\emptyset$ falsifies φ. We are now interested in *maximal* valuations falsifying φ, where valuations are compared using set inclusion.

Definition 4.29. *The* all-maximal-valuations *problem receives as input*

- *a monotone Boolean formula φ with propositional variables V, and*

- *a set $\mathcal{V}$ of maximal valuations falsifying φ.*

The question is then whether $\mathcal{V}$ is the set of all *maximal valuations falsifying φ.* ♢

As shown in [Nys09] (Lemma 6.13), the all-maximal-valuations problem is coNP-complete.

Proposition 4.30. *The problem of deciding whether a given $\mathcal{EL}$ GCI $C \sqsubseteq D'$ is a maximally strong weakening of the $\mathcal{EL}$ GCI $C \sqsubseteq D$ w.r.t. $\succ^{sub}$ is coNP-hard.* ♢

Proof. Given an instance $\varphi, \mathcal{V}$ of the all-maximal-valuations problem, we construct an instance of our problem as follows. For every subformula ψ of φ, we introduce a new concept name B_ψ. If ψ is not a propositional variable, we define the TBox:

$$\mathcal{T}_\psi := \begin{cases} \{B_{\psi_1} \sqcap B_{\psi_2} \sqsubseteq B_\psi\} & \psi = \psi_1 \wedge \psi_2, \\ \{B_{\psi_1} \sqsubseteq B_\psi, B_{\psi_2} \sqsubseteq B_\psi\} & \psi = \psi_1 \vee \psi_2. \end{cases}$$

Let V be the set of all propositional variables appearing in φ, and let $csub(\varphi)$ be the set of all subformulas of φ that are not in V.

We construct the ontology that has only one refutable axiom

$$A \sqsubseteq \exists r. \bigsqcap \{B_p \mid p \in V\},$$

and as static part the ontology

$$\mathcal{T}_s = \bigcup_{\psi \in csub(\varphi)} \mathcal{T}_\psi \cup \{\exists r. B_\varphi \sqsubseteq C\}.$$

Clearly, the refutable axiom is a justification for $A \sqsubseteq C$.

Given a set $\mathcal{W}$ of valuations, define the concept

$$X_{\mathcal{W}} := \bigsqcap_{W \in \mathcal{W}} \exists r. \bigsqcap \{B_p \mid p \in W\}.$$

It follows that $\{A \sqsubseteq X_{\mathcal{W}}\} \cup \mathcal{T}_s \not\models A \sqsubseteq C$ iff no valuation in $\mathcal{W}$ satisfies φ.

We *claim* that $\mathcal{V}$ is the set of all maximal valuations not satisfying φ iff $A \sqsubseteq X_{\mathcal{V}}$ is a maximally strong weakening of $A \sqsubseteq \exists r. \bigsqcap \{B_p \mid p \in V\}$.

First, assume that $\mathcal{V}$ is the set of all maximal valuations not satisfying φ. It implies that $\{A \sqsubseteq X_{\mathcal{V}}\} \cup \mathcal{T}_s \not\models A \sqsubseteq C$ and clearly $A \sqsubseteq \exists r. \bigsqcap \{B_p \mid p \in V\} \succ^{sub} A \sqsubseteq X_{\mathcal{V}}$. If $A \sqsubseteq X_{\mathcal{V}}$ is not maximally strong, then there is a concept E such that $\exists r. \bigsqcap \{B_p \mid p \in V\} \sqsubset E \sqsubset X_{\mathcal{V}}$ and $\{A \sqsubseteq E\} \cup \mathcal{T}_s \not\models A \sqsubseteq C$. The strict subsumption relationships imply the E contains a top-level conjunct $\exists r. \bigsqcap \{B_p \mid p \in U\}$ for a set $U \subseteq V$ such that U is incomparable w.r.t. set inclusion with all the sets in $\mathcal{V}$. Since $\mathcal{V}$ is the set of all maximal valuations not satisfying φ, this implies that U satisfies φ. Consequently, $\{A \sqsubseteq E\} \cup \mathcal{T}_s \models A \sqsubseteq C$, which yields a contradiction to our assumption that $A \sqsubseteq X_{\mathcal{V}}$ is not maximally strong.

Conversely, assume that $\mathcal{V}$ is not the set of all maximal valuations not satisfying φ, i.e., there is a maximal valuation U not satisfying φ such that $U \notin \mathcal{V}$. This implies that U is incomparable w.r.t. inclusion with any of the elements of $\mathcal{V}$, and thus $\exists r. \bigsqcap \{B_p \mid p \in V\} \sqsubset X_{\mathcal{V} \cup \{U\}} \sqsubset X_{\mathcal{V}}$. In addition, we know that $\{A \sqsubseteq X_{\mathcal{V} \cup \{U\}}\} \cup \mathcal{T}_s \not\models A \sqsubseteq C$, which shows that $A \sqsubseteq X_{\mathcal{V}}$ is not maximally strong. □

4.4.2 Syntactic Generalizations

In order to obtain a weakening relation that has better algorithmic properties than $\succ^{sub}$, we consider a syntactic approach for generalizing $\mathcal{EL}$ concepts. Basically, the concept D is a syntactic generalization of the concept C if D can be obtained from C by removing occurrences of subconcepts. To ensure that such a removal really generalizes the concept, we work here with reduced concepts.

Definition 4.31. *Let C, D be $\mathcal{EL}$ concepts. Then D is a* syntactic generalization *of C (written $C \sqsubset^{syn} D$) if it is obtained from the reduced form of C by replacing some occurrences of subconcepts $\neq \top$ with $\top$.* ♢

For example, the concept $C = A_1 \sqcap \exists r.(A_1 \sqcap A_2)$ is already reduced, and its syntactic generalizations include, among others, $\top \sqcap \exists r.(A_1 \sqcap A_2) \equiv^{\emptyset} \exists r.(A_1 \sqcap A_2)$, $A_1 \sqcap \exists r.(\top \sqcap A_2) \equiv^{\emptyset} A_1 \sqcap \exists r.A_2$, $\exists r.\top$, and $\top$.

Lemma 4.32. *If $C \sqsubset^{syn} D$, then $C \sqsubset D$, and the length of any $\sqsubset^{syn}$-chain issuing from C is linearly bounded by the size of C.*

Proof. We use a modified definition of size (called m-size) where only occurrences of concept and role names are counted. Reducing a concept preserves equivalence and never increases the m-size. Since the concept constructors of $\mathcal{EL}$ are monotonic, $C \sqsubset^{syn} D$ implies $C \sqsubseteq D$. In addition, the m-size of the reduced form of C is strictly larger than the m-size of the reduced form of D since concepts $\neq \top$ have an m-size > 0 whereas $\top$ has m-size 0. This shows $C \not\equiv^{\emptyset} D$ (and thus $C \sqsubset D$), since these reduced forms then cannot be equal up to associativity and commutativity of $\sqcap$. In addition, it clearly yields the desired linear bound on the length of $\sqsubset^{syn}$-chains. □

By Proposition 4.15, this linear bound implies that $\sqsubset^{syn}$ is one-step generated. In the corresponding one-step relation $\sqsubset_1^{syn}$, the replacements can be restricted to subconcepts that are concept names or existential restriction of the form $\exists r.\top$. For example, we have (modulo equivalence)

$$\exists r.(A_1 \sqcap A_2 \sqcap A_3) \sqsubset_1^{syn} \exists r.(A_1 \sqcap A_2) \sqsubset_1^{syn} \exists r.A_2 \sqsubset_1^{syn} \exists r.\top \sqsubset_1^{syn} \top.$$

However, not all such restricted replacements lead to single steps w.r.t. $\sqsubset^{syn}$. For example, consider the concept $C = \exists r.(A_1 \sqcap A_2) \sqcap \exists r.(A_2 \sqcap A_3)$. Then replacing A_3 by $\top$ leads to $D = \exists r.(A_1 \sqcap A_2) \sqcap \exists r.(A_2 \sqcap \top) \equiv \exists r.(A_1 \sqcap A_2)$, but we have $C \sqsubset^{syn} \exists r.(A_1 \sqcap A_2) \sqcap \exists r.A_3 \sqsubset^{syn} D$.

Before proving that every $\sqsubset_1^{syn}$-step can be realized by such restricted replacements, we use the fact that any $\mathcal{EL}$ concept can be written as a conjunction of concept names and existential restrictions to give a recursive characterization of $\sqsubset^{syn}$. Let C be an $\mathcal{EL}$ concept, and assume that its reduced form is

$$C' = A_1 \sqcap \ldots \sqcap A_k \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_\ell.C_\ell.$$

Then we have $A_i \neq A_j$ for all $i \neq j$ in $\{1,\ldots,k\}$ and $r_\mu \neq r_\nu$ or $C_\mu \not\sqsubseteq C_\nu$ for all $\nu \neq \mu$ in $\{1,\ldots,\ell\}$, since otherwise C' would not be reduced. Replacing some occurrences of subconcepts with $\top$ then corresponds (modulo equivalence) to

- removing some of the conjuncts of the form A_i,
- removing some of the conjuncts of the form $\exists r_\mu.C_\mu$,
- replacing some of the conjuncts of the form $\exists r_\nu.C_\nu$ with a conjunct of the form $\exists r_\nu.D_\nu$ where $C_\nu \sqsubset^{syn} D_\nu$

such that at least one of these actions is really taken. Thus, $C \sqsubset_1^{syn} D$ implies that D can be obtained from the reduced form of C by taking exactly one of these actions for exactly one conjunct. In fact, either taking several actions has the same effect as taking one of them, or taking the actions one after another leads to a sequence of several strict syntactic generalizations steps, which is precluded by the definition of $\sqsubset_1^{syn}$.

Lemma 4.33. *Let $C \not\equiv^\emptyset \top$ with reduced form $C' = A_1 \sqcap \ldots \sqcap A_k \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_\ell.C_\ell$, and assume that $C \sqsubset_1^{syn} D$. Then D is obtained (modulo equivalence) from C' by either*

1. *removing exactly one of the concept names A_i,*
2. *removing exactly one of the existential restrictions $\exists r_\mu.C_\mu$ for $C_\mu \equiv^\emptyset \top$, or*
3. *replacing exactly one of the existential restrictions $\exists r_\nu.C_\nu$ with $\exists r_\nu.D_\nu$ for $C_\nu \sqsubset_1^{syn} D_\nu$.*

Proof. As argued above, $C \sqsubset_1^{syn} D$ implies that D is obtained from C' by performing one of the following three actions:

- *Removing exactly one of the conjuncts of the form A_i:* in this case, we are done.
- *Removing exactly one of the conjuncts of the form $\exists r_\mu.C_\mu$:* in this case we are done if $C_\mu \equiv^\emptyset \top$. Thus, assume that $C_\mu \not\equiv^\emptyset \top$. Let D' be obtained from C' by replacing $\exists r_\mu.C_\mu$ with $\exists r_\mu.\top$. Then we either have $C \sqsubset^{syn} D' \sqsubset^{syn} D$ or $D' \equiv^\emptyset D$. The first case contradicts our assumption that $C \sqsubset_1^{syn} D$. The second case is dealt with below since $C_\mu \sqsubset^{syn} \top$.
- *Replacing exactly one of the conjuncts of the form $\exists r_\nu.C_\nu$ with a conjunct of the form $\exists r_\nu.D_\nu$ where $C_\nu \sqsubset^{syn} D_\nu$:* in this case we are done if $C_\nu \sqsubset_1^{syn} D_\nu$. Thus, assume that there is an $\mathcal{EL}$ concept D'_ν such that $C_\nu \sqsubset^{syn} D'_\nu \sqsubset^{syn} D_\nu$. Since we already know that $\sqsubset^{syn}$ is one-step generated, we can assume without loss of generality that $C_\nu \sqsubset_1^{syn} D'_\nu$. Let D' be obtained from C' by replacing $\exists r_\nu.C_\nu$ with $\exists r_\mu.D'_\nu$. Then we either have $C \sqsubset^{syn} D' \sqsubset^{syn} D$ or $D' \equiv^\emptyset D$. The first case contradicts our assumption that $C \sqsubset_1^{syn} D$. In the second case, we are done.

Since there are no other cases, this completes the proof of the lemma. □

Based on this lemma, the following proposition can now easily be shown by induction on the role depth of C.

Proposition 4.34. *Let C be an $\mathcal{EL}$ concept and C' its reduced form. If $C \sqsubset_1^{syn} D$, then D can be obtained (modulo equivalence) from C' by either replacing a concept name or a subconcept of the form $\exists r.\top$ by $\top$.* ♢

As an immediate consequence we obtain that $\sqsubset^{syn}$ is effectively linearly branching.

Corollary 4.35. *For a given $\mathcal{EL}$ concept C, the set $\{D \mid C \sqsubset_1^{syn} D\}$ has a cardinality that is linear in the size of C and it can be computed in polynomial time.*

Proof. That the cardinality of $\{D \mid C \sqsubset_1^{syn} D\}$ is linearly bounded by the size of C is an immediate consequence of Proposition 4.34. To compute the set, one first computes all concepts that can be obtained by replacing in the reduced form of C a concept name or a subconcept of the form $\exists r.\top$ by $\top$. The polynomially many concepts obtained this way contain all the elements of $\{D \mid C \sqsubset_1^{syn} D\}$. Additional elements in this set are obviously strictly subsumed by an element of $\{D \mid C \sqsubset_1^{syn} D\}$, and thus we can remove them by removing elements that are not subsumption minimal. □

Now, we define our new weakening relation, which syntactically generalizes the right-hand sides of GCIs:

$$C \sqsubseteq D \succ^{syn} C' \sqsubseteq D' \quad \text{if} \quad C = C', D \sqsubset^{syn} D' \text{ and } \{C' \sqsubseteq D'\} \not\models C \sqsubseteq D.$$

It is clear that syntactically generalizing the right-hand side of GCIs is a fragment of mechanisms to semantically generalize the right-hand side of GCIs that has been applied by $\succ^{sub}$, which means that $\succ^{syn} \subseteq \succ^{sub}$. Now, the following theorem is an easy consequence of the properties of $\sqsubset^{syn}$ and of Corollary 4.35.

Theorem 4.36. *The relation $\succ^{syn}$ on $\mathcal{EL}$ axiom is a linear, complete, one-step generated, and effectively linearly branching weakening relation.*

Due to fact that $\succ_1^{syn}$-steps do not increase the size of axioms, the linear bounds on the branching of $\succ_1^{syn}$ and the length of $\succ^{syn}$-chains imply that the algorithm described in the proof of Proposition 4.18 has an exponential search space.

Corollary 4.37. *All maximally strong weakenings w.r.t. $\succ^{syn}$ of an axiom in a justification can be computed in exponential time.*

The following example shows that there may be exponentially many maximally strong weakenings w.r.t. $\succ^{syn}$, and thus the exponential complexity stated above is optimal.

Example 4.38. *Let $\beta_i := P_i \sqcap Q_i \sqsubseteq B$ for $i = 1, \ldots, n$ and $\beta := A \sqsubseteq P_1 \sqcap Q_1 \sqcap \ldots \sqcap P_n \sqcap Q_n$. We consider the ontology $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$, where $\mathfrak{O}_{st} := \{\beta_i \mid 1 \leq i \leq n\}$ and $\mathfrak{O}_{rt} := \{\beta\}$. Then $J = \{\beta\}$ is a justification for the consequence $\alpha = A \sqsubseteq B$, and all axioms of the form $A \sqsubseteq X_1 \sqcap X_2 \sqcap \ldots \sqcap X_n$ with $X_i \in \{P_i, Q_i\}$ are maximally strong weakenings w.r.t. $\succ^{syn}$ of β in J. The same is true for $\succ^{sub}$ since in the absence of roles, these two weakening relations coincide.* ♢

A single maximally strong weakening can however be computed in polynomial time.

Proposition 4.39. *A single maximally strong weakening w.r.t. $\succ^{syn}$ can be computed in polynomial time.* ♢

Proof. The algorithm that computes a maximally strong weakening works as follows. Starting from the concept $D' := \top$, it looks at all possible ways of making one step in the direction of D using $\sqsupset_1^{syn}$, i.e., it considers all D'' where $D \sqsubseteq^{syn} D'' \sqsubset_1^{syn} D'$. The concepts D'' can be obtained by adding a concept name A or an existential restriction $\exists r.\top$ at a place where (the

reduced form of) D has such a concept or restriction. Obviously, there are only polynomially many such concepts D''. For each of them we check whether

$$\mathfrak{O}_{st} \cup (J \setminus \{C \sqsubseteq D\}) \cup \{C \sqsubseteq D''\} \models \alpha.$$

If this is the case for all D'', we return $C \sqsubseteq D'$. Otherwise, we choose an arbitrary D'' with $\mathfrak{O}_{st} \cup (J \setminus \{C \sqsubseteq D\}) \cup \{C \sqsubseteq D''\} \not\models \alpha$, and continue with $D' := D''$.

This algorithm terminates after linearly many iterations since in each iteration the size of D' is increased and it cannot get larger than D. In addition, $C \sqsubseteq D'$ is maximally strong since for every axiom $C \sqsubseteq E$ such that $C \sqsubseteq D \succ^{syn} C \sqsubseteq E \succ^{syn} C \sqsubseteq D'$ there is a sequence $E \sqsubset_1^{syn} \ldots \sqsubset_1^{syn} D'' \sqsubset_1^{syn} D'$. Consequently, $C \sqsubseteq D''$ has the consequence, and thus also $C \sqsubseteq E$. □

Nevertheless, we can show that deciding whether an axiom is a maximally strong weakening w.r.t. $\succ^{syn}$ is coNP-complete.

Proposition 4.40. *The problem of deciding whether a given $\mathcal{EL}$ GCI $C \sqsubseteq D'$ is a maximally strong weakening of the $\mathcal{EL}$ GCI $C \sqsubseteq D$ w.r.t. $\succ^{syn}$ is coNP-complete.* ♢

Proof. First, we show the coNP upper bound. Let $\mathfrak{O} = \mathfrak{O}_{st} \cup \mathfrak{O}_{rt}$, $J \subseteq \mathfrak{O}_{rt}$ a justification of the consequence α, $C \sqsubseteq D$ an element of J, and $C \sqsubseteq D'$ a GCI. Obviously, we can decide in polynomial time whether $C \sqsubseteq D \succ^{syn} C \sqsubseteq D'$ and whether $\mathfrak{O}_{st} \cup (J \setminus \{C \sqsubseteq D\}) \cup \{C \sqsubseteq D'\} \not\models \alpha$. To disprove that $C \sqsubseteq D'$ is maximally strong, we guess an $\mathcal{EL}$ concept D'' such that $D \sqsubset^{syn} D'' \sqsubset^{syn} D'$. This requires only polynomially many guesses: in fact, D' is obtained from D by replacing linearly many occurrences of subconcepts with $\top$, and we simply guess which of these replacements are not done when going from D to D''. We then check in polynomial time whether $C \sqsubseteq D''$ satisfies

- $\mathfrak{O}_{st} \cup (J \setminus \{C \sqsubseteq D\}) \cup \{C \sqsubseteq D''\} \not\models \alpha$, and
- $\{C \sqsubseteq D'\} \not\models C \sqsubseteq D''$.

If both tests succeed then $C \sqsubseteq D''$ is a counterexample to $C \sqsubseteq D'$ being maximally strong.

For the hardness proof, we use again the all-maximal-valuations problem. Given an instance $\varphi, \mathcal{V}$ of the all-maximal-valuations problem, we construct an instance of our problem as follows. For every subformula ψ of φ, we introduce a new concept name B_ψ. If ψ is not a propositional variable, we define the TBox:

$$\mathcal{T}_\psi := \begin{cases} \{B_{\psi_1} \sqcap B_{\psi_2} \sqsubseteq B_\psi\} & \psi = \psi_1 \wedge \psi_2 \\ \{B_{\psi_1} \sqsubseteq B_\psi, B_{\psi_2} \sqsubseteq B_\psi\} & \psi = \psi_1 \vee \psi_2. \end{cases}$$

Let V be the set of all propositional variables appearing in φ, and let $csub(\varphi)$ be the set of all subformulas of φ that are not in V. Define the concept

$$X_{\mathcal{V}} := \bigsqcap_{W \in \mathcal{V}} \exists r. \bigsqcap \{B_p \mid p \in W\}.$$

We construct the ontology that has only one refutable axiom

$$X_{\mathcal{V}} \sqsubseteq \exists r. \bigsqcap \{B_p \mid p \in V\},$$

and as static part the ontology

$$\mathcal{T}_s = \bigcup_{\psi \in csub(\varphi)} \mathcal{T}_\psi \cup \{\exists r.B_\varphi \sqsubseteq C\}$$

Clearly, the refutable axiom is the only justification for $X_\mathcal{V} \sqsubseteq C$.

For every valuation $W \subseteq V$, if W is a subset of some valuation in $\mathcal{V}$, then

$$X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in W\} \text{ is equivalent to } X_\mathcal{V} \sqsubseteq \top.$$

We *claim* that $X_\mathcal{V} \sqsubseteq \top$ is a maximally strong weakening w.r.t. $\succ^{syn}$ of the only refutable axiom iff $\mathcal{V}$ is the set of all maximal valuations not satisfying φ.

To prove this claim, first assume that $\mathcal{V}$ is not the set of all maximal valuations not satisfying φ, i.e., there is a maximal valuation W not satisfying φ such that $W \notin \mathcal{V}$. On the one hand, this implies that W is incomparable w.r.t. inclusion with any of the elements of $\mathcal{V}$, and thus $X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in W\}$ is not a tautology. On the other hand, we have

$$\mathcal{T}_s \cup \{X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in W\}\} \not\models X_\mathcal{V} \sqsubseteq C,$$

and $X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in V\} \succ^{syn} X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in W\}$. This shows that the tautology $X_\mathcal{V} \sqsubseteq \top$ is not a maximally strong weakening w.r.t. $\succ^{syn}$ of the only refutable axiom $X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in V\}$.

Conversely, assume that $\mathcal{V}$ is the set of all maximal valuations not satisfying φ, and that γ is a maximally strong weakening w.r.t. $\succ^{syn}$ of $X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in V\}$. If $\gamma = X_\mathcal{V} \sqsubseteq \top$, then we are done. Otherwise, there is a set $W \subseteq \mathcal{V}$ such that $\gamma = X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in W\}$. But then $\mathcal{T}_s \cup \{\gamma\} \not\models X_\mathcal{V} \sqsubseteq C$ implies that W does not satisfy φ, and thus W is a subset of some valuation in $\mathcal{V}$. Consequently, γ is a tautology and thus equivalent to $X_\mathcal{V} \sqsubseteq \top$. This shows that $X_\mathcal{V} \sqsubseteq \top$ is a maximally strong weakening w.r.t. $\succ^{syn}$ of $X_\mathcal{V} \sqsubseteq \exists r.\bigsqcap\{B_p \mid p \in V\}$. □

The standard reasoning procedures for $\mathcal{EL}$ first normalize the given TBox, where normalization breaks up large GCIs into smaller ones [BHL+17]. In some cases, applying classical repair to the normalized TBox also leads to more gentle repairs. For example, consider the refutable TBox $\mathcal{T} = \{A \sqsubseteq B_1 \sqcap B_2\}$, the strict ABox $\mathcal{A} = \{A(a)\}$, and the consequence $\alpha = (B_1 \sqcap B_2)(a)$. The TBox $\mathcal{T}$ is normalized to $\mathcal{T}' = \{A \sqsubseteq B_1, A \sqsubseteq B_2\}$, which has the two classical repairs $\mathcal{T}'_1 = \{A \sqsubseteq B_1\}$ and $\mathcal{T}'_2 = \{A \sqsubseteq B_2\}$. This is exactly what our gentle repair approach (using $\succ^{sub}$ or $\succ^{syn}$) would yield. However, normalization does not always do the job as illustrated by the following two examples. As a first example, consider the refutable TBox $\{A \sqsubseteq \exists r.B\}$, the strict ABox $\{A(a)\}$, and the consequence $\exists r.B(a)$. Here, the TBox is normalized, and classical repair removes the GCI. In contrast, our gentle repair approach can weaken the GCI to $A \sqsubseteq \exists r.\top$. Another problem with using normalization in this setting is that in general it introduces new concept names. As a second example, consider the refutable TBox $\{A \sqsubseteq \exists r.\exists r.B\}$ and the strict ABox $\{A(a)\}$, where the unwanted consequence is $\exists r.\exists r.B(a)$. Normalizing the TBox yields $\{A \sqsubseteq \exists r.X, X \sqsubseteq \exists r.B\}$; thus, classical repair yields as repairs the TBoxes consisting of $A \sqsubseteq \exists r.X$ or $X \sqsubseteq \exists r.B$. These two axioms do not make sense for the user since X is a name without meaning in the application. Thus, some post-processing that can get rid of the new names (similar to forgetting [NR14]) would be required. While an approach based on appropriate variants of normalization and forgetting may be able

to generate gentle repairs akin to what our approach produces using $\succ^{syn}$, it would not be able to deal with more sophisticated weakening relations such as $\succ^{sub}$. In addition, classical repair applied to the normalized TBox would not distinguish between more or less gentle repairs, and would also produce all classical repairs of the original TBox.

4.5 Weakening Relations for $\mathcal{ALC}$ Axioms

We have seen in Proposition 4.19 that $\succ^{g}$ is not one-step generated so that we cannot apply this weakening relation for $\mathcal{ALC}$ axioms. If we move to $\succ^{s}$ defined in Proposition 4.20, then we cannot use it either to specialize the left-hand side of $\mathcal{ALC}$ GCIs and unfortunately this weakening relation also cannot be employed for generalizing the right-hand size of $\mathcal{ALC}$ GCIs since the right-hand sides can be generalized infinitely as illustrated below

$$B \sqsubseteq A \succ^{s} B \sqsubseteq A \sqcup \forall r.\bot \succ^{s} B \sqsubseteq A \sqcup \forall r.\forall r.\bot \succ^{s} B \sqsubseteq A \sqcup \forall r.\forall r.\forall r.\bot \succ^{s} \ldots$$

To deal with this issue, we introduce weakening relations that have restrictions as defined in the remainder of this section. Note that we again focus only on $\mathcal{ALC}$ GCIs since $\mathcal{ALC}$ concept assertions can be treated similarly.

4.5.1 Generalizations and Specializations in $\mathcal{ALC}$ w.r.t. Role Depth

Here we define a weakening relation for $\mathcal{ALC}$ that generalizes (specializes) the concepts C such that the generalized (specialized) concepts are built over the signature of C and have role-depth bounded by the role-depth of C.

Definition 4.41. *Let C, D be $\mathcal{ALC}$ concepts. We say that*

- *D is a* bounded specialization *of C, denoted by $C \sqsupset^{bosp} D$, if $C \sqsupset D$, $rd(D) \leq rd(C)$, and D contains only concept and role names occurring in C.*
- *D is a* bounded generalization *of C, denoted by $C \sqsubset^{boge} D$, if $C \sqsubset D$, $rd(D) \leq rd(C)$, and D contains only concept and role names occurring in C.* ♢

Lemma 4.42. *The relations $\sqsupset^{bosp}$ and $\sqsubset^{boge}$ are well-founded.*

Proof. This is an easy consequence of the fact that, for fixed finite sets of concept names $\mathsf{N_C}$ and role names $\mathsf{N_R}$ and a fixed bound $n \geq 0$, there are only finitely many $\mathcal{ALC}$ concepts that are of role depth at most n and are only built over concept names and roles names from $\mathsf{N_C} \cup \mathsf{N_R}$. So, if we assume that $C_0 \sqsupset^{bosp} C_1 \sqsupset^{bosp} C_2 \sqsupset^{bosp} \ldots$, then there is an $i < j$ such that $C_i \equiv C_j$, which violates the fact that $C_i \sqsupset^{bosp} C_{i+1} \sqsupset^{bosp} \ldots \sqsupset^{bosp} C_j$. These arguments also work for $\sqsubset^{boge}$. □

Proposition 4.43. *Assume that our axioms are GCIs formulated in $\mathcal{ALC}$. If we define*

$$C \sqsubseteq D \succ^{b} C' \sqsubseteq D' \text{ if } C' \sqsupset^{bosp} C, D \sqsubset^{boge} D' \text{ and } \{C' \sqsubseteq D'\} \not\models C \sqsubseteq D,$$

then $\succ^{b}$ is a well-founded, complete, and one-step generated weakening relation. ♢

Proof. $\succ^b$ is well-founded due to Lemma 4.42. Then, the completeness follow from the fact that there is a concept $\top$ ($\bot$) as bounded generalization (specialization) of $\mathcal{ALC}$ concepts. Since this is bounded, by Proposition 4.15, $\succ^b$ is one-step generated. □

Note that weakening relations for $\mathcal{EL}$ axioms defined in Theorem 4.21 is contained in $\succ^b$. Consequently, it also follows that $\succ^b$ is not polynomial.

4.5.2 Syntactical Generalizations and Specializations in $\mathcal{ALC}$

In order to obtain a linear weakening relation for $\mathcal{ALC}$, we generate more general or more specific concepts by syntactically replacing subconcepts with $\top$ and $\bot$, depending on whether these subconcepts occur positively or negatively. Intuitively, a positive occurrence of a subconcept is under an even number of negations whereas a negative one is under an odd number of negations. To define this formally, let us consider the syntax trees of $\mathcal{ALC}$ concepts that are defined as follows.

Definition 4.44 (Syntax Tree of $\mathcal{ALC}$). *Let C be an $\mathcal{ALC}$ concept such that $C \not\equiv \top$ or $C \not\equiv \bot$. The* syntax tree *T_C of C is a tuple $(\mathcal{V}, \mathcal{E}, \varepsilon)$, where $\mathcal{V}$ is the set of concept names and constructors occurring in C, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, and ε is the root node. T_C is inductively defined as follows*

- *if $C = A \in \mathsf{N_C}$, then $\mathcal{E} = \emptyset$, and $\varepsilon = A$,*
- *if $C = \circ D$, where $\circ \in \{\neg\} \cup \{\forall r \exists r. \mid$ role name r occurs in $C\}$, then*
  - *$\mathcal{E} = \{(\neg, D)\} \cup \mathcal{E}'$ and*
  - *$\varepsilon = \circ$,*

  where $\mathcal{E}'$ is the set of edges in T_D, respectively, and
- *if $C = D_1 \circ D_2$, where $\circ \in \{\sqcap, \sqcup\}$, then*
  - *$\mathcal{E} = \{(\circ, D_1), (\circ, D_2)\} \cup \mathcal{E}_1 \cup \mathcal{E}_2$ and*
  - *$\varepsilon = \circ$,* ♢

  where $\mathcal{E}_1$ and $\mathcal{E}_2$ are the set of nodes and edges in T_{D_1} and T_{D_2}, respectively.

Occurrences of subconcepts correspond to nodes in the tree. Such a node is *positive* if on the path from root to this node, an even number of negations is encountered. Otherwise, it is *negative*. Occurrences of subconcepts are called positive (negative) if the corresponding nodes in the syntax tree are positive (negative).

Definition 4.45. *Let C, D be $\mathcal{ALC}$ concepts.*

1. *D is a* direct syntactic specialization *of C if it is obtained from C by replacing a positive occurrence of a subconcept $\neq \bot$ by $\bot$ or a negative occurrence of a subconcept $\neq \top$ by $\top$;*
2. *D is a* direct syntactic generalization *of C if it is obtained from C by replacing a positive occurrence of a subconcept $\neq \top$ by $\top$ or a negative occurrence of a subconcept $\neq \bot$ by $\bot$;*

The concept D is a syntactic specialization (generalization) *of C iff it is obtained from C by a finite sequence of direct syntactic specializations (generalizations).* ♢

The following lemma is an easy consequence of the fact that negation is anti-monotonic w.r.t. subsumption and all the other concept constructors of $\mathcal{ALC}$ are monotonic w.r.t. subsumption.

Lemma 4.46. *Let C, D be $\mathcal{ALC}$ concepts.*

1. *If D is a syntactic specialization of C, then $D \sqsubseteq C$.*
2. *if D is a syntactic generalization of C, then $C \sqsubseteq D$.*

We write $D \sqsubset^{syge} C$ to indicate that D is a syntactic specialization of C and $C \sqsupset^{sysp} D$ to indicate that D is a syntactic generalization of C.

Lemma 4.47. *The relations $\sqsubset^{syge}$ and $\sqsupset^{sysp}$ are well-founded and for every $\mathcal{ALC}$ concept C, the length of the longest $\sqsubset^{syge}$-chain($\sqsupset^{sysp}$-chain) issuing from C is linearly bounded by the size of C.*

Proof. First, consider $\sqsubset^{syge}$. We must show that there cannot be an infinite sequence $C_0 \sqsubset^{syge} C_1 \sqsubset^{syge} C_2 \sqsubset^{syge} \ldots$. For this purpose, we define the size $| C |$ of an $\mathcal{ALC}$ concept C by counting

- every occurrence of a concept constructor as 1,
- every occurrence of a concept name as 2,
- every positive occurrence of $\top$ and every negative occurrence of $\bot$ by 2, and
- every negative occurrence of $\top$ and every positive occurrence of $\bot$ by 1. □

If D is a direct syntactic specialization of C, then $| C |>| D |$. In fact, the replacement generates a new negative occurrence of $\top$ or a new positive occurrence of $\bot$, whose size is counted by 1. However, the subconcept that was replaced yields a contribution of at least 2 to the size of C. This shows well-foundedness of $\sqsubset^{syge}$.

More precisely, this argument shows that the length of the longest $\sqsubset^{syge}$-chain issuing from C is bounded by $| C |$. Note that our definition of $| C |$ is not the standard definition of $| C |$, but w.l.o.g. any reasonable definition for the size of C [5] is also linear in C. The relation $\sqsupset^{sysp}$ can also be treated similarly, but different definition of size should be considered first.

The following proposition is an easy consequence of the lemma above.

Proposition 4.48. *Assume that our axioms are GCIs formulated in $\mathcal{ALC}$. If we define*

$$C \sqsubseteq D \succ^{syt} C' \sqsubseteq D' \text{ if } C' \sqsubset^{syge} C, D \sqsupset^{sysp} D', \text{ and } \{C' \sqsubseteq D\} \not\models C \sqsubseteq D,$$

then $\succ^{syt}$ is a linear and complete weakening relation. ♢

Similar to relationships between $\succ^{sub}$ and $\succ^{syn}$, we can also clearly see that the mechanism to syntactically generalize (specialize) $\mathcal{ALC}$ concepts C based on positive or negative occurrences of subconcepts is a fragment of the procedure to generalize (specialize) $\mathcal{ALC}$

[5] e.g., one where the size of concept names and all constructors are counted as 1

concepts C based on the role-depth and the signature of C. This implies that $\succ^{syt} \subset \succ^{b}$, which means that $\succ^{syt}$ provides less iterations to reach a gentle repair than $\succ^{b}$.

However, the well-foundedness, boundedness, and the completeness properties are still not enough to make sure that we can always compute maximally strongest weakenings w.r.t. $\succ^{b}$ and $\succ^{syt}$. It will be interesting to see whether these two weakening relations are effectively finitely branching, which means that it might be the case that we need to find a procedure for looking upper neighbors or lower neighbors of $\mathcal{ALC}$ concepts w.r.t. $\succ^{b}$ and $\succ^{syt}$.

