TECHNISCHE UNIVERSITÄT DRESDEN

Technische Universität Dresden
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Axiomatization of General Concept Inclusions from Finite Interpretations

Daniel Borchmann Felix Distel Francesco Kriegel

LTCS-Report 15-13

Postal Address:
Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
01062 Dresden

http://lat.inf.tu-dresden.de

Visiting Address:
Nöthnitzer Str. 46
Dresden

Journal of Applied Non-Classical Logics
Vol. 00, No. 00, Month 201X, 1–42

RESEARCH ARTICLE

Axiomatization of General Concept Inclusions from Finite Interpretations

D. Borchmann[a†‡] and F. Distel[b§] and F. Kriegel[a]

[a] *Institute of Theoretical Computer Science, Technische Universität Dresden*
[b] *d-fine GmbH, Germany*

(*Received 00 Month 201X; final version received 00 Month 201X*)

Description logic knowledge bases can be used to represent knowledge about a particular domain in a formal and unambiguous manner. Their practical relevance has been shown in many research areas, especially in biology and the semantic web. However, the tasks of constructing knowledge bases itself, often performed by human experts, is difficult, time-consuming and expensive. In particular the synthesis of *terminological knowledge* is a challenge every expert has to face. Because human experts cannot be omitted completely from the construction of knowledge bases, it would therefore be desirable to at least get some support from machines during this process. To this end, we shall investigate in this work an approach which shall allow us to extract terminological knowledge in the form of *general concept inclusions* from factual data, where the data is given in the form of vertex and edge labeled graphs. As such graphs appear naturally within the scope of the Semantic Web in the form of sets of RDF triples, the presented approach opens up the possibility to extract terminological knowledge from the Linked Open Data Cloud. We shall also present first experimental results showing that our approach has the potential to be useful for practical applications.

Keywords: Description Logics, Formal Concept Analysis, Terminological Knowledge, Ontology Learning

1. Introduction

One of the main applications of logic in computer science today is to represent knowledge of application domains. Within the scope of this application, description logics [6] play an important role as a family of decidable fragments of first order logic that allow for varying expressivity and reasoning complexity. The practical relevance of description logics as knowledge representation formalisms is reflected by the fact that major biomedical knowledge bases are formulated in or can easily be translated to description logics knowledge bases [35], and by the fact that the languages used within the *Web Ontology Language* standard are based of description logics [23, 27].

The core notion of description logics is the one of a *knowledge base* (or simply *ontology*). Typically, such knowledge bases consist of two parts: a set of assertional axioms, called an $ABox$, and a set of terminological axioms, called a $TBox$. An example of an ontology

†Corresponding author. Email: borch@tcs.inf.tu-dresden.de

‡Partially supported by DFG Graduiertenkolleg 1763 (QuantLA) and by the Cluster of Excellence "Center for Advancing Electronics Dresden"

§Partially supported by the German Research Foundation (DFG) in the Collaborative Research Center 912 (HAEC)

is

$$\mathcal{K}_{\mathsf{MGM}} = (\{\, \mathsf{Cat}(\mathsf{tom}), \mathsf{Mouse}(\mathsf{jerry}) \,\}, \{\, \mathsf{Cat} \sqsubseteq \exists \mathsf{hunts}.\mathsf{Mouse} \,\}). \tag{1}$$

While we provide the formal semantics of a knowledge base later, we can still grasp the meaning of this example on an informal level. The first part of $\mathcal{K}$ denotes the ABox. Intuitively, it states that an individual called tom is an instance of the concept Cat, and that jerry is an instance of Mouse. In other words, we could read this ABox as stating that tom is a cat, and that jerry is a mouse. Thus, the ABox provides *factual information* about our domain.

The second part of $\mathcal{K}$ denotes the TBox. In contrast to the knowledge represented in the ABox, the TBox contains *terminological information*. In this specific example, the TBox contains the knowledge that every individual which is an instance of Cat is connected to another individual via a role named hunts, and that the latter individual is an instance of Mouse. In other words, the TBox states that every cat hunts some mouse.

While this example is not realistic, it still gives a feeling of the expressive power description logic ontologies can provide. This expressivity can even be increased if the underlying logic provides additional features not present in the above example. Studying the interplay between the power of expressiveness of the underlying description logic and the complexity of reasoning within such ontologies has been one of the main driving forces behind description logic research for the past 20 years.

However, this interplay is not within the focus of this work. Instead we are interested in the question of *how to obtain* such ontologies. This question is of high practical relevance, as constructing ontologies is a major undertaking normally requiring a lot of human expertise and time. Providing methods that aid during this process would improve the practicability and applicability of knowledge bases for real-world use cases.

In this work we want to focus on learning terminological knowledge. More specifically, we want to extract axioms of the form $C \sqsubseteq D$ from *description logic interpretations*. Axioms of the form $C \sqsubseteq D$ are called *general concept inclusions* (GCIs), and we have already seen an example of a GCI in (1). Interpretations are structures that serve to define semantics of description logics, and they can be best thought of as vertex and edge-labeled graphs. They are thus not very different from linked data [9], which can also be considered as such a graph. Therefore, what we want to consider in this work can be described as developing methods to obtain terminological knowledge from linked data.

To obtain such methods we consider connections between description logics and the theory of *formal concept analysis* [19]. Originally, formal concept analysis emerged as part of mathematical order theory, aiming at understanding ordered structures, and in particular lattices, as *hierarchies of concepts* of certain *contexts*. However, since its early days, formal concept analysis has developed into a rich theory connecting otherwise independent areas such as order theory, data mining and logic in a fruitful way.

Two of the most basic notions of formal concept analysis are the one of a *formal context* and that of an *implication*. While formal contexts can be roughly thought of as data tables, implications can be thought of as dependencies between attributes in those data tables. Then, formal concept analysis provides effective methods to extract *bases* of implications that are valid in a given formal context, optionally also with the constraint that the base is of minimal cardinality.

The principal approach of our methods to learn GCIs from interpretations is now to connect description logics and formal concept analysis such that learning GCIs from interpretations corresponds to extracting implications from formal contexts. Indeed, this idea is not far-fetched, as there are many similarities between interpretations and formal contexts, as well as between GCIs and implications. We shall discuss these similarities as soon as we have introduced all the necessary definitions.

Exploiting these similarities, we shall discuss methods to obtain terminological knowledge from interpretations. In particular, we shall investigate the idea of extracting *bases of valid GCIs* of such interpretations: as interpretations serve to define the semantics of description logics, we can define the notion of a GCI being valid in such an interpretation. Then a natural starting point for learning GCIs from finite interpretations would be to simply consider the set of all valid GCIs of a finite interpretation. Unfortunately, it can be seen easily that this set is infinite in general, and to make this approach practically relevant we shall resort to finding a *finite base* of all valid GCIs, i.e., a finite set of valid GCIs that already entails every other valid GCI of our given interpretation.

Additionally, we shall restrict our attention to GCIs which obey a certain *role-depth bound*, i.e., whose depth of nested quantifiers does not exceed a predefined limit. We do this for several reasons: firstly, GCIs extracted with our approach need to be validated by an external source, which is likely to be a human expert. Since human experts, even if trained in logic, have difficulties in understanding highly nested logical expressions, it seems wasteful to compute GCIs which a human expert cannot understand. Secondly, while it is possible to not employ a role-depth limit, it causes severe complications in both the underlying theory and possible implementations, which render the whole approach practically useless. We shall sketch those difficulties in a separate section.

We shall see that it is always possible to compute finite bases of GCIs whose quantifiers do not nest below a predefined limit, provided that the initial interpretation is finite. But we shall even go further by utilizing an algorithm from formal concept analysis that allows for the computation of *minimal* bases of valid implications of formal contexts. Here, a base is minimal if and only if the number of implications contained in this base is as small as possible. Using this algorithm we shall show that we can devise an algorithm that allows to compute a minimal base of all valid GCIs of a finite interpretation whose quantifiers depth is bounded by a given threshold.

The methods sketched in the previous paragraphs are effective, and we shall discuss a larger example where these methods are applied to data from the DBpedia project to obtain terminological knowledge about the child-relation between persons in the Wikipedia. While the resulting GCIs are quite promising, we can observe that due to errors in the data some GCIs are not learned although they would be interesting on their own. Even worse, errors in the data cause the GCIs we have learned to be quite complicated, simply due to the fact that they need to "circumvent" the errors in the data. We shall discuss these phenomena in detail when we consider this example.

This paper is structured as follows. At first, we shall review some existing related work in Section 2. Thereafter we shall introduce the necessary notions from description logics and formal concept analysis that are essential for our discussion. This will be done in Section 3. Then we shall introduce in Section 4 our approach of learning finite bases of GCIs with role-depth bound from finite interpretations. This approach is then applied to some real-world data-sets in Section 5. We conclude with an outlook on further results in Section 6.

The results presented here are partly included or are extensions of the work done in [15].

2. Related Work

The results we present in this article fall within the realm of *ontology learning* [26]. The main focus of this research area is to extract *formal representations of knowledge* from various forms of data, most notably from text and linked data, using methods and ideas from a multitude of fields, e.g., machine learning, inductive logical programming, or statistics. Additionally, the notion of an *ontology* is not fixed, but ranges from describing

lightweight collections of facts up to denoting completely formal description logics knowledge bases. Because of this, ontology learning itself is a diverse field consisting of many different lines of research.

In particular, there is plenty of prior work aiming at learning parts of description logic knowledge bases. One step in learning description logics knowledge bases is *concept learning* [25]. The problem here is to find a suitable concept description D in a given knowledge base such that all individuals of a set of *positive examples* satisfy D, and all individuals of a set of *negative examples* do not satisfy D. To find such a concept description D, methods from machine learning are employed, most notably *inductive logic programming*. In this approach, certain *refinement operators* are considered. Then, starting with a concept description D' that is not necessarily satisfied by all positive examples, or is satisfied by some negative examples, a suitable refinement operator ρ (depending on the target description logic) is applied to D' to obtain new candidates for the concept description D. Provided that the refinement operator ρ satisfies certain convergence properties, iterating this process and applying certain heuristics to choose among the candidates returned by ρ finally yields such a desired concept description D, if it exists.

A related approach is the one of *bottom-up construction* of description logics knowledge bases [8]. A natural top-down approach to constructing description logics knowledge bases is to first specify the ontology completely, and then use it to describe properties of individuals of the domain of interest. This approach, however, may not always be possible, as finding appropriate descriptions for all relevant concept descriptions of the application domain may be infeasible. Instead, in a bottom-up approach, the domain expert first specifies "typical" examples of a certain concept description to be defined, and from these examples a first concept description is inferred. This is done by first computing the *most specific concept description* of each of the given examples. Thereafter, the *least common subsumer* of these most specific concept descriptions is computed. This least common susumer C is then considered as a first proposal to describe all the examples given by the expert, and the expert can then refine and adapt C as necessary.

For ontology learning algorithms based on expert interaction, methods based on *formal concept analysis* have been particularly popular [31]. Here the main interest lies in adapting the algorithm of *attribute exploration* to the setting of description logics. In formal concept analysis, attribute exploration is a knowledge completion algorithm that uses expert interaction to decide newly found knowledge whose validity cannot be decided from the given data alone. A main obstacle in using formal concept analysis for ontology construction is that a *closed world* is always assumed, i.e., at any time all properties of the known individuals are completely available. Furthermore, the knowledge covered by attribute exploration can be expressed using definite Horn formulas, which is too inexpressive for description logics. Because of this, various approaches have been developed to extend the expressivity of attribute exploration. One of them is *relational exploration* [30], which provides a method to extract information from finite data-sets that allows to decide all subsumptions between $\mathcal{FLE}$-concept descriptions. However, the method itself does not directly yield terminological knowledge required to construct a knowledge base from the given data-set. The methods from [15] on which our argumentation is built are similar, but differ in the aspect that they directly yield terminological knowledge suitable for knowledge base construction. There also exists an extension of the latter methods to *ABox-exploration* [14] that allows the expert to provide counterexamples in an open-world fashion. Finally, attribute exploration has also been used to devise methods for *ontology completion* [7], in which expert interaction is used to ensure that the ontology at hand completely describes the application domain.