Chapter 5

Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ Instance Stores

We have seen that repairing ontologies can be an alternative way to eliminate unwanted consequences, or particularly consequences that should be hidden when we deal with a privacy setting. In this context, we may say that the ontology has been compliant with privacy constraints which will be called *privacy policies* in the following. However, if information about individuals contained in the ontology should be publicly published, then one needs to be aware that *compliance* property per se is not enough. In previous sections, we may assume that the background knowledge of attackers is also a part of the input ontology. Within the publication phase of data transactions, this may be the case that a possible attacker can also obtain relevant information from other sources, which together with the published information might violate the privacy policy. *Safety* requires that the combination of the published information with any other compliant information is again compliant. If both *compliance* and *safety* are not satisfied, then the ontology needs to be modified in a minimal way. The latter condition is ensured by the *optimality* property.

All three properties above are stated in [GK16; GK19] which lay the foundations of privacy-preserving data publishing (PPDP), well-studied in [FWC+10], in the context of Linked Data. However, the papers do not consider the case where the information is augmented by background knowledge or an ontology. In this chapter, we make a first step towards *privacy-preserving ontology publishing* (PPOP), which leads us to one of the goals aiming at building an ontology that satisfies three aforementioned properties. Instead of proposing a framework that can yield such ontologies, as what we did in the previous chapter for the topic on ontology repair, we realize an initial step of studying PPOP by considering a setting where the ontology consists of an ABox containing only concept assertions. In [HLT+04], such an ABox is called an *instance store*. In addition, we assume that the ontology is written in the Description Logic $\mathcal{EL}$ and is not augmented by TBoxes. As a consequence, we may assume that all information about individuals are given by an $\mathcal{EL}$ concept. [1]

A privacy policy is given by an instance query, i.e., by an $\mathcal{EL}$ concept D. A concept C (giving information about some individual a) is *compliant* with this policy, if it is not subsumed by D, i.e., if $C(a)$ does not imply $D(a)$. In our example, the policy could be formalized as the $\mathcal{EL}$ concept.

$$D = \mathsf{Patient} \sqcap \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology}),$$

[1] Since $\mathcal{EL}$ concepts are closed under conjunction, we can assume that the ABox contains only one assertion for a.

which says that one should not be able to find out who are the patients that are seen by a doctor that works for the oncology department. The concept

$$C = \mathsf{Patient} \sqcap \mathsf{Male} \sqcap \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \mathsf{Female} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology})$$

is not compliant with the policy D since $C \sqsubseteq D$. The concept

$$C' = \mathsf{Male} \sqcap \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \mathsf{Female} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology})$$

is a compliant generalization of C, i.e., $C \sqsubseteq C'$ and $C' \not\sqsubseteq D$.

However, this concept C' may not be safe to publish if there is an attacker using his own background knowledge, which together with C' may issue again the fact that a is an instance of D. For example, if the attacker's knowledge is represented by an $\mathcal{EL}$ concept and he knows that a is an instance of a concept Patient, then together with $C'(a)$, the hidden information D is revealed, i.e., $C' \sqcap \mathsf{Patient} \sqsubseteq D$. In contrast,

$$C'' = \mathsf{Male} \sqcap \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top),$$

is a safe generalization of C, though it is less obvious to see this. This concept is, however, not optimal since more information than necessary is removed. In fact, the concept

$$\begin{aligned} C''' = \mathsf{Male} \sqcap\ & \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top) \sqcap \\ & \exists \mathsf{seen\_by}.(\mathsf{Female} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology}) \end{aligned}$$

is a safe generalization of C that is more specific than C'', i.e. $C \sqsubseteq C''' \sqsubset C''$.

In another case, if we assume that the knowledge of the attacker is now encoded in differents DLs, such as $\mathcal{FL}_0$ or $\mathcal{FLE}$, then C''' is no longer safe. Suppose that the attacker knows that a is an instance of an $\mathcal{FL}_0$ concept

$$E = \mathsf{Patient} \sqcap \forall \mathsf{seen\_by}.(\forall \mathsf{works\_in}.\mathsf{Oncology}),$$

which is not subsumed by D. This implies that $C''' \sqcap E \sqsubseteq D$, which again reveals the fact that a is an instance of D. To deal with this issue, one can compute the following optimal safe generalization $\widehat{C}$ of C for D such that it is safe whenever it is conjoined with other compliant $\mathcal{FL}_0$ concepts, where

$$\widehat{C} = \mathsf{Patient} \sqcap \mathsf{Male} \sqcap \exists \mathsf{seen\_by}.(\mathsf{Doctor} \sqcap \mathsf{Female})$$

Nevertheless, this concept $\widehat{C}$ may still be not safe for $\mathcal{P}$ if there is an attacker whose knowledge is written in $\mathcal{FLE}$ that is more expressive than $\mathcal{EL}$ and $\mathcal{FL}_0$. Suppose the attacker knows that a is an instance of an $\mathcal{FLE}$ concept

$$F = \forall \mathsf{seen\_by}.(\exists \mathsf{works\_in}.\mathsf{Oncology}).$$

The conjunction of $\widehat{C} \sqcap F$ may again issue the fact that a is an instance of D. In this setting, the results shown in this chapter imply that $\widetilde{C} = \mathsf{Male}$ is the optimal safe generalization of C w.r.t. the policy D.

Relating these simple settings with the general setting we have in Chapter 4, we can apply the weakening relation $\succ^{sub}$ when weakening concept assertions $C(a)$ by generalizing C. The algorithm in Proposition 4.18 obviously can be used to get a maximally strong weakening of $C(a)$. However, since the algorithm has non-elementary complexity for $\succ^{sub}$, we will present algorithms for computing such optimal compliant (safe) generalizations that run only either in PTIME or EXPTIME depending on the formulation of attackers' knowledge.

Furthermore, the absence of TBoxes in all the settings above are justified by the following reasons. First, we may argue that when investigating a new inference problem in DL, it is usually quite hard to start with the most general situation. So, we explore this setting as a starting point of learning the problem of PPOP. A second reason is that, in a medical application that uses, for instance, SNOMED CT as an ontology, the TBox can be reduced away by expanding concept definitions since SNOMED CT is an acyclic TBox [Sun09]. In addition, patient data are usually annotated with SNOMED concepts, but not with SNOMED roles, which justifies considering an instance store rather than a general ABox. Finally, considering $\mathcal{FL}_0$ concepts as one of languages formulating the attacker's knowledge makes sense since in SNOMED CT roles have implicit typing constraints [Sun09], which are not explicitly stated using value restrictions, but which may be known to an attacker.

We will distribute the discussions on this topic to the following sections. In Section 5.1, we begin the study of PPOP with the formalization of sensitive information in $\mathcal{EL}$ instance store. In Section 5.2 and Section 5.3, we characterize the notion of compliant (safe) $\mathcal{EL}$ concepts as well as compute optimal compliant (safe) generalizations of $\mathcal{EL}$ concepts. In addition, Section 5.4 views the optimality as a decision problem and investigate its complexity. It is still assumed that the knowledge of possible attackers is written in $\mathcal{EL}$ from Section 5.2 to Section 5.4. Next, we consider that the attacker has knowledge encoded in the DL $\mathcal{FL}_0$. With respect to this new setting, we also provide a new characterization for the safety problem in Section 5.5 and construct a different algorithm to compute optimal safe generalizations of $\mathcal{EL}$ concept, which is then continued by showing the complexity of the optimality problem in Section 5.6. This setting is then extended again with a condition where the information owned by the attacker is now written in the DL $\mathcal{FLE}$ covering the expressiveness of $\mathcal{EL}$ and $\mathcal{FL}_0$, which will be explained in Section 5.7.

5.1 Formalizing Sensitive Information in $\mathcal{EL}$ Instance Stores

In this section, we introduce formal definitions for compliance, safety, and optimality. In particular, for safety and optimality, we introduce three variants of their definitions based on DLs that formalize the attacker's knowledge. Moreover, instead of only considering one policy concept as shown in the patient example above, we allow for a finite set of $\mathcal{EL}$ concepts as a policy. In addition, We assume that the concepts occurring in the policy are not equivalent to top since otherwise there would not be compliant concepts.

Definition 5.1. *A* policy *is a finite set $\mathcal{P} = \{D_1, \ldots, D_p\}$ of $\mathcal{EL}$ concepts such that $\top \not\equiv D_i$ for $i = 1, \ldots, p$. Let C be an $\mathcal{EL}$ concept, $Q \in \{\exists, \forall, \forall\exists\}$ and $\mathcal{L}_\exists = \mathcal{EL}, \mathcal{L}_\forall = \mathcal{FL}_0, \mathcal{L}_{\forall\exists} = \mathcal{FLE}$. We say that*

- *the $\mathcal{L}_Q$ concept C' is compliant with $\mathcal{P}$ if $C' \not\sqsubseteq D_i$ for all $i = 1, \ldots, p$ and*
- *the $\mathcal{EL}$ concept C' is*

Computation Problems	**Q** $= \exists$	**Q** $= \forall$	**Q** $= \forall\exists$
Optimal $\mathcal{P}$-Compliant Generalization	ExpTime (Thm. 5.13)		
Optimal $\mathcal{P}$-Safe^{Q} Generalization	ExpTime (Cor. 5.20)	ExpTime (Thm. 5.39)	PTime (Thm. 5.44)

Table 5.3: Complexity of computing One optimal $\mathcal{P}$-compliant (safe^{Q}) generalization

- *a $\mathcal{P}$-compliant generalization of C if $C \sqsubseteq C'$ and C' is compliant with $\mathcal{P}$;*
- *an optimal $\mathcal{P}$-compliant generalization of C if it is a $\mathcal{P}$-compliant generalization of C and there is no $\mathcal{P}$-compliant generalization C'' of C such that $C'' \sqsubset C'$;*
- safe^{Q} *for $\mathcal{P}$ if for all $\mathcal{L}_Q$ concepts C'' that are compliant with $\mathcal{P}$, $C' \sqcap C''$ is also compliant with $\mathcal{P}$, i.e., $C' \sqcap C'' \not\sqsubseteq D_i$ for all $i = 1, \ldots, p$;*
- *a $\mathcal{P}$-*safe^{Q} generalization *of C if $C \sqsubseteq C'$ and C' is safe^{Q} for $\mathcal{P}$;*
- *an* optimal $\mathcal{P}$-safe^{Q} generalization *of C if it is a $\mathcal{P}$-safe generalization of C and there is no $\mathcal{P}$-safe generalization C'' of C such that $C'' \sqsubset C'$.* ♢

The compliance *problem asks whether C' is compliant with $\mathcal{P}$. The* safety^{Q} *problem asks whether C' is safe^{Q} for $\mathcal{P}$. If $Q' \in \{\forall, \forall\exists\}$, then we say that the* optimality$^{Q'}$ *problem asks whether C' is an optimal $\mathcal{P}$-safe$^{Q'}$ generalization of C, while the* optimality$^{\exists}$ *problem asks whether C' is an optimal $\mathcal{P}$-compliant (safe$^{\exists}$) generalization of C.*

It is easy to see that safety^{Q} implies compliance since the top concept is always compliant: if C' is safe^{Q} for $\mathcal{P}$, then $\top \sqcap C' \equiv C'$ is compliant.

We call an $\mathcal{EL}$ policy $\mathcal{P}$ *redundancy-free* if $\mathcal{P}$ does not contain distinct concepts D, D' such that $D \sqsubseteq D'$. Particularly, when we discuss about the safety^{Q} problem, without loss of generality, we can restrict our attention to redundancy-free policies since removing redundant concepts (i.e., concepts $D' \in \mathcal{P}$ such that there is $D \in \mathcal{P} \setminus \{D'\}$ with $D \sqsubseteq D'$) does not change the sets of compliant and safe concepts. This is justified by the following lemma that is easy to prove.

Lemma 5.2. *Let $\mathcal{P}$ be a policy, $Q \in \{\exists, \forall, \forall\exists\}$, and assume that $D_i \in \mathcal{P}$ is redundant. Then the following holds for all $\mathcal{EL}$ concepts C:*

- *C is compliant with $\mathcal{P}$ iff C is compliant with $\mathcal{P} \setminus \{D_i\}$;*
- *C is safe^{Q} for $\mathcal{P}$ iff C is safe^{Q} for $\mathcal{P} \setminus \{D_i\}$.*

In the following sections, we will show how to compute an optimal compliant (safe^{Q}) generalization. The complexity of algorithms for computing those generalizations are summarized in Table 5.3. Apart from that, we also consider the problems written in Definition 5.1 as a decision problem, the corresponding results for each of them are shown in Table 5.4.

Decision Problems	**Q** = ∃	**Q** = ∀	**Q** = ∀∃
compliance	PTIME (Prop. 5.6)		
safe^Q	PTIME (Thm. 5.16)	PTIME (Thm. 5.31)	PTIME (Thm. 5.44)
optimality^Q	CONP (Cor. 5.27)	CONP (Lem. 5.41)	PTIME (Thm. 5.44)

Table 5.4: Complexity results of decision problems on PPOP for $\mathcal{EL}$ instance stores

5.2 Computing Optimal Compliant Generalizations

In this section, we characterize the concepts that are compliant with a given policy $\mathcal{P}$, and use this to develop an algorithm that computes all optimal $\mathcal{P}$-compliant generalizations of a given $\mathcal{EL}$ concept C.

But first, we need to introduce some more notations. Let us recall that $\mathtt{con}(C)$ is the set of all atoms occurring in the top-level conjunction of concept C. As a special case of Lemma 2.22, subsumption between atoms can be characterized as follows. If E, F are atoms, then $E \sqsubseteq F$ iff

- $E = F \in \mathsf{N_C}$ or
- E, F are existential restrictions of the form $E = \exists r.E', F = \exists r.F'$ such that $E' \sqsubseteq F'$.

Definition 5.5. *Let S, T be sets of atoms. Then we say that S* covers *T if for every $F \in T$ there is $E \in S$ such that $E \sqsubseteq F$.* ♢

With this notation, Lemma 2.22 can be reformulated as follows: $C \sqsubseteq D$ iff $\mathtt{con}(C)$ covers $\mathtt{con}(D)$. The following (polynomial-time decidable) characterization of compliance is thus an immediate consequence of Lemma 2.22.

Proposition 5.6. *The $\mathcal{EL}$ concept C' is compliant with the policy $\mathcal{P} = \{D_1, \ldots, D_p\}$ iff $\mathtt{con}(C')$ does not cover $\mathtt{con}(D_i)$ for any $i = 1, \ldots, p$, i.e., for every $i = 1, \ldots, p$, at least one of the following two properties holds:*

- *there is a concept name $A \in \mathtt{con}(D_i)$ such that $A \notin \mathtt{con}(C')$; or*
- *there is an existential restriction $\exists r.D \in \mathtt{con}(D_i)$ such that $C \not\sqsubseteq D$ for all existential restrictions of the form $\exists r.C \in \mathtt{con}(C')$.* ♢

Now assume that we are given an $\mathcal{EL}$ concept C and a policy $\mathcal{P} = \{D_1, \ldots, D_p\}$, and we want to construct a $\mathcal{P}$-compliant generalization C' of C. For C' to satisfy the condition of Proposition 5.6, there needs to exist for every $i = 1, \ldots, p$ an element of $\mathtt{con}(D_i)$ that is not covered by any element of $\mathtt{con}(C')$. In case $\mathtt{con}(C)$ contains elements covering such an atom, we need to remove or generalize them appropriately.

Definition 5.7. *We call $H \subseteq con(D_1) \cup \ldots \cup con(D_p)$ a* hitting set *of $con(D_1), \ldots, con(D_p)$ if $H \cap con(D_i) \neq \emptyset$ for every $i = 1, \ldots, p$. This hitting set is* minimal *if there is no other hitting set strictly contained in it.* ♢

Basically, the idea is now to choose a hitting set H of $\mathtt{con}(D_1), \ldots, \mathtt{con}(D_p)$ and use H to guide the construction of a compliant generalization of C. In order to make this generalization as specific as possible, we use minimal hitting sets. In case the policy contains concepts D_i with which C is already compliant (i.e., $C \not\sqsubseteq D_i$ holds), nothing needs to be done w.r.t. these concepts. This is why, in the following definition, $\mathtt{con}(D_i)$ does not take part in the construction of the hitting set if $C \not\sqsubseteq D_i$.

Definition 5.8. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. The set $SCG(C, \mathcal{P})$ of* specific compliant generalizations *of C w.r.t. $\mathcal{P}$ consists of the concepts that can be constructed from C as follows:*

- *If C is compliant with $\mathcal{P}$, then $SCG(C, \mathcal{P}) = \{C\}$.*
- *Otherwise, choose a minimal hitting set H of $con(D_{i_1}), \ldots, con(D_{i_q})$ where $i_1, \ldots, i_q$ are exactly the indices for which $C \sqsubseteq D_i$. Note that $q \geq 1$ since we are in the case where C is not compliant with $\mathcal{P}$. In addition, according to our definition of a policy, none of the concepts D_i is equivalent to $\top$, and thus the sets $con(D_{i_j})$ are non-empty. Consequently, at least one minimal hitting set exists. Each minimal hitting set yields a concept in $SCG(C, \mathcal{P})$ by removing or modifying atoms in the top-level conjunction of C in the following way:*
  - *For every concept name $A \in con(C)$, remove A from the top-level conjunction of C if $A \in H$;*
  - *For every existential restriction $\exists r_i.C_i \in con(C)$, consider the set*

$$\mathcal{P}_i := \{G \mid \text{ there is } \exists r_i.G \in H \text{ such that } C_i \sqsubseteq G\}.$$

    - *If $\mathcal{P}_i = \emptyset$ then leave $\exists r_i.C_i$ as it is.*
    - *If $\top \in \mathcal{P}_i$, then remove $\exists r_i.C_i$.*
    - *Otherwise, replace $\exists r_i.C_i$ with $\bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$.* ♢

We will show below that every element of $SCG(C, \mathcal{P})$ is a compliant generalization of C, and that all optimal compliant generalizations of C belong to $SCG(C, \mathcal{P})$. However, $SCG(C, \mathcal{P})$ may also contain compliant generalizations of C that are not optimal, as illustrated by the following example.

Example 5.9. *Let $C = \exists r.(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4)$ and $\mathcal{P} = \{D_1, D_2\}$, where*

$$D_1 = \exists r.A_1 \sqcap \exists r.(A_2 \sqcap A_3) \text{ and } D_2 = \exists r.A_2 \sqcap \exists r.A_4.$$

We have $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$, and thus C is not compliant with $\mathcal{P}$. Consequently, the elements of $SCG(C, \mathcal{P})$ are obtained by considering the minimal hitting sets of $\{\exists r.A_1, \exists r.(A_2 \sqcap A_3)\}$ and $\{\exists r.A_2, \exists r.A_4\}$.

If we take the minimal hitting set $H = \{\exists r.(A_2 \sqcap A_3), \exists r.A_2\}$ and consider the only existential restriction in $con(C)$, the corresponding set $\mathcal{P}_i$ consists of $A_2 \sqcap A_3$ and A_2. It is easy to see that $SCG(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4, \mathcal{P}_i) = \{A_1 \sqcap A_3 \sqcap A_4\}$ since the only minimal hitting set of $\{A_1, A_2\}$ and $\{A_2\}$ is $\{A_2\}$. Thus, we obtain $C' := \exists r.(A_1 \sqcap A_3 \sqcap A_4)$ as an element of $SCG(C, \mathcal{P})$.

However, if we take the minimal hitting set $H' = \{\exists r.A_1, \exists r.A_2\}$ instead, then the set $\mathcal{P}'_i$ corresponding to the only existential restriction in $\mathtt{con}(C)$ is $\{A_1, A_2\}$. Consequently, in this case $SCG(A_1 \sqcap A_2 \sqcap A_3 \sqcap A_4, \mathcal{P}'_i) = \{A_3 \sqcap A_4\}$ since the only minimal hitting set of $\{A_1\}$ and $\{A_2\}$ is $\{A_1, A_2\}$. This yields $C'' := \exists r.(A_3 \sqcap A_4)$ as another element of $SCG(C, \mathcal{P})$. Since $C' \sqsubset C''$, the element C'' cannot be optimal. ♢

Next, we show that the elements of $SCG(C, \mathcal{P})$ are compliant generalizations of C.

Proposition 5.10. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. If $C' \in SCG(C, \mathcal{P})$, then C' is a $\mathcal{P}$-compliant generalization of C.* ♢

Proof. In case C is already compliant with $\mathcal{P}$, then $C = C'$ and we are done. Thus, assume that C is not compliant with $\mathcal{P}$. We show that C' is a compliant generalization of C by induction on the role depth of C.

First, we show that C' is a generalization of C, i.e., $C \sqsubseteq C'$. This is an easy consequence of the fact that, when constructing C' from C, atoms from the top-level conjunction of C are left unchanged, are removed, or are replaced by a conjunction of more general atoms. The only non-trivial case is where we replace an existential restriction $\exists r_i.C_i$ with the conjunction $\bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$. By induction, we know that $C_i \sqsubseteq F$ for all $F \in SCG(C_i, \mathcal{P}_i)$, and thus $\exists r_i.C_i \sqsubseteq \bigsqcap_{F \in SCG(C_i, \mathcal{P}_i)} \exists r_i.F$.

Second, we show that C' is compliant with $\mathcal{P}$, i.e., $C' \not\sqsubseteq D_i$ holds for $i = 1, \ldots, p$. For the indices i with $C \not\sqsubseteq D_i$, we clearly also have $C' \not\sqsubseteq D_i$ since $C \sqsubseteq C'$. Now, consider one of the remaining indices $i_j \in \{i_1, \ldots, i_q\}$, where $i_1, \ldots, i_q$ are exactly the indices for which $C \sqsubseteq D_i$. The concept C' was constructed by taking some minimal hitting set H of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$. If the element in H hitting $\mathtt{con}(D_{i_j})$ is a concept name, then this concept name does not occur in $\mathtt{con}(C')$, and thus $C' \not\sqsubseteq D_{i_j}$. Thus, assume that it is an existential restriction $\exists r_i.G$. But then each existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ with $C_i \sqsubseteq G$ is either removed or replaced by a conjunction of existential restrictions $\exists r_i.F$ such that (by induction) $F \not\sqsubseteq G$. In addition, other existential restrictions are either removed or generalized. This clearly implies $C' \not\sqsubseteq D_{i_j}$ since $\exists r_i.G$ in $\mathtt{con}(D_{i_j})$ is not covered by any element of $\mathtt{con}(C')$. □

The next lemma states that every compliant generalization of C subsumes some element of $SCG(C, \mathcal{P})$.

Lemma 5.11. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. If C'' is a $\mathcal{P}$-compliant generalization of C, then there is $C' \in SCG(C, \mathcal{P})$ such that $C' \sqsubseteq C''$.*

Proof. If C is compliant with $\mathcal{P}$, then we have $C \in SCG(C, \mathcal{P})$ and $C \sqsubseteq C''$ since C'' is a generalization of C. Thus, assume that C is not compliant with $\mathcal{P}$, and let $i_1, \ldots, i_q$ be exactly the indices for which $C \sqsubseteq D_i$.

Now, let i_j be such an index. We have $C \sqsubseteq C'' \not\sqsubseteq D_{i_j}$ and $C \sqsubseteq D_{i_j}$. Since $C'' \not\sqsubseteq D_{i_j}$, there is an element $E_j \in \mathtt{con}(D_{i_j})$ that is not covered by any element of $\mathtt{con}(C'')$. Obviously, $H'' := \{E_1, \ldots, E_q\}$ is a hitting set of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$. Thus, there is a minimal hitting set H of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$ such that $H \subseteq H''$. Let C' be the element of $SCG(C, \mathcal{P})$ that was constructed using this hitting set H. We claim that $C' \sqsubseteq C''$. For this, it is sufficient to show that $\mathtt{con}(C')$ covers $\mathtt{con}(C'')$.

First, consider a concept name $A \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, we also have $A \in \mathtt{con}(C)$. If $A \notin H''$, then $A \notin H$, and thus A is not removed in the construction of C'. Consequently, $A \in \mathtt{con}(C')$ covers $A \in \mathtt{con}(C'')$. If $A \in H''$, then A is not covered by any element of $\mathtt{con}(C'')$ according to our definition of H'', which contradicts our assumption that $A \in \mathtt{con}(C'')$.

Second, consider an existential restriction $\exists r_i.E \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, there is an existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ such that $C_i \sqsubseteq E$. If this restriction is not removed or generalized when constructing C', then we are done since this restriction then belongs to $\mathtt{con}(C')$ and covers $\exists r_i.E$. Otherwise, $\mathcal{P}_i = \{G \mid \text{ there is } \exists r_i.G \in H \text{ such that } C_i \sqsubseteq G\}$ is non-empty.

If $\top \in \mathcal{P}_i$, then $\exists r_i.\top \in H \subseteq H''$. However, then $\exists r_i.E \in \mathtt{con}(C'')$ covers an element of H'', which is a contradiction.

Consequently, $\top \notin \mathcal{P}_i$, and thus $\exists r_i.C_i$ is replaced with $\bigsqcap_{F \in SCG(C_i,\mathcal{P}_i)} \exists r_i.F$ when constructing C' from C. According to our definition of H'' and the fact that $H \subseteq H''$, none of the existential restrictions $\exists r_i.G$ considered in the definition of $\mathcal{P}_i$ is covered by $\exists r_i.E \in \mathtt{con}(C'')$. This implies that E is a $\mathcal{P}_i$-compliant generalization of C_i. By induction (on the role depth) we can thus assume that there is an $F \in SCG(C_i, \mathcal{P}_i)$ such that $F \sqsubseteq E$. This shows that $\exists r_i.E \in \mathtt{con}(C'')$ is covered by $\exists r_i.F \in \mathtt{con}(C')$. □

As an easy consequence of this lemma, we obtain that all optimal compliant generalizations of C must belong to $SCG(C, \mathcal{P})$.

Proposition 5.12. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. If C'' is an optimal $\mathcal{P}$-compliant generalization of C, then $C'' \in SCG(C, \mathcal{P})$ (up to equivalence of concepts).* ♢

Proof. Let C'' be an optimal $\mathcal{P}$-compliant generalization of C. By Lemma 5.11, there is an element $C' \in SCG(C, \mathcal{P})$ such that $C' \sqsubseteq C''$. In addition, by Proposition 5.10, C' is a $\mathcal{P}$-compliant generalization of C. Thus, optimality of C'' implies $C'' \equiv C'$.

We are now ready to formulate and prove the main result of this section.

Theorem 5.13. *Let C be an $\mathcal{EL}$-concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a policy. Then the set of all optimal $\mathcal{P}$-compliant generalizations of C can be computed in time exponential in the size of C and $D_1, \ldots, D_p$.*

Proof. It is sufficient to show that the set $SCG(C, \mathcal{P})$ can be computed in exponential time. In fact, given $SCG(C, \mathcal{P})$, we can compute the set of all optimal $\mathcal{P}$-compliant generalizations of C by removing elements that are not minimal w.r.t. subsumption, which requires at most exponentially many subsumption tests. Each subsumption test takes at most exponential time since subsumption in $\mathcal{EL}$ is in PTIME, and the elements of $SCG(C, \mathcal{P})$ have at most exponential size, as shown below.

We show by induction on the role depth that $SCG(C, \mathcal{P})$ consists of at most exponentially many elements of at most exponential size. The at most exponential cardinality of $SCG(C, \mathcal{P})$ is an immediate consequence of the fact that there are at most exponentially many hitting sets of $\mathtt{con}(D_{i_1}), \ldots, \mathtt{con}(D_{i_q})$, and each yields exactly one element of $SCG(C, \mathcal{P})$ (see Definition 5.8). Regarding the size of these elements, note that we may assume by induction that an existential restriction may be replaced by a conjunction of at most exponentially many existential restrictions, where each is of at most exponential size. The overall size of the concept description obtained this way is thus also of at most exponential size. Given this, it is easy to see that the computation of these elements also takes at most exponential time.□

The following example shows that the exponential upper bounds can indeed by reached.

Example 5.14. *Let $C = P_1 \sqcap Q_1 \sqcap \ldots \sqcap P_n \sqcap Q_n$ and $\mathcal{P} = \{P_i \sqcap Q_i \mid 1 \leq i \leq n\}$. Then $SCG(C, \mathcal{P})$ contains 2^n elements since the sets $\{P_1, Q_1\}, \ldots, \{P_n, Q_n\}$ obviously have exponentially many hitting sets. To be more precise,*

$$SCG(C, \mathcal{P}) = \{X_1 \sqcap \ldots \sqcap X_n \mid X_i \in \{P_i, Q_i\} \text{ for } i = 1, \ldots, n\}.$$

This example can easily be modified to enforce an element of exponential size. Consider $\widehat{C} = \exists r.C$ and $\widehat{\mathcal{P}} = \{\exists r.(P_i \sqcap Q_i) \mid 1 \leq i \leq n\}$. Then $SCG(\widehat{C}, \widehat{\mathcal{P}}) = \{\bigsqcap_{F \in SCG(C,\mathcal{P})} \exists r.F\}$. We leave it to the reader to further modify the example in order to obtain exponentially many elements of exponential size. ♢

5.3 Computing Optimal Safe$^\exists$ Generalizations

As mentioned at the end of Section 5.1, since we will investigate the safety$^\exists$ problem, we assume that the policy used in this section is redundant-free. Then, we prove the following proposition which states a characterization for safety$^\exists$.

Proposition 5.15. *Let $\mathcal{P} = \{D_1, \ldots, D_p\}$ be a redundancy-free policy. The $\mathcal{EL}$ concept C' is safe$^\exists$ for $\mathcal{P}$ iff there is no pair of atoms (E, F) such that $E \in con(C')$, $F \in con(D_1) \cup \ldots \cup con(D_p)$, and $E \sqsubseteq F$.* ♢

Proof. First, assume that C' is not safe$^\exists$ for $\mathcal{P}$, i.e., there is an $\mathcal{EL}$ concept C'' that is compliant with $\mathcal{P}$, but for which $C' \sqcap C''$ is not compliant with $\mathcal{P}$. The latter implies that there is $D_i \in \mathcal{P}$ such that $C' \sqcap C'' \sqsubseteq D_i$, which is equivalent to saying that $\mathtt{con}(C') \cup \mathtt{con}(C'')$ covers $\mathtt{con}(D_i)$. On the other hand, we know that $\mathtt{con}(C'')$ does not cover $\mathtt{con}(D_i)$ since C'' is compliant with $\mathcal{P}$. Thus, there is an element $F \in \mathtt{con}(D_i)$ that is covered by an element E of $\mathtt{con}(C')$. This yields (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and $E \sqsubseteq F$.

Conversely, assume that there is a pair of atoms (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_i)$, and $E \sqsubseteq F$. Let C'' be the concept obtained from D_i by removing F from the top-level conjunction of D_i. Then we clearly have $D_i \sqsubseteq C''$. In addition, since D_i is normalized, we also have $C'' \not\sqsubseteq D_i$. Consider $D_j \in \mathcal{P}$ different from D_i, and assume that $C'' \sqsubseteq D_j$. But then $D_i \sqsubseteq C'' \sqsubseteq D_j$ contradicts our assumption that $\mathcal{P}$ does not contain redundant elements. Thus, we have shown that C'' is compliant with $\mathcal{P}$. In addition, $\mathtt{con}(C') \cup \mathtt{con}(C'')$ covers $\mathtt{con}(D_i)$. In fact, the elements of $\mathtt{con}(D_i) \setminus \{F\}$ belong to $\mathtt{con}(C'')$, and thus cover themselves. In addition, F is covered by $E \in \mathtt{con}(C')$. Thus $C' \sqcap C'' \sqsubseteq D_i$, which shows that C' is not safe$^\exists$ for $\mathcal{P}$. □

Clearly, the necessary and sufficient condition for safety$^\exists$ stated in this proposition can be decided in polynomial time. If needed, the policy can first be made redundancy-free, which can also be done in polynomial time.

Theorem 5.16. *Safety$^\exists$ of an $\mathcal{EL}$ concept for an $\mathcal{EL}$ policy is in P.*

We now consider the problem of computing optimal $\mathcal{P}$-safe generalizations of a given $\mathcal{EL}$ concept C. First note that, up to equivalence, there can be only one optimal $\mathcal{P}$-safe generalization of C. This is an immediate consequence of the fact that the conjunction of safe$^\exists$ concepts is again safe, which in turn is an easy consequence of Proposition 5.15.

Lemma 5.17. *Let C_1', C_2' be two $\mathcal{EL}$ concepts that are $\mathcal{P}$-safe$^\exists$ generalizations of C, where $\mathcal{P}$ is redundancy-free. Then $C_1' \sqcap C_2'$ is also a $\mathcal{P}$-safe$^\exists$ generalization of C.*

Thus there cannot be non-equivalent optimal $\mathcal{P}$-safe$^\exists$ generalizations of a given $\mathcal{EL}$ concept C since their conjunction would then be more specific, contradicting their optimality. This property is independent of whether the policy is redundancy-free or not since turning a policy into one that is redundancy-free preserves the set of concepts that are compliant with (safe$^\exists$ for) the policy.

Proposition 5.18. *If C_1', C_2' are optimal $\mathcal{P}$-safe$^\exists$ generalizations of the $\mathcal{EL}$ concept C, then $C_1' \equiv C_2'$.* ♢

The following theorem shows how an optimal safe$^\exists$ generalization of C can be constructed.

Theorem 5.19. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. We construct the concept C' from C by removing or modifying atoms in the top-level conjunction of C in the following way:*

- *For every concept name $A \in \mathit{con}(C)$, remove A from the top-level conjunction of C if $A \in \mathit{con}(D_1) \cup \ldots \cup \mathit{con}(D_p)$;*
- *For every existential restriction $\exists r_i.C_i \in \mathit{con}(C)$, consider the set of concepts*

$$\mathcal{P}_i := \{G \mid \text{there is } \exists r_i.G \in \mathit{con}(D_1) \cup \ldots \cup \mathit{con}(D_p) \text{ such that } C_i \sqsubseteq G\}.$$

  - *If $\mathcal{P}_i = \emptyset$ then leave $\exists r_i.C_i$ as it is.*
  - *If $\top \in \mathcal{P}_i$, then remove $\exists r_i.C_i$.*
  - *Otherwise, replace $\exists r_i.C_i$ with $\bigsqcap_{F \in OCG(C_i, \mathcal{P}_i)} \exists r_i.F$, where $OCG(C_i, \mathcal{P}_i)$ is the set of all optimal $\mathcal{P}_i$-compliant generalizations of C_i.*

Then C' is an optimal $\mathcal{P}$-safe$^\exists$ generalization of C.

Proof. Obviously $C \sqsubseteq C'$ since, when constructing C' from C, atoms from the top-level conjunction of C are left unchanged, are removed, or are replaced by a conjunction of more general atoms.

To show that C' is safe$^\exists$ for $\mathcal{P}$, we must show that the condition of Proposition 5.15 holds. Thus assume that it is violated, i.e., there is a pair of atoms (E, F) such that $E \in \mathtt{con}(C')$, $F \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and $E \sqsubseteq F$.

- First, we consider the case where $E = A$ is a concept name. Then $E \sqsubseteq F$ implies that $F = A$, and thus A is a concept name occurring in $\mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. However, all such concept names have been removed from the top-level conjunction of C when constructing C'. This contradicts our assumption that $E = A$ belongs to $\mathtt{con}(C')$.
- Second, assume that E is an existential restriction $E = \exists r_i.E'$. Then F is of the form $F = \exists r_i.G'$ and $E' \sqsubseteq G'$. In addition, there is an existential restriction $\exists r_i.C_i \in \mathtt{con}(C)$ from which $E = \exists r_i.E'$ was derived. By construction, $C_i \sqsubseteq E'$. In the construction of C', we consider the set $\mathcal{P}_i := \{G \mid \text{there is } \exists r_i.G \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ such that } C_i \sqsubseteq G\}$. Since $C_i \sqsubseteq E' \sqsubseteq G'$, this set is non-empty, and since $\exists r_i.E'$ is derived from $\exists r_i.C_i$, it does not contain $\top$. Consequently, we have $E' \in OCG(C_i, \mathcal{P}_i)$. However, $G' \in \mathcal{P}_i$ then implies that $E' \not\sqsubseteq G'$, which yields the desired contradiction.

It remains to show that C' is optimal. Thus assume that C'' is a $\mathcal{P}$-safe$^\exists$ generalization of C. It is sufficient to show that $C' \sqsubseteq C''$, i.e., that $\mathtt{con}(C')$ covers $\mathtt{con}(C'')$.

- Assume that $A \in \mathtt{con}(C'')$ is a concept name. Then $C \sqsubseteq C''$ implies that $A \in \mathtt{con}(C)$. Since C'' is safe$^\exists$ for $\mathcal{P}$, Proposition 5.15 implies that $A \notin \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. Thus, A is not removed in the construction of C', which yields $A \in \mathtt{con}(C')$.
- Second, consider an existential restriction $\exists r_i.E \in \mathtt{con}(C'')$. Since $C \sqsubseteq C''$, there is an existential restriction $\exists r_i.C_i$ in $\mathtt{con}(C)$ such that $C_i \sqsubseteq E$. If this restriction is not removed or generalized when constructing C', then we are done since this restriction then belongs to $\mathtt{con}(C')$ and covers $\exists r_i.E$. Otherwise,

  $$\mathcal{P}_i = \{G \mid \text{ there is } \exists r_i.G \ \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p) \text{ such that } C_i \sqsubseteq G\}$$

  is non-empty. If $\top \in \mathcal{P}_i$, then $\exists r_i.\top \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$. However, then $\exists r_i.E \in \mathtt{con}(C'')$ covers an element of $\mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, which is a contradiction to our assumption that C'' is safe$^\exists$ for $\mathcal{P}$. Consequently, $\top \notin \mathcal{P}_i$, and thus $\exists r_i.C_i$ is replaced with $\bigsqcap_{F \in OCG(C_i,\mathcal{P}_i)} \exists r_i.F$ when constructing C' from C. Since C'' is safe$^\exists$ for $\mathcal{P}$, none of the existential restrictions $\exists r_i.G$ considered in the definition of $\mathcal{P}_i$ is covered by $\exists r_i.E \in \mathtt{con}(C'')$. This implies that E is a $\mathcal{P}_i$-compliant generalization of C_i. Consequently, there is an $F \in OCG(C_i, \mathcal{P}_i)$ such that $F \sqsubseteq E$. This shows that $\exists r_i.E \in \mathtt{con}(C'')$ is covered by $\exists r_i.F \in \mathtt{con}(C')$. □

Since, by Theorem 5.13, $OCG(C_i, \mathcal{P}_i)$ can be computed in exponential time, the construction described in Theorem 5.19 can also be performed in exponential time.

Corollary 5.20. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. Then an optimal $\mathcal{P}$-safe$^\exists$ generalization of C can be computed in exponential time.*

Example 5.14 can easily be modified to provide an example that shows that this exponential bound can actually not be improved since there are cases where the safe$^\exists$ generalization is of exponential size.

5.4 Deciding Optimality$^\exists$ in $\mathcal{EL}$ Instance Stores

In this section, we consider *optimality$^\exists$ as a decision problem*, i.e., given $\mathcal{EL}$ concepts C, C' such that $C \sqsubseteq C'$ and a policy $\mathcal{P}$, decide whether C' is an optimal $\mathcal{P}$-compliant ($\mathcal{P}$-safe$^\exists$) generalization of C.

Theorem 5.13 and Corollary 5.20 show that the optimality problem is *in* ExpTime both for compliance and for safety. In fact, according to Theorem 5.13, given C and $\mathcal{P}$, we can compute the set of all optimal $\mathcal{P}$-compliant generalizations of C (up to equivalence) in exponential time. Consequently, this set contains at most exponentially many elements and each element has at most exponential size. This implies that we can test, in exponential time, whether a give concept C' is equivalent to one of the elements of this set. If this is the case, then C' is an optimal $\mathcal{P}$-compliant generalization of C, and otherwise not. The case of safety$^\exists$ can be treated similarly, using Corollary 5.20 instead of Theorem 5.13.

In the following, we show that this complexity upper bound can be improved to coNP. Actually, we will prove this upper bound not just for compliance and safety$^\exists$, but for a whole class of properties.

Definition 5.21. *Let F be a function that assigns a set of $\mathcal{EL}$ concepts to every input consisting of an $\mathcal{EL}$ concept C and a policy $\mathcal{P}$. We say that the function F defines a* polynomial, upward-closed property *if the following holds for every input $C, \mathcal{P}$:*

- *for every $\mathcal{EL}$ concept C', we can decide $C' \in F(C, \mathcal{P})$ in time polynomial in $C, C', \mathcal{P}$ (polynomiality);*
- *if $C' \in F(C, \mathcal{P})$ and $C' \sqsubseteq C''$, then $C'' \in F(C, \mathcal{P})$ (upward-closedness).*

We say that C' is an optimal F-generalization of C w.r.t. $\mathcal{P}$ *if $C \sqsubseteq C'$, $C' \in F(C, \mathcal{P})$, and there is no $C \sqsubseteq C'' \sqsubset C'$ such that $C'' \in F(C, \mathcal{P})$.* ♢

It is easy to see that compliance and safety$^\exists$ are polynomial, upward-closed properties. In fact, upward-closedness is an obvious consequence of the definition of compliance (safety$^\exists$). For compliance, polynomiality follows from the fact that subsumption in $\mathcal{EL}$ can be decided in polynomial time. For safety$^\exists$, it is stated in Corollary 5.16. In addition, the notion of optimality$^\exists$ introduced in the above definition coincides with the notion of optimality$^\exists$ introduced in Definition 5.1 for compliance and safety$^\exists$.

We will show that, for polynomial, upward-closed properties, the optimality$^\exists$ problem is in coNP, i.e., there is an NP-*algorithm* that, on input $C \sqsubseteq C'$ and $\mathcal{P}$, succeeds iff C' is *not* an optimal F-generalization of C w.r.t. $\mathcal{P}$. Basically, this algorithm proceeds as follows. It guesses a lower neighbor C'' of C' subsuming C, i.e., a concept C'' such that (i) $C \sqsubseteq C'' \sqsubseteq C'$ and (ii) there is no concept C''' with $C'' \sqsubset C''' \sqsubset C'$. If $C'' \in F(C, \mathcal{P})$, then the algorithm succeeds, and otherwise it fails.

In Section 4.4.1, we have seen that the relation $\sqsubseteq$ is *one-step generated*, i.e., the transitive closure of $\sqsubseteq_1$ is again $\sqsubseteq$ and defined the notion of *upper neighbor* which is obviously the converse of *lower neighbor*, i.e., if $C'' \sqsubset_1 C'$ then we call C' an *upper neighbor* of C'' and C'' a *lower neighbor* of C'. In the context of the optimality problem for polynomial, upward-closed properties, this implies the following: whenever there is a counterexample to the optimality of C' (i.e., a concept C'' such that $C \sqsubset C'' \sqsubset C'$ and $C'' \in F(C, \mathcal{P})$), then there is a lower neighbor of C' that provides such a counterexample. To see this, just note that $C'' \sqsubset C'$ implies that C' can be reached by a $\sqsubset_1$-chain from C''. The last element in this chain before C' is a lower neighbor of C', and it belongs to $F(C, \mathcal{P})$ since F is upward-closed. Then, it is also mentioned in Lemma 4.24 that a given $\mathcal{EL}$ concept has only polyomially many upper neighbors, each of which is polynomial size.

Regarding lower neighbors, it is sufficient for our purposes to show that they can be guessed in non-deterministic polynomial time. Thus, we are looking for an NP-algorithm that, given input concepts $C \sqsubseteq C'$, generates exactly the lower neighbors of C' that subsume C. Below, we sketch how an appropriate NP-algorithm can be obtained. A more detailed description as well as *proofs can be found in [Kri18].* First, note that the lower neighbors C'' of C' can be obtained by conjoining an atom not implied by C' to C'. In addition, $C \sqsubseteq C''$

implies that $sig(C'') \subseteq sig(C)$. Given an $\mathcal{EL}$ concept C' and a finite set Σ of concept and role names, the set of *lowering atoms* for C' w.r.t Σ is defined as

$$\begin{aligned} LA_\Sigma(C') \quad := \quad & \{A \in \Sigma \cap \mathsf{N_C} \mid A \notin \mathtt{con}(C')\} \cup \{\exists r.D \mid r \in \Sigma \cap \mathsf{N_R}, sig(D) \subseteq \Sigma, \\ & C' \not\sqsubseteq \exists r.D, \text{ and } C' \sqsubseteq \exists r.E \text{ for all } E \text{ with } D \sqsubset_1 E\}. \end{aligned}$$

Lemma 5.22. *Let C' be an $\mathcal{EL}$ concept and Σ a finite set of concept and role names with $sig(C') \subseteq \Sigma$. Then C'' is a lower neighbor of C' with $sig(C'') \subseteq \Sigma$ iff there is an atom $\mathsf{At} \in LA_\Sigma(C')$ such that $C'' \equiv C' \sqcap \mathsf{At}$.*

Intuitively, adding a single atom to the top-level conjunction of C' is sufficient to obtain a lower neighbor since adding two (non-redundant) atoms would step too far down in the subsumption hierarchy. The same is true for adding an existential restriction $\exists r.D$ for which $\exists r.E$ with $D \sqsubset_1 E$ does not subsume C' since then $C' \sqcap \exists r.D \sqsubset C' \sqcap \exists r.E \sqsubset C'$ would hold.

Example 5.23. *Let $\Sigma := \{r, A_1, A_2, B_1, B_2, C_1, C_2\}$ and*

$$C' := \exists r.(A_1 \sqcap A_2 \sqcap B_1 \sqcap B_2) \sqcap \exists r.(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2) \sqcap \exists r.(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2).$$

Then, for all $i, j, k \in \{1, 2\}$, the existential restriction $\exists r.D$ with $D := A_i \sqcap B_j \sqcap C_k$ belongs to $LA_\Sigma(C')$. In fact, $C' \not\sqsubseteq \exists r.D$ is obviously true, and since the upper neighbors of D are $A_i \sqcap B_j$, $B_j \sqcap C_k$, and $A_i \sqcap C_k$, we also have $C' \sqsubseteq \exists r.E$ for all E with $D \sqsubset_1 E$. Obviously, by using n instead of three pairs of concept names, we can produce a generalized version of this example that shows that the cardinality of $LA_\Sigma(C')$ can be exponential in the size of C' and Σ. ♢

In order to obtain an NP-algorithm that generates exactly the lower neighbors of C' that subsume C, it is sufficient to generate all lowering atoms for C' w.r.t $\Sigma := sig(C)$, and then remove the ones that do not subsume C. Unfortunately, the definition of lowering atoms given above Lemma 5.22 does not tell us directly how appropriate existential restrictions $\exists r.D$ can be found. The following necessary conditions follows from the characterization of lower neighbors given in [Kri18].

Lemma 5.24. *Let C' be reduced. If $\exists r.D \in LA_\Sigma(C')$, then there is a set of existential restrictions $\{\exists r.F'_1, \ldots, \exists r.F'_k\} \subseteq \mathtt{con}(C')$ and $F_1 \in LA_\Sigma(F'_1), \ldots, F_k \in LA_\Sigma(F'_k)$ such that $D \equiv F_1 \sqcap \ldots \sqcap F_k$.*

We illustrate this lemma using the lowering atom $D = A_i \sqcap B_j \sqcap C_k$ in Example 5.23. Here we take the set of all existential restrictions in $\mathtt{con}(C')$ and choose $C_k \in LA_\Sigma(A_1 \sqcap A_2 \sqcap B_1 \sqcap B_2)$, $B_j \in LA_\Sigma(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2)$, and $A_i \in LA_\Sigma(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2)$. Obviously, D is indeed equivalent to the conjunction of these three atoms.

In general, not all choices of subsets and lower neighbors yields an appropriate existential restriction. For instance, if we take a smaller set of existential restrictions in our example (e.g., $\{\exists r.(A_1 \sqcap A_2 \sqcap C_1 \sqcap C_2), \exists r.(B_1 \sqcap B_2 \sqcap C_1 \sqcap C_2)\}$), then the obtained conjunction of lowering atoms (e.g., $B_1 \sqcap A_2$) is not appropriate since the corresponding existential restriction (e.g., $\exists r.(B_1 \sqcap A_2)$) is subsumed by C'.

The *NP-algorithm* generating exactly the elements of $LA_\Sigma(C')$ works as follows: given a reduced concept C' and a finite set Σ of concept and role names such that $sig(C') \subseteq \Sigma$, it non-deterministically chooses one of the following two alternatives:

1. Choose a concept name $A \in \Sigma \setminus \mathtt{con}(C')$, and output A. If there is no such concept name, fail.
2. Choose $r \in \Sigma \cap \mathsf{N_R}$, a set of existential restrictions $\{\exists r.F'_1, \ldots, \exists r.F'_k\} \subseteq \mathtt{con}(C')$, and recursively guess elements $F_1 \in LA_\Sigma(F'_1), \ldots, F_k \in LA_\Sigma(F'_k)$. If for some $i, 1 \leq i \leq k$, the attempt to produce the atom $F_i \in LA_\Sigma(F'_i)$ fails, or if $C' \sqsubseteq \exists r.(F_1 \sqcap \ldots \sqcap F_k)$, or if $F_1 \sqcap \ldots \sqcap F_k$ has an upper neighbor E such that $C' \not\sqsubseteq \exists r.E$, then fail. Otherwise, output $\exists r.(F_1 \sqcap \ldots \sqcap F_k)$.

Lemma 5.25. *The algorithm described above runs in non-deterministic polynomial time, and its non-failing runs produce exactly the elements of $LA_\Sigma(C')$.*

Proof. Soundness of the algorithm is an immediate consequence of the fact that, in the second case, we explicitly test whether the conditions in the definition of lowering atoms are satisfied. Completeness is an easy consequence of Lemma 5.24. Finally, the choice of a concept name, a role name, and a subset of the existential restrictions in $\mathtt{con}(C')$, can clearly be achieved by making polynomially many binary choices. By induction on the role depth, we can assume that the algorithm can produce the elements $F_i \in LA_\Sigma(F'_i)$ in non-deterministic polynomial time, which shows that the overall algorithm runs in non-deterministic polynomial time. □

With this lemma in place, we can now show that the optimality problem for polynomial, upward-closed properties is in coNP.

Theorem 5.26. *Let F be a polynomial, upward-closed property. The problem of deciding, for a given input $C, C', \mathcal{P}$, whether C' is an optimal F-generalization of C w.r.t. $\mathcal{P}$ is in coNP.*

Proof. We show that non-optimality can be decided by an NP-algorithm, i.e., we describe an NP-algorithm that, given $C, C', \mathcal{P}$, succeeds iff C' is *not* an optimal F-generalization of C w.r.t. $\mathcal{P}$.

1. Check whether $C \sqsubseteq C'$ and $C' \in F(C, \mathcal{P})$. If this is not the case, then succeed. Otherwise, continue with the next step. Polynomiality of F and of subsumption in $\mathcal{EL}$ implies that this test can be done in polynomial time.
2. Set $\Sigma := sig(C)$ and guess a lowering atom $\mathsf{At} \in LA_\Sigma(C')$. If $C \not\sqsubseteq \mathsf{At}$, then fail. Otherwise, we know that $C'' := C' \sqcap \mathsf{At}$ is a lower neighbor of C' that subsumes C, and we continue with the next step. As shown above, the elements of $LA_\Sigma(C')$ can be generated by an NP-algorithm.
3. Check whether $C'' \in F(C, \mathcal{P})$. If this is the case, then succeed, and otherwise fail.

It is easy to see that this algorithm is correct and runs in non-deterministic polynomial time.□

Since compliance and safety are polynomial, upward-closed properties, the following corollary is an immediate consequence of this theorem.

Corollary 5.27. *The optimality$^\exists$ problem is in* coNP *for compliance and for safety.*

At the moment, we do not know whether these problems are also coNP-hard. We can show, however, that the Hypergraph Duality Problem [EG02] can be reduced to them. Note that this problem is in coNP, but conjectured to be neither in P nor coNP-hard [FK96; GM18]. Given two finite families of inclusion-incomparable sets $\mathcal{G}$ and $\mathcal{H}$, the *Hypergraph Duality Problem* (DUAL) asks whether $\mathcal{H}$ consists exactly of the minimal hitting sets of $\mathcal{G}$.

Proposition 5.28. *There is a polynomial reduction of* DUAL *to the optimality$^{\exists}$ problem that works both for compliance and for safety$^{\exists}$.*

Proof. Let $\mathcal{G} = \{G_1, \ldots, G_g\}, \mathcal{H} = \{H_1, \ldots, H_h\}$ be finite families of inclusion-incomparable sets and $G := G_1 \cup \ldots \cup G_g$. Since it can be checked in polynomial time whether a given set H is a minimal hitting set of $\mathcal{G}$, we can assume without loss of generality that all sets H_i are indeed minimal hitting sets of $\mathcal{G}$.

The problem to be decided by our reduction is thus whether $\mathcal{H}$ really contains *all* minimal hitting sets of $\mathcal{G}$. We view the elements of G as concept names, for $S \subseteq G$ write $\bigsqcap S$ for the conjunction of the concept names in S, and define

- $C := \exists r_1.\bigsqcap G$ and $\mathcal{P} := \{D_1 := \exists r_1.\bigsqcap G_1, \ldots, D_g := \exists r_1.\bigsqcap G_g\}$;
- $C' := \exists r_1.\bigsqcap(G \setminus H_1) \sqcap \ldots \sqcap \exists r_1.\bigsqcap(G \setminus H_h)$.

It is easy to see that C' is a $\mathcal{P}$-compliant and $\mathcal{P}$-safe$^{\exists}$ generalization of C.

According to Definition 5.8 and the proof of Theorem 5.13, C has exactly one optimal $\mathcal{P}$-compliant generalization, which is obtained as follows. First, note that the top-level conjunctions of C and $D_1, \ldots, D_g$ respectively consist of a single existential restriction for the same role r_1, and that the concepts D_i are pairwise incomparable. This implies that on this level only one hitting set is considered, which is $\mathcal{P}$. On the next role level, we have $\mathcal{P}_1 = \{\bigsqcap G_1, \ldots, \bigsqcap G_g\}$. The optimal $\mathcal{P}_1$-compliant generalizations of $C_1 := \bigsqcap G$ are obtained by considering all minimal hitting sets of $G_1, \ldots, G_g$, and removing their elements from the top-level conjunction of C_1. Consequently, the optimal $\mathcal{P}$-compliant generalization of C is given as

$$C'' := \bigsqcap_{H \text{ minimal hitting set of } \mathcal{G}} \exists r_1.\bigsqcap(G \setminus H).$$

A close look at Theorem 5.19 reveals that C'' is also the optimal $\mathcal{P}$-safe generalization of C. This shows that C' is optimal for compliance (safety$^{\exists}$) iff $\mathcal{H}$ contains all minimal hitting sets of $\mathcal{G}$. □

5.5 Characterizing Safety$^{\forall}$

We now turn our attention to the safety$^{\forall}$ problem in which knowledge of the attackers is represented as an $\mathcal{FL}_0$ concept. For this setting, we note that a value restriction can never imply an existential restriction. Thus, if C'' is an $\mathcal{FL}_0$ concept and D and $\mathcal{EL}$ concept of role depth > 0, then $C'' \not\sqsubseteq D$. This shows that an $\mathcal{FL}_0$ concept C'' is compliant with any $\mathcal{EL}$ policy that does not contain a concept of role depth 0.

Before characterizing safety$^{\forall}$, we need to characterize subsumption between an $\mathcal{FLE}$ concept and an $\mathcal{EL}$ concept. For this, we need the notion of $\texttt{filler}_r^{\forall}(C'')$, which is the set of all $\mathcal{FLE}$ concepts that becomes a filler of $\forall r$ in $\texttt{con}(C'')$. Formally, given an $\mathcal{FLE}$ concept C'', a role name $r \in N_R$, and a quantifier $\forall$, we define $\texttt{filler}_r^{\forall}(C'') := \{E \mid \forall r.E \in \texttt{con}(C'')\}$.

In Theorem 24 in [BKM99], it is written that the subsumption between two $\mathcal{FLE}$ concepts can be decided if there is a homomorphism function from the tree representation of one concept to the tree representation of another concept. Now, if one of the input is written as an $\mathcal{EL}$ concept, then the following lemma also works for characterizing subsumption between

an $\mathcal{FLE}$ concept and an $\mathcal{EL}$ concept and thus this lemma is an obvious consequence of that homomorphism characterization.

Proposition 5.29. *Let C'' be an $\mathcal{FLE}$ concept and D be an $\mathcal{EL}$ concept. It holds that $C'' \sqsubseteq D$ iff*

a.) for all $A \in \mathtt{con}(D)$, there is $A \in \mathtt{con}(C'')$ and

b.) for all $\exists r.D' \in \mathtt{con}(D)$, there is $\exists r.C' \in \mathtt{con}(C'')$ such that

$$C' \sqcap \bigsqcap \mathit{filler}_r^{\forall}(C'') \sqsubseteq D'.$$

It is obvious to see that the characterization above can be done in polynomial time. Now, we are ready to characterize the safety$^{\forall}$ problem.

Proposition 5.30. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P}$ a redundancy-free policy. Then, C is safe$^{\forall}$ for $\mathcal{P}$ iff the following two conditions hold for all $D \in \mathcal{P}$:*

1.) if $\mathtt{rd}(D) = 0$, then $\mathtt{con}(C) \cap \mathtt{con}(D) = 0$,

2.) if $\mathtt{rd}(D) > 0$, then there is $\exists r.D' \in \mathtt{con}(D)$ such that
   a.) if $\mathtt{rd}(D') = 0$, then there is no concept of the form $\exists r.C' \in \mathtt{con}(C)$,
   b.) if $\mathtt{rd}(D') > 0$, then for all $\exists r.C' \in \mathtt{con}(C)$, C' is safe$^{\forall}$ for $\{D'\}$. ♢

Proof. Assume that C is not safe$^{\forall}$ for $\mathcal{P}$. Then, there is an $\mathcal{EL}$ concept $D \in \mathcal{P}$ and an $\mathcal{FL}_0$ concept C'' that complies with $\mathcal{P}$ such that $C \sqcap C'' \sqsubseteq D$. Since $C \sqcap C''$ is an $\mathcal{FLE}$ concept, Proposition 5.29 applies to this subsumption. First, we consider the case where $\mathtt{rd}(D) = 0$. Proposition 5.29 implies that every concept name $A \in \mathtt{con}(D)$ is contained in $\mathtt{con}(C) \cup \mathtt{con}(C'')$. However, since C'' complies with $\mathcal{P}$, we have $C'' \not\sqsubseteq D$, and hence there must be an $A \in \mathtt{con}(D)$ that is not contained in $\mathtt{con}(C'')$. Consequently this A must belong to $\mathtt{con}(C)$, and thus property 1.) above is violated.

Now, let $\mathtt{rd}(D) > 0$, i.e., there is an existential restriction $\exists r.D' \in \mathtt{con}(D)$. By Proposition 5.29 and since C is an $\mathcal{EL}$ and C'' an $\mathcal{FL}_0$ concept, $C \sqcap C'' \sqsubseteq D$ implies that there is an existential restriction $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqcap \bigsqcap \mathtt{filler}_r^{\forall}(C'') \sqsubseteq D'$. If $\mathtt{rd}(D') = 0$, then this clearly violates 2a.). If $\mathtt{rd}(D') > 0$, then 2b.) is violated since $\bigsqcap \mathtt{filler}_r^{\forall}(C'')$ then cannot be subsumed by D', and thus $C' \sqcap \bigsqcap \mathtt{filler}_r^{\forall}(C'') \sqsubseteq D'$ shows that C' is not $\forall$-safe for $\{D'\}$.

To show the *only-if-direction*, we assume that one of the conditions 1.) or 2.) is violated, and prove that this implies that C is not safe$^{\forall}$ for $\mathcal{P}$.

First, assume that 1.) is violated, i.e., there is $D \in \mathcal{P}$ such that $\mathtt{rd}(D) = 0$ and there is $A \in \mathtt{con}(C) \cap \mathtt{con}(D)$. Then, $C'' := \bigsqcap(\mathtt{con}(D) \setminus \{A\})$ is an $\mathcal{FL}_0$ concept that complies with D, and satisfies $C \sqcap C'' \sqsubseteq D$. To conclude that C is not safe$^{\forall}$ for $\mathcal{P}$, it remains to shows that C'' also complies with all $\hat{D} \in \mathcal{P} \setminus \{D\}$. However, if we assume that $C'' \sqsubseteq \hat{D}$ for some $\hat{D} \in \mathcal{P} \setminus \{D\}$, then the fact that $D \sqsubseteq C''$ implies $D \sqsubseteq \hat{D}$, which contradicts our assumption that $\mathcal{P}$ is redundancy-free.

Second, assume that 2.) is violated. Then there is $D \in \mathcal{P}$ such that $\mathtt{rd}(D) > 0$ and for all $\exists r.D' \in \mathtt{con}(D)$, we have

- if $\mathtt{rd}(D') = 0$, then there is a concept of the form $\exists r.C' \in \mathtt{con}(C)$ and
- if $\mathtt{rd}(D') > 0$, then there is $\exists r.C' \in \mathtt{con}(C)$ such that C' is not safe$^{\forall}$ for $\{D'\}$.

We define the concept C'' as follows: $\underset{(1)}{\sqcap} A \sqcap \underset{(2)}{\sqcap} \forall r.D' \sqcap \underset{(3)}{\sqcap} \forall r.F$, where

(1) $A \in \mathtt{con}(D)$;

(2) $r \in N_R, \exists r.D' \in \mathtt{con}(D)$, and $\mathtt{rd}(D') = 0$;

(3) $r \in N_R, \exists r.D' \in \mathtt{con}(D), \mathtt{rd}(D') > 0$, and F is an $\mathcal{FL}_0$ concept complying with D', but $C' \sqcap F \sqsubseteq D'$. □

Note that C'' is an $\mathcal{FL}_0$ concept and is compliant with $\mathcal{P}$. To see the latter, assume that $\hat{D} \in \mathcal{P}$. If $\mathtt{rd}(\hat{D}) > 0$, then $C'' \not\sqsubseteq \hat{D}$ since an $\mathcal{FL}_0$ concept cannot imply an existential restriction. If $\mathtt{rd}(\hat{D}) = 0$, then $C'' \sqsubseteq \hat{D}$ would imply $D \sqsubseteq \hat{D}$, which contradicts our assumption that $\mathcal{P}$ is redundancy-free.

It remains to prove that $C \sqcap C'' \sqsubseteq D$, which we show using Proposition 5.29. First note that, by the construction of C'', each concept name $A \in \mathtt{con}(D)$ satisfies $A \in \mathtt{con}(C'')$, and thus $A \in \mathtt{con}(C \sqcap C'')$. Second, consider an existential restriction $\exists r.D' \in \mathtt{con}(D)$. If $\mathtt{rd}(D') = 0$, then there is $\exists r.C' \in \mathtt{con}(C)$, but also $\forall r.D' \in \mathtt{con}(C'')$. Thus, we have $C' \sqcap \bigsqcap \mathtt{filler}_r^{\forall}(C \sqcap C'') \sqsubseteq C' \sqcap D' \sqsubseteq D'$, as required by Proposition 5.29. If $\mathtt{rd}(D') > 0$, then we have $\exists r.C' \in \mathtt{con}(C)$ for an $\mathcal{EL}$ concept C' that is not safe$^{\forall}$ for $\{D'\}$. In addition, $\forall r D \in \mathtt{con}(C'')$, we here F is an $\mathcal{FL}_0$ concept such that $C' \sqcap F \sqsubseteq D'$. Consequently, we have $C' \sqcap \bigsqcap \mathtt{filler}_r^{\forall}(C \sqcap C'') \sqsubseteq C' \sqcap F \sqsubseteq D'$.

Obviously, the conditions for safety$^{\forall}$ stated above can be decided in polynomial time and it brings us to the following theorem.

Theorem 5.31. *The safety$^{\forall}$ problem can be decided in P.*

Since 1.) and 2.) in Proposition 5.30 are formulated for each $D \in \mathcal{P}$ separately, the following lemma is an immediate consequence of this proposition.

Lemma 5.32. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P}$ be a redundant-free $\mathcal{EL}$ policy. Then C is safe$^{\forall}$ for $\mathcal{P}$ iff C is $\forall$-safe for $\{D\}$ for all $D \in \mathcal{P}$.*

However, unlike safe$^{\exists}$ concepts that have the closed-under conjunction property, safe$^{\forall}$ concepts do not have such property. This is illustrated as follows.

Example 5.33. *Let $C_1 = \exists r.(A \sqcap B)$, $C_2 = \exists s.(A \sqcap B)$, and $\mathcal{P} = \{\exists r.A \sqcap \exists s.A\}$ be a policy. We have C_1 and C_2 as safe$^{\forall}$ concepts for $\mathcal{P}$, but $C_1 \sqcap C_2$ is not safe$^{\forall}$ for $\mathcal{P}$.* ◇

5.6 Optimal $\mathcal{P}$-safe$^{\forall}$ Generalizations

First, we plan to compute optimal $\mathcal{P}$-safe$^{\forall}$ generalizations of C. For this task, we consider the following notion. Given a concept D such that $\mathtt{rd}(D) > 0$, the set $\mathtt{con}^{\exists}(D)$ consists of all atoms that are of the form existential restrictions and occur in the top-level conjunctions of D. Then, one idea to compute an optimal $\mathcal{P}$-safe$^{\forall}$ generalization of C is removing all

concept names from $\mathtt{con}(C)$ that also occurs in the concepts $D \in \mathcal{P}$, where $\mathtt{rd}(D) = 0$, and then taking one existential restriction $\exists r.D'$ in each concept in $\mathcal{P}$, which has role depth, to subsequently remove or generalize the corresponding existential restriction in C that is not safe for $\exists r.D'$. For this, we again need the notion of hitting set that can guide us to choose those existential restrctions from each concept in $\mathcal{P}$. Now, we define a set that later we will show that it contains all optimal $\mathcal{P}$-safe$^\forall$ generalizations of C.

Definition 5.34. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P}$ be an $\mathcal{EL}$ policy. The set $SSG^\forall(C,\mathcal{P})$ of all* specific $\mathcal{P}$-safe$^\forall$ generalizations *C' of C consists of the concepts C' that are obtained from C by considering the following steps:*

- *if C is safe$^\forall$ for $\mathcal{P}$, then $SSG^\forall(C,\mathcal{P}) = \{C\}$.*
- *Otherwise, perform the following steps:*
  - *For all concept names $A \in con(C)$ such that $A \in con(D)$, where $D \in \{D_{i_1}, \ldots, D_{i_q}\}$ and $rd(D) = 0$, remove A from $con(C)$.*
  - *If $D_{j_1}, \ldots, D_{j_p}$ are all concepts in $\mathcal{P}$ such that $rd(D_{j_v}) > 0$, then construct a minimal hitting set H of $con^\exists(D_{j_1}), \ldots, con^\exists(D_{j_p})$ and do the following:*
    - *For all $\exists r.E \in con(C)$ such that there is a concept of the form $\exists r.D'$ in H with $rd(D') = 0$, remove them from $con(C)$.*
    - *Then, for each concept $\exists r_i.C_i \in con(C)$ that was not removed in the previous step, consider the set*

$$\mathcal{P}_i := \{D' \mid \exists r_i.D' \in H \text{ and } rd(D') > 0\}.$$

If $\mathcal{P}_i \neq \emptyset$, then replace $\exists r_i.C_i$ in $con(C)$ with $\bigsqcap \exists r_i.F$, where $F \in SSG^\forall(C_i,\mathcal{P}_i)$. If $\mathcal{P}_i$ is empty, then leave $\exists r_i.C_i$ as it is.

We show that each element in $\mathrm{SSG}^\forall(C,\mathcal{P})$ is a $\mathcal{P}$-safe$^\forall$ generalization of C.

Lemma 5.35. *Let C be an $\mathcal{EL}$ concept, $\mathcal{P}$ be an $\mathcal{EL}$ policy, and $C' \in SSG^\forall(C,\mathcal{P})$. It holds that C' is a $\mathcal{P}$-safe$^\forall$ generalization of C.*

Proof. First, we show that $C \sqsubseteq C'$. This is an easy consequence from the fact that, when constructing C' from C, atoms from the top-level conjunction of C are either kept unchanged or removed. The only non-trivial case is when $\exists r.C_i$ in $\mathtt{con}(C)$ is replaced with $\bigsqcap \exists r.F$, where $F \in \mathrm{SSG}^\forall(C_i,\mathcal{P}_i)$. By induction, we know that $C_i \sqsubseteq F$ and thus $\exists r.C_i \sqsubseteq \exists r.F$. This finally implies that $C \sqsubseteq C'$.

To prove that C' is safe for $\mathcal{P}$, we use the characterization given in Proposition 5.30. Thus, let $D \in \mathcal{P}$. If $\mathtt{rd}(D) = 0$, then $\mathtt{con}(D)$ is a set of concept names, and each of them has been removed in the construction of C'. Thus, $\mathtt{con}(C') \cap \mathtt{con}(D) = \emptyset$, as required by 1.) in Proposition 5.30.

If $\mathtt{rd}(D) > 0$, then the minimal hitting set H used in the construction of C' contains an existential restriction $\exists r.\hat{D} \in \mathtt{con}(D)$. If $\mathtt{rd}(\hat{D}) = 0$, then all existential restrictions for the role r are removed from the top-level conjunction of C, and thus 2a.) of Proposition 5.30 is satisfied. Finally, consider the case where $\mathtt{rd}(\hat{D}) > 0$. If $\exists r.E \in \mathtt{con}(C')$, then there is $\exists r_i.C_i \in \mathtt{con}(C)$ such that

$$\mathcal{P}_i = \{D' \mid \exists r_i.D' \in H \text{ and } \mathtt{rd}(D') > 0\} \neq \emptyset,$$

and $r = r_i$ and $E \in \text{SSG}^\forall(C_i, \mathcal{P}_i)$. Note that $\hat{D} \in \mathcal{P}_i$, and thus $\mathcal{P} = \emptyset$ is not possible for an existential restriction $\exists r_i.C_i \in \text{con}(C)$ with $r_i = r$. Induction (over the role depth) yields that E is safe$^\forall$ for $\mathcal{P}_i$, and thus for its subset $\{\hat{D}\}$. Hence, 2b.) of Proposition 5.30 is satisfied. □

However, $\text{SSG}^\forall(C, \mathcal{P})$ may also contain $\mathcal{P}$-safe$^\forall$ generalizations C' of C that are not optimal. Let us consider the following example.