Finally, learning ontological knowledge from text or web documents has been one of the most prominent lines of research in ontology learning [33]. Here the focus is usually

not to construct fully formalized knowledge bases, but merely extract *taxonomies* of concept names or even only facts from textually represented data. However, there have been some approaches to also learn *concept definitions* from text [16]. Those concept definitions can be seen as a special form of terminological knowledge, and can be used in description logics knowledge bases.

3. Preliminaries

In the introduction we have already encountered some notions which are relevant for the purpose of this work. Up to now, we have used these notions rather intuitively, without proper formal foundations. It is the purpose of this section to remedy this deficiency, and to provide formal definitions as far as they are needed here.

To this end, we shall introduce in Section 3.1 the necessary notions of description logics [6] crucial for this work. In particular, we shall introduce the description logic $\mathcal{EL}^\bot$, its syntax and semantics, as well as the notion of *general concept inclusions*. Moreover, as we shall make use of results from the field of formal concept analysis [19], we shall introduce in Section 3.2 basic notions from this area, including *formal contexts*, *implications*, and the *canonical base*.

3.1 *The Description Logic $\mathcal{EL}^\bot$*

The description logic $\mathcal{EL}^\bot$ is one of the least expressive description logics considered in the literature [4, 5], and yet this logic has practical relevance for representing knowledge. In particular, some large ontologies from the domain of medicine and bio-medicine can be represented in $\mathcal{EL}^\bot$ or slight extensions thereof, examples being SNOMED-CT [28], GALEN [29] and the Gene Ontology [1].

As a logic, $\mathcal{EL}^\bot$ consists of syntax and semantics. To this end, we need to fix a background vocabulary which consists of two disjoint sets N_C and N_R of *concept names* and *role names*, respectively. Then an *$\mathcal{EL}$-concept description* C (over N_C and N_R) is defined according to the rule

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

where $A \in N_C$ is a concept name and $r \in N_R$ a role name. Sometimes we call $C \sqcap D$ a *conjunction*, $\exists r.C$ an *existential restriction*, and $\top$ the *top concept*. An *$\mathcal{EL}^\bot$-concept description* (over N_C and N_R) is then either an $\mathcal{EL}$-concept description or the special constructor $\bot$, and we refer to $\bot$ as the *bottom concept*. We shall occasionally denote the set of all $\mathcal{EL}^\bot$-concept descriptions for the signature N_C and N_R by $\mathcal{EL}^\bot(N_C, N_R)$. The *role-depth* $\mathrm{rd}(C)$ of an $\mathcal{EL}^\bot$-concept description C is inductivelly defined as follows:

$$\begin{aligned}
\mathrm{rd}(\bot) &:= 0,\\
\mathrm{rd}(\top) &:= 0,\\
\mathrm{rd}(A) &:= 0 \quad (A \in N_C),\\
\mathrm{rd}(\exists r.C) &:= 1 + \mathrm{rd}(C) \quad (r \in N_R, C \in \mathcal{EL}^\bot(N_C, N_R)).
\end{aligned}$$

The set of all $\mathcal{EL}^\bot$-concept descriptions (over N_C and N_R) with a role-depth at most d is denoted by $\mathcal{EL}^\bot(N_C, N_R)_d$.

As an example we can consider the sets $N_C := \{\, \mathsf{Cat}, \mathsf{Mouse} \,\}$ and $N_R := \{\, \mathsf{hunts} \,\}$.

Examples for $\mathcal{EL}^\bot$ descriptions over this vocabulary are then

$$\mathsf{Cat},\ \exists\mathsf{hunts}.\mathsf{Mouse},\ \mathsf{Cat} \sqcap \mathsf{Mouse},\ \bot.$$

Note that, by convention, concept names are often denoted with capitalized words, where role names are denoted by lower-case words.

To give semantics to $\mathcal{EL}^\bot$-concept descriptions we shall introduce the notion of an *interpretation* $\mathcal{I}$ (over the vocabulary N_C and N_R). Those are structures $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ such that $\Delta^\mathcal{I}$ is a non-empty set, which is called *domain* and whose elements are called *individuals*. Moreover, $\cdot^\mathcal{I}$ is a mapping from $N_C \cup N_R$ to $\mathfrak{P}(\Delta^\mathcal{I}) \cup \mathfrak{P}(\Delta^\mathcal{I} \times \Delta^\mathcal{I})$ satisfying

$$A^\mathcal{I} \subseteq \Delta^\mathcal{I} \quad \text{and} \quad r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$$

for $A \in N_C$ and $r \in N_R$. The mapping $\cdot^\mathcal{I}$ can be extended easily to the set $\mathcal{EL}^\bot(N_C, N_R)$ of all $\mathcal{EL}^\bot$-concept descriptions over N_C and N_R:

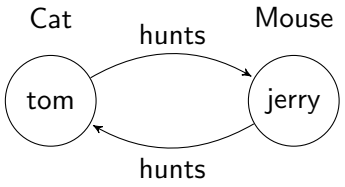
$$\begin{aligned}
\top^\mathcal{I} &:= \Delta^\mathcal{I} \\
\bot^\mathcal{I} &:= \emptyset \\
(C \sqcap D)^\mathcal{I} &:= C^\mathcal{I} \cap D^\mathcal{I} \\
(\exists r.C)^\mathcal{I} &:= \{\, x \in \Delta^\mathcal{I} \mid \exists y \in \Delta^\mathcal{I} \colon (x, y) \in r^\mathcal{I} \text{ and } y \in C^\mathcal{I} \,\}
\end{aligned}$$

where $C, D \in \mathcal{EL}^\bot(N_C, N_R)$ and $r \in N_R$. If C is an $\mathcal{EL}^\bot$-concept description, then we shall call $C^\mathcal{I}$ its *extension* in $\mathcal{I}$, and shall say for every individual $x \in \Delta^\mathcal{I}$ that it *satisfies* C in $\mathcal{I}$ if and only if $x \in C^\mathcal{I}$.

Let us consider an example interpretation for our signature $N_C = \{\, \mathsf{Cat}, \mathsf{Mouse} \,\}, N_R = \{\, \mathsf{hunts} \,\}$. Define $\mathcal{I}_\mathsf{MGM} = (\{\, \mathsf{tom}, \mathsf{jerry} \,\}, \cdot^{\mathcal{I}_\mathsf{MGM}})$ by

$$\begin{aligned}
\mathsf{Cat}^{\mathcal{I}_\mathsf{MGM}} &:= \{\, \mathsf{tom} \,\}, \\
\mathsf{Mouse}^{\mathcal{I}_\mathsf{MGM}} &:= \{\, \mathsf{jerry} \,\}, \\
\mathsf{hunts}^{\mathcal{I}_\mathsf{MGM}} &:= \{\, (\mathsf{tom}, \mathsf{jerry}), (\mathsf{jerry}, \mathsf{tom}) \,\}.
\end{aligned}$$

It is not hard to see that $\mathcal{I}_\mathsf{MGM}$ can also be represented as a graph, which may give more insight into its structure:



For this interpretation, we can compute extensions of certain concept descriptions:

$$\begin{aligned}
\exists\mathsf{hunts}.\mathsf{Mouse}^{\mathcal{I}_\mathsf{MGM}} &= \{\, \mathsf{tom} \,\} = \mathsf{Cat}^{\mathcal{I}_\mathsf{MGM}}, \\
\mathsf{Cat} \sqcap \mathsf{Mouse}^{\mathcal{I}_\mathsf{MGM}} &= \emptyset = \bot^{\mathcal{I}_\mathsf{MGM}}.
\end{aligned}$$

As already indicated in the introduction, it is possible for some interpretation $\mathcal{I}$ and two $\mathcal{EL}^\bot$-concept descriptions C, D that whenever an individual satisfies C in $\mathcal{I}$ then it also

satisfies D in $\mathcal{I}$, i.e.,

$$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}. \tag{2}$$

This implication-like dependency between concept descriptions can be lifted to the logical level by considering *general concept inclusions* (GCIs). These are expressions of the form $C \sqsubseteq D$, where C and D are $\mathcal{EL}^{\bot}$-concept descriptions over N_C and N_R. Such a GCI is then said to *hold* in $\mathcal{I}$ (is *valid* in $\mathcal{I}$) if and only if (2) holds, i.e., $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. We write $\mathcal{I} \models (C \sqsubseteq D)$ in this case. If $C \sqsubseteq D$ is valid in every interpretation, then we say that C is *subsumed by* D, and simply write $C \sqsubseteq D$ in this case. The set of GCIs valid in $\mathcal{I}$ will be denoted by $\mathrm{Th}_{\mathcal{EL}^{\bot}(N_C,N_R)}(\mathcal{I})$. If the underlying logic and vocabulary are clear from the context, we shall omit the subscript and write $\mathrm{Th}(\mathcal{I})$ instead. Analogously, the set of all $\mathcal{EL}^{\bot}$-GCIs $C \sqsubseteq D$ (over N_C and N_R) that hold in $\mathcal{I}$ and satisfy $\mathrm{rd}(C), \mathrm{rd}(D) \leq d$ is denoted by $\mathrm{Th}_{\mathcal{EL}^{\bot}(N_C,N_R)_d}(\mathcal{I})$. We may abbreviate this set by $\mathrm{Th}^d(\mathcal{I})$.

As an example, let us consider the GCI from (1) again, namely

$$\mathsf{Cat} \sqsubseteq \exists \mathsf{hunts}.\mathsf{Mouse}.$$

Then since $\mathsf{Cat}^{\mathcal{I}_{\mathsf{MGM}}} = \exists \mathsf{hunts}.\mathsf{Mouse}^{\mathcal{I}_{\mathsf{MGM}}}$, this GCI is valid in $\mathcal{I}_{\mathsf{MGM}}$.

General concept inclusions allow us to express *terminological knowledge*, i.e., knowledge about dependencies between concept descriptions. However, as we have already indicated in the introduction, it is also possible to express *assertional knowledge*, i.e., facts about individuals, using concept descriptions. For this we extend the vocabulary by a set N_I of *individual names* which is disjoint to both N_C and N_R. Then a *complex concept assertion* is of the form $C(a)$, where C is an $\mathcal{EL}^{\bot}$-concept description over N_C and N_R, and $a \in N_I$ is an individual name. A *role assertion* is of the form $r(a, b)$ for $r \in N_R$ and $a, b \in N_I$, and a *concept assertion* is of the form $A(a)$ for $A \in N_C$ and $a \in N_I$.

To give semantics to concept assertions we shall extend the notion of interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ to include an interpretation for the individual names. To this end, we simply demand that the mapping $\cdot^{\mathcal{I}}$ injectively assigns to individual names $a \in N_I$ individuals in $\Delta^{\mathcal{I}}$, i.e., $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ for each $a \in N_I$ and $a^{\mathcal{I}} = b^{\mathcal{I}}$ implies $a = b$. Using this extension we can now say that the assertions $C(a)$ and $r(a, b)$ *hold* in $\mathcal{I}$ if and only if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ are true, respectively.

Assertional and terminological knowledge can be combined into a *knowledge base*. Formally, this is a pair $\mathcal{K} = (\mathcal{A}, \mathcal{T})$ of an *ABox* $\mathcal{A}$ and a *TBox* $\mathcal{T}$. Here an ABox (for "assertional box") is a collection of concept and role assertions, and a TBox (for "terminological box") is a collection of GCIs. An interpretation $\mathcal{I}$ is then a *model* for $\mathcal{K}$ if and only if all assertions in $\mathcal{A}$ and all GCIs in $\mathcal{T}$ are valid in $\mathcal{I}$.