Example 5.36. *Let $C = \exists r_1.(A \sqcap B) \sqcap \exists r_2.B \sqcap \exists r_3.A$ and $\mathcal{P} = \{D_1, D_2\}$, where*

$$D_1 = \exists r_1.A \sqcap \exists r_2.\top \text{ and } D_2 = \exists r_1.B \sqcap \exists r_3.\top.$$

We have $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$, and thus C is not even compliant, let alone safe$^\forall$, for $\mathcal{P}$. Applying the construction of Definition 5.34 to C and $\mathcal{P}$, we first construct the minimal hitting set $H_1 = \{\exists r_1.A, \exists r_1.B\}$ of $\mathit{con}^\exists(D_1)$ and $\mathit{con}^\exists(D_2)$. Since $\mathit{rd}(A) = 0 = \mathit{rd}(B)$, we remove the atom $\exists r_1.(A \sqcap B)$ from $\mathit{con}(C)$, which yields the concept $C'_1 = \exists r_2.B \sqcap \exists r_3.A \in \mathit{SSG}^\forall(C, \mathcal{P})$.

If we take the minimal hitting set $H_2 = \{\exists r_1.A, \exists r_3.\top\}$ instead, then we need to remove the atoms $\exists r_1.(A \sqcap B)$ and $\exists r_3.A$ from $\mathit{con}(C)$, which yields $C'_2 = \exists r_2.B \in \mathit{SSG}^\forall(C, \mathcal{P})$. Since $C'_1 \sqsubset C'_2$, the concept C'_2 cannot be optimal.

The next lemma states every $\mathcal{P}$-safe$^\forall$ generalization of C subsumes some element of $\text{SSG}^\forall(C, \mathcal{P})$.

Lemma 5.37. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P}$ be an $\mathcal{EL}$ policy. For all $\mathcal{P}$-safe$^\forall$ generalizations C'' of C, there is $C' \in \mathit{SSG}^\forall(C, \mathcal{P})$ such that $C' \sqsubseteq C''$.*

Proof. If C is safe$^\forall$ for $\mathcal{P}$, then obviously $C \in \text{SSG}^\forall(C, \mathcal{P})$ and we have $C \sqsubseteq C'$. Now, let us assume that C is not safe$^\forall$ for $\mathcal{P}$. Since C'' is a $\mathcal{P}$-safe$^\forall$ generalization of C, we have $C \sqsubseteq C''$ and C'' satisfies the properties 1.) and 2.) in Proposition 5.30. Due to 1.), $\text{con}(C'')$ contains no concept name A such that $A \in \text{con}(D)$ for some $D \in \mathcal{P}$ with $\text{rd}(D) = 0$. In addition, for all $D_j \in \mathcal{P}$ such that $\text{rd}(D_j) > 0$, there is $\exists r.G_j \in \text{con}(D_j)$ such that 2a.) or 2b.) of Proposition 5.30 holds. The set $H' := \{G_{j_1}, \ldots, G_{j_p}\}$ is a hitting set of the sets $\text{con}^\exists(D_{j_1}), \ldots \text{con}^\exists(D_{j_p})$ considered in Definition 5.34. Thus, there is a minimal hitting set H of $\text{con}^\exists(D_{j_1}), \ldots \text{con}^\exists(D_{j_p})$ such that $H \sqsubseteq H'$. Let C' be the element of $\text{SSG}^\forall(C, \mathcal{P})$ that is constructed by using H. We shows that $C' \sqsubseteq C''$ using Proposition 5.29.

First, consider a concept name $A \in \text{con}(C'')$. Since $C \sqsubseteq C''$, we know that $A \in \text{con}(C)$. In addition, as mentioned above, $\text{con}(C'')$ contains no concept name A such that $A \in \text{con}(D)$ for some $D \in \mathcal{P}$ with $\text{rd}(D) = 0$. Consequently, when constructing C' from C, the concept name A is not removed, which yields $A \in \text{con}(C')$.

Second, consider an existential restriction $\exists r.E \in \text{con}(C'')$. Since $C \sqsubseteq C''$, there is $\exists r.C_i \in \text{con}(C)$ such that $C_i \sqsubseteq E$. If $\exists r.C_i$ is not removed or generalized when constructing C', then $\exists r.C_i \in \text{con}(C')$, and we are done. If $\exists r.C_i$ is removed from $\text{con}(C)$ to construct C', then there is $\exists r.D' \in H \subseteq H'$ such that $\text{rd}(D') = 0$. By the definition of H', we thus know that $\exists r.D'$ must satisfy 2a.) of Proposition 5.30. But then, $\exists r.E \in \text{con}(C'')$ would not be possible.

Finally, if $\exists r.C_i$ is generalized in the construction of C' from C by replacing it with $\bigsqcap_{F \in \text{SSG}^\forall(C_i, \mathcal{P}_i)} \exists r.F$, then we know that $\mathcal{P}_i$ is non-empty. Now, consider an element D' of $\mathcal{P}_i$. Then, $\exists r.D' \in H \subseteq H'$ and $\text{rd}(D') > 0$ imply that $\exists r.D'$ satisfies 2b.) of Proposition 5.30. Since $\exists r.E \in \text{con}(C'')$, we thus know that E is safe$^\forall$ for $\{D'\}$. Since this is true for all

element D' of $\mathcal{P}_i$, Lemma 5.32 yields that E is safe$^\forall$ for $\mathcal{P}_i$ and thus induction yields that there is $F \in \text{SSG}^\forall(C_i, \mathcal{P}_i)$ such that $F \sqsubseteq E$. Since $\exists r.F \in \text{con}(C')$, this concludes our proof that $C' \sqsubseteq C''$. □

The following proposition states that all optimal $\mathcal{P}$-safe$^\forall$ generalizations of C for $\mathcal{P}$ are contained in $\text{SSG}^\forall(C, \mathcal{P})$.

Proposition 5.38. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. If C'' is an optimal $\mathcal{P}$-safe$^\forall$ generalization of C, then $C'' \in SSG^\forall(C, \mathcal{P})$ (up to equivalence).* ♢

The following theorem is an easy consequence of this proposition and the definition of $\text{SSG}^\forall(C, \mathcal{P})$.

Theorem 5.39. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ a redundancy-free policy. The cardinality of the set of all optimal $\mathcal{P}$-safe$^\forall$ generalization of C is at most exponential and each of its elements has exponential size in the size of C. Additionally, the set of all optimal $\mathcal{P}$-safe$^\forall$ generalizations of C can be computed in exponential time in the size of C and $D_1, \ldots, D_p$.*

Proof. It is sufficient to show that the set $\text{SSG}^\forall(C, \mathcal{P})$ satisfies the properties stated above. The cardinality of $\text{SSG}^\forall(C, \mathcal{P})$ is at most exponential due to exponentially many of hitting sets that should be constructed in the Definition 5.34 and each of them yields exactly one element of $\text{SSG}^\forall(C, \mathcal{P})$. Moreover, the size of each element C' in $\text{SSG}^\forall(C, \mathcal{P})$ may become exponential since during constructing C', we may also need to compute a conjunction of at most exponentially many existential restrictions and each of them also has at most exponential size by induction. To compute the set of all optimal $\mathcal{P}$-safe$^\forall$ generalizations of C, we need to remove all concepts in $\text{SSG}^\forall(C, \mathcal{P})$ that are not minimal w.r.t. subsumption. This implies that there are exponentially many subsumption tests that need to be done. □

The example below shows that the exponential time of the complexity of the algorithm above are indeed optimal.

Example 5.40. *Let us define an $\mathcal{EL}$ concept*

$$C = \exists r_1.(\exists s_1.\top \sqcap \exists s_2.\top) \sqcap \ldots \sqcap \exists r_n.(\exists s_1.\top \sqcap \exists s_2.\top)$$

and an $\mathcal{EL}$ policy

$$\mathcal{P} = \{\exists r_i.\exists s_1.\top \sqcap \exists r_i.\exists s_2.\top \mid 1 \leq i \leq n\}.$$

Then, using the algorithm defined in Theorem 5.39, there are 2^n optimal $\mathcal{P}$-safe$^\forall$ generalizations of C, that are of the form $\exists r_1.\exists s_{j_1}.\top \sqcap \ldots \sqcap \exists r_n.\exists s_{j_n}.\top$, where $j_i \in \{1, 2\}$ for all $i = 1, \ldots, n$. Thus all of them belong to $SSG^\forall(C, \mathcal{P})$ ♢

Next, we turn our attention to show the complexity of the optimality$^\forall$ problem.

Lemma 5.41. *The optimality$^\forall$ problem can be decided in coNP.*

Proof. According to Definition 5.21, safety$^\forall$ is polynomial and upward-closed property. Polynomiality follows from Lemma 5.30, whereas upward-closedness is an obvious consequence of the definition of safety$^\forall$. By Theorem 5.26, it is stated that deciding whether C' is an optimal F-generalization of C w.r.t. $\mathcal{P}$ is in coNP. It implies that the optimality$^\forall$ problem is also in coNP. □

Similar to the optimality$^{\exists}$ problem, we do not know whether the optimality$^{\forall}$ problem is also coNP-hard. But then, we can also show that the Hypergraph Duality Problem can be reduced to the optimality$^{\forall}$ problem.

Proposition 5.42. *There is a polynomial time reduction of* DUAL *to the optimality$^{\forall}$ problem.*$\diamondsuit$

Proof. Let $\mathcal{G} = \{G_1, \ldots, G_g\}, \mathcal{H} = \{H_1, \ldots, H_h\}$ be finite families of inclusion-incomparable sets and $G := G_1 \cup \ldots \cup G_g$. We again assume without loss of generality that H_i is indeed a minimal hitting set. We view the elements in G as distinct unqualified existential restrictions $\exists s_i.\top$ and then for every $S \subseteq G$, we write $\bigsqcap S$ for the conjunction of the unqualified existential restrictions in S. Now, we define

- $C := \exists r.\bigsqcap G$ and $\mathcal{P} := \{D_1 := \exists r.\bigsqcap G_1, \ldots, D_g := \exists r.\bigsqcap G_g\}$;
- $C' = \exists r.\bigsqcap(G \setminus H_1) \sqcap \ldots \sqcap \exists r.\bigsqcap(G \setminus H_h)$

Since each H_i is a minimal hitting set, it is easy to see that C' is a $\mathcal{P}$-safe$^{\forall}$ generalization of C. We will show that there is only one optimal $\mathcal{P}$-safe$^{\forall}$ generalization of C. According to Definition 5.34, we only consider all $D_{i_1}, \ldots, D_{i_q} \in \mathcal{P}$, where these are all indices for which C is not safe$^{\forall}$ for D_i. Since all concepts in $\mathcal{P}$ do not have concept names in their top-level conjunction, we just consider the step where $rd(D_i) > 0$. It means that we construct a minimal hitting set H of $\mathtt{con}^{\exists}(D_{i_1}), \ldots, \mathtt{con}^{\exists}(D_{i_q})$ and there is only one H that can be constructed and it is equal to $\mathcal{P}$. Then, for each $\exists r.\bigsqcap G_j$, we have $\mathtt{rd}(G_j) > 0$, so that we go to the next step that consider the set $\mathcal{P}_i$ for each concept $\exists r_i.C_i \in \mathtt{con}(C)$. On this level, we only have one $C_1 = \bigsqcap G$ and $\mathcal{P}_1 = \{\bigsqcap G_1, \ldots, \bigsqcap G_g\}$. Then, we compute all optimal $\mathcal{P}$-safe$^{\forall}$ generalizations of C_1 that is obtained by computing all minimal hitting sets of $G_1, \ldots, G_g$ and for each of this hitting set, we remove exactly one $\exists s_i.\top$ from the top-level conjunction of C_1. As a consequence, the optimal $\mathcal{P}$-safe$^{\forall}$ generalization of C is

$$C'' := \bigsqcap_{H \text{ minimal hitting set of } \mathcal{G}} \exists r.\bigsqcap(G \setminus H)$$

It is easy to see that $\mathcal{H}$ contains all minimal hitting sets of $\mathcal{G}$ iff C'' is the optimal $\mathcal{P}$-safe$^{\forall}$ generalization of C. $\square$

5.7 Characterizing Safety$^{\forall\exists}$ and Optimality$^{\forall\exists}$

As in the case of characterizing safety$^{\exists}$ and safety$^{\forall}$, we also require the policy $\mathcal{P}$ to be *redundant-free*. Now, the following lemma shows the characterization for safe$^{\forall\exists}$ concepts.

Lemma 5.43. *Let C be an $\mathcal{EL}$ concept and $\mathcal{P} = \{D_1, \ldots, D_p\}$ be a redundancy-free $\mathcal{EL}$ policy. C is safe$^{\forall\exists}$ for $\mathcal{P}$ iff*

1.) *for all concept names $A \in con(D_1) \cup \ldots \cup con(D_p)$, $A \notin con(C)$ and*

2.) *for all existential restrictions $\exists r.D' \in con(D_1) \cup \ldots \cup con(D_p)$, there is no concept of the form $\exists r.E$ in $con(C)$.*

Proof. First, assume that C is not safe$^{\forall\exists}$ for $\mathcal{P}$. Hence, there is $D_i \in \mathcal{P}$ and an $\mathcal{FLE}$ concept C'' such that C'' complies with $\mathcal{P}$, but $C \sqcap C'' \sqsubseteq D_i$. This subsumption implies that $A \in \mathtt{con}(C) \cup \mathtt{con}(C'')$ holds for all $A \in \mathtt{con}(D_i)$. If there is an $A \in \mathtt{con}(D_i)$ such that $A \in \mathtt{con}(C)$, then property 1.) is violated. Otherwise, all $A \in \mathtt{con}(D_i)$ belong to $\mathtt{con}(C'')$. But then $C'' \not\sqsubseteq D_i$ can only be due to the fact that there is $\exists r.D' \in \mathtt{con}(D_i)$ such that, for all $\exists r.C' \in \mathtt{con}(C'')$, we have $C' \bigsqcap \mathtt{filler}_r^{\forall}(C'') \not\sqsubseteq D'$. Applying Proposition 5.29 again to the subsumption $C \sqcap C'' \sqsubseteq D_i$ thus yields that that there is $\exists r.E \in \mathtt{con}(C)$ such that $E \bigsqcap \mathtt{filler}_r^{\forall}(C'') \sqsubseteq D'$. Consequently, property 2.) is violated.

To show the other direction, assume that condition 1.) or 2.) is violated. If 1.) is violated, then there are $D_i \in \mathcal{P}$ and a concept name A such that $A \in \mathtt{con}(C) \cap \mathtt{con}(D_i)$. We modify D_i to C'' by removing A from the top-level conjunction of D_i. Then C'' is an $\mathcal{EL}$ concept, and thus also an $\mathcal{FLE}$ concept, such that $C'' \not\sqsubseteq D_i$ and $C \sqcap C'' \equiv C \sqcap D_i \sqsubseteq D_i$. Given $D \in \mathcal{P} \setminus \{D_i\}$ we have $C'' \not\sqsubseteq D$ since otherwise $D_i \sqsubseteq C'' \sqsubseteq D$ would contradict our assumption that $\mathcal{P}$ is redundancy-free. Thus C is not safe$^{\forall\exists}$ for $\mathcal{P}$.

If condition 2.) is violated, then there are $D_i \in \mathcal{P}$ and existential restrictions $\exists r.D' \in \mathtt{con}(D_i)$ and $\exists r.E \in \mathtt{con}(C)$. Let C'' be obtained from D_i by replacing every existential restriction $\exists r.F$ from the top-level conjunction of D_i with the corresponding value restriction $\forall r.F$. To show that $C \sqcap C'' \sqsubseteq D_i$, it is sufficient to show that $C \sqcap C'' \sqsubseteq \exists r.F$ for all $\exists r.F \in \mathtt{con}(D_i)$. This is the case since $C \sqcap C'' \sqsubseteq \exists r.E \sqcap \forall r.F \sqsubseteq \exists r.(E \sqcap F) \sqsubseteq \exists r.F$.

It remains to show that C'' is compliant with $\mathcal{P}$, i.e., for all $D \in \mathcal{P}$ we have $C'' \not\sqsubseteq D$. If D contains an existential restriction for r, then this holds since C'' does not contain an existential restriction for r. In particular, this covers the case where $D = D_i$. If D does not contain an existential restriction for r, then the changes we made when going from D_i to C'' are not relevant for D, i.e., we have $C'' \sqsubseteq D$ iff $D_i \sqsubseteq D$. Since $\mathcal{P}$ is redundancy-free, this yields $C'' \not\sqsubseteq D$. □

Due to the simplicity of the conditions 1.) and 2.) in this proposition, it is now easy to show that all relevant computation or decision problems for safety$^{\forall\exists}$ are tractable.

Theorem 5.44. *Given $\mathcal{EL}$ concepts C, C'' and an $\mathcal{EL}$ policy $\mathcal{P}$ that is redundant-free, we can*

- *decide whether C is safe$^{\forall\exists}$ for $\mathcal{P}$,*
- *compute the optimal $\mathcal{P}$-safe$^{\forall\exists}$ generalization of C, and*
- *decide whether C'' is an optimal $\mathcal{P}$-safe$^{\forall\exists}$ of C*

in polynomial time.

Proof. First, note that the characterization of safety$^{\forall\exists}$ given in Proposition 5.43 can obviously be checked in polynomial time. Secondly, to obtain the optimal safe$^{\forall\exists}$ generalization of C for $\mathcal{P}$, we simply remove from $\mathtt{con}(C)$ all concept names A with $A \in \mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$, and all existential restrictions $\exists r.E$ such that $\mathtt{con}(D_1) \cup \ldots \cup \mathtt{con}(D_p)$ contains an existential restriction for the role r. This can clearly be done in polynomial time. Finally, to decide whether C'' is an optimal safe$^{\forall\exists}$ generalization of C for $\mathcal{P}$, apply the procedure just described to C, and check whether the resulting concept C' is equivalent to C''. Since the subsumption problem is polynomial in $\mathcal{EL}$, this yields a polynomial-time decision procedure for the optimality problem. □

Chapter 6

Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ ABoxes

In this chapter, we extend the privacy setting defined in the previous chapter by assuming that the information about individuals as well as the knowledge of attackers are given by $\mathcal{EL}$ ABoxes consisting of concept and role assertions, while the privacy policy is either an instance query ($\mathcal{EL}$ concept) or a conjunctive query. We set the $\mathcal{EL}$ ABoxes in this chapter to contain axioms stating information about known and anonymous individuals which correspond to constants and nulls, respectively, in relational datasets formulated by [GK16; GK19]. If constants are treated in that papers as an object whose information needs to be protected, then here we assume that the individuals whose information is not allowed to be disclosed are the known ones. In this setting, we still do not include (general) TBoxes as a part of the input in this setting. However, as argued before in the previous chapter, ontologies that use acyclic TBoxes, such as SNOMED CT, NCI, or GeneOntology, still can make use of the results presented in this chapter since the TBoxes can be reduced away by unfolding concept definitions [Sun09].

Analogous to the formalization of sensitive information in $\mathcal{EL}$ instance stores, here we define *compliance* if an ABox does not reveal any sensitive answer to the policy and *safety* if the combination of the ABox and any other policy-compliant $\mathcal{EL}$ ABox does not disclose any sensitive answer to the policy. If the given ABox does not satisfy such two properties, an anonymization operator, called *anonymizer*, is applied to it such that the anonymized ABox fulfill the two privacy requirements. The *optimality* property is also mentioned here to guarantee that the modified ABox, which is compliant and safe, still preserves as much information from the original ABox as possible. To modify $\mathcal{EL}$ ABoxes, we will weaken the concept assertions $C(a)$ by semantically generalizing C as we performed in Chapter 4 and 5, and additionally, we rename individuals occurring in the concept or role assertions. We call the combination of concept generalization and individuals renaming an *anonymization* approach applied to $\mathcal{EL}$ ABoxes.

This sort of approach also sounds common in the privacy area as a mainstream technique to prevent the disclosure of sensitive data in information systems. Some popular and recent anonymization techniques in databases or linked data can be found in [Swe02a; MW04; GK16], which commonly either replace constants (or null values) with new null values in the context of RDF graph or suppress characters of values of databases' attributes with a new anonymous character $*$ in the context of relational databases. As an illustration of our anonymization approach, we extend the previous medical example in the beginning of Chapter 5 with some role assertions.

We define an $\mathcal{EL}$ ABox $\mathcal{A}_0$ consisting of the following assertions:

$$\begin{aligned}\mathcal{A}_0 = \{&(\mathsf{Male} \sqcap \mathsf{Patient} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}))(\mathsf{BOB}),\\ &(\mathsf{Female} \sqcap \mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology})(\mathsf{DIANA}),\\ &\mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA})\}\end{aligned}$$

Basically, $\mathcal{A}_0$ says that BOB is a male patient who suffers from a disease which has coughing and fatigue as its symptoms. Then, $\mathcal{A}_0$ additionally states that BOB is seen by DIANA who is a female doctor working in an oncology department. The following $\mathcal{EL}$ concept D is a policy such that one should not be able to find out the following information from any individual

$$\begin{aligned}D = \mathsf{Patient} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}) \sqcap\\ \exists \mathsf{seen\_by}(\mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology}).\end{aligned}$$

Nonetheless, the ABox $\mathcal{A}_0$ entails the axiom $D(\mathsf{BOB})$. According to the formal definition of compliance in this chapter, $\mathcal{A}_0$ is not compliant with the policy D since there is an individual, which is BOB, that belongs to D w.r.t. the given ABox. Now, we construct $\mathcal{A}_1$, which is an anonymization of $\mathcal{A}_0$ and compliant with D.

$$\begin{aligned}\mathcal{A}_1 = \{&(\mathsf{Male} \sqcap \mathsf{Patient} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}))(\mathsf{BOB}),\\ &(\mathsf{Female} \sqcap \mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology})(y),\\ &\mathsf{seen\_by}(\mathsf{BOB}, x)\}\end{aligned}$$

The anonymization operation works in this example by renaming all occurrences of the known individual DIANA with anonymous individuals x and y. We still can make this anonymization as optimal as possible by taking the ABox $\mathcal{A}_0$ and sufficiently renaming the known individual DIANA in one of the assertions, but not generalizing any concept as described as follows.

$$\begin{aligned}\mathcal{A}_2 = \{&(\mathsf{Male} \sqcap \mathsf{Patient} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}))(\mathsf{BOB}),\\ &(\mathsf{Female} \sqcap \mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology})(\mathsf{DIANA}),\\ &\mathsf{seen\_by}(\mathsf{BOB}, x)\}\end{aligned}$$

However, this ABox $\mathcal{A}_2$ still can be attacked by other users who know that BOB is seen by DIANA, where the assertion $\mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA})$ is compliant with D, but

$$\mathcal{A}_2 \cup \{\mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA})\} \models D(\mathsf{BOB}).$$

To alleviate this issue, the safety property is considered and thus the following anonymization $\mathcal{A}_3$ of $\mathcal{A}_0$ is safe for D such that for each $\mathcal{EL}$ ABox $\mathcal{A}'$ complying with D, the union $\mathcal{A}_3 \cup \mathcal{A}'$ also complies with D.

$$\begin{aligned}\mathcal{A}_3 = \{&(\mathsf{Male} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}))(\mathsf{BOB}),\\ &(\mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top)(\mathsf{DIANA}),\\ &\mathsf{seen\_by}(\mathsf{BOB}, x)\}\end{aligned}$$

Despite being safe for D, the ABox $\mathcal{A}_3$ is still not optimal in keeping more information from $\mathcal{A}_0$ as much as possible. In fact, the following $\mathcal{EL}$ ABox $\mathcal{A}_4$ is also a safe anonymization w.r.t. D and more informative than $\mathcal{A}_3$.

$$\begin{aligned}\mathcal{A}_4 = \{&\mathsf{Male} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough}) \sqcap \\ &\exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}) \sqcap \\ &\exists \mathsf{suffer}.(\exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}))(\mathsf{BOB}), \\ &(\mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top)(\mathsf{DIANA}), \\ &\mathsf{seen\_by}(\mathsf{BOB}, x)\}\end{aligned}$$

Motivated by the illustration above, within this chapter, we will show how to address the following decision problems:

- Is a given $\mathcal{EL}$ ABox compliant with a policy? (*compliance*)
- Is a given $\mathcal{EL}$ ABox safe for a policy? (*safety*)
- Does a given ABox anonymizer yield an anonymization that is not only compliant with (safe for) a policy, but also preserves information from the original ABox as much as possible? (*optimality*)

Before providing the formal definitions for each problem and giving the corresponding algorithms for each of them, in Section 6.1, we discuss how to characterize logical entailments between $\mathcal{EL}$ ABoxes that contain information about anonymous individuals. Then, in Section 6.2, we introduce an approach to anonymize $\mathcal{EL}$ ABoxes and then provide a small illustration explaining that our approach has more features for performing anonymization than the approach in [GK16; GK19]. Further, in Section 6.3, we present formal definitions for each privacy property, which are *compliance*, *safety*, and *optimality* for $\mathcal{EL}$ ABoxes, and then define the corresponding decision problems asking whether each property is guaranteed. This is followed by introducing characterizations for compliance and safety problems in Section 6.4, which afterwards provide us complexity results for both problems. Finally, we close this chapter in Section 6.5, by defining algorithms for deciding the optimality of $\mathcal{EL}$ ABoxes and presenting reductions from an existing problem in graph theory to our optimality problems.

6.1 Logical Entailments in $\mathcal{EL}$ ABoxes with Anonymous Individuals

As stated in the beginning of this chapter that in this setting the $\mathcal{EL}$ ABoxes are specified to have axioms about known and anonymous individuals. Within this section, we explain how to characterize the entailment between ABoxes and the conjunctive query entailment problem when the input ABoxes include anonymous individuals in their axioms. When talking about signature in this kind of $\mathcal{EL}$ ABoxes $\mathcal{A}$, we only restrict twithin o concept names, role names, and known individuals occurring in $\mathcal{A}$ as the elements of $sig(\mathcal{A})$. To clarify the symbol we use for these two types of individuals, we write x, y to denote anonymous individuals, a, b, c for known individuals, and u to symbolize individual names in general regardless of the type. In this privacy setting, we also assume that the set N_{KI} of known individuals contains *at least one* known individual that does not occur in the given ABox $\mathcal{A}$. Differing with how

we treat anonymous individuals in Chapter 3 as an object that hides the 'real name' of a known individual, in this chapter we treat them as an *existentially quantified object*, which means that these anonymous individuals cannot be interpreted under the standard notion of interpretation. Consequently, in the semantics sense, this gives us more notions to define the semantics of $\mathcal{EL}$ ABoxes with anonymous individuals.

Given an interpretation $\mathcal{I}$, an *assignment* θ w.r.t. $\mathcal{I}$ is a function $\theta : \mathsf{N_{AI}} \to \Delta^{\mathcal{I}}$. For every $u \in \mathsf{N_I}$, if $\mathcal{I}$ is an interpretation and θ is an assignment w.r.t. $\mathcal{I}$, then we have $u^{\mathcal{I},\theta} = u^{\mathcal{I}}$ if $u \in \mathsf{N_{KI}}$ and $u^{\mathcal{I},\theta} = \theta(u)$ if $u \in \mathsf{N_{AI}}$. Given an $\mathcal{EL}$ ABox $\mathcal{A}$, an interpretation $\mathcal{I}$ and an assignment θ w.r.t. $\mathcal{I}$ *satisfies* $\mathcal{A}$, denoted by $(\mathcal{I}, \theta) \models \mathcal{A}$ iff for all concept and role assertions $C(u)$ and $r(u_1, u_2)$, respectively, in $\mathcal{A}$, we have $u^{\mathcal{I},\theta} \in C^{\mathcal{I}}$ and $(u_1^{\mathcal{I},\theta}, u_2^{\mathcal{I},\theta}) \in r^{\mathcal{I}}$, respectively. Then, we call an interpretation $\mathcal{I}$ a *model* of $\mathcal{A}$ iff there exists an assignment θ w.r.t. $\mathcal{I}$ such that $(\mathcal{I}, \theta) \models \mathcal{A}$.

Another consequence of having anonymous individuals as an existentially quantified object is if we are given two different $\mathcal{EL}$ ABoxes $\mathcal{A}$ and $\mathcal{A}'$, then we assume that anonymous individuals in $\mathcal{A}$ and $\mathcal{A}'$ are renamed apart so that every anonymous individual in $\mathcal{A}$ cannot be linked to other anonymous individuals in $\mathcal{A}$'. Now, given $\mathcal{EL}$ ABoxes $\mathcal{A}$ and $\mathcal{A}'$, one may be interested in knowing whether an ABox $\mathcal{A}'$ is entailed by another ABox $\mathcal{A}$. Formally, $\mathcal{A}$ *entails* $\mathcal{A}'$, denoted by $\mathcal{A} \models \mathcal{A}'$, iff every model of $\mathcal{A}$ is a model of $\mathcal{A}'$. We say that two $\mathcal{EL}$ ABoxes $\mathcal{A}, \mathcal{A}'$ are *equivalent* iff $\mathcal{A} \models \mathcal{A}'$ and $\mathcal{A}' \models \mathcal{A}$.

Another important reasoning task over $\mathcal{EL}$ ABoxes that becomes a basis for reasoning problems considered in this chapter is the CQ entailment problem w.r.t. $\mathcal{EL}$ ABoxes. To characterize this problem for $\mathcal{EL}$ ABoxes with anonymous individuals, we can emulate what the authors in [GK16; GK19] do to their conjunctive queries and datasets. For this reason, we first need to know how these relational datasets are formally defined.

Definition 6.1. *Let* Const *and* Null *be pairwise disjoint sets of constants and nulls, respectively, and* Rel *be a set of first-order predicates with n-arity, where* $n > 0$. *A* dataset $\mathcal{D}$ *is a finite set of atomic formulas that are built over* Rel *and* Const $\cup$ Null.

In the following, we describe a *decomposition process* applied to $\mathcal{EL}$ ABoxes such their representation is similar to a dataset consisting of atomic assertions only. As stated in Subsection 2.1.3, an $\mathcal{EL}$ ABox $\mathcal{A}$ is semantically equivalent to its first-order representation $\pi(\mathcal{A})$. Note that all variables in $\pi(\mathcal{A})$ are existentially quantified. Since anonymous individuals are also existentially quantified, we may replace the variable names w in $\pi(\mathcal{A})$ with new anonymous individuals x. Without loss of generality, we may now see this first-order representation as $\exists \vec{z}. \bigwedge \mathcal{A}_d$, where $\mathcal{A}_d$ is the set of all atomic formulas of the form of unary and binary predicates occurring in $\pi(\mathcal{A})$ with arguments from known and anonymous individuals and $\vec{z}$ are anonymous individuals in $\mathcal{A}_d$. We can also view $\mathcal{A}_d$ is essentially an $\mathcal{EL}$ ABox consisting of atomic concept assertions $A(u)$ and role assertions $r(u, u')$, where $u, u' \in \mathsf{N_I}$. We will call $\mathcal{A}_d$ the *decomposed ABox* of $\mathcal{A}$ afterwards. Using this representation, $\exists \vec{z}. \bigwedge \mathcal{A}_d$ is basically the same with the first order formula used to represent datasets in [GK16; GK19]. Then, our decomposed ABox is the same with a dataset built over Const, Null, unary, and binary predicates only. Note that this decomposition process can be readily performed in linear time. Nevertheless, if it is known from the context that all assertions in the original ABox $\mathcal{A}$ are already in the form $A(u)$ or $r(u, u')$, then we do not need to do the decomposition process above.

$$
\begin{aligned}
\mathcal{A}_{0_d} = \{ & \mathsf{Male}(\mathsf{BOB}) && (\alpha_1)\\
& \mathsf{Patient}(\mathsf{BOB}) && (\alpha_2)\\
& \mathsf{suffer}(\mathsf{BOB}, x_1) && (\alpha_3)\\
& \mathsf{Disease}(x_1) && (\alpha_4)\\
& \mathsf{symptom}(x_1, x_2) && (\alpha_5)\\
& \mathsf{symptom}(x_1, x_3) && (\alpha_6)\\
& \mathsf{Cough}(x_2) && (\alpha_7)\\
& \mathsf{Fatigue}(x_3) && (\alpha_8)\\
& \mathsf{Female}(\mathsf{DIANA}) && (\alpha_9)\\
& \mathsf{Doctor}(\mathsf{DIANA}) && (\alpha_{10})\\
& \mathsf{works\_in}(\mathsf{DIANA}, x_4) && (\alpha_{11})\\
& \mathsf{Oncology}(x_4) && (\alpha_{12})\\
& \mathsf{seen\_by}(\mathsf{BOB}, \mathsf{DIANA})\} && (\alpha_{13})
\end{aligned}
$$

Figure 6.3: The decomposed ABox $\mathcal{A}_{0_d}$ obtained from the decomposition applied to $\mathcal{A}_0$

Example 6.2. *As an example, let us consider the ABox $\mathcal{A}_0$ defined in the beginning of this chapter. If the decomposition process explained above is applied to it, then we obtain the decomposed ABox $\mathcal{A}_{0_d}$ depicted in Figure 6.3. In particular, the concept assertion containing* BOB *is decomposed into assertions $(\alpha_1), \ldots, (\alpha_8)$, the concept assertion containing* DIANA *is decomposed into assertions $(\alpha_9), \ldots, (\alpha_1 2)$, and the role assertion* seen_by(BOB, DIANA) *does not need to be decomposed since it is already atomic.* ♢

We have explained our assumptions for $\mathcal{EL}$ ABoxes with anonymous individuals above. Now, for the conjunctive queries we consider in this setting it is important to note that it may only make sense if individual names in the arguments are only restricted to known individuals since the anonymous ones can also be semantically replaced by existentially quantified variables in q. This is again different with how we treat conjunctive queries in the view-based identity problem in Chapter 3 since anonymous individuals are not necessarily treated as an existentially quantified object.

Using all these representations and referring to the notion of homomorphism between the body of CQs and the datasets in [GK16; GK19], we are ready to characterize the CQ entailment problem for $\mathcal{EL}$ ABoxes with anonymous individuals. Given a CQ $q(\vec{v}) \leftarrow \exists \vec{w}.\phi(\vec{v}, \vec{w})$ and an $\mathcal{EL}$ ABox $\mathcal{A}$, a *homomorphism* from the body of q to the decomposed ABox $\mathcal{A}_d$ of $\mathcal{A}$ is a mapping $h : \mathsf{N_{KI}} \cup \vec{v} \cup \vec{w} \rightarrow \mathsf{N_{KI}} \cup \mathsf{N_{AI}}$ such that

$$h(a) = a \text{ for all known individuals and } h(\phi(\vec{v}, \vec{w})) \subseteq \mathcal{A}_d.$$

We say that $\vec{u}$ is an answer to $q(\vec{v})$ w.r.t. $\mathcal{A}$ iff there exists a homomorphism from the body of q to $\mathcal{A}_d$ such that $h(\vec{v}) = \vec{u}$.