An example of such a knowledge base $\mathcal{K}_{\mathsf{MGM}}$ has been given in (1), and $\mathcal{I}_{\mathsf{MGM}}$ is model of $\mathcal{K}$.

Knowledge bases allow for a variety of *reasoning tasks*, based on their semantics. These tasks include *consistency*, *satisfiability*, *subsumption*, *equivalence* and *instance checking*:

Consistency Given a knowledge base $\mathcal{K}$, does there exist a model for $\mathcal{K}$? (*Is $\mathcal{K}$ consistent?*)

Satisfiability Given a knowledge base $\mathcal{K}$ and a concept description C, does there exist a model for $\mathcal{K}$ such that $C^{\mathcal{I}} \neq \emptyset$? (*Is C satisfiable with respect to $\mathcal{K}$?*)

Subsumption Given a knowledge base $\mathcal{K}$ and two concept descriptions C, D, is it true for all models $\mathcal{I}$ of $\mathcal{K}$ that $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$? (*Does $\mathcal{K}$ entail $C \sqsubseteq D$?*)

Equivalence Given a knowledge base $\mathcal{K}$ and two concept descriptions C, D, is it true for all models $\mathcal{I}$ of $\mathcal{K}$ that $C^{\mathcal{I}} = D^{\mathcal{I}}$? (*Does $\mathcal{K}$ entail $C \equiv D$?*)

Instance Checking Given a knowledge base $\mathcal{K}$, a concept description C and an indi-

vidual name $a \in N_I$, is it true for all models $\mathcal{I}$ of $\mathcal{K}$ that $a^{\mathcal{I}} \in C^{\mathcal{I}}$? (*Does* $\mathcal{K}$ *entail* $C(a)$*?*)

Since $\mathcal{I}_{\mathsf{MGM}}$ is model of $\mathcal{K}_{\mathsf{MGM}}$, we know that $\mathcal{K}_{\mathsf{MGM}}$ is consistent. Moreover, as $\mathsf{Cat}^{\mathcal{I}_{\mathsf{MGM}}} \neq \emptyset \neq \mathsf{Mouse}^{\mathcal{I}_{\mathsf{MGM}}}$, we know that both Cat and Mouse are satisfiable with respect to $\mathcal{K}_{\mathsf{MGM}}$.

As $\mathsf{Cat} \sqsubseteq \exists\mathsf{hunts}.\mathsf{Mouse}$ is explicitly mentioned in $\mathcal{K}_{\mathsf{MGM}}$, it is also entailed by it, as is

$$\exists\mathsf{hunts}.\mathsf{Cat} \sqsubseteq \exists\mathsf{hunts}.\exists\mathsf{hunts}.\mathsf{Mouse}.$$

However, $\exists\mathsf{hunts}.\mathsf{Mouse} \sqsubseteq \mathsf{Cat}$ is not entailed by $\mathcal{K}_{\mathsf{MGM}}$, and therefore $\mathcal{K}_{\mathsf{MGM}}$ does not entail

$$\mathsf{Cat} \equiv \exists\mathsf{hunts}.\mathsf{Mouse}.$$

Finally, $\mathcal{K}_{\mathsf{MGM}}$ clearly entails $\mathsf{Cat}(\mathsf{tom})$ and $\exists\mathsf{hunts}.\mathsf{Mouse}(\mathsf{tom})$, but neither $\mathsf{Mouse}(\mathsf{tom})$ nor $\exists\mathsf{hunts}.\mathsf{Mouse}(\mathsf{jerry})$.

When considering different description logics, one of the first questions is how complex these reasoning tasks are. One of the many advantages of $\mathcal{EL}^{\bot}$ is that all of these reasoning tasks can be decided in polynomial time.

3.2 *Formal Concept Analysis*

Formal concept analysis started in the 1980s as a branch of mathematical order theory, and has since evolved into a wide theory, with applications in data mining, knowledge representation, and even psychology. The original concern of formal concept analysis was to study a mathematical connection between *complete lattices*, a particular form of ordered sets, on the one hand, and *formal contexts* on the other. More precisely, formal concept analysis allowed to understand arbitrary complete lattices as *conceptual hierarchies*, with the notion of a *concept* defined in a corresponding formal context.

This original line of research is not immediately relevant for our course of argumentation. Instead, we shall exploit another connection which formal concept analysis makes explicit. This connection is concerned with *closure systems* on finite sets, which in turn are always finite complete lattices. Such closure systems can be described in terms of *implications* in a suitable formal context, and formal concept analysis provides well-established means to study and extract implications from formal contexts. We shall see in the course of this paper that we can exploit those means to our advantage when learning valid GCIs from a given data set. It is the purpose of this section to introduce the necessary definitions to argue how this can be done.

We start with introducing formal contexts. To this end, let G, M be two sets, and let $I \subseteq G \times M$. Then a *formal context* $\mathbb{K}$ is just a triple $\mathbb{K} = (G, M, I)$. When talking about formal contexts, we shall call the set G the set of *objects*, the set M the set of *attributes*, and we shall say that an object $g \in G$ *has* an attribute $m \in M$ if and only if $(g, m) \in I$. In this case, we shall also write $g\ I\ m$ instead of $(g, m) \in I$.

For a set $A \subseteq G$ of objects, we can form the set A' of all attributes that all objects in A have in common. More precisely, we shall define the *derivation*

$$A' = \{\, m \in M \mid \forall g \in A\colon\ g\ I\ m \,\}.$$

Likewise, for a set $B \subseteq M$, the set of objects sharing all attributes in B is defined as

$$B' = \{\, g \in G \mid \forall m \in M\colon\ g\ I\ m \,\}.$$

We sometimes also write $A'_{\mathbb{K}}$ to emphasize that the derivation is done in $\mathbb{K}$.

Formal contexts can be thought of, among others, as *data sets*, where we record for the objects of interest all the attributes they have. It is then quite natural to ask whether certain combination of attributes *imply* certain other combinations of attributes. More precisely, given two sets $X, Y \subseteq M$, is it true that every object that has all attributes from X also has all attributes from Y, i.e.,

$$\forall g \in G\colon\ g \in X' \implies g \in Y' \quad ?$$

In formal concept analysis this notion is formalized with the notion of an *implication*. More precisely, an implication over some set M is an expression $A \to B$ with $A, B \subseteq M$. We say that an implication $A \to B$ over some set M *holds* in some formal context with attribute set M if and only if every object that has all attribute from A also has all attributes from B, i.e.,

$$A' \subseteq B'.$$

It can be shown that this condition is equivalent to $B \subseteq A''$. An implication which holds in a formal context $\mathbb{K}$ is also said to be a *valid* implication of $\mathbb{K}$. The set of all valid implications of $\mathbb{K}$ is denoted with $\mathrm{Th}(\mathbb{K})$. If $\mathcal{L}$ is a set of valid implications of $\mathbb{K}$ then we shall also call $\mathbb{K}$ a *model* of $\mathcal{L}$.

For a set $X \subseteq M$ of attributes and a set $\mathcal{L}$ of implications over M we can ask which other attributes *follow* from X and $\mathcal{L}$. We shall denote with $\mathcal{L}(X) \subseteq M$ the set of all attributes that follow from X and $\mathcal{L}$, and define it as follows

- $\mathcal{L}^1(X) = X \cup \bigcup\{\, B \mid (A \to B) \in \mathcal{L}, A \subseteq X \,\}$,
- $\mathcal{L}^{i+1}(X) = \mathcal{L}^i(\mathcal{L}^1(X))$,
- $\mathcal{L}(X) = \bigcup_{i \in \mathbb{N} \setminus \{\,0\,\}} \mathcal{L}^i(X)$.

We say that X is *closed* under $\mathcal{L}$ if and only if $X = \mathcal{L}(X)$.

A formal context $\mathbb{K} = (G, M, I)$ can have an exponential number of valid implications, where the size of $\mathbb{K}$ is defined to be $|G| \cdot |M|$. For practical purposes, it is desirable to find *smaller* sets of valid implications which nevertheless contain all information of the whole set of all implications that hold in $\mathbb{K}$. Such sets are called *bases*. Formally, a set $\mathcal{B} \subseteq \mathrm{Th}(\mathbb{K})$ is called a *base* of $\mathbb{K}$ if for all implications $(X \to Y) \in \mathrm{Th}(\mathbb{K})$ we have that $\mathcal{B}$ *entails* $X \to Y$. Here, $\mathcal{B}$ *entails* $X \to Y$ if and only if for each formal context $\mathbb{L}$ with attribute set M such that $\mathcal{B} \subseteq \mathrm{Th}(\mathbb{L})$ it is also true that $(X \to Y) \in \mathrm{Th}(\mathbb{L})$. We shall write $\mathcal{B} \models (X \to Y)$ in this case. The base $\mathcal{B}$ is called *non-redundant* (or *irredundant*) if no proper subset of $\mathcal{B}$ is a base of $\mathbb{K}$. $\mathcal{B}$ is called *minimal*, if it has minimal cardinality among all bases of $\mathbb{K}$, i.e., if there does not exists a base of $\mathbb{K}$ with fewer elements than $\mathcal{B}$.

A base can be considered with respect to some *background knowledge* $\mathcal{S} \subseteq \mathrm{Th}(\mathbb{K})$. More precisely, a *base with background knowledge* $\mathcal{S}$ is a set $\mathcal{B} \subseteq \mathrm{Th}(\mathbb{K})$ such that $\mathcal{B} \cup \mathcal{S}$ is a base of $\mathbb{K}$. Such a base $\mathcal{B}$ is called *non-redundant* (*irredundant*) if no proper subset of $\mathcal{B}$ is a base of $\mathbb{K}$ with background knowledge $\mathcal{S}$. $\mathcal{B}$ is called *minimal* if the cardinality of $\mathcal{B}$ is minimal among all bases of $\mathbb{K}$ with background knowledge $\mathcal{S}$.

A main line of research of formal concept analysis is concerned with developing fast algorithms for computing bases of given formal contexts, possibly with some background knowledge. Particular interest has been generated by the so-called *canonical base* $\mathrm{Can}(\mathbb{K}, \mathcal{S})$ of $\mathbb{K}$ with background knowledge $\mathcal{S}$ [21, 32]. This base is a minimal base with background knowledge $\mathcal{S}$, and for whose computation practical algorithms are available. To describe $\mathrm{Can}(\mathbb{K}, \mathcal{S})$ we need to introduce the notion of $\mathcal{S}$*-pseudo-intents* of $\mathbb{K}$. These are sets $P \subseteq M$ such that

- $P \neq P''$,
- $P = \mathcal{S}(P)$, and
- for all $\mathcal{S}$-pseudo-intents $Q \subsetneq P$ it is true that $Q'' \subseteq P$.

With this we have

$$\mathrm{Can}(\mathbb{K}, \mathcal{S}) \coloneqq \{\, P \to P'' \mid P \text{ an } \mathcal{S}\text{-pseudo-intent of } \mathbb{K} \,\}.$$

4. Exact Mining of General Concept Inclusions

With all necessary definitions at hand we are now ready to discuss our method of learning terminological knowledge from a finite data set. For this we assume this data set to be represented as a finite interpretation $\mathcal{I}$, i.e., as an interpretation whose set of elements is finite. What we then want is to compute a *finite base* of all $\mathcal{EL}^\bot$-GCIs that are valid in $\mathcal{I}$, and whose quantifiers do not nest deeper than a chosen depth $d \in \mathbb{N}$. In other words, we want to compute a finite set $\mathcal{B}$ of valid GCIs of $\mathcal{I}$ such that every other valid $\mathcal{EL}^\bot$-GCI of $\mathcal{I}$ follows from $\mathcal{B}$, and such that all concept descriptions occurring in $\mathcal{B}$ are $\mathcal{EL}^\bot$-concept descriptions with quantifier depth at most d.