6.2 Anonymizing $\mathcal{EL}$-ABoxes

Now, we introduce our anonymization function that will be applied to $\mathcal{EL}$ ABoxes. In this context, we proceed anonymizations by renaming known or anonymous individuals with

new anonymous individuals both in concept and role assertions. In addition to individual renaming, our anonymization function may generalize the concept C occurring in concept assertions $C(u) \in \mathcal{A}$.

When generalizing concepts C, intuitively this anonymization function will map C to a more general concept C' such that $C \sqsubseteq C'$, where concept names and role names occurring in C' obviously occur in C, or, in general, $sig(C') \subseteq sig(\mathcal{A})$. For this reason, given an ABox $\mathcal{A}$, we define $\mathcal{C}_{\mathcal{EL}}^{\mathcal{A}}$ as the set of $\mathcal{EL}$ concepts that are built over the signature of $\mathcal{A}$, and then we put $\mathcal{C}_{\mathcal{EL}}^{\mathcal{A}}$ as a part of the range of our anonymization function.

Similar to [GK16; GK19], we need the notion of *position* that basically represents an occurrence of a concept or an individual in an ABox $\mathcal{A}$. A *position* ρ in $\mathcal{A}$ is a pair $\langle \beta, j \rangle$ for a role or a concept assertion β in $\mathcal{A}$ and $j \in \{1, 2\}$. Then, the *value* $val(s, \mathcal{A})$ of ρ in $\mathcal{A}$ is the j-th argument of β. If β is a role assertion $r(a, b)$, then a and b are the first and the second arguments of β, respectively. Likewise, if β is a concept assertion $C(a)$, then C and a are the first and the second arguments of β, respectively.

Definition 6.4. *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox and $\mathcal{C}_{\mathcal{EL}}^{\mathcal{A}}$ be the set of all $\mathcal{EL}$ concepts that are built over the signature Σ of $\mathcal{A}$. An* $\mathcal{A}$-anonymizer *is a function f mapping positions in $\mathcal{A}$ to $\mathsf{N_{KI}} \cup \mathsf{N_{AI}} \cup \mathcal{C}_{\mathcal{EL}}^{\mathcal{A}}$ such that for all positions ρ and ρ' in $\mathcal{A}$,*

- *if $val(\rho, \mathcal{A}), val(\rho', \mathcal{A}) \in \mathsf{N_{AI}} \cup \mathsf{N_{KI}}$, then*
  - *$f(\rho) \in \mathsf{N_{KI}} \cup \mathsf{N_{AI}}$,*
  - *$f(\rho) \in \mathsf{N_{KI}}$ implies $val(\rho, \mathcal{A}) = f(\rho)$, and*
  - *$f(\rho) = f(\rho')$ implies $val(\rho, \mathcal{A}) = val(\rho', \mathcal{A})$.*
- *if $val(\rho, \mathcal{A})$ is an $\mathcal{EL}$ concept C, then $f(\rho)$ is an $\mathcal{EL}$ concept $C' \in \mathcal{C}_{\mathcal{EL}}^{\mathcal{A}}$ such that $C \sqsubseteq C'$.*

Given an $\mathcal{A}$-anonymizer f, an ABox $\mathcal{A}'$ is an anonymization of $\mathcal{A}$ w.r.t. f, *written $f(\mathcal{A}) = \mathcal{A}'$, iff $\mathcal{A}'$ is obtained from $\mathcal{A}$ by*

- *for each $\beta = r(a_1, a_2) \in \mathcal{A}$, replace them with $r(u_1, u_2)$ such that $u_1 = f(\langle \beta, 1 \rangle)$ and $u_2 = f(\langle \beta, 2 \rangle)$, and*
- *for each $\beta = C(a) \in \mathcal{A}$, replace them with $C'(u)$, where $f(\langle \beta, 1 \rangle) = C'$ and $f(\langle \beta, 2 \rangle) = u$.*

$\diamondsuit$

The following example illustrates how to anonymize an $\mathcal{EL}$ ABox using an anonymizer.

Example 6.5. *Let us consider again the ABox $\mathcal{A}_0$ and $\mathcal{A}_4$ defined in the beginning of this chapter. We will show how to obtain $\mathcal{A}_4$ from an anonymizer applied to $\mathcal{A}_0$. From the representation of $\mathcal{A}_0$, there are six positions $\beta_1, \ldots, \beta_6$ in $\mathcal{A}_0$, where*

- $val(\beta_1, \mathcal{A}) = \mathsf{Male} \sqcap \mathsf{Patient} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue})$,
- $val(\beta_2, \mathcal{A}) = \mathsf{BOB}$,
- $val(\beta_3, \mathcal{A}) = \mathsf{Female} \sqcap \mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology}$,
- $val(\beta_4, \mathcal{A}) = \mathsf{DIANA}$,
- $val(\beta_5, \mathcal{A}) = \mathsf{BOB}$, *and*

– $val(\beta_6, \mathcal{A}) = \mathsf{DIANA}$.

We construct an $\mathcal{A}$-anonymizer f such that f maps each position to the following individuals or concepts and $f(\mathcal{A}_0) = \mathcal{A}_4$:

– $f(\beta_1) = \mathsf{Male} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Cough}) \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue}) \sqcap \exists \mathsf{suffer}.(\exists \mathsf{symptom}.\mathsf{Cough} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue})$,

– $f(\beta_2) = \mathsf{BOB}$,

– $f(\beta_3) = \mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top$,

– $f(\beta_4) = \mathsf{DIANA}$,

– $f(\beta_5) = \mathsf{BOB}$, *and*

– $f(\beta_6) = x$.

Additionally, Figure 6.6 depicts a graphical illustration of anonymizing $\mathcal{A}_0$ to $\mathcal{A}_4$ via f. Note that in the figure the black nodes represent a known individual, the gray nodes represent an anonymous individual, and the white nodes represent a filler of an existential restriction occurring as a subconcept in an ABox. ♢

One may see this $\mathcal{A}$-anonymizer as a weakening operator applied to each axiom in $\mathcal{A}$ since it weakens each axiom by either renaming individuals in assertion or generalizing the concept. The next question is whether the anonymizers also weaken the whole ABox. The following lemma shows that an anonymization of $\mathcal{A}$ w.r.t. an $\mathcal{A}$-anonymizer f is indeed a logical consequence of $\mathcal{A}$.

Lemma 6.7. *Let $\mathcal{A}, \mathcal{A}'$ be $\mathcal{EL}$ ABoxes and f be an $\mathcal{A}$-anonymizer such that $f(\mathcal{A}) = \mathcal{A}'$. It holds that $\mathcal{A}$ entails $\mathcal{A}'$.*

Proof. Let $\mathcal{I}$ be a model of $\mathcal{A}$. Then, there is an assignment θ w.r.t. $\mathcal{I}$ such that $(\mathcal{I}, \theta) \models \mathcal{A}$. To prove this lemma, we only need to show that $\mathcal{I}$ is also a model of $\mathcal{A}'$. Note that for all $C(u) \in \mathcal{A}$ and $C'(u') \in \mathcal{A}$, where $val(\rho, \mathcal{A}) = C$ and $f(\rho) = C'$, we have $C^{\mathcal{I}} \subseteq C'^{\mathcal{I}}$. Now, let us construct an assignment θ' w.r.t. $\mathcal{I}$ such that

- for all anonymous individuals x occurring in $\mathcal{A}'$ and $\mathcal{A}$, we have $\theta'(x) := \theta(x)$ and
- for all anonymous individuals y occurring in $\mathcal{A}'$, but it does not occur in $\mathcal{A}$, and $f(\rho) = y$, we have $\theta'(y) := u^{\mathcal{I},\theta}$, where $val(\rho, \mathcal{A}) = u$.

By taking the interpretation $\mathcal{I}$ and the construction of θ' into account, we have

$$u^{(\mathcal{I},\theta)} \in C^{\mathcal{I}} \text{ implies } u'^{(\mathcal{I},\theta')} \in C'^{\mathcal{I}} \text{ for all } C(u) \in \mathcal{A},$$

where $val(\beta_1, \mathcal{A}) = C, val(\beta_2, \mathcal{A}) = u$, $f(\beta_1) = C', f(\beta_2) = u'$, and $C'(u') \in \mathcal{A}'$. In addition to that, we have

$$(u_1^{(\mathcal{I},\theta)}, u_2^{(\mathcal{I},\theta)}) \in r^{\mathcal{I}} \text{ implies } (u_1^{(\mathcal{I},\theta')}, u_2^{(\mathcal{I},\theta')}) \in r^{\mathcal{I}} \text{ for all } r(u_1, u_2) \in \mathcal{A},$$

where $val(\beta_3, \mathcal{A}) = u_1, val(\beta_4, \mathcal{A}) = u_2$, $f(\beta_3) = u_1, f(\beta_4) = u_2$, and $r(u_1, u_2) \in \mathcal{A}'$. This implies that $(\mathcal{I}, \theta')$ satisfies $\mathcal{A}'$, and thus $\mathcal{I}$ is a model of $\mathcal{A}'$. □

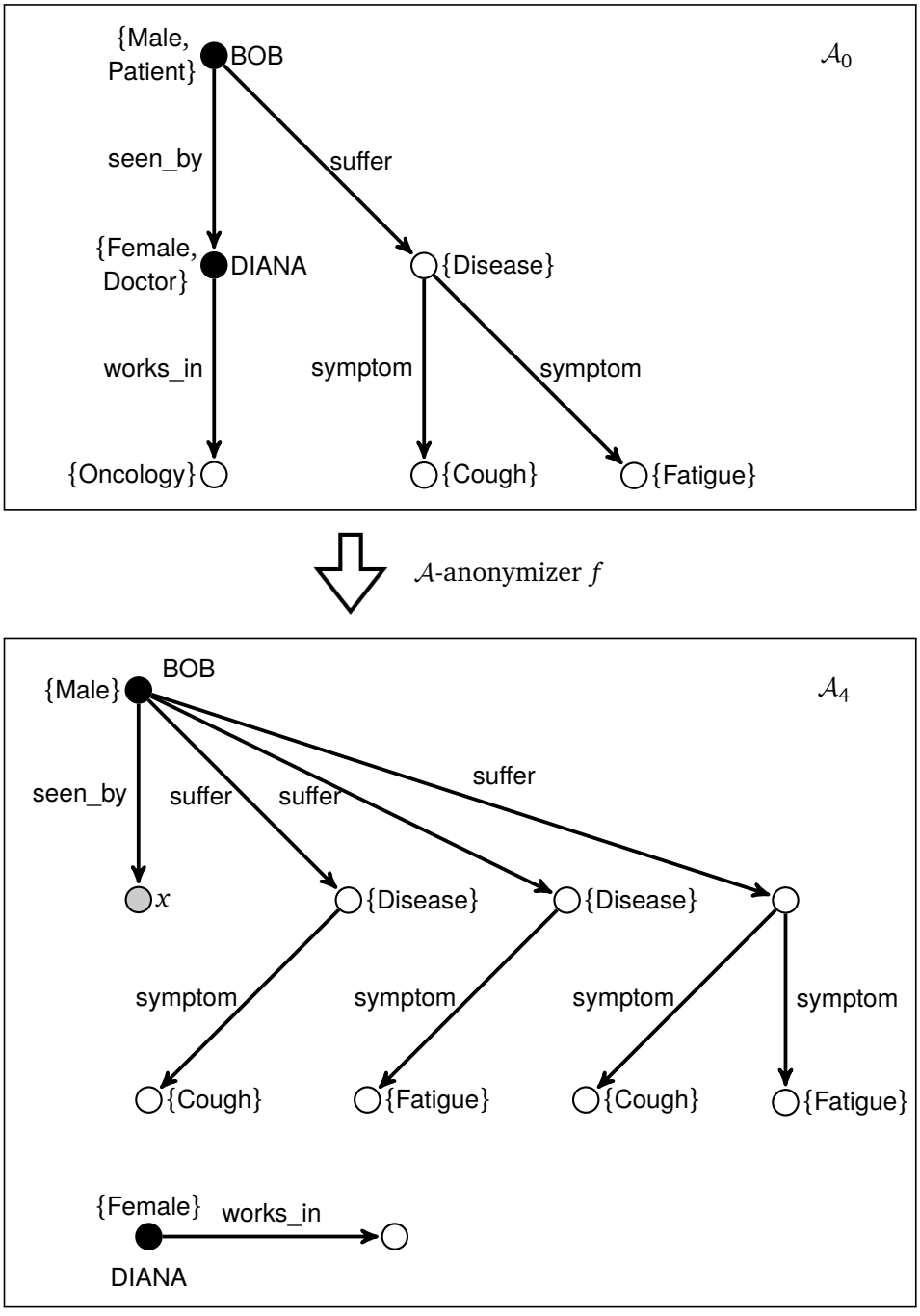


Figure 6.6: An illustration of anonymizing $\mathcal{A}_0$ to $\mathcal{A}_4$ via an $\mathcal{A}_0$-anonymizer f

Unlike the approach we used in Chapter 5 that we apply to strictly weaken $\mathcal{EL}$ instance stores by generalizing $\mathcal{EL}$ concepts, here we show that the anonymizers we defined do not *strictly* weaken the original ABox in general.

Example 6.8. *Consider the $\mathcal{EL}$ ABox $\mathcal{A} = \{\exists r.\top(a), r(a,b)\}$. If we apply an $\mathcal{A}$-anonymizer f that does not rename any individual, but only generalizes $\exists r.\top$ to $\top$, then we have the following anonymized ABox $\mathcal{A}' = \{\top(a), r(a,b)\}$. It can be easily seen that $\mathcal{A}$ is equivalent to $\mathcal{A}'$.* ♢

Next, one may also wonder whether for every $\mathcal{EL}$ ABoxes $\mathcal{A}$ and $\mathcal{A}'$ such that $\mathcal{A} \models \mathcal{A}'$, we can always construct an $\mathcal{A}$-anonymizer f such that $f(\mathcal{A}) = \mathcal{A}'$. This kind of desiderata, however, is countered by the following example.

Example 6.9. *Consider the $\mathcal{EL}$ ABoxes $\mathcal{A} = \{r(a,a)\}$ and $\mathcal{A}' = \{\exists r.\exists r.\exists r.\top(a)\}$. It holds that $\mathcal{A} \models \mathcal{A}'$. However, there is no $\mathcal{A}$-anonymizer f that can weaken $r(a,a)$ to a concept assertion $\exists r.\exists r.\exists r.\top(a)$ since f can only rename a or leave it as it is.* ♢

Now, we will compare our anonymizers with suppressors defined in [GK16; GK19]. By making use of the decomposition process we described in Section 6.1 to obtain $\mathcal{A}_d$ from a given $\mathcal{EL}$ ABox, we will show that our anonymizers actually provide more features to 'split' labeled nulls and 'add' more atoms when generalizing a concept than the suppressors defined in [GK16; GK19]. This is more clearly illustrated in the following toy example.

Example 6.10. *Suppose that we have an $\mathcal{EL}$ ABox $\mathcal{A} = \{(\exists r.(B_1 \sqcap B_2))(a)\}$. Then, we translate $\mathcal{A}$ to a dataset*

$$\mathcal{D}_{\mathcal{A}} = \{r(a,w), B_1(w), B_2(w)\},$$

where a and w are constant and null, respectively. Applying suppressors defined in Definition 1 of [GK16] to $\mathcal{D}_{\mathcal{A}}$, we are only able to rename a, or rename w, or just keep them as what they are. For instance, we want to hide the information that a has an r-successor that is also an instance of $B_1 \sqcap B_2$. One way to achieve this is by renaming w with a new anonymous one in one of its occurrences.

$$\hat{f}(\mathcal{D}_{\mathcal{A}}) = \{(r(a,w), B_1(w), B_2(w'))\}.$$

Using this dataset, the information we want to hide is exactly not inferred from the dataset. Likewise, if we rename w in the other occurrence, i.e., change $B_1(w)$ with $B_1(w')$. However, as an implication, the consequence stating that a has a relation to an instance of B_2 is lost with respect to $\hat{f}(\mathcal{D}_{\mathcal{A}})$. But then, using our anonymizers defined in Definition 6.4, we can preserve that consequence by generalizing $\exists r.(B_1 \sqcap B_2)$ to $\exists r.B_1 \sqcap \exists r.B_2$ such that if f is an $\mathcal{A}$-anonymizer, then we may have

$$f(\mathcal{A}) = \{(\exists r.B_1 \sqcap \exists r.B_2)(a)\}.$$

Now, if we transform $f(\mathcal{A})$ above to the corresponding dataset $\mathcal{D}_{f(\mathcal{A})}$, then we have

$$\mathcal{D}_{f(\mathcal{A})} = \{(r(a,w_1), r(a,w_2), B_1(w_1), B_2(w_2))\}.$$

By looking at the representation of $\mathcal{D}_{f(\mathcal{A})}$, our anonymizer has a sort of feature to 'split' null w to w_1 and w_2, and thus $\mathcal{D}_{f(\mathcal{A})}$ explicitly says that w_1 and w_2 are also instances of B_1 and B_2 respectively. This kind of feature does not belong the suppressors defined in [GK16; GK19].

In the context of preserving privacy of information of individuals as well as publishing them to the external Web, the notions of information privacy and information availability becomes a trade-off considered frequently. One needs to ensure that the information about individuals are protected in terms of their confidentiality but remains practically useful. To guarantee this property, similar to [GK16; GK19], we introduce an order on anonymizers. Given anonymizers f_1 and f_2, intuitively f_1 is *more informative than* f_2 if it can be obtained from f_2 by keeping more known individuals, identifying more distinct anonymous individuals, or specializing more $\mathcal{EL}$ concepts.

Definition 6.11. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f_1, f_2 be $\mathcal{A}$-anonymizers. The function f_1 is* more informative than *f_2, written $f_1 \geq f_2$, if and only if for all positions ρ and ρ' in $\mathcal{A}$,*

1.) *if $val(\rho, \mathcal{A}), val(\rho', \mathcal{A}) \in \mathsf{N_{AI}} \cup \mathsf{N_{KI}}$, then*
   a.) *if $f_2(\rho) \in \mathsf{N_{KI}}$, then $f_1(\rho) = f_2(\rho)$ and*
   b.) *if $f_2(\rho) = f_2(\rho')$, then $f_1(\rho) = f_1(\rho')$.*

2.) *if $val(\rho, \mathcal{A})$ is an $\mathcal{EL}$ concept, then $f_1(\rho) \sqsubseteq f_2(\rho)$.* ♢

We write $f_1 > f_2$ if $f_1 \geq f_2$, but $f_2 \geq f_1$ does not hold. If $f_1 \geq f_2$ and $f_2 \geq f_1$, then we have $f_1 \simeq f_2$, which means that f_1 is as informative as f_2.

One obvious consequence which happens due to the informativeness order defined above is that if f_1, f_2 are $\mathcal{A}$-anonymizers, then $f_1 \geq f_2$ implies $\mathcal{A}_1 \models \mathcal{A}_2$, where $f_1(\mathcal{A}) = \mathcal{A}_1$ and $f_2(\mathcal{A}) = \mathcal{A}_2$.

6.3 Formalizing Sensitive Information in $\mathcal{EL}$ ABoxes

As mentioned before that in this setting we plan to formalize the sensitive information into policies which are given either as an instance query ($\mathcal{EL}$ concept) or a conjunctive query. For the first type of policy, we emphasize that the *sensitive answers* for a given $\mathcal{EL}$ concept D w.r.t. an ABox are all known individuals, while for the second one, we consider that an answer of a conjunctive query q w.r.t. a given ABox $\mathcal{A}$ is *sensitive* if it is a tuple $\vec{a}$ of known individuals. Now, we are ready to define privacy properties that should be satisfied in PPOP in $\mathcal{EL}$ ABoxes.

Definition 6.12. *A* policy *P is either an $\mathcal{EL}$ concept D such that $D \not\equiv \top$ or a conjunctive query q. Given an $\mathcal{EL}$ ABox $\mathcal{A}$, an $\mathcal{A}$-anonymizer f, the $\mathcal{EL}$ ABox $\mathcal{A}$ is*

- compliant with D *iff* $\mathcal{A} \not\models D(a)$*, for all* $a \in \mathsf{N_{KI}}$,
- compliant with q *iff* $\mathcal{A} \not\models q(\vec{a})$ *for all tuples* $\vec{a}$ *of known individuals, and*
- safe for P *iff* $\mathcal{A} \cup \mathcal{A}'$ *complies with* P *for all* $\mathcal{A}'$ *complying with* P.

The $\mathcal{A}$-anonymizer f is an optimal P-compliant (safe) anonymizer of $\mathcal{A}$ *iff*

- $f(\mathcal{A})$ *is compliant with (safe for)* P*, and*
- *there is no $\mathcal{A}$-anonymizer f' such that $f'(\mathcal{A})$ is compliant with (safe for) P, where $f' > f$.* ♢

	X = IQ	**X = CQ**
COMPLIANCE$_X$	PTIME (Cor. 6.14)	CONP-complete (Cor. 6.14)
SAFETY$_X$	PTIME (Thm. 6.23)	Π_2^p and DP-hard (Thm. 6.23)
OPTIMAL-COMPLIANCE$_X$	CONP (Thm. 6.34)	Π_2^p and DP-hard (Thm. 6.34)
OPTIMAL-SAFETY$_X$	CONP (Thm. 6.34)	Π_3^p and DP-hard (Thm. 6.34)

Table 6.13: Complexity Results on PPOP in $\mathcal{EL}$ ABoxes

The standardization apart between anonymous individual names in $\mathcal{A}$ and $\mathcal{A}'$ also implies that the anonymous individuals are *first renamed apart before* constructing the set-theoretic union between $\mathcal{A}$ and $\mathcal{A}'$.

We are now ready to define the decision problems that will be investigated throughout this chapter. Let D be an $\mathcal{EL}$ concept, $\mathcal{A}$ be an $\mathcal{EL}$ ABox, and f be an $\mathcal{A}$-anonymizer. The formal definitions for the decision problems mentioned in the beginning of this chapter are as follows:

- The COMPLIANCE$_{IQ}$ *problem* asks whether $\mathcal{A}$ is compliant with D
- The SAFETY$_{IQ}$ *problem* asks if $\mathcal{A}$ is safe for D, and
- The OPTIMAL-COMPLIANCE$_{IQ}$ *(*OPTIMAL-SAFETY$_{IQ}$*) problem* asks whether f is an optimal D-compliant (safe) anonymizer of $\mathcal{A}$.

We define COMPLIANCE$_{CQ}$, SAFETY$_{CQ}$, OPTIMAL-COMPLIANCE$_{CQ}$, and OPTIMAL-SAFETY$_{CQ}$ analogously by requiring the policy to be a conjunctive query. The complexity results of all these problems are summarized in Table 6.13.

6.4 Compliance and Safety for $\mathcal{EL}$-ABoxes

First, we focus on the characterization of compliance in $\mathcal{EL}$ ABoxes. One may easily see that solving the compliance problem where the policy is an $\mathcal{EL}$ concept is the same as solving the complement of the instance problem in $\mathcal{EL}$ as characterized in Lemma 2.23, and thus by Lemma 2.24, the complexity is in in PTIME. Meanwhile, the compliance problem, where the policy is now a CQ, is the complement of the CQ entailment problem, described in Lemma 2.25, and thus the complexity of it is CONP-complete.

Corollary 6.14. *The* COMPLIANCE$_{IQ}$ *problem is in* PTIME *and the* COMPLIANCE$_{CQ}$ *problem is* CONP*-complete.*

Now, we move our attention to SAFETY$_{IQ}$. Before we can characterize this problem, we need to assume that D should be reduced. It is easy to see that $\mathcal{A}$ is safe for D iff $\mathcal{A}$ is safe for the reduced form of D. Moreover, during the proof of a characterization for SAFETY$_{IQ}$,

written in Lemma 6.16, a condition stating that $\mathcal{A}$ is *safe* for a reduced $\mathcal{EL}$ concept D' and an anonymous individual x will be taken into account to support the arguments within the proof. Thus, a mechanism to check whether this condition holds should be considered first.

Given an $\mathcal{EL}$ ABox $\mathcal{A}$, a reduced $\mathcal{EL}$ concept D', and an anonymous individual x occurring in $\mathcal{A}$, the ABox $\mathcal{A}$ is *safe for D' and x* iff for all $\mathcal{A}'$, we have $\mathcal{A} \cup \mathcal{A}' \not\models D'(x)$. In the following lemma, we show how to characterize this problem.

Lemma 6.15. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox, D' be a reduced $\mathcal{EL}$ concept, and $x \in \mathsf{N}_{\mathsf{AI}}$ occurring in $\mathcal{A}$. The ABox $\mathcal{A}$ is safe for D' and x iff $D' \not\equiv \top$ and one of the following conditions holds:*

1.) *there is $A \in \mathit{con}(D')$ such that for all $F(x) \in \mathcal{A}$, $A \notin \mathit{con}(F)$ or*

2.) *there is $\exists r.D'' \in \mathit{con}(D')$ such that*
   a.) *for all $F(x) \in \mathcal{A}$ and all $\exists r.F' \in \mathit{con}(F)$, we have $F' \not\sqsubseteq D''$ and*
   b.) *for all role assertions $r(x,u) \in \mathcal{A}$, we have $u \notin \mathsf{N}_{\mathsf{KI}}$ and $\mathcal{A}$ is safe for D'' and u.*

Proof. First, we assume that $\mathcal{A}$ is not safe for D' and x. We show that $D' \equiv \top$ or the two conditions above are violated. If $\mathcal{A}$ is not safe for D' and x, then there is $\mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A}' \models D'(x)$. Note that anonymous individuals in $\mathcal{A}$ and $\mathcal{A}'$ are renamed apart, and thus x does not occur in $\mathcal{A}'$. However, first consider that $\mathcal{A}' \models D'(x)$. If this the case and x does not occur in $\mathcal{A}'$, then it implies that D' is $\top$ since $\top(x)$ is a tautology and entailed by any ABox.

Now, consider that $\mathcal{A}' \not\models D'(x)$, but $\mathcal{A} \cup \mathcal{A}' \models D'(x)$. The first item in Lemma 2.23 implies that for all $A \in \mathtt{con}(D')$, there is $F(x) \in \mathcal{A} \cup \mathcal{A}'$ such that $A \in \mathtt{con}(F)$, while the second item in Lemma 2.23 implies that for all $\exists r.D'' \in \mathtt{con}(D')$, there is $F(x) \in \mathcal{A} \cup \mathcal{A}'$ and $\exists r.F' \in \mathtt{con}(F)$ such that $F' \sqsubseteq D''$ or there is $r(x,u) \in \mathcal{A} \cup \mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A}' \models D''(u)$. However, the assertions $F(x)$ and $r(x,u)$ only exist in $\mathcal{A}$ since $x \in \mathsf{N}_{\mathsf{AI}}$ and all anonymous individuals in $\mathcal{A}'$ and $\mathcal{A}$ are named differently. Since $F(x) \in \mathcal{A}$, this implies that the conditions 1.) and 2a.) are violated. Now, let $r(x,u) \in \mathcal{A}$. Then, there are two possibilities, which is either $u \in \mathsf{N}_{\mathsf{KI}}$ or $u \in \mathsf{N}_{\mathsf{AI}}$. If $u \in \mathsf{N}_{\mathsf{KI}}$, then this violates the condition 2b.) and if $u \in \mathsf{N}_{\mathsf{AI}}$, then it directly implies that $\mathcal{A}$ is not safe for D'' and u, which also violates the condition 2b.)

To show the converse direction, we assume that $D' \equiv \top$ or all the conditions above are false. If the former assumption holds, then it is easy to see that $\mathcal{A}$ is not safe for D' and x. Now, let the conditions 1.) and 2.) be violated. Suppose that all atoms in $\mathtt{con}(D')$ satisfy one of the following two conditions, which are the complement of conditions 1.) and 2a.), respectively:

- for all $A \in \mathtt{con}(D')$, there is $F(x) \in \mathcal{A}$ such that $A \in \mathtt{con}(F)$ and
- for all $\exists r.D'' \in \mathtt{con}(D')$, there is $F(x) \in \mathcal{A}$ and $\exists r.F' \in \mathtt{con}(F)$ such that $F' \sqsubseteq D''$.

Then, it directly implies that $\mathcal{A}$ is not safe for D' and x since $\mathcal{A} \models D'(x)$. However, if there are existential restrictions $\exists r.D'' \in \mathtt{con}(D')$ that satisfies the converse of the condition 2b.), which are

*.) there is $r(x,u) \in \mathcal{A}$, where $u \in \mathsf{N}_{\mathsf{KI}}$ or

**.) there is $r(x,u) \in \mathcal{A}$, where $u \in \mathsf{N}_{\mathsf{AI}}$ and $\mathcal{A}$ is not safe for D'' and u,

then the following considerations should be taken into account. First, let us consider all existential restrictions $\exists r_1.D_1'', \ldots, \exists r_p.D_p'' \in \mathtt{con}(D')$ which satisfy conditions *) or **). Without loss of generality, we may assume that there are two groups of these existential restrictions such that the first group consists of $\exists r_j.D_j''$ satisfying condition *.) for all $j = 1, \ldots, i$ and the second group consists of $\exists r_k.D_k''$ satisfying condition **.) for all $k = i+1, \ldots, p$. For each $\exists r_j.D_j''$ in the first group, we take one $u \in \mathsf{N_{KI}}$ such that $r(x,u) \in \mathcal{A}$, and then we construct an $\mathcal{EL}$ ABox $\mathcal{A}_j = \{D_j''(u)\}$. Then, for each $\exists r_k.D_k''$ in the second group, we take one $r(x,u) \in \mathcal{A}$, where $u \in \mathsf{N_{AI}}$, and then take one $\mathcal{EL}$ ABox $\mathcal{A}_k$ not implying $D_k''(u)$, but $\mathcal{A} \cup \mathcal{A}_k \models D_k''(u)$.

If we construct the union of all $\mathcal{A}_j$ and $\mathcal{A}_k$, then we have

$$\bigcup_{1 \le j \le i} \mathcal{A}_j \cup \bigcup_{i+1 \le k \le q} \mathcal{A}_k \not\models D(x).$$

This is because x does not occur as an individual in $\mathcal{A}_j$ which only consists of known individuals and, additionally, since $\mathcal{A}_k$ is constructed when making a union of $\mathcal{A}$ and $\mathcal{A}_k$, all anonymous individuals in $\mathcal{A}_k$ do not occur in $\mathcal{A}$, which means that x also does not occur in $\mathcal{A}_k$. Nevertheless, we have

$$\mathcal{A} \cup \bigcup_{1 \le j \le i} \mathcal{A}_j \cup \bigcup_{i+1 \le k \le q} \mathcal{A}_k \models D(x).$$

This can be shown by sufficiently considering all those existential restrictions atoms satisfying only conditions *.) or **.). For those atoms $\exists r_j.D_j''$ fulfilling condition *.), there is $r_j(x,u) \in \mathcal{A}$ and $D_j''(u) \in \mathcal{A}_j$ such that it finally holds that

$$\mathcal{A} \cup \bigcup_{1 \le j \le i} \mathcal{A}_j \models \exists r.D_j''(x).$$

Then, for those $\exists r_k.D_k'' \in \mathtt{con}(D')$ fulfilling condition **.), there is $r_k(x,u) \in \mathcal{A}$ and $\mathcal{A}_k$, where $\mathcal{A} \cup \mathcal{A}_k \models D_k(u)$, such that

$$\mathcal{A} \cup \bigcup_{i+1 \le k \le q} \mathcal{A}_k \models \exists r_i.D_k''(x).$$

To summarize, since the conditions 1.) and 2a.) are violated in $\mathcal{A}$ and either the condition *) or **) is satisfied by the some existential restrictions in $\mathtt{con}(D')$, this finally shows that $\mathcal{A}$ is not safe for D' and $x \in \mathsf{N_{AI}}$. □

Using the characterization above, clearly, it can be decided in polynomial time to check whether an ABox is safe for a reduced $\mathcal{EL}$ concept D' and $x \in \mathsf{N_{AI}}$. Besides considering this problem, we also require the notion of a *tree-shaped ABox of an $\mathcal{EL}$ concept*, that will also be used within the proof of Lemma 6.16.

Let D be a reduced $\mathcal{EL}$ concept. Referring to a translation from DL concepts to a first-order formula in Subsection 2.1.2, we construct $\pi_w(D)$. Then, we replace all variables in $\pi_w(D)$ with new anonymous individuals, and thus we have the following formula $\exists \vec{x}. \bigwedge \mathcal{A}_D$, where $\mathcal{A}_D$ is an $\mathcal{EL}$ ABox consisting of all assertions of the form $A(x_1)$ or $r(x_1, x_2)$ and $\vec{x}$ are anonymous individuals occurring in $\mathcal{A}_D$. From now on, we call $\mathcal{A}_D$ the *tree-shaped ABox of D*.

Note that every anonymous individual in $\mathcal{A}_D$ corresponds to an occurrence of a subconcept of D.

Further, for the anonymous individual x replacing the universally quantified variable w which occurs in $\pi_w(D)$, we call x the *individual root* of $\mathcal{A}_D$. An individual y in $\mathcal{A}_D$ is called *leaf* if it does not have any role assertion $r(y, y')$ for all $r \in \mathsf{N_R}$. The *depth* of an individual x' in $\mathcal{A}_D$, denoted by $\mathtt{dep}(x', \mathcal{A}_D)$, is $m \in \mathbb{N}$ iff there is a chain of role assertions $r_1(x', x_1), \ldots r_m(x_{m-1}, y)$ starting from x' to a leaf y and for all chains of role assertions $r_1(x', x_1), \ldots r_{m'}(x_{m'-1}, y')$ starting from x' to leaves y', we have $m' \leq m$.

Now, we are ready to characterize $\textsc{Safety}_{IQ}$ described in the following lemma.

Lemma 6.16. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and D be a reduced $\mathcal{EL}$ concept such that $D \not\equiv \top$. The ABox $\mathcal{A}$ is safe for D iff for all $a \in \mathsf{N_{KI}}$, the following holds*

1.) if $C(a) \in \mathcal{A}$ and $E \in sub(D)$, then C is safe$^{\exists}$ for $\{E\}$ and

2.) if $r(a, u) \in \mathcal{A}$ and $\exists r.D' \in sub(D)$, then $u \notin \mathsf{N_{KI}}$ and $\mathcal{A}$ is safe for D' and u.

Proof. First, we show the 'if direction' by assuming that $\mathcal{A}$ is not safe for D such that there are $c \in \mathsf{N_{KI}}$ and an ABox $\mathcal{A}'$ where $\mathcal{A}'$ complies with D, but $\mathcal{A} \cup \mathcal{A}' \models D(c)$. Instead of only considering the known individual c to prove that one of the conditions 1.) and 2.) above is violated, we use the following claim, that is more general, saying that if $u \in \mathsf{N_I}$ is a known individual or an anonymous individual occurring in $\mathcal{A}'$ and $\mathcal{A}' \not\models D(u)$, but $\mathcal{A} \cup \mathcal{A}' \models D(u)$, then either the condition 1.) or 2.) is not satisfied.

Claim 6.17. *Let $\mathcal{A}, \mathcal{A}'$ be $\mathcal{EL}$ ABoxes, D be a reduced $\mathcal{EL}$ concept, and $u \in \mathsf{N_I}$ be a known individual or an anonymous individual occurring in $\mathcal{A}'$. If $\mathcal{A}' \not\models D(u)$, but $\mathcal{A} \cup \mathcal{A}' \models D(u)$, then there is $b \in \mathsf{N_{KI}}$ such that*

**.) there are $C(b) \in \mathcal{A}$ and $E \in sub(D)$ such that C is not safe$^{\exists}$ for $\{E\}$ or*

***.) there are $r(b, u') \in \mathcal{A}$ and $\exists r.D' \in sub(D)$ such that $u' \in \mathsf{N_{KI}}$ or $\mathcal{A}$ is not safe for D' and u'.*

To show the claim above, we distinguish the following two cases step by step.

Case 1: Let $u \in \mathsf{N_{KI}}$. Since $\mathcal{A}' \not\models D(u)$, one of the following two conditions should be considered:

- there is $A \in \mathtt{con}(D)$ such that for all $C(u) \in \mathcal{A}'$, $A \notin \mathtt{con}(C)$ or
- there is $\exists r.D' \in \mathtt{con}(D)$ such that
  - for all $C(u) \in \mathcal{A}'$ and all $\exists r.C' \in \mathtt{con}(C)$, we have $C' \not\sqsubseteq D'$, and
  - for all $r(u, u') \in \mathcal{A}'$, we have $\mathcal{A}' \not\models D'(u')$.

In contrast, $\mathcal{A} \cup \mathcal{A}' \models D(u)$. The only possible situations that make this entailment holds are:

- For all concept names A that are imposed by the former condition above, there is $C(u) \in \mathcal{A}$ such that $A \in \mathtt{con}(C)$ or
- For all existential restrictions $\exists r.D'$ that satisfy the latter condition above,
  - there is $C(u) \in \mathcal{A}$ such that there is $\exists r.C' \in \mathtt{con}(C)$ and $C' \sqsubseteq \exists r.D'$ or
  - there is $r(u, u') \in \mathcal{A} \cup \mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A} \models D'(u')$.

If such assertions $C(u)$ occur in $\mathcal{A}$, then this directly implies that there is $u \in \mathsf{N_{KI}}$ such that there are $C(u) \in \mathcal{A}$ and atoms of the form A or $\exists r.D'$ in $\mathtt{con}(D) \subseteq sub(D)$ such that C is not safe$^{\exists}$ for either $\{A\}$ or $\{\exists r.D'\}$. Consequently, the condition *.) is fulfilled. Now, we consider the situation where there are $\exists r.D' \in \mathtt{con}(D)$ and $r(u,u') \in \mathcal{A} \cup \mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A} \models D'(u')$. If $r(u,u') \in \mathcal{A}$ and u' is a known individual, then the condition **.) is satisfied. Likewise, if $r(u,u') \in \mathcal{A}$ and u' is an anonymous individual, which directly means that u' occur in $\mathcal{A}$, but does not occur in $\mathcal{A}'$, then $\mathcal{A}$ is not safe for D' and u' and thus the condition **.) is satisfied. Now, let $r(u,u') \in \mathcal{A}'$ and u' be a known individual or an anonymous individual occurring in $\mathcal{A}'$. Since $\mathcal{A}' \not\models D'(u')$, $\mathcal{A} \cup \mathcal{A}' \models D'(u')$, and D' is reduced, by induction on the role depth of D, it shows that one of the conditions *.) and **.) holds. Finally, all possible situations that make $\mathcal{A} \cup \mathcal{A}' \models D(u)$ indeed ensure that one of the conditions *.) and **.) holds in the case that $u \in \mathsf{N_{KI}}$

Case 2: Let u be an anonymous individual occurring in $\mathcal{A}'$. Since $\mathcal{A}' \not\models D(u)$ and $\mathcal{A} \cup \mathcal{A}' \models D(u)$, it is easy to see that the only possible cause why $\mathcal{A}' \not\models D(u)$, but $\mathcal{A} \cup \mathcal{A}'$ entails $D(u)$ is that there is $\exists r.D' \in \mathtt{con}(D)$ such that

- for all $C(u) \in \mathcal{A}'$ and all $\exists r.C' \in \mathtt{con}(C)$, we have $C' \not\sqsubseteq D'$ and
- for all $r(u,u_1) \in \mathcal{A}'$, $\mathcal{A}' \not\models D'(u_1)$, but there is $r(u,u_2) \in \mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A}' \models D'(u_2)$.

Note that we only consider role assertions $r(u,u_2) \in \mathcal{A}'$ since u is an anonymous occurring in $\mathcal{A}'$ and thus it is not possible that $r(u,u_2) \in \mathcal{A}$. Let u_2 be a known individual or an anonymous individual occurring in $\mathcal{A}'$. Since $\mathcal{A}' \not\models D'(u_2)$, $\mathcal{A} \cup \mathcal{A}' \models D'(u_2)$, and D' is reduced, by induction on the role depth of D, it shows that one of the conditions *.) and **.) holds.

We have proved the claim above and now we are back to our assumption that $\mathcal{A}$ is not safe for D because there are $c \in \mathsf{N_{KI}}$ and an ABox $\mathcal{A}'$, where $\mathcal{A}'$ complies with D, but $\mathcal{A} \cup \mathcal{A}' \models D(c)$. Since c is a known individual, Claim 6.17 finally helps us show that one of the conditions 1.) and 2.) is violated.

To show the 'only if direction', we assume that one of the conditions 1.) and 2.) is violated, and then prove that $\mathcal{A}$ is not safe for D. This means that there are three possible conditions that are able to make $\mathcal{A}$ being not safe for D as described as follows:

*.) there are $a \in \mathsf{N_{KI}}$, $C(a) \in \mathcal{A}$, and $E \in sub(D)$ such that C is not safe$^{\exists}$ for $\{E\}$, or

**.) there are $a \in \mathsf{N_{KI}}$, $r(a,u) \in \mathcal{A}$, and $\exists r.D' \in sub(D)$ such that $u \in \mathsf{N_{KI}}$, or

***.) there are $a \in \mathsf{N_{KI}}$, $r(a,u) \in \mathcal{A}$, and $\exists r.D' \in sub(D)$ such that $u \notin \mathsf{N_{KI}}$ and $\mathcal{A}$ is not safe for D' and u.

First, let us consider the condition *). Since C is not safe$^{\exists}$ for $\{E\}$, as stated in Proposition 5.15, there are atoms $F_1 \in \mathtt{con}(E)$ and $F_2 \in \mathtt{con}(C)$ such that $F_2 \sqsubseteq F_1$. Then, we construct C' that is obtained from E by removing F_1 from $\mathtt{con}(E)$. This implies that $C' \not\sqsubseteq E$, but $C \sqcap C' \sqsubseteq E$. Moreover, we construct a syntactic generalization G of D that is obtained from D by replacing one occurrence of E in D with $\top$.

For this condition, we will construct two $\mathcal{EL}$ ABoxes $\mathcal{A}_{C'}$ and $\mathcal{A}_{\widehat{G}}$ such that $\mathcal{A}_{C'} = \{C'(a)\}$ and $\mathcal{A}_{\widehat{G}}$ is obtained from the tree-shaped $\mathcal{EL}$ ABox $\mathcal{A}_G$ of G by the following considerations:

- If the individual root x in $\mathcal{A}_G$ is the occurrence of E in D that is replaced by $\top$, then replace x with a.

- Otherwise, replace x with $c \in \mathsf{N_{KI}}$ such that c does not occur in $\mathcal{A}$ and for the individual x' in $\mathcal{A}_G$, where x' is the occurrence of E in D that is replaced by $\top$, we replace x' with a.

Note that if G is $\top$, then the corresponding tree-shaped $\mathcal{EL}$ ABox $\mathcal{A}_G$ is empty and $\mathcal{A}_{\widehat{G}}$ is also empty. Now, we show that the union $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ complies with D. Since the only known individuals occurring in $\mathcal{A}_{\widehat{G}}$ are c and a, it follows that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \not\models D(b)$, where $b \in \mathsf{N_{KI}} \setminus \{c, a\}$. To show whether $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ entails neither $D(a)$ nor $D(c)$, but $\mathcal{A} \cup \mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ entails either $D(a)$ or $D(c)$, the following cases are distinguished.

Case 1. Let $\mathcal{A}_{\widehat{G}}$ be an empty set. This implies that $D = E$ and the only known individual occurring in $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ is a. However, $C' \not\sqsubseteq E$ and thus $\mathcal{A}_{C'} \not\models D(a)$, which implies that $\mathcal{A}_{C'}$ complies with D. But, in this case we have $\mathcal{A} \cup \mathcal{A}_{C'} \models D(a)$ since the occurrence of $C(a) \in \mathcal{A}$ implies $C \sqcap C' \sqsubseteq D$, and thus a is an instance of D w.r.t. $\mathcal{A} \cup \mathcal{A}_{C'}$.

Case 2. Let $\mathcal{A}_{\widehat{G}}$ be not empty and the individual root of $\mathcal{A}_{\widehat{G}}$ be a. This implies that the only known individual occurring in $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ is a and E occurs in the top-level conjunction of D. Now, assume that $D = E \sqcap E'$, which implies that $G = E'$ and $\mathcal{A}_{\widehat{G}} = \{E'(a)\}$. Note that E and E' are incomparable w.r.t. subsumption order since D is reduced. However, for the atom $F_1 \in \mathtt{con}(E) \subset \mathtt{con}(D)$, since $C' \not\sqsubseteq E$, and E and E' are also not subsumed by each other, we know that there is no atom $F_2 \in \mathtt{con}(C') \cup \mathtt{con}(E')$, where $C'(a) \in \mathcal{A}_{C'}$, such that $F_2 \sqsubseteq F_1$. This consequently means that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \not\models D(a)$ and thus $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ is compliant with D. However, we have $\mathcal{A} \cup \mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(a)$ since again we have $C(a) \in \mathcal{A}$, which implies that $C \sqcap C' \sqsubseteq D$, and thus a is an instance of D w.r.t. $\mathcal{A} \cup \mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$.

Case 3. Let $\mathcal{A}_{\widehat{G}}$ be not empty and the individual root of $\mathcal{A}_{\widehat{G}}$ be c. This implies that the known individuals occurring in $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ are c and a. Then, since a is not the individual root of $\mathcal{A}_{\widehat{G}}$ and a is the occurrence of E in D that is replaced by $\top$, we know that E is a subconcept of D that does not occur in the top-level conjunction of D and thus $\mathtt{rd}(D) > 0$.

- Assume that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(a)$. Since $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ is tree-shaped, $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(a)$ implies that $\mathtt{dep}(a, \mathcal{A}_{\widehat{G}}) \geq \mathtt{rd}(D)$ or $\mathtt{rd}(C') \geq \mathtt{rd}(D)$. However, the individual a itself is an occurrence of E and E does not occur in the top-level conjunction of D, and thus $\mathtt{dep}(a, \mathcal{A}_{\widehat{G}}) < \mathtt{rd}(D)$. Meanwhile, by the construction of C', the concept C' itself is a subconcept of E, and hence $\mathtt{rd}(C') < \mathtt{rd}(D)$. This obviously contradicts our assumption that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(a)$.
- Assume that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$. Let $\mathcal{A}^a_{\widehat{G}} \subset \mathcal{A}_{\widehat{G}}$ such that $\mathcal{A}^a_{\widehat{G}}$ is tree-shaped and all individuals in $\mathcal{A}^a_{\widehat{G}}$ are reachable from a in $\mathcal{A}_{\widehat{G}}$. Note that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ iff $\mathcal{A}_{C'} \models E(a)$ or $\mathcal{A}^a_{\widehat{G}} \models E(a)$. However, the former, which is $\mathcal{A}_{C'} \models E(a)$, does not hold since $C' \not\sqsubseteq E$ and the latter, which is $\mathcal{A}^a_{\widehat{G}} \models E(a)$, does not hold either due to the similar arguments as written in Case 1 and 2 above. This consequently implies that our assumption saying that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ is wrong.

These two assumptions above finally yield a consequence that $\mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$ is compliant with D. However, it is easy again to see that $\mathcal{A} \cup \mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ due to the existence of $C(a) \in \mathcal{A}$, which implies that $C \sqcap C' \sqsubseteq D$. As the same as the arguments in the previous cases, we finally see that c is an instance of D w.r.t. $\mathcal{A} \cup \mathcal{A}_{C'} \cup \mathcal{A}_{\widehat{G}}$.

Using three cases above, we are able to show that the condition *.) makes $\mathcal{A}$ being not safe for D. Now, we move to the condition **.) and show that $\mathcal{A}$ is also not safe for D. For this condition, we have $a \in \mathsf{N_{KI}}$, $r(a,u) \in \mathcal{A}$, and $\exists r.D' \in \mathit{sub}(D)$ such that $u \in \mathsf{N_{KI}}$. We construct again a syntactic generalization G of D that is obtained by replacing an occurrence of $\exists r.D'$ in D with $\top$. Next, we construct two ABoxes $\mathcal{A}_{\widehat{G}}$, as analogously defined before for the condition *.), and $\mathcal{A}_{D'}$ consisting of a single assertion $D'(u)$ only. This construction implies that there are only c, a, and u as known individuals in $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_{D'}$ (u may be the same or different with a). However, it is easy to see that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_{D'} \not\models D(u)$ since $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_{D'} \models D(u)$ iff $\mathcal{A}_{D'} \models D(u)$, but we know that D is reduced and $D' \in \mathit{sub}(D)$ does not occur in the top-level conjunction of D, and thus $D' \not\sqsubseteq D$ and $\mathcal{A}_{D'} \not\models D(u)$. It remains to show that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_{D'}$ entails neither $D(a)$ nor $D(c)$.

We skip the case where $\mathcal{A}_{\widehat{G}}$ is empty or the case where $\mathcal{A}_{\widehat{G}}$ is not empty and the individual root of $\mathcal{A}_{\widehat{G}}$ is a since the arguments within these cases to show that $\mathcal{A}$ is not safe for D are similar to Case 1 and Case 2 for the condition *.). We focus on the case where $\mathcal{A}_{\widehat{G}}$ is not empty and c is the individual root of $\mathcal{A}_{\widehat{G}}$. If we assume that $\mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \models D(a)$, then this implies that $\mathtt{dep}(a, \mathcal{A}_{\widehat{G}}) \geq \mathtt{rd}(D)$ or $\mathtt{rd}(D') \geq \mathtt{rd}(D)$. However, as written in Case 3 above, we know that $\mathtt{dep}(a, \mathcal{A}_{\widehat{G}}) < \mathtt{rd}(D)$, whereas D' is a subconcept of D not occurring in the top-level conjunction of D and thus $\mathtt{rd}(D') < \mathtt{rd}(D)$. This contradiction shows that $\mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \not\models D(a)$. Further, it is easy to see that $\mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ iff $\mathcal{A}_{D'} \cup \mathcal{A}^{a}_{\widehat{G}} \models \exists r.D'(a)$. However, it does not hold since $D' \not\sqsubseteq \exists r.D'$ and there is no $r(a,x) \in \mathcal{A}^{a}_{\widehat{G}}$ such that $\mathcal{A}^{a}_{\widehat{G}} \models D'(x)$ as analogously argued in Case 1 or 2 for the condition *.) above. This consequently implies that $\mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \not\models D(c)$.

Now, the arguments written in the previous paragraph show that $\mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}}$ is compliant with D, but it is easy to see that $\mathcal{A} \cup \mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ since there is $r(a,u) \in \mathcal{A}$, where $u \in \mathsf{N_{KI}}$, and $\mathcal{A} \cup \mathcal{A}_{D'} \cup \mathcal{A}_{\widehat{G}} \models D'(u)$. This finally shows that $\mathcal{A}$ is not safe for D by assuming that the condition **.) holds.

We turn our attention now to the last condition ***.). This condition consists of an assumption that there are $r(a,u) \in \mathcal{A}$ and $\exists r.D' \in \mathit{sub}(D)$ such that $u \in \mathsf{N_{AI}}$ and $\mathcal{A}$ is not safe for D' and u. This implies that there is $\mathcal{A}'$ such that $\mathcal{A} \cup \mathcal{A}' \models D'(u)$. In the case $D' \equiv \top$ or, otherwise, $D' \not\equiv \top$, but $\mathtt{rd}(D') = 0$, then $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ iff for all $A \in \mathtt{con}(D')$, there is $C(u) \in \mathcal{A}$ such that $A \in \mathtt{con}(C)$. For this case, we have $\mathcal{A} \models D'(u)$. Now, we need to construct a syntactic generalization G of D that is obtained from D by replacing an occurrence of $\exists r.D'$ in D with $\top$, and then construct $\mathcal{A}_{\widehat{G}}$ as defined before for the conditions *.) and **.). Similar to the condition **.), $\mathcal{A}_{\widehat{G}}$ does not entail both $D(a)$ and $D(c)$, which implies that $\mathcal{A}_{\widehat{G}}$ complies with D. But, it is easy to see that $\mathcal{A} \cup \mathcal{A}_{\widehat{G}} \models D(c)$ iff there are $r(a,u) \in \mathcal{A}$ and $\mathcal{A} \models D'(u)$. Thus, we know that $\mathcal{A}$ is not safe for D if the condition ***.) holds, where $\mathtt{rd}(D') = 0$ or $D' \equiv \top$.

Now, we go to the assumption that $\mathtt{rd}(D') > 0$. By Lemma 2.23, $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ holds because

- for all $A \in \mathtt{con}(D')$, there is $C(u) \in \mathcal{A}$ such that $A \in \mathtt{con}(C)$ and
- for all $\exists r.D'' \in \mathtt{con}(D')$, there is $C(u) \in \mathcal{A}$ and $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqsubseteq D''$ or there is $r(u,u') \in \mathcal{A}$ such that $\mathcal{A} \cup \mathcal{A}' \models D''(u)$.

Intuitively, one reason why $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ is because there are known individuals reachable from u in $\mathcal{A}$ and additionally, there are also some axioms speaking about these known

individuals in $\mathcal{A}'$ such that if $\mathcal{A}$ and $\mathcal{A}'$ is combined, then $D'(u)$ is revealed w.r.t. $\mathcal{A} \cup \mathcal{A}'$. For this reason, we construct a sort of *ground ABox* $\mathcal{A}_\chi$ that consists of only known individuals and is obtained from $\mathcal{A} \cup \mathcal{A}'$ based on a function χ formally defined as follows.

Let $\mathcal{A}, \mathcal{A}'$ be $\mathcal{EL}$ ABoxes, D' be an $\mathcal{EL}$ concept such that $\mathtt{rd}(D') > 0$, and $u \in \mathsf{N_{AI}}$ occur in $\mathcal{A}$ such that $\mathcal{A} \cup \mathcal{A}' \models D'(u)$. The function $\chi : \mathtt{con}^{\exists}(D') \to \mathsf{N_I}$ is called an *entailment function induced by* $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ if

$$\chi(\exists r.D'') = u \text{ iff } r(u, u') \in \mathcal{A} \text{ and } \mathcal{A} \cup \mathcal{A}' \models D''(u'). \tag{6.1}$$

Then, given $\mathcal{A}, \mathcal{A}'$, an $\mathcal{EL}$ concept D', an individual $u \in \mathsf{N_{AI}}$ occurring in $\mathcal{A}$, and an entailment function χ induced by $\mathcal{A} \cup \mathcal{A}' \models D'(u)$, we construct a *ground ABox* $\mathcal{A}_\chi$ *induced by* $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ by performing the following steps:

- First, we initialize $\mathcal{A}_\chi := \emptyset$.
- Second, for each $\exists r.D'' \in \mathtt{con}^{\exists}(D')$ such that $\chi(\exists r.D'') = b \in \mathsf{N_{KI}}$, we add a concept assertion $D''(b)$ to $\mathcal{A}_\chi$.
- Last, for each $\exists r.D'' \in \mathtt{con}^{\exists}(D')$ such that $\chi(\exists r.D'') = x \in \mathsf{N_{AI}}$, we take an entailment function χ' induced by $\mathcal{A} \cup \mathcal{A}' \models D''(x)$ mapping existential restrictions from $\mathtt{con}^{\exists}(D'')$ to $\mathsf{N_I}$, and then add $\mathcal{A}_{\chi'}$ to $\mathcal{A}_\chi$.

Using this sort of ABox, we make the following claim.

Claim 6.18. *A ground ABox* $\mathcal{A}_\chi$ *induced by* $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ *does not entail* $D'(u)$*, but* $\mathcal{A} \cup \mathcal{A}_\chi \models D'(u)$*.*

The first conjecture stating that $\mathcal{A}_\chi \not\models D'(u)$ is obvious since there is no information about anonymous individuals in $\mathcal{A}_\chi$. The second statement saying that $\mathcal{A} \cup \mathcal{A}_\chi \models D'(u)$ is justified as follows:

- Since $\mathcal{A} \cup \mathcal{A}' \models D'(u)$ and u only occur in $\mathcal{A}$, it is obvious to see that for all $A \in \mathtt{con}(D')$, there is $C(u) \in \mathcal{A}$ such that $A \in \mathtt{con}(C)$.
- Then, due to $\mathcal{A} \cup \mathcal{A}' \models D'(u)$, it also implies that there are $\exists r.D''' \in \mathtt{con}(D')$ that have $C(u) \in \mathcal{A}$ and $\exists r.C' \in \mathtt{con}(C)$ such that $C' \sqsubseteq D'''$. But, there are also $\exists r.D'' \in \mathtt{con}(D')$ that do not have such $C(u) \in \mathcal{A}$ and $\exists r.C' \in \mathtt{con}(C)$. However, since $\mathcal{A}_\chi$ is constructed based on the function χ satisfying Equation 6.1, we have $r(u, u') \in \mathcal{A}$ such that $\mathcal{A} \cup \mathcal{A}' \models D''(u')$. If $u' \in \mathsf{N_{KI}}$, then there is $D''(u') \in \mathcal{A}_\chi$ and thus we have $\mathcal{A} \cup \mathcal{A}_\chi \models D''(u')$. If $u' \in \mathsf{N_{AI}}$, then by induction on the role depth of D', there is $\mathcal{A}_{\chi'} \subseteq \mathcal{A}_\chi$ such that $\mathcal{A} \cup \mathcal{A}_{\chi'} \models D''(u')$.

Now, we are back to our aim to show that $\mathcal{A}$ is not safe for D if the condition ***.) holds and $\mathtt{rd}(D') > 0$. Again, we skip the case where $\mathcal{A}_{\widehat{G}}$ is empty or the case where $\mathcal{A}_{\widehat{G}}$ is not empty and the individual root of $\mathcal{A}_{\widehat{G}}$ is a since these can be proved by following arguments in Case 1 and Case 2 for the condition *.). We jump directly to the case where $\mathcal{A}_{\widehat{G}}$ is not empty and the individual root of $\mathcal{A}_{\widehat{G}}$ is c. Note that using Claim 6.18, we can say that if there is $\mathcal{A}'$ such that $\mathcal{A}' \not\models D'(u)$, then there is a ground ABox $\mathcal{A}_\chi$ w.r.t. $\mathcal{A} \cup \mathcal{A}' \models D$ such that $\mathcal{A}_\chi \not\models D'(u)$, but $\mathcal{A} \cup \mathcal{A}_\chi \models D'(u)$. Thus, it remains to show that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi$ is compliant with D, but $\mathcal{A} \cup \mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi$ is not compliant with D.

It is obvious to see that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \models D(b)$ iff $\mathcal{A}_\chi \models D(b)$ for all $b \in \mathsf{N_{KI}} \setminus \{c, a\}$. However, by the construction of $\mathcal{A}_\chi$, the ABox $\mathcal{A}_\chi$ only consists of concept assertions whose concept is a subconcept of D not occurring in $\mathtt{con}(D)$, and thus all concept assertions, of which b is an instance in $\mathcal{A}$, do not belong to $\mathtt{con}(D)$, which shows that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \not\models D(b)$. It is also clear to see that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \models D(a)$ iff $\mathcal{A}^a_{\widehat{G}} \cup \mathcal{A}_\chi \models D(a)$. However, all concept assertions speaking about a in $\mathcal{A}^a_{\widehat{G}} \cup \mathcal{A}_\chi$ contains concepts that are subconcepts of D and not occurring in $\mathtt{con}(D)$, and thus again we have $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \not\models D(a)$. Last, $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \models D(c)$ iff $\mathcal{A}^a_{\widehat{G}} \cup \mathcal{A}_\chi \models \exists r.D'(a)$ and yet there is no $r(a, x) \in \mathcal{A}^a_{\widehat{G}}$ such that $\mathcal{A}^a_{\widehat{G}} \models D'(x)$ as analogously stated in Case 1 for the condition *.) above and, additionally, for all concept assertions $C(a) \in \mathcal{A}_\chi$, by the construction of $\mathcal{A}_\chi$, we can see that $C \in sub(D')$ does not occur in the top-level conjunction of D'. Nevertheless, we can show that $\mathcal{A} \cup \mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \models D(c)$ because there is $r(a, u) \in \mathcal{A}$ such that $\mathcal{A} \cup \mathcal{A}_\chi \models D'(u)$ and $\mathcal{A} \cup \mathcal{A}_\chi \models \exists r.D'(a)$, which implies that $\mathcal{A} \cup \mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi \models D(c)$. Therefore, we show that $\mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi$ is compliant with D, but $\mathcal{A} \cup \mathcal{A}_{\widehat{G}} \cup \mathcal{A}_\chi$ is not compliant with D.

Finally, by assuming that one of the conditions *.), **.), and ***.) holds and looking all the arguments above, we finally show that $\mathcal{A}$ is not safe for D, which brings us to a conclusion that these three conditions are sound and we are finally done to prove this lemma. □

To be more illustrative on the conditions stated in Lemma 6.16, let us consider the medical example we showed in the beginning of this chapter.

Example 6.19. *As stated before that the $\mathcal{EL}$ ABox $\mathcal{A}_3$ is safe for D. This is justified by the following conditions:*

- *The concepts* $\mathsf{Male} \sqcap \exists \mathsf{suffer}.(\mathsf{Disease} \sqcap \exists \mathsf{symptom}.\mathsf{Fatigue})$ *and* $\mathsf{Female} \sqcap \exists \mathsf{works\_in}.\top$*, of which the individuals* BOB *and* DIANA *are instances of, respectively, are safe$^\exists$ for any subconcept of D.*
- *Then, the role assertion* $\mathsf{seen\_by}(\mathsf{BOB}, x)$ *asserts that* BOB *does not have* seen_by*-successors that are known individuals.*
- *Last, for each anonymous individuals, which are only x in this case, it is easy to see that $\mathcal{A}_3$ is safe for* $\mathsf{Doctor} \sqcap \exists \mathsf{works\_in}.\mathsf{Oncology}$. ♢

Now, we turn our attention to the problem of deciding whether an $\mathcal{EL}$ ABox is safe for a conjunctive query. Recall the definition of $\textsc{Safety}_{CQ}$, given an $\mathcal{EL}$ ABox $\mathcal{A}$ and a CQ q, we need to consider all $\mathcal{EL}$ ABoxes $\mathcal{A}'$ that are compliant with q, but $\mathcal{A} \cup \mathcal{A}' \not\models q$. However, there may be infinitely many such $\mathcal{A}'$ with arbitrary size. Therefore, the following lemma defines a characterization that can allow us to only check finitely many q-compliant $\mathcal{EL}$ ABoxes whose size is bounded by the size of a query in q.

Lemma 6.20. *Given an $\mathcal{EL}$ ABox $\mathcal{A}$ and a CQ q, deciding whether $\mathcal{A}$ is safe for q is in Π^p_2.*

Proof. Let us assume that $\mathcal{A}$ is not safe for q. This implies that there is a tuple $\vec{a}$ of known individuals, and an $\mathcal{EL}$ ABox $\mathcal{A}'$ such that $\mathcal{A}'$ complies with q, but $\mathcal{A} \cup \mathcal{A}' \models q(\vec{a})$. It implies that there is a homomorphism h from q to $\mathcal{A}_d \cup \mathcal{A}'_d$. Now, let $\mathcal{A}'' \subseteq \mathcal{A}'_d$ be the homomorphic image of h over $q(\vec{a})$. It is obvious to see that $\mathcal{A}'' \not\models q(\vec{a})$ and $\mathcal{A}''$ has the size bounded by q. It is also clear that $sig(\mathcal{A}'')$ is a subset of the set of all known individuals, concept names,

and role names occurring in q. Thus, to check whether $\mathcal{A}$ is safe for q, first we check finitely many $\mathcal{A}''$ whose size is at most $|\, q \,|$ and the signature is a subset of the symbols occurring in q. Then, we check either $\mathcal{A}''$ does not comply with q or $\mathcal{A} \cup \mathcal{A}''$ complies with q, both of which can be done by calling an NP oracle as stated by Corollary 6.14. □

Unfortunately, the precise complexity of SAFETY_{CQ} still remains open. This is justified by the following two lemmas describing hardness results for this problem. We start with a lemma stating that SAFETY_{CQ} is at least as hard as the GRAPHHOMOMORPHISM problem.

Lemma 6.21. *There is a polynomial reduction from* GRAPHHOMOMORPHISM *to* SAFETY_{CQ}.

Proof. The input of GRAPHHOMOMORPHISM consists of two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$. Then, without loss of generality, we assume that $\mathcal{G}_1$ is weakly connected, which means that there is an undirected path from any node to any node, and both $\mathcal{G}_1$ and $\mathcal{G}_2$ do not have the same nodes. We construct an $\mathcal{EL}$ ABox $\mathcal{A} := \{r(a, b)\}$, where $a, b \in \mathsf{N}_{\mathsf{I}}$ and a CQ q, where q is a Boolean conjunctive query consisting of the following conjuncts:

- $s(w_{\mu_1}, w_{\mu_2})$ for all edges (μ_1, μ_2) of $\mathcal{G}_1$ and $\mathcal{G}_2$,
- $r(w_{\hat{\mu}}, w)$ for an arbitrary chosen node $\hat{\mu}$ of $\mathcal{G}_1$, and
- $r(w_\nu, w_\nu)$ for each node ν of $\mathcal{G}_2$.

It is obvious to see that the construction above can be done in polynomial time. Now, we show that $\mathcal{A}$ is safe for q iff there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$.

Now, we start with the forward direction. Since $\mathcal{A}$ is safe for q, it is known that if the union of $\mathcal{A}$ with an $\mathcal{EL}$ ABox $\mathcal{A}'$ does not comply with q, then $\mathcal{A}'$ does not comply with q either. Let h be a function that sends $w_{\hat{\mu}}$ and w to a and b, respectively and all other variables to other fresh known individuals. We apply this function to all atoms in q and thus we have an $\mathcal{EL}$ ABox $\mathcal{A}'' = h(q)$, where $\mathcal{A} \subseteq \mathcal{A}''$. It means that h is a homomorphism from q to $\mathcal{A}''$. Since $\mathcal{A}$ is safe for q, this implies that there is a homomorphism h' from q to $\mathcal{A}'' \setminus \{r(a, b)\}$. By the construction of $\mathcal{A}''$, there is no node μ of $\mathcal{G}_1$ and a known individual c such that $r(h(w_\mu), c)$ in $\mathcal{A}'' \setminus \{r(a, b)\}$. This implies that for the node $\hat{\mu}$ of $\mathcal{G}_1$, we have $h'(w_{\hat{\mu}}) = h(w_\nu)$, where ν is a node of $\mathcal{G}_2$. Since $\mathcal{G}_1$ is weakly connected and $\mathcal{G}_1$ and $\mathcal{G}_2$ do not have any common node, for any node μ of $\mathcal{G}_1$, there is a node ν of $\mathcal{G}_2$ such that $h'(w_\mu) = h(w_\nu)$. Thus, h' also defines a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$.

For the converse direction, we assume that there is a homomorphism h'' from $\mathcal{G}_1$ to $\mathcal{G}_2$. Let us take an $\mathcal{EL}$ ABox $\mathcal{A}'$ such that there is a homomorphism h from q to $\mathcal{A} \cup \mathcal{A}'$. We need to show that there is also a homomorphism from q to $\mathcal{A}'$. Since any atom $r(w_\nu, w_\nu)$ in q, where ν is a node of $\mathcal{G}_2$, cannot be mapped to $r(a, b)$, we define a function h' that maps

- w_μ to $h(w_{h''(\mu)})$ for all nodes μ of $\mathcal{G}_1$ and
- w to $h(w_{h''(\hat{\mu})})$, and
- w_ν to $h(w_\nu)$ for all nodes ν of $\mathcal{G}_2$.

This can be seen that h' is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$. □

Since it is known that GRAPHHOMOMORPHISM is NP-complete, it implies that SAFETY_{CQ} is NP-hard. However, the lemma above uses a reduction that relies on an $\mathcal{EL}$ ABox with a single

assertion. In the following lemma, we show that SAFETY$_{CQ}$ has a better lower bound, which is in DP-hard, obtained from the reduction of HOMOMORPHISM-NOHOMOMORPHISM that is known in DP-complete. In this reduction, we use an $\mathcal{EL}$ ABox that has multiple assertions.

Lemma 6.22. *There is a polynomial reduction from* HOMOMORPHISM-NOHOMOMORPHISM *to* SAFETY$_{CQ}$.

Proof. The input of HOMOMORPHISM-NOHOMOMORPHISM consists of four connected directed graphs $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}'_1, \mathcal{G}'_2$ and then check whether there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$. We use these four graphs in order to construct an $\mathcal{EL}$ ABox $\mathcal{A}$ as well as a CQ q before checking where $\mathcal{A}$ is safe for q.

First we define $\mathcal{A} := \{r(x,x)\} \cup \{s(x_\mu, x_\nu) \mid (\mu, \nu)$ is an edge in $\mathcal{G}'_2\}$, where $x, x_u, x_v \in \mathsf{N_{AI}}$ for all nodes μ, ν of $\mathcal{G}'_2$. Then, we define a Boolean CQ q consisting of the following parts:

1. A direct representation of graph $\mathcal{G}_1$ over binary predicate name r and fresh existentially quantified variables for all nodes of $\mathcal{G}_1$. Let E_1 be the set of all edges in $\mathcal{G}_1$. This first part is formally represented as

$$\exists \vec{w}.\phi(\vec{w}) = \exists \vec{w}. \bigwedge_{(\mu,\nu)\in E_1} r(w_\mu, w_\nu).$$

2. A direct representation of graph $\mathcal{G}_2$ over role name r and fresh existentially quantified variables for all nodes of $\mathcal{G}_2$ and, in addition, every variable occur in atom $A(x)$. If V_2 and E_2 be the set of all nodes and edges in $\mathcal{G}_2$, respectively, then this second part is formally represented below.

$$\exists \vec{w'}.\phi'(\vec{w'}) = \exists \vec{w'}. \bigwedge_{(\mu,\nu)\in E_2} r(w'_\mu, w'_\nu) \wedge \bigwedge_{\mu\in V_2} A(w'_\mu).$$

3. A direct representation of graph $\mathcal{G}'_1$ over role name s and fresh existentially quantified variables for all nodes of $\mathcal{G}'_1$. Let E'_1 be the set of all edges in $\mathcal{G}'_1$. The third part is formally represented as

$$\exists \vec{w''}.\phi''(\vec{w''}) = \exists \vec{w''}. \bigwedge_{(\mu,\nu)\in E'_1} s(w''_\mu, w''_\nu).$$

Now, we claim that there is a homomorphism from G_1 to G_2, but there is no homomorphism from G'_1 to G'_2 iff $\mathcal{A}$ is safe for q.