The choice of representing data sets as finite interpretations does not impose a severe restriction on the applicability of our approach. In fact, as we have already sketched in the introduction, every finite interpretation can be seen as a vertex- and edge-labeled graph, and data sets representable as graphs can be obtained easily, for example from RDFS graphs [12]. Thus our approach allows us, at least in principle, to automatically construct $\mathcal{EL}^\bot$-TBoxes from the linked open data cloud.

Our argumentation to obtain bases of valid $\mathcal{EL}^\bot$-GCIs of finite interpretations makes use of ideas from formal concept analysis. More precisely, as we shall see shortly, we can exploit similarities between formal concept analysis and description logics to reformulate methods of constructing bases of finite formal contexts as methods to compute finite bases of valid GCIs. We shall even show that we can associate to every finite interpretation $\mathcal{I}$ a finite formal context $\mathbb{K}_\mathcal{I}$ such that all finite bases of $\mathbb{K}_\mathcal{I}$ can easily be transformed into finite bases of $\mathcal{I}$. In this way we can, without further modifications, utilize algorithms from formal concept analysis for computing implicational bases to compute finite bases of finite interpretations.

A major obstacle in finding a finite base of a finite interpretation $\mathcal{I}$ is the fact that the number of valid $\mathcal{EL}^\bot$-GCIs of $\mathcal{I}$ is infinite in general. This is because if $C \sqsubseteq D$ holds in $\mathcal{I}$, and if $r \in N_R$, then $\exists r.C \sqsubseteq \exists r.D$ holds in $\mathcal{I}$ as well.

This section is structured as follows. In Section 4.1 we shall introduce *model-based most-specific concept-descriptions*, which we shall employ in Section 4.2 to compute finite bases of valid GCIs of finite interpretations. As model-based most-specific concept descriptions turn out to be crucial for our purposes, we shall discuss in Section 4.3 efficient ways for their computation. Finally, we shall describe in Section 4.4 a base of valid GCIs with minimal cardinality.

The argumentation described in this section is an extension of [15], but we shall restrict our attention to $\mathcal{EL}^\bot$-GCIs with bounded quantifier depth. This restriction has the advantage of being easier to follow, as it requires less theoretical particularities. To give an impression of the original argumentation we shall give a brief overview of it in Section 4.5.

4.1 *Model-Based Most-Specific Concept Descriptions*

There are astonishing similarities between formal concept analysis and description logics. For example, an interpretation $\mathcal{I}$ is "similar" to a formal context $\mathbb{K} = (G, M, I)$ in the

sense that elements in $\mathcal{I}$ can satisfy certain properties (concept descriptions) in the same spirit as objects in $\mathbb{K}$ satisfy certain properties (attributes). Likewise, the extension function $\cdot^{\mathcal{I}}$ maps every concept description $C^{\mathcal{I}}$ to the elements in $\mathcal{I}$ that satisfy C, much like the derivation operator $\cdot'$ in $\mathbb{K}$ maps every subset $A \subseteq M$ to the set of objects in $\mathbb{K}$ that satisfy all attributes in M.

In description logics there is, however, a missing counterpart to the other derivation operator from formal concept analysis that maps a set A of objects to the set A' of all attributes that all objects in A have. To transfer ideas from formal concept analysis to descriptions logics an analogue to this mapping is necessary, and such an analogue should map a set X of elements of $\mathcal{I}$ to a concept description that contains all properties shared by all elements of X. We introduce this kind of concept descriptions as *model-based most-specific concept descriptions*.

4.1 Definition (Model-Based Most-Specific Concept Description)**.** Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a finite interpretation over the signature (N_C, N_R), and let $d \in \mathbb{N}$. Then for $X \subseteq \Delta^{\mathcal{I}}$ an $\mathcal{EL}^{\bot}$-concept description C over N_C and N_R is a (role-depth-bounded) *model-based most-specific concept description* of X in $\mathcal{I}$ with role-depth at most d if and only if

(i) $\mathrm{rd}(C) \leq d$,
(ii) $X \subseteq C^{\mathcal{I}}$, and
(iii) for all $\mathcal{EL}^{\bot}$-concept descriptions D over N_C and N_R with $\mathrm{rd}(D) \leq d$ and $X \subseteq D^{\mathcal{I}}$, it is true that $C \sqsubseteq D$.

Note that role-depth-bounded model-based most-specific concept descriptions always exist. This is because the set $\mathcal{EL}^{\bot}(N_C, N_R)_d$ of all $\mathcal{EL}^{\bot}$-concept descriptions over N_C and N_R with role-depth at most d is finite up to equivalence, and is closed under $\sqcap$. In other words, if $\mathcal{T}$ is a set of representatives of $\mathcal{EL}^{\bot}(N_C, N_R)_d$ with respect to the equivalence relation $\equiv$, then $\mathcal{T}$ is finite, and a model-based most-specific concept description of $X \subseteq \Delta^{\mathcal{I}}$ can be obtained as

$$\bigsqcap \{\, C \in \mathcal{T} \mid X \subseteq C^{\mathcal{I}} \,\}.$$

Of course, this way of computing model-based most-specific concept descriptions is not efficient. We shall discuss a more practical method in 4.3.

Note that by their very definition, model-based most-specific concept descriptions are unique up to equivalence among all concept descriptions with role-depth at most d, and it is therefore save to talk about *the* model-based most-specific concept description of X. We shall denote this concept description by $X^{\mathcal{I}^d}$, to stress the similarity to the corresponding derivation operator from formal concept analysis.

One of the most important structural properties of the mappings $(\cdot)^{\mathcal{I}}$ and $(\cdot)^{\mathcal{I}^d}$ is that they satisfy the main property of an *isotone Galois connection*.

4.2 Lemma. *For all $\mathcal{EL}^{\bot}$-concept descriptions D with $\mathrm{rd}(D) \leq d$, and all $X \subseteq \Delta^{\mathcal{I}}$, it is true that*

$$X^{\mathcal{I}^d} \sqsubseteq D \iff X \subseteq D^{\mathcal{I}}. \tag{3}$$

Proof. Let $X \subseteq D^{\mathcal{I}}$. Then by definition of the model-based most-specific concept descriptions, $X^{\mathcal{I}^d} \sqsubseteq D$, because $\mathrm{rd}(D) \leq d$.

If $X^{\mathcal{I}^d} \sqsubseteq D$, then because of $X \subseteq X^{\mathcal{I}^d\mathcal{I}}$ we obtain $X \subseteq D^{\mathcal{I}}$ as required. □

With $(\cdot)^{\mathcal{I}}$ and $(\cdot)^{\mathcal{I}^d}$ satisfying (3), we immediately obtain some useful statements about the interplay of these two mappings, some of which are obvious on their own.

4.3 Lemma. *For all $\mathcal{EL}^{\bot}$-concept descriptions C, D with $\mathrm{rd}(C), \mathrm{rd}(D) \leq d$, and all $X, Y \subseteq \Delta^{\mathcal{I}}$ it is true that*

(i) $X \subseteq Y \implies X^{\mathcal{I}^d} \sqsubseteq Y^{\mathcal{I}^d}$
(ii) $C \sqsubseteq D \implies C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
(iii) $X \subseteq X^{\mathcal{I}^d\mathcal{I}}$
(iv) $C^{\mathcal{I}\mathcal{I}^d} \sqsubseteq C$
(v) $X^{\mathcal{I}^d\mathcal{I}\mathcal{I}^d} \equiv X^{\mathcal{I}^d}$
(vi) $C^{\mathcal{I}\mathcal{I}^d\mathcal{I}} = C^{\mathcal{I}}$

Proof. To see (i), we obtain with (3) from $Y^{\mathcal{I}^d} \sqsubseteq Y^{\mathcal{I}^d}$ that $Y \subseteq Y^{\mathcal{I}^d\mathcal{I}}$ (which already shows (iii)). Since $X \subseteq Y$, $X \subseteq Y^{\mathcal{I}^d\mathcal{I}}$, and another application of (3) yields $X^{\mathcal{I}^d} \sqsubseteq Y^{\mathcal{I}^d}$, as desired.

Statement (ii) is clear from the definition of $\mathcal{I}$. Applying (3) to $C^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ immediately yields $C^{\mathcal{I}\mathcal{I}^d} \sqsubseteq C$, which shows (iv).

For (v) we observe that $X \subseteq X^{\mathcal{I}^d\mathcal{I}}$ implies $X^{\mathcal{I}^d} \sqsubseteq X^{\mathcal{I}^d\mathcal{I}\mathcal{I}^d}$ by (i). On the other hand, Statement (iv) with $C \coloneqq X^{\mathcal{I}^d}$ shows $X^{\mathcal{I}^d\mathcal{I}\mathcal{I}^d} \sqsubseteq X^{\mathcal{I}^d}$.

Finally, Statement (vi) can be shown by first observing that with $X \coloneqq C^{\mathcal{I}}$, Statement (iii) yields $C^{\mathcal{I}} \subseteq C^{\mathcal{I}\mathcal{I}^d\mathcal{I}}$. On the other hand, $C^{\mathcal{I}\mathcal{I}^d} \sqsubseteq C$ entails $C^{\mathcal{I}\mathcal{I}^d\mathcal{I}} \subseteq C^{\mathcal{I}}$ by (ii). □

4.4 Proposition. *For all $\mathcal{EL}^{\bot}$-concept descriptions C, D over N_C and N_R and all $r \in N_R$ it is true that*

$$(C^{\mathcal{I}\mathcal{I}^d} \sqcap D)^{\mathcal{I}} = (C \sqcap D)^{\mathcal{I}},$$
$$(\exists r.C^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}} = (\exists r.C)^{\mathcal{I}}.$$

Proof. For the first equation we obtain

$$\begin{aligned}(C^{\mathcal{I}\mathcal{I}^d} \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}\mathcal{I}^d\mathcal{I}} \cap D^{\mathcal{I}} \\ &= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\ &= (C \sqcap D)^{\mathcal{I}}.\end{aligned}$$

For the second equation we can compute

$$\begin{aligned}x \in (\exists r.C^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}} &\iff \exists y \in C^{\mathcal{I}\mathcal{I}^d\mathcal{I}} \colon (x,y) \in r^{\mathcal{I}} \\ &\iff \exists y \in C^{\mathcal{I}} \colon (x,y) \in r^{\mathcal{I}} \\ &\iff x \in (\exists r.C)^{\mathcal{I}}\end{aligned}$$

which shows the claim. □

4.2 *Bases of GCIs*

In this section we shall show how we can use model-based most-specific concept descriptions to adapt the argumentation from formal concept analysis to obtain finite bases of finite interpretations.

4.5 Definition (Base)**.** Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a finite interpretation over N_C and N_R, and let $d \in \mathbb{N}$. A *base* of all GCIs of role-depth at most d is a finite set $\mathcal{B}$ of GCIs with role-depth at most d such that

(i) all $(C \sqsubseteq D) \in \mathcal{B}$ are valid in $\mathcal{I}$, i.e., $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, and
(ii) for all GCIs $E \sqsubseteq F$ that are valid in $\mathcal{I}$ and satisfy $\mathrm{rd}(E), \mathrm{rd}(F) \leq d$ it is true that $E \sqsubseteq F$ follows from $\mathcal{B}$, i.e., $\mathcal{B} \models (E \sqsubseteq F)$.