We first prove the completeness of this characterization. We show that if $\mathcal{A}$ is safe for q, then for all $\mathcal{A}'$, either there is a homomorphism from q to $\mathcal{A}'$ or there is no homomorphism from q to $\mathcal{A} \cup \mathcal{A}'$. Let $\mathcal{A}'$ coincide with the second and the third part of q, except now we have only known individuals in place of variables. In formal way, we define

$$\mathcal{A}' = \{A(a_\mu) \mid \mu \in V_2\} \cup \{r(a_\mu, a_\nu) \mid (\mu,\nu) \in E_2\} \cup \{s(a_{\mu'}, a_{w_{\nu'}}) \mid (\mu', \nu') \in E'_1\}.$$

If all variables in part 1 of q are mapped to x in $\mathcal{A}$ and all other variables are mapped to their counterparts in $\mathcal{A}'$, then there is a homomorphism from q to $\mathcal{A} \cup \mathcal{A}'$. Since $\mathcal{A}$ is safe for $\mathcal{P}_{CQ}$, it implies that there is a homomorphism from q to $\mathcal{A}'$ and thus there is a homomorphism

from $\mathcal{G}_1$ to $\mathcal{G}_2$. Now, we consider the representation of $\mathcal{A}'$ coincides with part 1 and part 2 of q and again we only have known individuals in $\mathcal{A}'$ in place of variables. This is formally defined as

$$\begin{aligned}\mathcal{A}' =& \{r(a_\mu, a_\nu) \mid (\mu, \nu) \in E_1\} \cup \\ & \{r(a_{\mu'}, a_{\nu'}) \mid (\mu', \nu') \in E_2\} \cup \{A(a_{\mu'}) \mid \mu' \in V_2\}\end{aligned}$$

Since $\mathcal{A}'$ does not have role name s, there is no homomorphism from q to $\mathcal{A}'$. However, since $\mathcal{A}$ is safe for q, we know that there is no homomorphism from q to $\mathcal{A} \cup \mathcal{A}'$. This is due to the fact we can send all parts 1 and 2 of q will be $\mathcal{A}'$, but then there is no homomorphism from part 3 to the part of $\mathcal{A} \cup \mathcal{A}'$ over s and thus there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$.

To check the soundness of this characterization, we assume that there is a homomorphism from G_1 to G_2, but there is no homomorphism from G'_1 to G'_2. Then, we assume by contradiction that $\mathcal{A}$ is not safe for q, which implies that there is $\mathcal{A}'$ such that there is no homomorphism from q to $\mathcal{A}'$, but there is a homomorphism from q to $\mathcal{A} \cup \mathcal{A}'$. Since $\mathcal{A}$ does not contain any concept assertion and there is a homomorphism from q to $\mathcal{A} \cup \mathcal{A}'$, this implies that there is a homomorphism from the second part of q to $\mathcal{A}'$. Since there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and both are the representations of the first and the second parts of q, respectively, it means that we have a homomorphism from the first part of q to $\mathcal{A}'$. Since there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$ and all graphs are connected, this implies that there is no homomorphism from the third part of q to $\mathcal{A}$ over the role name s, which consequently says that there is a homomorphism from the third part of q to $\mathcal{A}'$. Due to all previous arguments, it can be inferred that there is a homomorphism from q to $\mathcal{A}'$, which violates our assumption that $\mathcal{A}'$ is not safe for q. Hence, $\mathcal{A}$ is safe for q. □

It is easy to see that the conditions written in both Lemma 6.15 and Lemma 6.16 can be checked in polynomial time. Then, Lemma 6.20 and 6.21 also provide the upper-bound and the lower-bound for the SAFETY_{CQ}. This finally lead us to the following theorem stating the complexity results for SAFETY_{IQ} and SAFETY_{CQ}.

Theorem 6.23. SAFETY_{IQ} *is in* PTIME*, whereas* SAFETY_{CQ} *is in* Π_2^p *and* DP*-hard.*

6.5 Optimal Anonymizers

In this section, we plan to investigate $\text{OPTIMAL-COMPLIANCE}_X$ and OPTIMAL-SAFETY_X, where $X \in \{IQ, CQ\}$, and then presents the upper bounds and lower bounds for both problems in $\mathcal{EL}$ ABoxes. Recall that these problems ask whether a given $\mathcal{A}$-anonymizer f is an optimal P-compliant (safe) anonymizer of an ABox $\mathcal{A}$, where P is either an $\mathcal{EL}$ concept or a conjunctive query. One idea to solve these problems is first to check whether $f(\mathcal{A})$ itself is already compliant or safe with respect to P. If it is the case, then we look at all functions f' that are *adjacent* to f w.r.t. $\mathcal{A}$ and then check whether $f'(\mathcal{A})$ is also compliant or safe with respect to P. Intuitively, this adjacency says that f' is strictly more informative than f and there is no other f'' that lies in between f' and f w.r.t. the informativeness order. Formally, given $\mathcal{A}$-anonymizers f and f', we say that f' *adjacent* to f w.r.t. $\mathcal{A}$, written $f' >_1 f$, iff $f' > f$ and there is no f'' such that $f' > f'' > f$. As a simple illustration, the following example is provided.

Example 6.24. *We consider $\mathcal{EL}$ ABoxes $\mathcal{A}_0$ and $\mathcal{A}_4$ presented in the beginning of this chapter and an $\mathcal{A}_0$-anonymizer f such that $f(\mathcal{A}_0) = \mathcal{A}_4$. The following two $\mathcal{A}_0$-anonymizers f', f'' are adjacent to f such that f' is obtained from f by only replacing x to* DIANA *and f'' is obtained from f by only adding* Patient *to the top-level conjunction of the concept of which the individual* BOB *is the instance in $\mathcal{A}_0$.* ♢

Next, we present the definition for a set of functions that eventually will be proved as the set of all of adjacent functions of a given $\mathcal{A}$-anonymizer f. Given $\mathcal{EL}$ concepts C and C', the following definition requires the set $LA_\Sigma(C')$ of all lowering atoms for C' w.r.t. Σ, where $\Sigma = sig(C)$.

Definition 6.25. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f be an $\mathcal{A}$-anonymizer. We define $\mathtt{adj}(f, \mathcal{A})$ as the set of all functions f' that can be obtained from f by performing exactly one of the following operations.*

1. *Take $x \in \mathsf{N_{AI}}$ occurring in $f(\mathcal{A})$ and all positions $\rho_1, \ldots, \rho_n$, where*

$$f(\rho_i) = x \text{ and } val(\rho_i, \mathcal{A}) = a \in \mathsf{N_{KI}} \text{ for each } i = 1, \ldots, n.$$

   Then, for all those positions $\rho_1, \ldots, \rho_n$, we have $f'(\rho_i) = a$, for each $i = 1, \ldots, n$, and for all positions σ that are not equal to each ρ_i, we have $f'(\sigma) = f(\sigma)$.

2. *Take two distinct individuals $x, y \in \mathsf{N_{AI}}$ occurring in $f(\mathcal{A})$ and all positions $\rho_1, \ldots, \rho_i$ and $\rho_{i+1}, \ldots, \rho_n$, where*

$$f(\rho_j) = x, f(\rho_k) = y, \text{ and } val(\rho_j, \mathcal{A}) = val(\rho_k, \mathcal{A}),$$

   for each $j = 1, \ldots, i$ and $k = i{+}1, \ldots, n$. Then, for each ρ_j, ρ_k, we have $f'(\rho_j) = f'(\rho_k) = x$, and additionally, for all positions σ' that are not equal to each ρ_j, ρ_k, we have $f'(\sigma) = f(\sigma)$.

3. *Take one position ρ in $\mathcal{A}$ such that $val(\rho, \mathcal{A})$ and $f(\rho)$ are $\mathcal{EL}$ concepts C and C', respectively, where $C \sqsubset C'$. Then, set $\Sigma = sig(C)$, guess $\mathsf{At} \in LA_\Sigma(C')$ such that $C \sqsubseteq \mathsf{At}$, $C'' := C' \sqcap \mathsf{At}$ is a lower neighbor of C', and $f'(\rho) = C''$. Meanwhile, for all positions σ that are not ρ, we have $f'(\sigma) = f(\sigma)$.* ♢

The next lemma explains why each function $f' \in \mathtt{adj}(f, \mathcal{A})$ is strictly more informative than f, i.e., $f' > f$.

Lemma 6.26. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f be an $\mathcal{A}$-anonymizer. It holds that all $f' \in \mathtt{adj}(f, \mathcal{A})$ are $\mathcal{A}$-anonymizers and $f' > f$.*

Proof. It is easy to see that each operation basically either replaces one anonymous individual with a known individuals, identifies two anonymous individual, or replaces a concept with its lower neighbor. The latter is justified by Lemma 5.22 saying that if there is an atom $\mathsf{At} \in LA_\Sigma(C')$ such that $C'' \equiv C' \sqcap \mathsf{At}$, then C'' is a lower neighbor of C' w.r.t. Σ. This obviously implies that the constructed function f' is more informative than f, i.e., $f' \geq f$.

To see that this informativeness order is strict, it remains to show that $f \not\geq f'$. If the first operation is taken to yield f', then f does not satisfy the condition 1a.) since there are positions ρ_i, where $f'(\rho_i) \in \mathsf{N_{KI}}$, but $f(\rho) \in \mathsf{N_{AI}}$. If the second operation is now used to

construct f', then f does not satisfy the condition 1b.) since there are positions ρ_j and ρ_k, where $f'(\rho_j) = f'(\rho_k)$, but $f(\rho_j) \neq f(\rho_k)$. Last, if the third operation is used to yield f', then f does not satisfy the condition 2 because there is a position ρ such that $f(\rho) \not\sqsubseteq f'(\rho)$.□

Next, we prove that every $\mathcal{A}$-anonymizer f'', where $f'' > f$, is more informative than a function from $\mathtt{adj}(f, \mathcal{A})$.

Lemma 6.27. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f be an $\mathcal{A}$-anonymizer. For all $\mathcal{A}$-anonymizers f'' such that $f'' > f$, there is $f' \in \mathtt{adj}(f, \mathcal{A})$, where $f'' \geq f'$.*

Proof. To show this lemma, given $\mathcal{A}$-anonymizers f and f'', we consider all positions ρ_i, ρ_j in $\mathcal{A}$ satisfying one of the following conditions:

1.) $val(\rho_i, \mathcal{A}) \in \mathsf{N_{KI}}$, $f''(\rho_i) = val(\rho_i, \mathcal{A})$, and $f(\rho_i) \in \mathsf{N_{AI}}$,

2.) $val(\rho_i, \mathcal{A}), val(\rho_j, \mathcal{A}) \in \mathsf{N_{KI}} \cup \mathsf{N_{AI}}$, $f''(\rho_i) = f''(\rho_j)$, but $f(\rho_i) \neq f(\rho_j)$,

3.) $val(\rho_i, \mathcal{A})$ is an $\mathcal{EL}$ concept C, $f(\rho_i) = C'$, and $f''(\rho_i) = C'''$ such that $C \sqsubseteq C''' \sqsubset C'$.

Now, we build a function f' by using one of the following operations:

- Take two individuals $x \in \mathsf{N_{AI}}, a \in \mathsf{N_{KI}}$, and all positions ρ_i satisfying the condition 1.) above, such that for each ρ_i, we have $f(\rho_i) = x$ and $f''(\rho_i) = val(\rho_i, \mathcal{A}) = a$. Then, for all ρ_i, we have $f'(\rho_i) = a$ and for all σ that are not ρ_i, we have $f'(\sigma) = f(\sigma)$.
- Take two individuals $x, y \in \mathsf{N_{AI}}$ and all positions $\rho_1, \ldots, \rho_n$ such that for each $i = 1, \ldots, k$ and $j = k+1, \ldots, n$, we have that ρ_i and ρ_j satisfy the condition 2.) above, and additionally
$$f(\rho_i) = x, f(\rho_j) = y, \text{ and } val(\rho_i, \mathcal{A}) = val(\rho_j, \mathcal{A}).$$
Then, for each ρ_i, ρ_j, we have $f'(\rho_i) = f'(\rho_j) = x$, and additionally, for all positions σ' that are not equal to each ρ_i, ρ_j, we have $f'(\sigma) = f(\sigma)$.
- Take one position ρ in $\mathcal{A}$ satisfying the condition 3.). Then, set $\Sigma := sig(C''')$, guess $\mathsf{At} \in LA_\Sigma(C')$ such that $C''' \sqsubseteq \mathsf{At}$, and define $C'' := C' \sqcap \mathsf{At}$ as well as $f'(\rho) = C''$. Meanwhile, for all positions σ that are not ρ, we have $f'(\sigma) = f(\sigma)$.

Obviously, $f' \in \mathtt{adj}(f, \mathcal{A})$. It remains to show that $f'' \geq f'$. Let f' be generated by the first operation. Since for each position ρ_i, we have $f''(\rho_i) = f'(\rho_i) = a$, obviously the condition 1a.) in Definition 6.11 is satisfied. Then, for all positons σ that are not ρ_i, since $f'(\sigma) = f(\sigma)$ and $f'' > f$, it implies that each position $\sigma_i, \sigma_{i'}$ satisfies all conditions in Definition 6.11. This implies that $f'' \geq f'$.

Now, let f' be constructed by the second operation. Since for each ρ_j, ρ_k, we have $f''(\rho_j) = f''(\rho_k)$ and $f'(\rho_j) = f'(\rho_k)$, this implies that all these positions satisfy the condition 1b.) in Definition 6.11. For all positions $\sigma_i, \sigma_{i'}$ that are not both ρ_j and ρ_k, the same arguments as in the previous case also hold, which implies that $f'' \geq f'$.

Finally, let f' be constructed by the third operation. Since $f''(\rho) = C'''$ and $f'(\rho) = C''$ such that C'' is a lower neighbor of C' that subsumes C''', by the definition of lower neighbor, it implies that the position ρ satisfies the condition 2.) in Definition 6.11. The same with previous arguments, for all positions σ that are not ρ, we have $f'(\sigma) = f(\sigma)$ and since $f'' > f$, we know that all these positions clearly satisfy all conditions in Definition 6.11 and thus $f'' \geq f'$. □

Next, we show that all distinct functions in $\mathtt{adj}(f, \mathcal{A})$ are not comparable w.r.t. the informativeness order.

Lemma 6.28. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f be an $\mathcal{A}$-anonymizer. If f_1 and f_2 are different functions in $\mathtt{adj}(f, \mathcal{A})$, then $f_1 \not\geq f_2$.*

Proof. If both f_1 and f_2 are generated by taking different operations in Definition 6.25, then it is easy to see that $f_1 \not\geq f_2$. Even, if both f_1 and f_2 are generated by using the same operation, in particular, either using operation 1.) or 2.), then it is also obvious that $f_1 \not\geq f_2$. Now, it remains to show that if both f_1 and f_2 are constructed using the same operation 3.), then $f_1 \not\geq f_2$. But then, if the position ρ that is taken to yield f_1 is different with the positon ρ' that is considered to output f_2, then clearly we have $f_1 \not\geq f_2$.

Thus, let us consider that the positions, taken in operation 3.) to yield both f_1 and f_2, are the same. Assume that we take the position ρ and $f(\rho) = C'$. Since in this operation we replace $f(\rho)$ with a lower neighbor of $f(\rho)$, we have $f_1(\rho) = C_1$ and $f_2(\rho) = C_2$ such that $C_1 \sqsubset_1 C'$ and $C_2 \sqsubset_1 C'$. Since f_1 and f_2 are different, C_1 and C_2 are distinct, too. However, C_1 and C_2 are generated by conjoining two different lowering atoms At_1, At_2 to C', respectively. By the definition of $LA_\Sigma(C')$, every two atoms is incomparable to each other w.r.t. subsumption, and thus $C_1 \not\sqsubseteq C_2$ and $C_2 \not\sqsubseteq C_1$. This implies that $f_1 \not\geq f_2$. □

Now, we are ready to prove that all elements in $\mathtt{adj}(f, \mathcal{A})$ are $\mathcal{A}$-anonymizers that are adjacent to f w.r.t. $\mathcal{A}$.

Proposition 6.29. *Let $\mathcal{A}$ be an $\mathcal{EL}$ ABox and f be an $\mathcal{A}$-anonymizer. Then, the following holds:*

1.) *Every function f'' that is adjacent to f w.r.t. $\mathcal{A}$ is as informative as one of the anonymizers in $\mathtt{adj}(f, \mathcal{A})$.*

2.) *Every anonymizer in $\mathtt{adj}(f, \mathcal{A})$ is adjacent to f w.r.t. $\mathcal{A}$.*

3.) *The cardinality of $\mathtt{adj}(f, \mathcal{A})$ is not polynomial in the size $\mathcal{A}$ and $\mathcal{A}'$ in general.* ♢

Proof. For the claim 1.), let f'' be adjacent to f w.r.t. $\mathcal{A}$. By Lemma 6.27, we have $f' \in \mathtt{adj}(f, \mathcal{A})$ such that $f'' \geq f'$. But then, $f'' \geq f' > f$ and $f'' > f$ imply $f'' \simeq f'$, and thus f'' is as informative as one of the elements from $\mathtt{adj}(f, \mathcal{A})$.

For the claim 2.), suppose that $f' \in \mathtt{adj}(f, \mathcal{A})$. Then, Lemma 6.26 yields $f' > f$. To show that f' is adjacent to f w.r.t. $\mathcal{A}$, then in contrast we assume that there is an $\mathcal{A}$-anonymizer f'' such that $f' > f'' > f$. Then, Lemma 6.27 shows that there is $f''' \in \mathtt{adj}(f, \mathcal{A})$ such that $f' > f'' \geq f''' > f$. But then, f' and f''' are two different $\mathcal{A}$-anonymizers, which means that, by Lemma 6.28, we know that this informativeness order contradicts Lemma 6.28.

For the last claim, we show that the cardinality of $\mathtt{adj}(f, \mathcal{A})$ is not polynomial in the size $\mathcal{A}$ by referring to Example 5.23. This implies there may be exponentially many lower neighbors of an $\mathcal{EL}$ concept written in that example, which implies that, in general, the number of $\mathcal{A}$-anonymizers that are adjacent to f w.r.t. $\mathcal{A}$ is not polynomial. □

The fact that the cardinality of $\mathtt{adj}(f, \mathcal{A})$ is not polynomial only happen if the third operation is used to find a lower neighbor of an $\mathcal{EL}$ concept C occurring in the concept

assertion $C(a) \in \mathcal{A}$. Meanwhile, the first and second operation only do polynomially many replacements or identifications for anonymous individuals. However, By Lemma 5.25, we are able to generate an element of the set $LA_\Sigma(C)$ of lowering atoms using an NP algorithm. This implies that this is also not hard to see that the operations defined in Definition 6.25 can also perform in non-deterministic polynomial time.

Now, we move to the problem of deciding whether f is an optimal D-compliant (safe) anonymizer of $\mathcal{A}$, where D is a reduced $\mathcal{EL}$ concept. We present the following NP-algorithm to solve the complements of OPTIMAL-COMPLIANCE$_{IQ}$ and OPTIMAL-SAFETY$_{IQ}$, which relies on guessing an anonymization function.

Proposition 6.30. *Given an $\mathcal{EL}$ ABox $\mathcal{A}$, an $\mathcal{A}$-anonymizer f, and a reduced $\mathcal{EL}$ concept D, it holds that there is an* NP*-algorithm to decide whether f is not an optimal D-compliant (safe) anonymizer of $\mathcal{A}$.* ♢

Proof. We describe the following NP-algorithm that, given $\mathcal{A}, f, D$, succeeds iff f is *not* an optimal D-compliant (safe) anonymizer of $\mathcal{A}$.

- Check whether $f(\mathcal{A})$ is compliant with or safe for D. By Corollary 6.14 and Theorem 6.23, we know that this test can be done in polynomial time. If this is the case, then continue with the next step. Otherwise, the algorithm succeeds.
- Guess an adjacent function $f' \in \mathtt{adj}(f, \mathcal{A})$ and then check if $f'(\mathcal{A})$ is compliant with (safe for) D. If this is the case, then the algorithm succeeds, otherwise fail.

It is easy to see that the algorithm is correct and runs in non-deterministic polynomial time.□

The following theorem is an immediate consequence of the proposition above.

Theorem 6.31. OPTIMAL-COMPLIANCE$_{IQ}$ *and* OPTIMAL-SAFETY$_{IQ}$ *are in* coNP.

Next, we move to the complement of the decision problems OPTIMAL-COMPLIANCE$_{CQ}$ and OPTIMAL-SAFETY$_{CQ}$, and then show that there is an algorithm, which basically check whether $\mathcal{A}$ is compliant with (safe for) q, then generate an element f' of $\mathtt{adj}(f, \mathcal{A})$ in NP, and subsequently call an oracle for check whether $f'(\mathcal{A})$ is compliant with (safe for) q.

Proposition 6.32. *Given an $\mathcal{EL}$ ABox $\mathcal{A}$, an $\mathcal{A}$-anonymizer f, and a conjunctive query q, it holds*

- *there is a Σ_2^p-algorithm to check that f is* not *an optimal q-compliant anonymizer of $\mathcal{A}$ and*
- *there is a Σ_3^p-algorithm to check that f is* not *an optimal q-safe anonymizer of $\mathcal{A}$.* ♢

Proof. We present the following algorithm that, given $\mathcal{A}, f, q$, succeeds iff f is not an optimal q-compliant (safe) anonymizer of $\mathcal{A}$.

- Check whether $f(\mathcal{A})$ is compliant with (safe for) q. If this is not the case, then the algorithm succeeds. Otherwise, we continue with the next step.
- Guess an adjacent function $f' \in \mathtt{adj}(f, \mathcal{A})$ and then check if $f'(\mathcal{A})$ is compliant with (safe for) q. If this the case, then the algorithm succeeds, otherwise fail.

It is easy to see that this algorithm is correct. For compliance, the test in the first step can be done in non-deterministic polynomial time by Corollary 6.14, while the second step guess an adjacent function $f' \in \mathtt{adj}(f, \mathcal{A})$ and call an NP oracle to check whether $f'(\mathcal{A})$ complies with q. This implies that the algorithm above runs in Σ_2^p to solve the complement of OPTIMAL-COMPLIANCE$_{CQ}$. Meanwhile, for safety, the test in the first step can be done in Σ_2^p by Lemma 6.20 and additionally the second step an adjacent function $f' \in \mathtt{adj}(f, \mathcal{A})$ and then call an Π_2^p oracle check whether $f'(\mathcal{A})$ complies with q. This finally shows that the algorithm above runs in Σ_3^p to solve the complement of OPTIMAL-SAFETY$_{CQ}$. □

The proposition above directly provides the upper bounds for both OPTIMAL-COMPLIANCE$_{CQ}$ and OPTIMAL-SAFETY$_{CQ}$, which are in Π_2^p and Π_3^p, respectively. Unfortunately, the precise complexity for these problems still remain open, but the following proposition shows that we can reduce the HOMOMORPHISM-NOHOMOMORPHISM problem to both OPTIMAL-COMPLIANCE$_{CQ}$ and OPTIMAL-SAFETY$_{CQ}$.

Proposition 6.33. *There are polynomial reductions from* HOMOMORPHISM-NOHOMOMORPHISM *to* OPTIMAL-COMPLIANCE$_{CQ}$ *and to* OPTIMAL-SAFETY$_{CQ}$*, respectively.*

Proof. As written in Lemma 6.22, the input of HOMOMORPHISM-NOHOMOMORPHISM consists of four connected directed graphs $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_1', \mathcal{G}_2'$ and then check whether there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}_1'$ to $\mathcal{G}_2'$. We employ these four graphs to construct $\mathcal{EL}$ ABoxes $\mathcal{A}, \mathcal{A}'$, an $\mathcal{A}$-anonymizer f such that $f(\mathcal{A}) = \mathcal{A}'$, and a boolean query q.

We start with the *first reduction* from HOMOMORPHISM-NOHOMOMORPHISM to the problem of OPTIMAL-COMPLIANCE$_{CQ}$. First, we define $\mathcal{A}$ as

$$\begin{aligned}\mathcal{A} := \{r(a_\mu, a_\nu) \mid (\mu, \nu) \text{ is an edge of } \mathcal{G}_2\} \cup \\ \{s(a_{\mu'}, a_{\nu'}) \mid (\mu', \nu') \text{ is an edge of } \mathcal{G}_2'\} \cup \{s(x, x)\},\end{aligned}$$

where $a_\mu, a_\nu, a_{\mu'}, a_{\nu'}$ are known individuals and z is an anonymous individual. Then, we define a boolean CQ q consisting of the atoms $r(w_{\mu_1}, w_{\mu_2})$ for each edge (μ_1, μ_2) of $\mathcal{G}_2$ and $s(w_{\nu_1}, w_{\nu_2})$ for each edge (ν_1, ν_2) of $\mathcal{G}_2'$, where each $w_{\mu_1}, w_{\mu_2}, w_{\nu_1}, w_{\nu_2}$ is an existentially quantified variable. Next, we define an $\mathcal{A}$-anonymizer f that simply replaces the value of the positions ρ_1, ρ_2 over x in the first and the second argument of $s(c, c)$ with x_1 and x_2, respectively, and then for each position σ that are not ρ, we have $f(\sigma) = val(\sigma, \mathcal{A})$. Thus, we have $\mathcal{A}' = f(\mathcal{A})$ as the anonymization of $\mathcal{A}$ w.r.t. f. Now, we prove the following claim: $\mathcal{A}'$ is an optimal q-compliant anonymization of $\mathcal{A}$ iff there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}_1'$ to $\mathcal{G}_2'$.

Note that the only $\mathcal{A}$-anonymizer that is adjacent to f is the function f' such that f' maps all individuals to itself, i.e., $f'(\mathcal{A}) = \mathcal{A}$. Suppose that $\mathcal{A}'$ is an optimal q-compliant anonymization of $\mathcal{A}$, which implies that $\mathcal{A}'$ is compliant with q, but $\mathcal{A}$ is not compliant with q. Since $\mathcal{A}$ is not compliant with q, it implies that there is a homomorphism h from q to $\mathcal{A}$. All individuals in atoms with the role r in q will be mapped by f to all individuals in assertions with the role r in $\mathcal{A}$. This implies that there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$. Then, the existence of a homomorphism from q to $\mathcal{A}$ also implies that for the individuals occurring in the atoms over the role s, there are two possibilities which are either they are mapped to the

representation of $\mathcal{G}_2$ in $\mathcal{A}$ or to the individual x in the atom $s(x,x)$. If they are mapped to the former, then there is also a homomorphism from q to $\mathcal{A}'$, which is a contradiction, and thus the only possibility is that they are mapped to the latter. This implies that there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$.

Conversely, suppose that there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$. This implies that there is a homomorphism h from q to $\mathcal{A}$, which maps all individuals x in the atoms with role r to the representation of $\mathcal{G}_2$ in $\mathcal{A}$ and additionally, maps all individuals y in the representation of $\mathcal{G}'_1$ to z. Now, suppose that there is a homomorphism h' from q to $\mathcal{A}'$. Then, h' will map y to x_1 or x_2 in $\mathcal{A}'$. Since x_1 and x_2 are not connected to any individual in $\mathcal{A}'$ and G'_1 is also connected, then there is only one atom $s(w_{v_1}, w_{v_2})$ in q with the role s. However, since $\mathcal{G}'_2$ is not empty, then there is also a homomorphism from individuals w_{v_1}, w_{v_2} to some individual in the representation of $\mathcal{G}'_2$ in $\mathcal{A}'$. However, this is again a contradiction since we assume that there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$.

Next, we move to the *second reduction* that reduces HOMOMORPHISM-NOHOMOMORPHISM to OPTIMAL-SAFETY$_{CQ}$. For this setting, we construct the query q in the same way as constructing q in Lemma 6.22. Then, we define

$$\mathcal{A} := \{r(c,c)\} \cup \{s(x_\mu, x_\nu) \mid (\mu, \nu) \text{ is an edge in } \mathcal{G}'_2\},$$

where c is a known individual and all individuals x_μ, x_ν are anonymous. We construct an $\mathcal{A}$-anonymizer f, that maps all anonymous individuals to itself and the known individual c in each position to the anonymous x, such that $f(\mathcal{A}) = \mathcal{A}'$. We prove the claim that $\mathcal{A}'$ is an optimal q-safe anonymization of $\mathcal{A}$ iff there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$.

It is easy to see that the only $\mathcal{A}$-anonymizer function that is adjacent to f is f', such that $f'(\mathcal{A}) = \mathcal{A}$, and f' is obtained from f by replacing the anonymous x with the known c. To show that there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$, we assume that $\mathcal{A}'$ is an optimal q-safe anonymization of $\mathcal{A}$. Since $f'(\mathcal{A}) = \mathcal{A}$, it is enough to assume that $\mathcal{A}'$ is safe for q and $\mathcal{A}$ is not safe for q. Note that the ABox $\mathcal{A}'$ we have in this setting is the same as the ABox $\mathcal{A}$ in Lemma 6.22. According to that lemma, if $\mathcal{A}'$ is safe for q, then it holds that there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$. Thus, we are done with the 'if direction' Now, we show the converse direction. Assume that there is a homomorphism from $\mathcal{G}_1$ to $\mathcal{G}_2$ and there is no homomorphism from $\mathcal{G}'_1$ to $\mathcal{G}'_2$. By Lemma 6.22, it implies that $\mathcal{A}'$ is safe for q. It remains to show that $\mathcal{A}$ is not safe for q. However, it is easy to see that we may construct an ABox $\mathcal{A}''$ consisting of the part 3 of q, which is over known individuals, and in addition to that, a concept assertion $A(c)$, where $A \in \mathsf{N_C}$. Obviously, $\mathcal{A}''$ complies with q, but there is a homomorphism h from q to $\mathcal{A} \cup \mathcal{A}''$, where h maps all individuals in the parts 1 and 2 of q to c and then since $\mathcal{A}''$ consists of the part 3 of, all anonymous individuals in the part 3 of q are mapped to their counterparts in $\mathcal{A}''$. This implies that $\mathcal{A}$ is not safe for q. □

Given Propositions 6.32 and 6.33, the following theorem describes the complexity results for OPTIMAL-COMPLIANCE$_{CQ}$ and OPTIMAL-SAFETY$_{CQ}$.

Theorem 6.34. OPTIMAL-COMPLIANCE$_{CQ}$ *is in* Π_2^p *and hard for* DP, *whereas* OPTIMAL-SAFETY$_{CQ}$ *is in* Π_3^p *and also hard for* DP.

Chapter 7

Conclusions

7.1 Main Results

In Chapter 3, we have provided some initial definitions and results regarding the identity problem in DL ontologies, i.e., the question whether the ontology implies that a given anonymous individual is equal to a known individual. We have also considered a more involved rôle-based access control scenario where users can access parts of the ontology depending on their rôle. In a setting where users can change rôles dynamically, the question is then whether, by changing rôles and asking queries in these rôles, the user can find out the identity of an anonymous individual although this may not be possible for a single rôle. We have shown how to use the identity problem to address this question. We also had a look at the k-hiding problem motivated from a case where an attacker probably does not need to know the real identity of this anonymous individual, but he only wants to deduce whether its identity is one of k known individuals w.r.t. a given ontology. We show that this problem is actually not harder than the identity problem in most of DLs with equality power.

If this is the case that one is able to deduce the confidentiality information about individuals in DL-based ontologies, then in Chapter 4, we proposed a framework for repairing DL-based ontologies that is used to get rid of unwanted or secret consequences from the original ontologies. The repair approach itself is based on weakening axioms rather than deleting them. Additionally, we introduced the notion of maximally strong weakenings of an axiom that can be obtained using restricted weakening relations. We then showed how to instantiate this framework for the DLs $\mathcal{EL}$ and $\mathcal{ALC}$ using appropriate weakening relations. For $\mathcal{EL}$ GCIs, we introduced weakening relations that semantically generalizes the right-hand size of GCIs. However, for the sake of finding maximally strong weakenings w.r.t. this relation, the algorithm we described in the proof of Proposition 4.18 may have non-elementary complexity in general. To get a weakening relation that has better algorithmic properties than before, we introduced a weakening relation for $\mathcal{EL}$ GCIs that syntactically generalize the right-hand side of GCIs and w.r.t. this, a single maximally strong weakening can always be computed in polynomial time. For $\mathcal{ALC}$ GCIs, we considered two weakening relations that guarantee the well-foundedness of subsumption and inverse subsumption. The first one specializes and generalizes $\mathcal{ALC}$ concepts using a finite set of fixed signature and a fixed role-depth bounded, whereas the second one syntactically replace positive or negative occurrences of subconcepts in C with either $\top$ or $\bot$. We also showed that the second one has indeed better properties in terms of complexity.

In the privacy context, repairing ontologies for getting rid of unwanted consequences is still not sufficient since it might be that a possible attacker owns relevant information from other sources, which together with the repaired ontologies, the privacy policy is still violated.

To handle this issue, in Chapter 5, we have introduced the notions of compliance with and safety for a policy in a simple setting where both the knowledge about individuals and the policy are given by $\mathcal{EL}$ concepts and knowledge of the attackers about the individuals are stated as either $\mathcal{EL}$, $\mathcal{FL}_0$, or $\mathcal{FLE}$ concepts. In a setting where knowledge of attackers is encoded in $\mathcal{EL}$, we have shown that compliance and safety can be decided in PTIME. Then, we have shown that there are exponentially many optimal compliant generalizations, each of them has exponential size and can be computed in EXPTIME Meanwhile, the optimal safe generalization is unique and of exponential size and can also be computed in EXPTIME. When the setting is changed with a condition where the knowledge of attackers are written as $\mathcal{FL}_0$ concepts, it has been shown that the complexity of safety are the same as in $\mathcal{EL}$ case, but there are exponentially many optimal safe generalizations and each of them can also be computed in EXPTIME. For both cases in $\mathcal{EL}$ and $\mathcal{FL}_0$, the optimality problem can be decided in CONP and we observed that there is a polynomial reduction from the Hypergraph Duality problem to the optimality problem in these two logics. However, when the background information owned by attackers is given by an $\mathcal{FLE}$ concept, computing optimal safe generalization and deciding both safety and optimality becomes tractable, which can be performed in PTIME.

In Chapter 6, we extended the problem setting in Chapter 5 by considering that the information about individuals as well as the knowledge of attackers are given by $\mathcal{EL}$ ABoxes and the privacy policy is provided as an $\mathcal{EL}$ concept or a conjunctive query. If one privacy policy is violated w.r.t. a given ABox, we proposed an anonymization approach using a function, called *anonymizer*, that renames individuals with new anonymous individuals and generalizes concepts occurring in ABox assertions. This is then followed by presenting characterizations for compliance and safety checking w.r.t. $\mathcal{EL}$ ABoxes and by designing an algorithm for deciding whether a given anonymizer is optimal, i.e., it preserves information from the original ABox as much as possible. If the policy is an $\mathcal{EL}$ concept, then the compliance and safety properties can be decided in polynomial time, whereas the optimality problem can be decided in CONP. However, if the policy is now a conjunctive query, then the compliance and safety are in CONP and Π_2^p, respectively. As a consequence, the complexity of the optimality problem lies on the second and third level of polynomial hierarchy for compliance and safety, respectively. Although the complexity results we have are still not tight yet in the case where the policy is a conjunctive query, we showed that there is a reduction from some known DP problems to our safety and optimality problems, which ultimately provides us hardness results.