Of course, the set

$$\{\, C \sqsubseteq D \mid C, D \in \mathcal{EL}^{\bot}(N_C, N_R)_d, C^{\mathcal{I}} \subseteq D^{\mathcal{I}} \,\} \tag{4}$$

is a base of $\mathcal{I}$ that is, up to equivalence, even a finite set. Regrettably, this base can be quite large, as the number of concept descriptions grows non-elementary with the role-depth d. More precisely, for role-depth 0 there are $2^{|N_C|}+1$ different $\mathcal{EL}^\bot$-concept descriptions over the signature (N_C, N_R), since such a concept description is either $\bot$ or may contain at most $|N_C|$ concept names as conjuncts. For a role-depth $d>0$ every $\mathcal{EL}^\bot$-concept description may furthermore contain at most $|N_R|\cdot|\mathcal{EL}^\bot(N_C,N_R)_{d-1}|$ existential restrictions as conjuncts, i.e.

$$|\mathcal{EL}^\bot(N_C,N_R)_d| = \big(2^{|N_C|}+1\big)\cdot 2^{|N_R|\cdot|\mathcal{EL}^\bot(N_C,N_R)_{d-1}|}$$

holds. It follows that the number of $\mathcal{EL}^\bot$-concept descriptions with role-depth $\leq d$ is d-exponential in the size of the signature, i.e.,

$$|\mathcal{EL}^\bot(N_C,N_R)_d| = \mathcal{O}\big(2^{2^{\cdots^{|N_C|\cdot|N_R|}}}\big).$$

Therefore, for practical purposes, the base (4) is useless, and finding a smaller base is desirable.

A first idea in this direction is to use the following fact from formal concept analysis and transfer it into the realm of description logics: if $\mathbb{K}$ is a formal context and $A\to B$ is a valid implication of $\mathbb{K}$, then

$$\{\, A\to A''\,\} \models (A\to B).$$

An analogous result also holds in the case of GCIs.

4.6 Lemma. *If $\mathcal{I}\models(C\sqsubseteq D)$, $\mathrm{rd}(C),\mathrm{rd}(D)\leq d$, then $C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d}$ holds in $\mathcal{I}$, and*

$$\{\, C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d}\,\}\models C\sqsubseteq D.$$

Proof. By Lemma 4.3 we know that $C^\mathcal{I} = C^{\mathcal{I}\mathcal{I}^d\mathcal{I}}$, and in particular $C^\mathcal{I}\subseteq (C^{\mathcal{I}\mathcal{I}^d})^\mathcal{I}$, i.e., $C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d}$ holds in $\mathcal{I}$.

Let $\mathcal{J}$ be an interpretation such that $\mathcal{J}\models(C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d})$. Then $C^\mathcal{J}\subseteq (C^{\mathcal{I}\mathcal{I}^d})^\mathcal{J}$, and by (3)

$$C^{\mathcal{J}\mathcal{J}^d}\sqsubseteq C^{\mathcal{I}\mathcal{I}^d}. \tag{5}$$

Since $C\sqsubseteq D$ holds in $\mathcal{I}$, we have $C^\mathcal{I}\subseteq D^\mathcal{I}$ and, using (3) again, $C^{\mathcal{I}\mathcal{I}^d}\sqsubseteq D$. Together with (5) we therefore obtain $C^{\mathcal{J}\mathcal{J}^d}\sqsubseteq D$, and, using (3) once again,

$$C^\mathcal{J}\subseteq D^\mathcal{J},$$

which shows that $\mathcal{J}\models(C\sqsubseteq D)$. Since $\mathcal{J}$ had been chosen arbitrarily, we have shown $(C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d})\models(C\sqsubseteq D)$ as desired. □

From this lemma we easily obtain our first base.

4.7 Corollary. Let $\mathcal{I}$ be a finite interpretation over (N_C,N_R), and let $d\in\mathbb{N}$. Then the set

$$\mathcal{B}_0 := \{\, C\sqsubseteq C^{\mathcal{I}\mathcal{I}^d} \mid C\in\mathcal{EL}^\bot(N_C,N_R), C\neq\bot, \mathrm{rd}(C)\leq d\,\}$$

is sound and complete for $\mathrm{Th}^d(\mathcal{I})$.

Proof. Let $(E \sqsubseteq F) \in \mathrm{Th}^d(\mathcal{I})$. Then $\mathrm{rd}(E), \mathrm{rd}(F) \leq d$, and therefore

$$(E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d}) \in \mathcal{B}_0.$$

By 4.6 we obtain

$$(E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d}) \models (E \sqsubseteq F),$$

and therefore $\mathcal{B}_0 \models (E \sqsubseteq F)$ as required. □

The base $\mathcal{B}_0$ is still too large, as we need to consider all concept descriptions in $\mathcal{EL}^\bot(N_C, N_R)_d$. To further reduce the size of the base we shall make use of a particular choice of concept description we shall show later to be sufficient for our purposes. More precisely, we set

$$M_{\mathcal{I},d} \coloneqq N_C \cup \{ \bot \} \cup \{ \exists r.X^{\mathcal{I}^{d-1}} \mid X \subseteq \Delta^{\mathcal{I}}, X \neq \emptyset \}.$$

Then the first thing we shall show is that every model-based most-specific concept description is *expressible in terms of* $M_{\mathcal{I},d}$, i.e., for each such concept description C there exists $N \subseteq M_{\mathcal{I},d}$ such that $C \equiv \bigsqcap N$, where

$$\bigsqcap N \coloneqq \begin{cases} \bigsqcap_{D \in N} D & \text{if } N \neq \emptyset \\ \top & \text{otherwise.} \end{cases}$$

For the purpose of showing that all model-based most-specific concept descriptions are expressible in terms of $M_{\mathcal{I},d}$, we define for an $\mathcal{EL}^\bot$-concept description $C \neq \bot$ the *lower approximation* $\mathrm{approx}_{\mathcal{I},d}(C)$ of C in $M_{\mathcal{I},d}$ as follows. Let $U \subseteq N_C$ and $\Pi \subseteq N_R \times \mathcal{EL}^\bot(N_C, N_R)$ such that

$$C = \bigsqcap U \sqcap \bigsqcap_{(r,E) \in \Pi} \exists r.E.$$

Then

$$\mathrm{approx}_{\mathcal{I},d}(C) \coloneqq \bigsqcap U \sqcap \bigsqcap_{(r,E) \in \Pi} \exists r.E^{\mathcal{I}\mathcal{I}^{d-1}}.$$

If $C = \bot$, then define $\mathrm{approx}_{\mathcal{I},d}(C) \coloneqq \bot$.

4.8 Proposition. *If* $\mathrm{rd}(C) \leq d$*, then it is true that*

$$C^{\mathcal{I}\mathcal{I}^d} \sqsubseteq \mathrm{approx}_{\mathcal{I},d}(C) \sqsubseteq C.$$

Proof. The claim is clearly true for $C = \bot$. Therefore, assume that

$$C = \bigsqcap U \sqcap \bigsqcap_{(r,E) \in \Pi} \exists r.E.$$

We know by Lemma 4.3 that $E^{\mathcal{I}\mathcal{I}^{d-1}} \sqsubseteq E$ is true for all $(r, E) \in \Pi$, and thus $\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}} \sqsubseteq \exists r.E$. Therefore,

$$\begin{aligned}\operatorname{approx}_{\mathcal{I},d}(C) &= \bigsqcap U \sqcap \bigsqcap_{(r,E)\in\Pi} \exists r.E^{\mathcal{I}\mathcal{I}^{d-1}} \\ &\sqsubseteq \bigsqcap U \sqcap \bigsqcap_{(r,E)\in\Pi} \exists r.E \\ &= C.\end{aligned}$$

Furthermore, it is true that

$$\begin{aligned}C^{\mathcal{I}} &= \Big(\bigsqcap U \sqcap \bigsqcap_{(r,E)\in\Pi} \exists r.E\Big)^{\mathcal{I}} \\ &= \Big(\bigsqcap U \sqcap \bigsqcap_{(r,E)\in\Pi} \exists r.E^{\mathcal{I}\mathcal{I}^{d-1}}\Big)^{\mathcal{I}} \\ &= \big(\operatorname{approx}_{\mathcal{I},d}(C)\big)^{\mathcal{I}}\end{aligned}$$

using Proposition 4.4. In particular, we obtain $C^{\mathcal{I}} \subseteq \big(\operatorname{approx}_{\mathcal{I},d}(C)\big)^{\mathcal{I}}$, and by (3)

$$C^{\mathcal{I}\mathcal{I}^d} \sqsubseteq \operatorname{approx}_{\mathcal{I},d}(C),$$

as desired. □

4.9 Lemma. *For every $X \subseteq \Delta^{\mathcal{I}}$ the concept description $X^{\mathcal{I}^d}$ is expressible in terms of $M_{\mathcal{I},d}$.*

Proof. By Proposition 4.8 we have

$$(X^{\mathcal{I}^d})^{\mathcal{I}\mathcal{I}^d} \sqsubseteq \operatorname{approx}_{\mathcal{I},d}(X^{\mathcal{I}^d}) \sqsubseteq X^{\mathcal{I}^d}.$$

By Lemma 4.3, $X^{\mathcal{I}^d\mathcal{I}\mathcal{I}^d} \equiv X^{\mathcal{I}^d}$, and therefore the previous statement specializes to

$$X^{\mathcal{I}^d\mathcal{I}\mathcal{I}^d} \equiv \operatorname{approx}_{\mathcal{I},d}(X^{\mathcal{I}^d}) \equiv X^{\mathcal{I}^d},$$

and since $\operatorname{approx}_{\mathcal{I},d}(X^{\mathcal{I}^d})$ is expressible in terms of $M_{\mathcal{I},d}$, so is $X^{\mathcal{I}^d}$. □

We are now ready to describe a finite base of $\mathcal{I}$ which is "only" exponential in the size of $M_{\mathcal{I},d}$, which in turn can be exponential in the size of $\mathcal{I}$. Compared to the base in (4) this is still a huge improvement. However, we shall see later in Section 4.4 a base of $\mathcal{I}$ that has even minimal cardinality.

4.10 Theorem. *Let $\mathcal{I}$ be a finite interpretation over the signature (N_C, N_R), and let $d \in \mathbb{N}$. Then the set*

$$\mathcal{B}_2 := \{\, \bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}^d} \mid U \subseteq M_{\mathcal{I},d} \,\}$$

is a finite base of $\mathcal{I}$ w.r.t. role-depth d.

Proof. Since $\mathcal{B}_2 \subseteq \mathcal{B}_0$, and $\mathcal{B}_0$ is sound for $\mathcal{I}$, so is $\mathcal{B}_2$. Furthermore, $\mathcal{B}_2$ is finite because $M_{\mathcal{I},d}$ is finite.

By Lemma 4.6, to show that $\mathcal{B}_2$ is complete for $\mathrm{Th}^d(\mathcal{I})$ it is enough to show for all $\mathcal{EL}^\bot$-concept descriptions D over N_C and N_R with $\mathrm{rd}(D) \leq d$ that

$$\mathcal{B}_2 \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d}).$$

We shall show this claim by induction over the structure of D.

Base Case: $D = \bot$ or $D = A \in N_C$. If $D = \bot$, then $D = \bigsqcap \emptyset$, and since $\emptyset \subseteq M_{\mathcal{I},d}$, it is true that $(D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d}) \in \mathcal{B}_2$. In particular, $\mathcal{B}_2 \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d})$. If $D = A$, then $D = \bigsqcap \{ A \}$, and since $A \in M_{\mathcal{I},d}$, $\{ A \} \subseteq M_{\mathcal{I},d}$, and again $(D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d}) \in \mathcal{B}_2$ and thus $\mathcal{B}_2 \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d})$.

Step Case $D = E \sqcap F$. Let $\mathcal{J}$ be an interpretation such that $\mathcal{J} \models \mathcal{B}_2$. Then

$$D^\mathcal{J} = (E \sqcap F)^\mathcal{J} = E^\mathcal{J} \cap F^\mathcal{J}.$$

By induction hypothesis, $\mathcal{B}_2 \models (E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d})$ and $\mathcal{B}_2 \models (F \sqsubseteq F^{\mathcal{I}\mathcal{I}^d})$, and thus $E^\mathcal{J} \subseteq (E^{\mathcal{I}\mathcal{I}^d})^\mathcal{J}, F^\mathcal{J} \subseteq (F^{\mathcal{I}\mathcal{I}^d})^\mathcal{J}$. Therefore

$$\begin{aligned} D^\mathcal{J} &= E^\mathcal{J} \cap F^\mathcal{J} \\ &\subseteq (E^{\mathcal{I}\mathcal{I}^d})^\mathcal{J} \cap (F^{\mathcal{I}\mathcal{I}^d})^\mathcal{J} \\ &= (E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d})^\mathcal{J}. \end{aligned}$$

By Lemma 4.9, $E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d}$ is expressible in terms of $M_{\mathcal{I},d}$. Therefore, $\mathcal{B}_2$ contains $(E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d}) \sqsubseteq (E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}\mathcal{I}^d}$ up to equivalence. Since $\mathcal{J}$ is a model of $\mathcal{B}_2$ it follows

$$\begin{aligned} (E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d})^\mathcal{J} &\subseteq (E^{\mathcal{I}\mathcal{I}^d} \sqcap F^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}\mathcal{I}^d\mathcal{J}} \\ &= (E \sqcap F)^{\mathcal{I}\mathcal{I}^d\mathcal{J}} \\ &= D^{\mathcal{I}\mathcal{I}^d\mathcal{J}}, \end{aligned}$$

using Proposition 4.4. Therefore, $D^\mathcal{J} \subseteq D^{\mathcal{I}\mathcal{I}^d\mathcal{J}}$, and thus $\mathcal{J} \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d})$. Since $\mathcal{J}$ had been chosen arbitrarily we obtain $\mathcal{B}_2 \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^d})$.

Step Case $D = \exists r.E$. Again, let $\mathcal{J}$ be an interpretation such that $\mathcal{J} \models \mathcal{B}_2$. By the definition of the semantics of existential restrictions, we have

$$\begin{aligned} x \in D^\mathcal{J} &\iff x \in (\exists r.E)^\mathcal{J} \\ &\iff \exists y \in E^\mathcal{J} : (x, y) \in r^\mathcal{J}. \end{aligned}$$

By induction hypothesis, $\mathcal{B}_2 \models (E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d})$ holds. Of course we have $E^{\mathcal{I}\mathcal{I}^d} \sqsubseteq E^{\mathcal{I}\mathcal{I}^{d-1}}$, and thus we furthermore obtain

$$\begin{aligned} x \in D^\mathcal{J} &\implies \exists y \in E^{\mathcal{I}\mathcal{I}^{d-1}\mathcal{J}} : (x, y) \in r^\mathcal{J} \\ &\iff x \in (\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}})^\mathcal{J}. \end{aligned}$$

Since $(\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}}) \in M_{\mathcal{I},d}$, it is true that $(\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}} \sqsubseteq (\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}})^{\mathcal{I}\mathcal{I}^{d-1}}) \in \mathcal{B}_2$, so

$$\begin{aligned} (\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}})^\mathcal{J} &\subseteq (\exists r.E^{\mathcal{I}\mathcal{I}^{d-1}})^{\mathcal{I}\mathcal{I}^{d-1}\mathcal{J}} \\ &= (\exists r.E)^{\mathcal{I}\mathcal{I}^{d-1}\mathcal{J}} \end{aligned}$$

$$= D^{\mathcal{I}\mathcal{I}^{d-1}\mathcal{J}}$$

using Proposition 4.4. Putting it all together we obtain $D^{\mathcal{J}} \subseteq D^{\mathcal{I}\mathcal{I}^{d}\mathcal{J}}$, and since $\mathcal{J}$ had been chosen arbitrarily, we have shown $\mathcal{B}_2 \models (D \sqsubseteq D^{\mathcal{I}\mathcal{I}^{d}})$. This completes the proof of the induction step, and thus the proof of this theorem. □

For the computation of the set $M_{\mathcal{I},d}$ we in particular need to compute all model-based most-specific concept descriptions of role-depth $d-1$. A naive approach would be to simply compute those role-depth-bounded model-based most-specific concept descriptions for all subsets of the interpretation's domain. A faster solution uses the NextClosure algorithm [18], which is able to compute closures of a closure operator c on a set M. This algorithm is applicable in our setting as the mapping

$$X \mapsto X^{\mathcal{I}^{d}\mathcal{I}}$$

is a closure operator by Lemma 4.3. We shall not discuss the details of the NextClosure algorithm here, and refer the interested reader to the given literature.

4.3 *Computation of Model-Based Most-Specific Concept Descriptions*

In Section 4.1 we have introduced the notion of model-based most-specific concept descriptions in an abstract way, and have shown that it can be used to obtain finite bases of finite interpretations. Back then we have not discussed how to actually compute model-based most-specific concept descriptions, and hence the results of Section 4.2 were rather ineffective.

The purpose of this section is to remedy this ineffectiveness by providing methods to compute model-based most-specific concept descriptions. To this end we shall use the notions of *description graphs* and *least common subsumers* as they have been used in [2, 3, 8]. We shall see how we can combine these notions to obtain an effective algorithm to compute model-based most specific concept descriptions. The argumentation of this section follows the corresponding argumentation in [15].

We start by introducing description graphs. These graphs provide a representation of both $\mathcal{EL}$-concept descriptions and interpretations by means of directed, edge- and vertex-labeled graphs. As we shall see later we then can use the structure of these graphs to decide subsumption of concept descriptions as well as the question whether an element belongs to the extension of a concept description in a given finite interpretation.

4.11 Definition (Description Graphs)**.** An *$\mathcal{EL}$-description graph* over N_C and N_R is a tuple $\mathcal{G} = (V, E, L, v)$ consisting of a set V, a set $E \subseteq V \times N_R \times V$, a function $L\colon V \to \mathfrak{P}(N_C)$, and some $v \in V$. V is called the set of *vertices* of $\mathcal{G}$, E is called the set of (*labeled*) *edges* of $\mathcal{G}$, L is called the *labeling function* of $\mathcal{G}$, and v is called the *root* of $\mathcal{G}$.

Let C be an $\mathcal{EL}$-concept description over N_C and N_R. Then the *description graph* $\mathcal{G}_C = (V_C, E_C, L_C, v_C)$ of C is inductively defined as follows. Let

$$C = P_1 \sqcap \cdots \sqcap P_k \sqcap \exists r_1.D_1 \sqcap \cdots \sqcap \exists r_\ell.D_\ell,$$

where $\{ P_1, \ldots, P_k \} \subseteq N_C$, $\{ r_1, \ldots, r_\ell \} \subseteq N_R$ and $D_1, \ldots, D_\ell$ are $\mathcal{EL}$-concept descriptions over N_C and N_R. Assume inductively that $\mathcal{G}_{D_i} = (V_{D_i}, E_{D_i}, L_{D_i}, v_{D_i})$ are the $\mathcal{EL}$-description graphs of D_i, where without loss of generality all the V_{D_i} are disjoint. Let

v_C be some element not in any V_{D_i}. Then the description graph of C is defined via

$$V_C \coloneqq \{\, v_C \,\} \cup \bigcup_{i=1}^{\ell} V_{D_i},$$
$$E_C \coloneqq \big\{\, (v_C, r_i, v_{D_i}) \;\big|\; 1 \leq i \leq \ell \,\big\} \cup \bigcup_{i=1}^{\ell} E_i,$$
$$L_C \coloneqq \big\{\, (v_C, \{\, P_1, \ldots, P_k \,\}) \,\big\} \cup \bigcup_{i=1}^{\ell} L_{D_i}.$$

Let $\mathcal{I}$ be an interpretation over N_C and N_R, and let $x \in \Delta^{\mathcal{I}}$. Then the *description graph* $\mathcal{G}_{\mathcal{I}}^{x}$ of $\mathcal{I}$ rooted at x is defined as $\mathcal{G}_{\mathcal{I}}^{x} \coloneqq (V_{\mathcal{I}}, E_{\mathcal{I}}, L_{\mathcal{I}}, x)$, where

$$V_{\mathcal{I}} \coloneqq \Delta^{\mathcal{I}},$$
$$E_{\mathcal{I}} \coloneqq \big\{\, (x, r, y) \;\big|\; (x, y) \in r^{\mathcal{I}}, r \in N_R \,\big\},$$
$$L_{\mathcal{I}}(x) \coloneqq \{\, A \in N_C \mid x \in A^{\mathcal{I}} \,\} \quad (x \in \Delta^{\mathcal{I}}).$$

If a description graph $\mathcal{G}$ is a directed tree with root v, then we say that $\mathcal{G}$ is a *description tree*.

It is quite easy to see that a description tree $\mathcal{G} = (V, E, L, v)$ corresponds canonically to an $\mathcal{EL}$-concept description. For this denote for $w \in V$ with $\mathcal{G}(w)$ the directed subtree of $\mathcal{G}$ with root w. In other words, $\mathcal{G}(w) = (W, F, H, w)$ contains all vertices W from V which are reachable in $\mathcal{G}$ via a directed path that starts in w, and F and H arise from the restriction of E and L to W, respectively.

Let $(v, r_1, w_1), \ldots, (v, r_\ell, w_\ell)$ be all edges from E originating at v. Assuming inductively that the $\mathcal{EL}$-concept descriptions $C_{\mathcal{G}(w_1)}, \ldots, C_{\mathcal{G}(w_\ell)}$ correspond to the description graphs $\mathcal{G}(w_1), \ldots, \mathcal{G}(w_\ell)$, we define

$$C_{\mathcal{G}} \coloneqq P_1 \sqcap \ldots P_k \sqcap \exists r_1.C_{\mathcal{G}(w_1)} \sqcap \ldots \exists r_\ell.C_{\mathcal{G}(w_\ell)},$$

where $L(v) = \{\, P_1, \ldots, P_k \,\}$. With this definition we have for all $\mathcal{EL}$-concept descriptions C

$$C \equiv C_{\mathcal{G}_C},$$

which is why we can say the description graph $\mathcal{G}$ corresponds canonically to the $\mathcal{EL}$-concept description $C_{\mathcal{G}}$. Note that $\mathcal{G} = \mathcal{G}_{C_{\mathcal{G}}}$ holds for all $\mathcal{EL}$-description graphs $\mathcal{G}$.

Analogously, there is a one-to-one correspondence between interpretations and $\mathcal{EL}$-description graphs. Assume that $\mathcal{G} = (V, E, L, v)$ is an $\mathcal{EL}$-description graph over N_C and N_R. Then the interpretation $\mathcal{I}_{\mathcal{G}}$ (over N_C and N_R) is defined as follows:

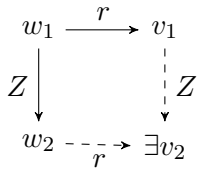
$$\Delta^{\mathcal{I}_{\mathcal{G}}} \coloneqq V,$$
$$A^{\mathcal{I}_{\mathcal{G}}} \coloneqq \{\, v \in V \mid A \in L(v) \,\} \quad (A \in N_C),$$
$$r^{\mathcal{I}_{\mathcal{G}}} \coloneqq \{\, (v, w) \in V \times V \mid (v, r, w) \in E \,\} \quad (r \in N_R).$$

It can be readily verified that $\mathcal{I} = \mathcal{I}_{\mathcal{G}_{\mathcal{I}}^{x}}$ holds for all interpretations $\mathcal{I}$ where $x \in \Delta^{\mathcal{I}}$ is an arbitrary individual, and $\mathcal{G} = \mathcal{G}_{\mathcal{I}_{\mathcal{G}}}^{v}$ holds for all $\mathcal{EL}$-description graphs $\mathcal{G} = (V, E, L, v)$.

As already mentioned above, description graphs can be used to decide the reasoning tasks of subsumption and elementhood. To achieve this we shall introduce the notion of *simulations* between description graphs [3, 17]. Later on we shall see that we can replace the use of simulations by the easier notion of *homomorphisms* between description trees.