7.2 Future Work

For the *identity problem in DL ontologies and its variants*, we note that our complexity results are until now measured in terms of the size of the whole input. In data-intensive applications, one may only consider data complexity, where the complexity is measured in terms of the size of the data (i.e., ABox) only. Then, in most practical scenarios, for instance in a probabilistic setting as addressed by [BBG+17], the information about known and anonymous individuals can also be assumed to only hold with a certain probability, e.g., using ontologies with subjective probability as introduced in [GJL+17]. In this setting, equality can also only be derived with a certain probability, and one might want to keep the probability of derived identities low enough. Another interesting direction is that actually to extend the identity

problem in the k-anonymity setting, a well-known confidentiality criteria in database area. The original definition of k-anonymity requires you to check the information for each person contained in a database table cannot be distinguished from at least k-1 individuals whose information also appears in the table. To capture this definition, we need to obviously change our goal from 'finding at most k-1 known individuals ...' to 'finding at least k-1 known individuals ...', and further we need to define what it means by 'the information for each person' whose occurrences we want to compare.

Apart from that, during the work on *repairing Description Logic ontologies*, we have seen that computing maximally strong weakenings is analogous to the black-box approach for computing justifications. It would be interesting to see whether a glass-box approach that modifies an $\mathcal{EL}$ reasoning procedure can also be used for this purpose, similar to the way a tableau-based algorithm for $\mathcal{ALC}$ was modified in [LSP+08]. Our weakening relations can also be used in the setting where the ontology is first modified, and then repaired using the classical approach as in [DQF14]. In fact, for effectively finitely branching and well-founded weakening relations such as $\succ^{sub}$ and $\succ^{syn}$, we can add for each axiom all (or some of) its finitely many weakenings w.r.t. the given relation, and then apply the classical repair approach. In contrast to the gentle repair approach proposed in this paper, a single axiom could then be replaced by several axioms, which might blow up the size of the ontology.

So, our complexity results for deciding whether an axiom is a maximally strong weakening w.r.t. $\succ^{sub}$ is coNP-hard. However, the upper bound still remains open. One idea to achieve this is by probably looking at the notion of lower neighbors that already helped us to obtain the upper-bound of the optimality$^{\exists}$ problem. Moreover, until now our definition of what it means that an axiom is β is weaker than γ does not involve the given ontology. In other words, we require $\mathsf{Con}(\{\beta\}) \subset \mathsf{Con}(\{\gamma\})$ rather than $\mathsf{Con}(\{\beta \cup \mathfrak{O}\}) \subset \mathsf{Con}(\{\gamma\mathfrak{O}\})$. Since the ontology implies unwanted consequences, it makes not sense to work with the whole ontology. Instead, we can restrict our attention to $\mathfrak{O}_{st}$ that is considered to be static and correct. Thus, in this setting, it is reasonable to work with $\mathfrak{O}_{st}$.

For our work on *privacy-preserving ontology publishing for $\mathcal{EL}$ instance stores*, we consider to add an $\mathcal{EL}$ TBox as additional information about concepts to which the individuals belong that will be publicly published. However, the optimal repairs w.r.t. this setting need no longer exist, in general. To this end, one may also start looking at another restricted type of general TBoxes that satisfies certain cycle-restrictions (see, e.g., [BBM12]). In the setting of $\mathcal{EL}$ instance stores, another interesting investigation is now to look at safety and optimality problems that take $\mathcal{ALC}$ into account as knowledge of attackers. Since this logic contains negations, disjunctions, and bottom concept, we need to be careful with the situation, where the knowledge of attacker is actually the negation of the concept C that will be published. If this is the case, then the combination of this knowledge and the concept C is a bottom concept, which is vacuously subsumed by all concepts in $\mathcal{P}$. Considering such situation, this may also make sense if a condition, which emphasizes that C is safe for $\mathcal{P}$ if for all concepts C'' that are compliant with $\mathcal{P}$, $C \sqcap C'' \not\equiv \bot$, can be added to Definition 5.1.

Similar to the work for $\mathcal{EL}$ instance stores, on *privacy-preserving ontology publishing for $\mathcal{EL}$ ABoxes*, we also envision that including $\mathcal{EL}$ TBoxes will be a good direction for future work. Likewise, considering compliant and safe generalizations that are obtained by looking at $\mathcal{EL}$ TBoxes satisfying certain cycle-restrictions are things we need to take into account in the beginning before going further. Another subtle investigation coming out of this work is extending the definition of privacy policies, which currently is still of the form of an $\mathcal{EL}$ concept.

If we broaden the form to a set of $\mathcal{EL}$ concepts, then we conjecture that the characterization for safety presented in Lemma 6.16 will no longer be applicable. Furthermore, our complexity result for SAFETY_{CQ} is not tight yet. As analyzed by [GK19], this unmatched complexity results happens since a similar, well-known safety problem in the database theory, called *critical tuple problem* [MS07], which has the same spirit with the safety problem in this thesis and in [GK16; GK19], still does not have a tight complexity either. This might be interesting to also first determine the precise complexity for the *critical tuple problem*. The same issue for unmatched complexity result is also still found in $\text{OPTIMAL-COMPLIANCE}_X$, where $X \in \{IQ, CQ\}$, that provides us a future work to find the lower bound for them. Last, so far we just considered optimality as a decision problem in this work. Similar to the case of $\mathcal{EL}$ instance stores, it is also interesting to look for algorithms which *compute* an optimal compliance (safety) anonymizer of an ABox, w.r.t. an $\mathcal{EL}$ concept or a conjunctive query.

Nevertheless, the assumed background knowledge of the attackers we defined in Chapter 5 and 6 still arguably lacks of expressiveness. According to the work from [BS13], our defined background knowledge of attackers only captures the object-level of the attackers' knowledge. In other words, the attackers know parts of the domain knowledge that is not axiomatized in our published information, even though what the attackers know complies with the given privacy policy. Besides that, one also needs to consider the meta-level of the attackers' knowledge. In [BS13], using his meta-level knowledge, the attackers may assume that the published information has complete knowledge about a certain set of axioms or the attackers know very well the signature occurring in the published information. For the former case, attacks to complete knowledge have a connection with a *closed-world* setting. We can start learning this setting by referring to the work of [NOS16; SFB09] where the notion of DLs with closed predicates are used. Attacks to complete knowledge may also have a connection with a setting where knowledge of the attackers is equipped with integrity constraints or functional dependencies, where examples of such knowledge have been investigated before in [TW13; LM04].

Bibliography

[AHV95] Serge Abiteboul, Richard Hull, and Victor Vianu, editors: *Foundations of Databases: The Logical Level*. 1st. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1995. ISBN: 0201537710 (cited on page 28).

[ACK+09] Alessandro Artale, Diego Calvanese, Roman Kontchakov, and Michael Zakharyaschev: 'The DL-Lite Family and Relations'. In *J. Artif. Intell. Res.* **36**: 2009, pages 1–69 (cited on pages 7, 30, 31, 35–37, 49).

[BBN17a] Franz Baader, Daniel Borchmann, and Adrian Nuradiansyah: 'Preliminary Results on the Identity Problem in Description Logic Ontologies'. In *Proceedings of the 30th International Workshop on Description Logics, Montpellier, France, July 18-21, 2017*. Edited by Alessandro Artale, Birte Glimm, and Roman Kontchakov. 2017 (cited on page 11).

[BBN17b] Franz Baader, Daniel Borchmann, and Adrian Nuradiansyah: 'The Identity Problem in Description Logic Ontologies and Its Application to View-Based Information Hiding'. In *Semantic Technology - 7th Joint International Conference, JIST 2017, Gold Coast, QLD, Australia, November 10-12, 2017, Proceedings*. Edited by Zhe Wang, Anni-Yasmin Turhan, Kewen Wang, and Xiaowang Zhang. 2017, pages 102–117 (cited on page 11).

[BBM12] Franz Baader, Stefan Borgwardt, and Barbara Morawska: 'Extending Unification in EL Towards General TBoxes'. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012*. Edited by Gerhard Brewka, Thomas Eiter, and Sheila A. McIlraith. 2012 (cited on page 129).

[BBL05] Franz Baader, Sebastian Brandt, and Carsten Lutz: 'Pushing the EL Envelope'. In *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*. 2005, pages 364–369 (cited on pages 6, 25, 27, 30, 35).

[BCM+03] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors: *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003 (cited on pages 2, 5, 15, 31).

[BH95] Franz Baader and Bernhard Hollunder: 'Embedding Defaults into Terminological Knowledge Representation Formalisms'. In *J. Autom. Reasoning* **14**(1): 1995, pages 149–180 (cited on page 9).

[BHL+17] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler: *An Introduction to Description Logic*. Cambridge University Press, 2017 (cited on pages 2, 5, 15, 21, 24, 28, 31, 37, 38, 72).

[BKP12] Franz Baader, Martin Knechtel, and Rafael Peñaloza: 'Context-dependent views to axioms and consequences of Semantic Web ontologies'. In *J. Web Semant.* **12**: 2012, pages 22–40 (cited on page 7).

[BKN19] Franz Baader, Francesco Kriegel, and Adrian Nuradiansyah: 'Privacy-Preserving Ontology Publishing for $\mathcal{EL}$ Instance Stores'. In *16th European Conference on Logics in Artificial Intelligence, JELIA 2019, Rende, Italy, May 8-10, 2019, Proceedings*. Edited by Francesco Calimeri, Nicola Leone, and Marco Manna. Lecture Notes in Computer Science. Springer, 2019 (cited on page 12).

[BKN+18a] Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza: 'Making Repairs in Description Logics More Gentle'. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*. 2018, pages 319–328 (cited on page 12).

[BKN+18b] Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza: 'Making Repairs in Description Logics More Gentle (Extended Abstract)'. In *Proceedings of the 31st International Workshop on Description Logics, Tempe, Arizona, US, October 27th - to - 29th, 2018*. Edited by Magdalena Ortiz and Thomas Schneider. 2018 (cited on page 12).

[BKN+18c] Franz Baader, Francesco Kriegel, Adrian Nuradiansyah, and Rafael Peñaloza: 'Repairing Description Logic Ontologies by Weakening Axioms'. In *CoRR* **abs/1808.00248**: 2018 (cited on page 12).

[BKM99] Franz Baader, Ralf Küsters, and Ralf Molitor: 'Computing Least Common Subsumers in Description Logics with Existential Restrictions'. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 99, Stockholm, Sweden, July 31 - August 6, 1999. 2 Volumes, 1450 pages*. Edited by Thomas Dean. 1999, pages 96–103 (cited on pages 24, 91).

[BM10] Franz Baader and Barbara Morawska: 'Unification in the Description Logic EL'. In *Logical Methods in Computer Science* **6**(3): 2010 (cited on pages 25, 64).

[BN18] Franz Baader and Adrian Nuradiansyah: 'Towards Privacy-Preserving Ontology Publishing'. In *Proceedings of the 31st International Workshop on Description Logics, Tempe, Arizona, US, October 27th - to - 29th, 2018*. Edited by Magdalena Ortiz and Thomas Schneider. 2018 (cited on page 12).

[BN19] Franz Baader and Adrian Nuradiansyah: 'Mixing Description Logics in Privacy-Preserving Ontology Publishing'. In *KI 2019: Advances in Artificial Intelligence - 42nd German Conference on AI, Kassel, Germany, September 23-26, 2019, Proceedings*. Edited by Christoph Benzmüller and Heiner Stuckenschmidt. Lecture Notes in Computer Science. To appear. Springer, 2019 (cited on page 12).

[BPS07] Franz Baader, Rafael Peñaloza, and Boontawee Suntisrivaraporn: 'Pinpointing in the Description Logic EL^+'. In *KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, KI 2007, Osnabrück, Germany, September 10-13, 2007, Proceedings*. 2007, pages 52–67 (cited on page 59).

[BS08] Franz Baader and Boontawee Suntisrivaraporn: 'Debugging SNOMED CT Using Axiom Pinpointing in the Description Logic EL+'. In *Proceedings of the Third International Conference on Knowledge Representation in Medicine, Phoenix, Arizona, USA, May 31st - June 2nd, 2008*. 2008 (cited on page 9).

[BO15] Meghyn Bienvenu and Magdalena Ortiz: 'Ontology-Mediated Query Answering with Data-Tractable Description Logics'. In *Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures*. Edited by Wolfgang Faber and Adrian Paschke. 2015, pages 218–307 (cited on page 27).

[BB04] Joachim Biskup and Piero A. Bonatti: 'Controlled query evaluation for enforcing confidentiality in complete information systems'. In *Int. J. Inf. Sec.* **3**(1): 2004, pages 14–27 (cited on pages 1, 7).

[BBG+17] Joachim Biskup, Piero A. Bonatti, Clemente Galdi, and Luigi Sauro: 'Inference-proof Data Filtering for a Probabilistic Setting'. In *Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 22, 2017*. Edited by Christopher Brewster, Michelle Cheatham, Mathieu d'Aquin, Stefan Decker, and Sabrina Kirrane. 2017 (cited on page 128).

[BTW10] Joachim Biskup, Cornelia Tadros, and Lena Wiese: 'Towards Controlled Query Evaluation for Incomplete First-Order Databases'. In *Foundations of Information and Knowledge Systems, 6th International Symposium, FoIKS 2010, Sofia, Bulgaria, February 15-19, 2010. Proceedings*. Edited by Sebastian Link and Henri Prade. 2010, pages 230–247 (cited on page 1).

[BW08] Joachim Biskup and Torben Weibert: 'Keeping secrets in incomplete databases'. In *Int. J. Inf. Sec.* **7**(3): 2008, pages 199–217 (cited on page 1).

[BS13] Piero A. Bonatti and Luigi Sauro: 'A Confidentiality Model for Ontologies'. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*. Edited by Harith Alani, Lalana Kagal, Achille Fokoue, Paul T. Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha F. Noy, Chris Welty, and Krzysztof Janowicz. 2013, pages 17–32 (cited on pages 7, 130).

[BGG97] Egon Börger, Erich Grädel, and Yuri Gurevich: *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer, 1997 (cited on page 36).

[Bra04] Sebastian Brandt: 'Polynomial Time Reasoning in a Description Logic with Existential Restrictions, GCI Axioms, and - What Else?' In *Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004*. 2004, pages 298–302 (cited on page 25).

[CDL+07] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati: 'Tractable Reasoning and Efficient Query Answering in Description Logics: The *DL-Lite* Family'. In *J. Autom. Reasoning* **39**(3): 2007, pages 385–429 (cited on pages 7, 30, 31, 49).

[CDL+13] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati: 'Data complexity of query answering in description logics'. In *Artif. Intell.* **195**: 2013, pages 335–360 (cited on page 6).

[CDL+12] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati: 'View-based query answering in Description Logics: Semantics and complexity'. In *J. Comput. Syst. Sci.* **78**(1): 2012, pages 26–46 (cited on pages 7, 29).

[CN10] Stephen Cook and Phuong Nguyen: *Logical Foundations of Proof Complexity*. 1st. New York, NY, USA: Cambridge University Press, 2010 (cited on page 34).

[Dal77] T. Dalenius: 'Towards a methodology for statistical disclosure control'. In *Statistik Tidskrift* **15**(429-444): 1977, pages 2–1 (cited on page 10).

[DP05] Alin Deutsch and Yannis Papakonstantinou: 'Privacy in Database Publishing'. In *Database Theory - ICDT 2005, 10th International Conference, Edinburgh, UK, January 5-7, 2005, Proceedings*. 2005, pages 230–245 (cited on pages 1, 7).

[DLN+92] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, Bernhard Hollunder, Werner Nutt, and Alberto Marchetti-Spaccamela: 'The Complexity of Existential Quantification in Concept Languages'. In *Artif. Intell.* **53**(2-3): 1992, pages 309–327 (cited on page 24).

[DQF14] Jianfeng Du, Guilin Qi, and Xuefeng Fu: 'A Practical Fine-grained Approach to Resolving Incoherent OWL 2 DL Terminologies'. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*. 2014, pages 919–928 (cited on pages 3, 9, 129).

[Dwo06] Cynthia Dwork: 'Differential Privacy'. In *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*. Edited by Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener. 2006, pages 1–12 (cited on pages 8, 10).

[EG02] Thomas Eiter and Georg Gottlob: 'Hypergraph Transversal Computation and Related Problems in Logic and AI'. In *Logics in Artificial Intelligence, European Conference, JELIA 2002*. Edited by Sergio Flesca, Sergio Greco, Nicola Leone, and Giovambattista Ianni. Volume 2424. Lecture Notes in Computer Science. Springer, 2002, pages 549–564 (cited on page 90).

[FAK+16] Simone Fischer-Hübner, Julio Angulo, Farzaneh Karegar, and Tobias Pulls: 'Transparency, Privacy and Trust - Technology for Tracking and Controlling My Data Disclosures: Does This Work?' In *Trust Management X - 10th IFIP WG 11.11 International Conference, IFIPTM 2016, Darmstadt, Germany, July 18-22, 2016, Proceedings*. Edited by Sheikh Mahbub Habib, Julita Vassileva, Sjouke Mauw, and Max Mühlhäuser. 2016, pages 3–14 (cited on page 3).

[FK96] Michael L. Fredman and Leonid Khachiyan: 'On the Complexity of Dualization of Monotone Disjunctive Normal Forms'. In *J. Algorithms* **21**(3): 1996, pages 618–628 (cited on page 90).

[FR07] Barry A. Friedman and Lisa J. Reed: 'Workplace Privacy: Employee Relations and Legal Implications of Monitoring Employee E-mail Use'. In *Employee Responsibilities and Rights Journal* **19**(2): 2007, pages 75–83 (cited on page 1).

[FWC+10] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu: 'Privacy-preserving Data Publishing: A Survey of Recent Developments'. In *ACM Comput. Surv.* **42**(4): June 2010 (cited on pages 1, 3, 10, 77).

[Gal15] Jean H. Gallier: *Logic for Computer Science: Foundations of Automatic Theorem Proving, Second Edition*. Dover, 2015 (cited on page 31).

[GJ90] Michael R. Garey and David S. Johnson: *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co., 1990. ISBN: 0716710455 (cited on page 59).

[GM18] Georg Gottlob and Enrico Malizia: 'Achieving New Upper Bounds for the Hypergraph Duality Problem through Logic'. In *SIAM J. Comput.* **47**(2): 2018, pages 456–492 (cited on page 90).

[Gra10] Bernardo Cuenca Grau: 'Privacy in ontology-based information systems: A pending matter'. In *Semantic Web* **1**(1-2): 2010, pages 137–141 (cited on pages 2, 3, 8).

[GH08] Bernardo Cuenca Grau and Ian Horrocks: 'Privacy-Preserving Query Answering in Logic-based Information Systems'. In *ECAI 2008 - 18th European Conference on Artificial Intelligence, Patras, Greece, July 21-25, 2008, Proceedings*. 2008, pages 40–44 (cited on pages 7, 29).

[GHM+08] Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter F. Patel-Schneider, and Ulrike Sattler: 'OWL 2: The next step for OWL'. In *J. Web Semant.* **6**(4): 2008, pages 309–322 (cited on page 6).

[GKK+15] Bernardo Cuenca Grau, Evgeny Kharlamov, Egor V. Kostylev, and Dmitriy Zheleznyakov: 'Controlled Query Evaluation for Datalog and OWL 2 Profile Ontologies'. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. Edited by Qiang Yang and Michael J. Wooldridge. 2015, pages 2883–2889 (cited on page 7).

[GK16] Bernardo Cuenca Grau and Egor V. Kostylev: 'Logical Foundations of Privacy-Preserving Publishing of Linked Data'. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*. 2016, pages 943–949 (cited on pages 10, 11, 77, 99, 101–104, 107, 108, 130).

[GK19] Bernardo Cuenca Grau and Egor V. Kostylev: 'Logical Foundations of Linked Data Anonymisation'. In *J. Artif. Intell. Res.* **64**: 2019, pages 253–314 (cited on pages 10, 11, 77, 99, 101–104, 107, 108, 130).

[GM12] Bernardo Cuenca Grau and Boris Motik: 'Reasoning over Ontologies with Hidden Content: The Import-by-Query Approach'. In *J. Artif. Intell. Res.* **45**: 2012, pages 197–255 (cited on page 7).

[GMK09] Bernardo Cuenca Grau, Boris Motik, and Yevgeny Kazakov: 'Import-by-Query: Ontology Reasoning under Access Limitations'. In *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*. Edited by Craig Boutilier. 2009, pages 727–732 (cited on page 7).

[GJL+17] Víctor Gutiérrez-Basulto, Jean Christoph Jung, Carsten Lutz, and Lutz Schröder: 'Probabilistic Description Logics for Subjective Uncertainty'. In *J. Artif. Intell. Res.* **58**: 2017, pages 1–66 (cited on page 128).

[HSG15] Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos: 'The role of ontologies in biological and biomedical research: a functional perspective'. In *Briefings in Bioinformatics* **16**(6): 2015, pages 1069–1080 (cited on page 5).

[HB91] Bernhard Hollunder and Franz Baader: 'Qualifying Number Restrictions in Concept Languages'. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Cambridge, MA, USA, April 22-25, 1991*. 1991, pages 335–346 (cited on page 30).

[Hor11] Matthew Horridge: 'Justification based explanation in ontologies'. PhD thesis. University of Manchester, UK, 2011 (cited on pages 3, 9).

[HKS06] Ian Horrocks, Oliver Kutz, and Ulrike Sattler: 'The Even More Irresistible SROIQ'. In *Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006*. Edited by Patrick Doherty, John Mylopoulos, and Christopher A. Welty. 2006, pages 57–67 (cited on page 6).

[HLT+04] Ian Horrocks, Lei Li, Daniele Turi, and Sean Bechhofer: 'The Instance Store: DL Reasoning with Large Numbers of Individuals'. In *Proceedings of the 2004 International Workshop on Description Logics (DL2004), Whistler, British Columbia, Canada, June 6-8, 2004*. 2004 (cited on page 77).

[HPH03] Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen: 'From SHIQ and RDF to OWL: the making of a Web Ontology Language'. In *J. Web Semant.* **1**(1): 2003, pages 7–26 (cited on page 5).

[KPH+07] Aditya Kalyanpur, Bijan Parsia, Matthew Horridge, and Evren Sirin: 'Finding All Justifications of OWL DL Entailments'. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*. 2007, pages 267–280 (cited on page 9).

[KPS+06a] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, and Bernardo Cuenca Grau: 'Repairing Unsatisfiable Concepts in OWL Ontologies'. In *The Semantic Web: Research and Applications, 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings*. Edited by York Sure and John Domingue. 2006, pages 170–184 (cited on pages 9, 53).

[KPS+06b] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau, and James A. Hendler: 'Swoop: A Web Ontology Editing Browser'. In *J. Web Semant.* **4**(2): 2006, pages 144–153 (cited on page 9).

[KKS12] Yevgeny Kazakov, Markus Kroetzsch, and Frantisek Simancik: 'Practical Reasoning with Nominals in the $\mathcal{EL}$ Family of Description Logics'. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012*. 2012 (cited on pages 30, 48).

[Kri18] Francesco Kriegel: 'The Distributive, Graded Lattice of $\mathcal{EL}$ Concept Descriptions and its Neighborhood Relation (Extended Version)'. LTCS-Report 18-10. Dresden, Germany: Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, 2018 (cited on pages 65, 88, 89).

[Küs01] Ralf Küsters: *Non-Standard Inferences in Description Logics*. Volume 2100. Lecture Notes in Computer Science. Springer, 2001 (cited on page 25).

[LSP+08] Joey Sik Chun Lam, Derek H. Sleeman, Jeff Z. Pan, and Wamberto Weber Vasconcelos: 'A Fine-Grained Approach to Resolving Unsatisfiable Ontologies'. In *J. Data Semantics* **10**: 2008, pages 62–95 (cited on pages 3, 9, 54, 57, 129).

[LH10] Jens Lehmann and Pascal Hitzler: 'Concept learning in description logics using refinement operators'. In *Machine Learning* **78**(1-2): 2010, pages 203–250 (cited on page 9).

[LB87] Hector J. Levesque and Ronald J. Brachman: 'Expressiveness and tractability in knowledge representation and reasoning'. In *Computational Intelligence* **3**: 1987, pages 78–93 (cited on page 25).

[Lev96] Alon Y. Levy: 'Obtaining Complete Answers from Incomplete Databases'. In *VLDB'96, Proceedings of 22th International Conference on Very Large Data Bases, September 3-6, 1996, Mumbai (Bombay), India*. Edited by T. M. Vijayaraman, Alejandro P. Buchmann, C. Mohan, and Nandlal L. Sarda. 1996, pages 402–412 (cited on page 1).

[LLV07] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian: 't-Closeness: Privacy Beyond k-Anonymity and l-Diversity'. In *Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007*. Edited by Rada Chirkova, Asuman Dogac, M. Tamer Özsu, and Timos K. Sellis. 2007, pages 106–115 (cited on page 10).

[LM04] Carsten Lutz and Maja Milicic: 'Description Logics with Concrete Domains and Functional Dependencies'. In *Proceedings of the 16th Eureopean Conference on Artificial Intelligence, ECAI'2004, including Prestigious Applicants of Intelligent Systems, PAIS 2004, Valencia, Spain, August 22-27, 2004*. Edited by Ramón López de Mántaras and Lorenza Saitta. 2004, pages 378–382 (cited on page 130).

[LW10] Carsten Lutz and Frank Wolter: 'Deciding inseparability and conservative extensions in the description logic EL'. In *J. Symb. Comput.* **45**(2): 2010, pages 194–228 (cited on page 25).

[MKG+07] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam: '*L*-diversity: Privacy beyond *k*-anonymity'. In *TKDD* **1**(1): 2007, page 3 (cited on page 10).

[MKM+07] David J. Martin, Daniel Kifer, Ashwin Machanavajjhala, Johannes Gehrke, and Joseph Y. Halpern: 'Worst-Case Background Knowledge for Privacy-Preserving Data Publishing'. In *Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007*. Edited by Rada Chirkova, Asuman Dogac, M. Tamer Özsu, and Timos K. Sellis. 2007, pages 126–135 (cited on page 10).

[MLB+06] Thomas Andreas Meyer, Kevin Lee, Richard Booth, and Jeff Z. Pan: 'Finding Maximally Satisfiable Terminologies for the Description Logic ALC'. In *Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA*. 2006, pages 269–274 (cited on page 9).

[MW04] Adam Meyerson and Ryan Williams: 'On the Complexity of Optimal K-Anonymity'. In *Proceedings of the Twenty-third ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 14-16, 2004, Paris, France*. Edited by Catriel Beeri and Alin Deutsch. 2004, pages 223–228 (cited on pages 8, 99).

[MS07] Gerome Miklau and Dan Suciu: 'A formal analysis of information disclosure in data exchange'. In *J. Comput. Syst. Sci.* **73**(3): 2007, pages 507–534 (cited on pages 1, 7, 130).

[MHS09] Boris Motik, Ian Horrocks, and Ulrike Sattler: 'Bridging the gap between OWL and relational databases'. In *J. Web Semant.* **7**(2): 2009, pages 74–89 (cited on page 1).

[NOS16] Nhung Ngo, Magdalena Ortiz, and Mantas Simkus: 'Closed Predicates in Description Logics: Results on Combined Complexity'. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016*. Edited by Chitta Baral, James P. Delgrande, and Frank Wolter. 2016, pages 237–246 (cited on page 130).

[NR14] Nadeschda Nikitina and Sebastian Rudolph: '(Non-)Succinctness of uniform interpolants of general terminologies in the description logic EL'. In *Artif. Intell.* **215**: 2014, pages 120–140 (cited on page 72).

[Nys09] Rafael Peñaloza Nyssen: 'Axiom pinpointing in description logics and beyond'. PhD thesis. Dresden University of Technology, 2009 (cited on page 67).

[OS12] Magdalena Ortiz and Mantas Simkus: 'Reasoning and Query Answering in Description Logics'. In *Reasoning Web. Semantic Technologies for Advanced Query Answering - 8th International Summer School 2012, Vienna, Austria, September 3-8, 2012. Proceedings*. Edited by Thomas Eiter and Thomas Krennwallner. 2012, pages 1–53 (cited on page 43).

[Pap07] Christos H. Papadimitriou: *Computational complexity*. Academic Internet Publ., 2007. ISBN: 978-1-4288-1409-7 (cited on page 27).

[PSK05] Bijan Parsia, Evren Sirin, and Aditya Kalyanpur: 'Debugging OWL ontologies'. In *Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*. 2005, pages 633–640 (cited on page 9).

[Pra05] Ian Pratt-Hartmann: 'Complexity of the Two-Variable Fragment with Counting Quantifiers'. In *J. of Logic, Lang. and Inf.* **14**(3): June 2005 (cited on page 39).

[Rei87] R Reiter: 'A Theory of Diagnosis from First Principles'. In *Artif. Intell.* **32**(1): Apr. 1987, pages 57–95. ISSN: 0004-3702 (cited on pages 9, 56).

[Ros07] Riccardo Rosati: 'On Conjunctive Query Answering in EL'. In *Proceedings of the 2007 International Workshop on Description Logics (DL2007), Brixen-Bressanone, near Bozen-Bolzano, Italy, 8-10 June, 2007*. Edited by Diego Calvanese, Enrico Franconi, Volker Haarslev, Domenico Lembo, Boris Motik, Anni-Yasmin Turhan, and Sergio Tessaris. 2007 (cited on pages 6, 27).

[Ros11] Jeffrey Rosen: 'The right to be forgotten'. In *Stan. L. Rev. Online* **64**: 2011, page 88 (cited on page 3).

[SS98] Pierangela Samarati and Latanya Sweeney: 'Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression'. Technical report. 1998 (cited on pages 6, 8).

[SCF+96] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, and Charles E. Youman: 'Role-Based Access Control Models'. In *IEEE Computer* **29**(2): 1996, pages 38–47 (cited on pages 6, 30).

[Sch94] Andrea Schaerf: 'Reasoning with Individuals in Concept Languages'. In *Data Knowl. Eng.* **13**(2): 1994, pages 141–176 (cited on pages 30, 37).

[Sch91] Klaus Schild: 'A Correspondence Theory for Terminological Logics: Preliminary Report'. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence. Sydney, Australia, August 24-30, 1991*. 1991, pages 466–471 (cited on pages 6, 37).

[Sch05a] Stefan Schlobach: 'Debugging and Semantic Clarification by Pinpointing'. In *The Semantic Web: Research and Applications, Second European Semantic Web Conference, ESWC 2005, Heraklion, Crete, Greece, May 29 - June 1, 2005, Proceedings*. Edited by Asunción Gómez-Pérez and Jérôme Euzenat. 2005, pages 226–240 (cited on page 53).

[Sch05b] Stefan Schlobach: 'Diagnosing Terminologies'. In *Proceedings, The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference, July 9-13, 2005, Pittsburgh, Pennsylvania, USA*. Edited by Manuela M. Veloso and Subbarao Kambhampati. 2005, pages 670–675 (cited on page 53).

[SC03] Stefan Schlobach and Ronald Cornet: 'Non-Standard Reasoning Services for the Debugging of Description Logic Terminologies'. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*. 2003, pages 355–362 (cited on page 9).

[SHC+07] Stefan Schlobach, Zhisheng Huang, Ronald Cornet, and Frank van Harmelen: 'Debugging Incoherent Terminologies'. In *J. Autom. Reasoning* **39**(3): 2007, pages 317–349 (cited on page 9).

[SS91] Manfred Schmidt-Schaubß and Gert Smolka: 'Attributive Concept Descriptions with Complements'. In *Artif. Intell.* **48**(1): Feb. 1991, pages 1–26. ISSN: 0004-3702 (cited on pages 6, 15).

[SFB09] Inanç Seylan, Enrico Franconi, and Jos de Bruijn: 'Effective Query Rewriting with Ontologies over DBoxes (Extended Abstract)'. In *Proceedings of the 22nd International Workshop on Description Logics (DL 2009), Oxford, UK, July 27-30, 2009*. Edited by Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, and Ulrike Sattler. 2009 (cited on page 130).

[SC79] Larry J. Stockmeyer and Ashok K. Chandra: 'Provably Difficult Combinatorial Games'. In *SIAM J. Comput.* **8**(2): 1979, pages 151–174 (cited on page 28).

[SS06] Phiniki Stouppa and Thomas Studer: 'A Formal Model of Data Privacy'. In *Perspectives of Systems Informatics, 6th International Andrei Ershov Memorial Conference, PSI 2006, Novosibirsk, Russia, June 27-30, 2006. Revised Papers*. 2006, pages 400–408 (cited on page 40).

[SS09] Phiniki Stouppa and Thomas Studer: 'Data Privacy for Knowledge Bases'. In *Logical Foundations of Computer Science, International Symposium, LFCS 2009, Deerfield Beach, FL, USA, January 3-6, 2009. Proceedings*. 2009, pages 409–421 (cited on pages 7, 29, 40).

[Sun09] Boontawee Suntisrivaraporn: 'Polynomial time reasoning support for design and maintenance of large-scale biomedical ontologies'. PhD thesis. Dresden University of Technology, Germany, 2009 (cited on pages 79, 99).

[Swe00] Latanya Sweeney: 'Simple Demographics Often Identify People Uniquely'. Working paper. 2000 (cited on page 1).

[Swe02a] Latanya Sweeney: 'Achieving k-Anonymity Privacy Protection Using Generalization and Suppression'. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(5): 2002, pages 571–588 (cited on pages 1, 8, 99).

[Swe02b] Latanya Sweeney: '*k*-Anonymity: A Model for Protecting Privacy'. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(5): 2002, pages 557–570 (cited on page 8).

[TSH10] Jia Tao, Giora Slutzki, and Vasant G. Honavar: 'Secrecy-Preserving Query Answering for Instance Checking in *EL*'. In *Web Reasoning and Rule Systems - Fourth International Conference, RR 2010, Bressanone/Brixen, Italy, September 22-24, 2010. Proceedings*. 2010, pages 195–203 (cited on page 7).

[Tob00] Stephan Tobies: 'The Complexity of Reasoning with Cardinality Restrictions and Nominals in Expressive Description Logics'. In *J. Artif. Intell. Res.* **12**: 2000, pages 199–217 (cited on page 39).

[Tob01] Stephan Tobies: 'Complexity results and practical algorithms for logics in knowledge representation'. PhD thesis. RWTH Aachen University, Germany, 2001 (cited on page 37).

[TW13] David Toman and Grant E. Weddell: 'Conjunctive Query Answering in $\mathcal{CFD}_{nc}$) : A PTIME Description Logic with Functional Constraints and Disjointness'. In *AI 2013: Advances in Artificial Intelligence - 26th Australasian Joint Conference, Dunedin, New Zealand, December 1-6, 2013. Proceedings*. 2013, pages 350–361 (cited on pages 30, 32, 34, 49, 130).

[TCG+18] Nicolas Troquard, Roberto Confalonieri, Pietro Galliani, Rafael Peñaloza, Daniele Porello, and Oliver Kutz: 'Repairing Ontologies via Axiom Weakening'. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, February 2-7, 2018*. 2018 (cited on pages 3, 9, 54).

[Wes76] Alan F. Westin: 'Computers, Health Records, and Citizen Rights.' Tech. rep., 1976 (cited on page 1).