4.12 Definition (Simulation)**.** Let $\mathcal{G}_1 = (V_1, E_1, L_1, v_1)$ and $\mathcal{G}_2 = (V_2, E_2, L_2, v_2)$ be two $\mathcal{EL}$-description graphs. A binary relation $Z \subseteq V_1 \times V_2$ is a *simulation* from $\mathcal{G}_1$ to $\mathcal{G}_2$, written $Z\colon \mathcal{G}_1 \rightrightarrows \mathcal{G}_2$, if and only if the following conditions are satisfied:

(S1) $(v_1, v_2) \in Z$,
(S2) $(w_1, w_2) \in Z$ implies $L_1(w_1) \subseteq L_2(w_2)$, and
(S3) whenever $(w_1, w_2) \in Z$ and $(w_1, r, w_1') \in E_1$, then there exists $w_2' \in V_2$ such that $(w_2, r, w_2') \in E_2$ and $(w_1', w_2') \in Z$.



It can be easily verified that the class of simulations is closed under composition, i.e., if $Z_1\colon \mathcal{G}_1 \rightrightarrows \mathcal{G}_2$ and $Z_2\colon \mathcal{G}_2 \rightrightarrows \mathcal{G}_3$ are simulations, then the product

$$Z_1 \circ Z_2 \coloneqq \{\, (w_1, w_3) \mid \exists w_2 \in V_2\colon (w_1, w_2) \in Z_1, (w_2, w_3) \in Z_2 \,\}$$

is a simulation from $\mathcal{G}_1$ to $\mathcal{G}_3$.

The statement and the proof of the following proposition are a special case of [17, Proposition 18], adapted to the needs of this paper.

4.13 Proposition. *Let $\mathcal{I}$ be an interpretation over N_C and N_R, let C be an $\mathcal{EL}$-concept description over N_C and N_R, and let $\mathcal{G}_C = (V_C, E_C, L_C, v_C)$ be the $\mathcal{EL}$-description graph of C. Then for every $x \in \Delta^{\mathcal{I}}$ the following statements are equivalent:*

(i) $x \in C^{\mathcal{I}}$,
(ii) there exists a simulation $Z\colon \mathcal{G}_C \rightrightarrows \mathcal{G}_{\mathcal{I}}^x$.

Proof. (i) $\implies$ (ii). Suppose $x \in C^{\mathcal{I}}$. Define

$$Z = \{\, (v, y) \in V_C \times \Delta^{\mathcal{I}} \mid y \in (C_{\mathcal{G}_C(v)})^{\mathcal{I}} \,\}.$$

We show that Z is a simulation. Since $x \in C^{\mathcal{I}}$ we have $(v_C, x) \in Z$. Let $(v, y) \in Z$, and let

$$C_{\mathcal{G}_C(v)} = P_1 \sqcap \cdots \sqcap P_k \sqcap \exists r_1.D_1 \sqcap \cdots \sqcap \exists r_\ell.D_\ell.$$

(S2) It is true that $L_{\mathcal{G}_C}(v) = \{\, P_1, \ldots, P_k \,\}$. Since $y \in (C_{\mathcal{G}_C(v)})^{\mathcal{I}}$ we have $y \in P_i^{\mathcal{I}}$ for $1 \leq i \leq k$. Therefore, $\{\, P_1, \ldots, P_k \,\} \subseteq L_{\mathcal{G}_{\mathcal{I}}}(y)$.
(S3) Let $(v, r, v') \in E_{\mathcal{G}_C}$. Then $r = r_i$ for some $1 \leq i \leq \ell$, and $D_i = C_{\mathcal{G}_C(v')}$. Since $y \in (C_{\mathcal{G}_C(v)})^{\mathcal{I}}$ it is true that $y \in (\exists r_i.D_i)^{\mathcal{I}}$. Therefore, there exists $y_i \in \Delta^{\mathcal{I}}$ such that $(y, y_i) \in r^{\mathcal{I}}$ and $y_i \in D_i^{\mathcal{I}}$. But then $(y, r_i, y_i) \in E_{\mathcal{G}_{\mathcal{I}}}$ and $(v', y_i) \in Z$, as required.

(ii) $\implies$ (i). Let $Z\colon \mathcal{G}_C \rightrightarrows \mathcal{G}_{\mathcal{I}}^x$ be a simulation. For $v \in V_C$ denote with $\mathrm{h}(v)$ the maximal length of a path from v to some leaf in $\mathcal{G}_C$. We show by induction over $\mathrm{h}(v)$ that

$$(v, y) \in Z \implies y \in (C_{\mathcal{G}_C(v)})^{\mathcal{I}}. \tag{6}$$

Since $(v_C, x) \in Z$, we obtain from this that $x \in (C_{\mathcal{G}_C(v_C)})^{\mathcal{I}} = C^{\mathcal{I}}$ as desired.

Let $(v, y) \in Z$ and assume that (6) holds for each $w \in V_C$ with $\mathrm{h}(w) < \mathrm{h}(v)$. Again let

$$C_{\mathcal{G}_C(v)} = P_1 \sqcap \cdots \sqcap P_k \sqcap \exists r_1.D_1 \sqcap \cdots \sqcap \exists r_\ell.D_\ell.$$

Since Z is a simulation, we have $\{ P_1, \ldots, P_k \} = L_{\mathcal{G}_C}(v) \subseteq L_{\mathcal{G}_\mathcal{I}^x}(y)$, and therefore

$$y \in (P_1 \sqcap \cdots \sqcap P_k)^\mathcal{I}.$$

Now let $\{ (v, r_i, v_i) \mid 1 \leq i \leq \ell \} \subseteq E_C$ be all outgoing edges of v in $\mathcal{G}_C$. Then $D_i = C_{\mathcal{G}_C(v_i)}$ for each $i \in \{ 1, \ldots, \ell \}$. Since Z is a simulation, for each v_i there exists y_i such that $(v_i, y_i) \in Z$ and $(y, r_i, y_i) \in E_{\mathcal{G}_\mathcal{I}^x}$. Since $\mathrm{h}(v_i) < \mathrm{h}(v)$, the induction hypothesis yields $y_i \in (C_{\mathcal{G}_C(v_i)})^\mathcal{I} = D_i^\mathcal{I}$ for each $i \in \{ 1, \ldots, \ell \}$. Moreover, since $(y, r_i, y_i) \in E_{\mathcal{G}_\mathcal{I}^x}$, it is true that $(y, y_i) \in r_i^\mathcal{I}$ and thus $y \in (\exists r_i.D_i)^\mathcal{I}$ for each $i \in \{ 1, \ldots, \ell \}$. All in all we obtain

$$y \in (\exists r_1.D_1 \sqcap \cdots \sqcap \exists r_\ell.D_\ell)^\mathcal{I}$$

and thus $y \in (C_{\mathcal{G}_C(v)})^\mathcal{I}$ as required. □

4.14 Definition (Homomorphism)**.** Let $\mathcal{G}_1 = (V_1, E_1, L_1, v_1)$ and $\mathcal{G}_2 = (V_2, E_2, L_2, v_2)$ be two $\mathcal{EL}$-description graphs. A mapping $\varphi \colon \mathcal{G}_1 \to \mathcal{G}_2$ is called a *homomorphism* from $\mathcal{G}_1$ to $\mathcal{G}_2$ if and only if the following conditions are satisfied:

(i) $\varphi(v_1) = v_2$,
(ii) $L_1(v) \subseteq L_2(\varphi(v))$ for all $v \in V_1$, and
(iii) $(\varphi(v), r, \varphi(w)) \in E_2$ for all $(v, r, w) \in E_1$.

In analogy to simulations it is true that the class of homomorphisms is closed under composition, i.e., whenever φ is a homomorphism to $\mathcal{G}$ and ψ is a homomorphism from $\mathcal{G}$, then $\psi \circ \varphi$ is a homomorphism as well.

The following proposition relates the existence of simulations to the existence of homomorphisms.[1]

4.15 Proposition. *Let $\mathcal{G}_1 = (V_1, E_1, L_1, v_1)$ be a description tree, and let $\mathcal{G}_2 = (V_2, E_2, L_2, v_2)$ be a description graph. Then there exists a simulation $Z \colon \mathcal{G}_1 \precsim \mathcal{G}_2$ if and only if there exists a homomorphism $\varphi \colon \mathcal{G}_1 \to \mathcal{G}_2$.*

Proof. If there exists a homomorphism $\varphi \colon \mathcal{G}_1 \to \mathcal{G}_2$, then

$$Z \coloneqq \{ (x, \varphi(x)) \mid x \in V_1 \}$$

is clearly a simulation $Z \colon \mathcal{G}_1 \precsim \mathcal{G}_2$.

Conversely, if $Z \colon \mathcal{G}_1 \precsim \mathcal{G}_2$ is a simulation, then we can inductively define a homomorphism $\varphi \colon \mathcal{G}_1 \to \mathcal{G}_2$ with $\varphi \subseteq Z$ as follows. Set $\varphi(v_1) \coloneqq v_2$, and suppose inductively that φ has already been defined for all nodes with depth in $\mathcal{G}_1$ of at most n. Let v' be a node of depth $n + 1$. Then there exists $v \in V_1$ and $r \in N_R$ such that $(v, r, v') \in E_1$. Since v has depth n, $w = \varphi(v)$ is already defined and $(v, w) \in Z$. Since Z is a simulation, there exists $w' \in V_2$ such that $(v', w') \in Z$ and $(w, r, w') \in E_2$. Set $\varphi(v') \coloneqq w'$. Then $\varphi \colon \mathcal{G}_1 \to \mathcal{G}_2$ is a homomorphism by construction. □

Checking for the existence of a simulation between two given finite description graphs can be done in polynomial time [22], and thus the proposition yields that the existence

[1] The authors thank Prof. Baader for hinting at this simple yet helpful connection.

of a homomorphism from a finite description tree to a finite description graph can be decided in polynomial time as well.

Since $\mathcal{EL}$-description graphs of $\mathcal{EL}$-concept descriptions are always trees we immediately obtain the following result.

4.16 Corollary. Let $\mathcal{I}$ be an interpretation over N_C and N_R, C be an $\mathcal{EL}$-concept description over N_C and N_R, and $x \in \Delta^{\mathcal{I}}$. Then $x \in C^{\mathcal{I}}$ if and only if there exists a homomorphism $\varphi\colon \mathcal{G}_C \to \mathcal{G}_{\mathcal{I}}^x$.

The following characterization of subsumption by means of homomorphisms was introduced in [8]. The proof of this statement is an adaption of the proof of [17, Theorem 19].

4.17 Proposition. *Let C, D be two $\mathcal{EL}$-concept descriptions. Then $C \sqsubseteq D$ if and only if there exists a homomorphism $\varphi\colon \mathcal{G}_D \to \mathcal{G}_C$.*

Proof. First assume that there is a homomorphism $\varphi\colon \mathcal{G}_D \to \mathcal{G}_C$. Let $\mathcal{I}$ be an interpretation and let $x \in C^{\mathcal{I}}$. We need to show that $x \in D^{\mathcal{I}}$. By Corollary 4.16, $x \in C^{\mathcal{I}}$ implies that there is an homomorphism $\psi\colon \mathcal{G}_C \to \mathcal{G}_{\mathcal{I}}^x$. But then $\psi \circ \varphi\colon \mathcal{G}_D \to \mathcal{G}_{\mathcal{I}}^x$ is a homomorphism, and therefore $x \in D^{\mathcal{I}}$, again by Corollary 4.16.

Conversely assume $C \sqsubseteq D$. Consider the description graph $\mathcal{G}_C$ as an interpretation $\mathcal{I}_{\mathcal{G}_C}$. Then since $\mathrm{id}\colon \mathcal{G}_C \to \mathcal{G}_C$ is a homomorphism, Corollary 4.16 yields $v_C \in C^{\mathcal{I}_{\mathcal{G}_C}}$, where v_C is again the root of $\mathcal{G}_C$. But then $v_C \in D^{\mathcal{I}_{\mathcal{G}_C}}$, and therefore there exists a homomorphism $\varphi\colon \mathcal{G}_D \to \mathcal{G}_{\mathcal{I}_{\mathcal{G}_C}}$. Since $\mathcal{G}_{\mathcal{I}_{\mathcal{G}_C}}$ is the same as $\mathcal{G}_C$, $\varphi\colon \mathcal{G}_D \to \mathcal{G}_C$ is a homomorphism as required. □

Model-based most-specific concept descriptions can be obtained from the description graphs of the underlying finite interpretation by *unravelling* this interpretation starting from a certain individual. The following definition makes precise what an unravelling of a description graph is, and the following theorem then shows how this unravelling can be used to compute model-based most-specific concept descriptions of singleton sets.

For the definition of unravellings we first introduce the notion of a *path* w in a description graph $\mathcal{G} = (V, E, L, x)$. This is a sequence $w = v_0 r_1 v_1 r_2 \dots r_n v_n$, where $v_0, \dots, v_n \in V$, $r_1, \dots, r_n \in N_R$ and $(v_{i-1}, r_i, v_i) \in E$ for all $i \in \{1, \dots, n\}$. We say that w has *length* n, that w *starts* at v_0, and denote the last element v_n by $\delta(w)$.

4.18 Definition (Unravelling)**.** Let $\mathcal{G} = (V, E, L, x)$ be a description graph, and let $d \in \mathbb{N}$. Then the *unravelling* of $\mathcal{G}$ up to depth d is the description graph $\mathcal{G}\restriction_d = (V_d, E_d, L_d, x)$ defined as follows. The set V_d is the set of all paths in $\mathcal{G}$ starting at x having length at most d. The set E_d is defined as

$$E_d \coloneqq \{\, (w, r, wrv) \mid w \in V_d, r \in N_R, v \in V, wrv \in V_d \,\},$$

i.e., two paths are connected in $\mathcal{G}\restriction_d$ via an r-edge if and only if the second arises from the first by appending an r-edge from $\mathcal{G}$. Finally, L_d is defined via $L_d(w) = L(\delta(w))$.

4.19 Theorem. *Let $\mathcal{I}$ be an interpretation, $d \in \mathbb{N}$ and $x \in \Delta^{\mathcal{I}}$. Then $C_{\mathcal{G}_{\mathcal{I}}^x\restriction_d}$ is the model-based most-specific concept description of depth d of $\{x\}$ in $\mathcal{I}$ (up to equivalence).*

Proof. Let $C \coloneqq C_{\mathcal{G}_{\mathcal{I}}^x\restriction_d}$. Obviously, C has a role-depth of at most d. Furthermore, we have to show two claims:

(i) $x \in C^{\mathcal{I}}$, and
(ii) for each $\mathcal{EL}$-concept description D with $\mathrm{rd}(D) \leq d$ and $x \in D^{\mathcal{I}}$, it is true that $C \sqsubseteq D$.

For the first claim let $\mathcal{G}_C = (V_C, E_C, L_C, v_C)$ and $\mathcal{G}_\mathcal{I}^x = (V_\mathcal{I}, E_\mathcal{I}, L_\mathcal{I}, x)$. Then $\mathcal{G}_C$ is canonically isomorphic to $\mathcal{G}_\mathcal{I}^x\restriction_d$, thus we assume that they are the same. The function δ that maps a path to its last vertex is clearly a homomorphism from $\mathcal{G}_C$ to $\mathcal{G}_\mathcal{I}^x$. By Proposition 4.13 we obtain $x \in C^\mathcal{I}$ as required.

For the second claim let D be an $\mathcal{EL}$-concept description such that $x \in D^\mathcal{I}$ and $\mathrm{rd}(D) \leq d$. By Proposition 4.13, there exists a homomorphism $\varphi\colon \mathcal{G}_D \to \mathcal{G}_\mathcal{I}^x$. Then we define the mapping $\hat{\varphi}\colon \mathcal{G}_D \to \mathcal{G}_\mathcal{I}^x\restriction_d$ as follows: let $v \in V_\mathcal{I}$ and let $v_C r_1 v_1 r_2 \dots r_n v$ be the unique path in $\mathcal{G}_D$ from v_D to v. Note that $n \leq d$ since $\mathrm{rd}(D) \leq d$. We set

$$\hat{\varphi}(v) \coloneqq \varphi(v_D) r_1 \varphi(v_1) r_2 \dots r_n \varphi(v).$$

It is easily seen that $\hat{\varphi}$ is a homomorphism. Since $\mathcal{G}_\mathcal{I}^x\restriction_d$ is the description graph of C, Proposition 4.17 yields $C \sqsubseteq D$. □

Computing model-based most-specific concept descriptions for arbitrary, non-empty sets $X \subseteq \Delta^\mathcal{I}$ of individuals is achieved by computing the *least common subsumer* of the model-based most-specific concept descriptions of all $\{ x \}, x \in X$.

4.20 Definition (Least Common Subsumer)**.** Let $C_1, \dots, C_n$ be $\mathcal{EL}$-concept descriptions. Then an $\mathcal{EL}$-concept description C is a *least common subsumer* of $C_1, \dots, C_n$ (in $\mathcal{EL}$) if the following conditions are satisfied:

(i) $C_i \sqsubseteq C$ for all $i \in \{ 1, \dots, n \}$, and
(ii) every $\mathcal{EL}$-concept description D satisfying $C_i \sqsubseteq D$ for all $i \in \{ 1, \dots, n \}$ also satisfies $C \sqsubseteq D$.

We write $C = \mathrm{lcs}\{ C_1, \dots, C_n \}$ if C is the least common subsumer of $C_1, \dots, C_n$.

Note that the least common subsumer is unique up to equivalence, so using the notation $\mathrm{lcs}\{ C_1, \dots, C_n \}$ does not cause any problems.

It can be shown that least common subsumers always exist in $\mathcal{EL}$, and that they can effectively be computed by means of *products* of description trees. For the following definition recall for a description tree $\mathcal{G} = (V, E, L, v)$ and $v' \in V$ that $\mathcal{G}(v')$ denotes the subtree of $\mathcal{G}$ with root v'.

4.21 Definition (Product of Description Trees)**.** Let $\mathcal{G}_1 = (V_1, E_1, L_1, v_1), \mathcal{G}_2 = (V_2, E_2, L_2, v_2)$ be two $\mathcal{EL}$-description trees. Then the *product* $\mathcal{G}_1 \times \mathcal{G}_2 = (V, E, L, (v_1, v_2))$ is a description tree which is inductively defined as follows. The root of $\mathcal{G}_1 \times \mathcal{G}_2$ is the pair (v_1, v_2), which is labeled via L by $L_1(v_1) \cap L_2(v_2)$. Then for each $r \in N_R$, $(v_1, r, v_1') \in E_1$ and $(v_2, r, v_2') \in E_2$, it is true that $((v_1, v_2), r, (v_1', v_2')) \in E$, and $(\mathcal{G}_1 \times \mathcal{G}_2)(v_1', v_2') = \mathcal{G}_1(v_1') \times \mathcal{G}_2(v_2')$.

4.22 Theorem (Theorem 2 of [8])**.** *Let C, D be two $\mathcal{EL}$-concept descriptions, and $\mathcal{G}_C, \mathcal{G}_D$ their $\mathcal{EL}$-description trees. Then $C_{\mathcal{G}_C \times \mathcal{G}_D}$ is the least common subsumer of C and D.*

The definition of the product of description trees can be extended to an arbitrary number of description trees in the obvious way. In analogy to Theorem 4.22, it can be proven that $C_{\prod_{i=1}^n \mathcal{G}_{C_i}}$ is the least common subsumer of the $\mathcal{EL}$-concept descriptions $C_1, \dots, C_n$.

4.23 Corollary. Let $X \subseteq \Delta^\mathcal{I}, X \neq \emptyset$, and let $d \in \mathbb{N}$. Then

$$X^{\mathcal{I}^d} \equiv \mathrm{lcs}\{ \{ x \}^{\mathcal{I}^d} \mid x \in X \}.$$

Proof. Let $C \coloneqq \mathrm{lcs}\big\{ \{ x \}^{\mathcal{I}^d} \;\big|\; x \in X \big\}$. By Theorem 4.22 we know that C exists and that

$$C \equiv C_{\mathcal{G}_X},$$

where $\mathcal{G}_X := \prod_{x \in X} \mathcal{G}_{\{x\}^{\mathcal{I}^d}}$ is the product of all description graphs of the concept descriptions $\{x\}^{\mathcal{I}^d}, x \in X$. To show that $C \equiv X^{\mathcal{I}^d}$, we need to show that

(i) $X \subseteq C^{\mathcal{I}}$,
(ii) $\mathrm{rd}(C) \leq d$, and
(iii) for all $\mathcal{EL}$-concept descriptions D with $\mathrm{rd}(D) \leq d$ and $X \subseteq D^{\mathcal{I}}$, it is true that $C \sqsubseteq D$.

By definition, $\{x\}^{\mathcal{I}^d} \sqsubseteq C$, and by Lemma 4.2 it is true that $x \in C^{\mathcal{I}}$ for all $x \in X$. Therefore, $X \subseteq C^{\mathcal{I}}$, which shows the first claim.

The second claim is also immediately clear: the description graphs of all $\{x\}^{\mathcal{I}^d}, x \in X$ have depth at most d, and thus the product $\mathcal{G}_X$ of these description graphs has also depth at most d. Thus, $\mathrm{rd}(C) \leq d$.

For the last claim let D as described. Then since $X \subseteq D^{\mathcal{I}}$, it is true that $\{x\} \subseteq D^{\mathcal{I}}$ for all $x \in X$. Using Lemma 4.2 again we obtain $\{x\}^{\mathcal{I}^d} \sqsubseteq D$. By definition of the least common subsumer

$$C = \mathrm{lcs}\{\{x\}^{\mathcal{I}^d} \mid x \in X\} \sqsubseteq D,$$

as required. □

4.4 *Reducing the Size of the Base*

We have seen in Theorem 4.10 how to obtain a finite base of all GCIs with bounded role-depth. Our motivation for this theorem was to find a smaller base than the set of all possible GCIs with role-depth not exceeding a given bound. However, this theorem does not really satisfy this motivation, as it does not tell anything about whether the size of the base is "small" or not.

We shall remedy this deficit of Section 4.2 by discussing in this section means to reduce the size of the base. Indeed, we shall even show that it is possible to obtain a base of minimal cardinality, again using methods from formal concept analysis.

We shall start by introducing *induced contexts*, which provide a means to associate a formal context to a given set of concept descriptions and a given interpretation.

4.24 Definition (Induced Context)**.** Let $\mathcal{I}$ be an interpretation over the signature (N_C, N_R) and let $M \subseteq \mathcal{EL}^{\bot}(N_C, N_R)$. Then the *induced context* $\mathbb{K}_{M,\mathcal{I}}$ of M and $\mathcal{I}$ is defined as

$$\mathbb{K}_{M,\mathcal{I}} := (\Delta^{\mathcal{I}}, M, \nabla),$$

where $(d, C) \in \nabla$ iff $d \in C^{\mathcal{I}}$.

Induced contexts allow us to express the similarities between description logics and formal concept analysis in a clear and formal way. The following two statements are contained in [15], and are given here without proof.

The first statement relates one of the derivation operators in the induced context to the extension function in the original interpretation. This statement can be seen as a formalization of our previous remark about the similarities between interpretations and formal contexts.

4.25 Proposition (Lemma 4.10 of [15])**.** *Let M be a set of $\mathcal{EL}^{\bot}$-concept descriptions. Then for each $U \subseteq M$ it is true that*

$$U'_{\mathbb{K}_{M,\mathcal{I}}} = (\bigsqcap U)^{\mathcal{I}}.$$

The next statement can be seen as the dual to the previous one, as it relates the extension function of the given interpretation to the corresponding derivation operator in the induced context. For this we need to introduce another notion. For M being a set of $\mathcal{EL}^\bot$-concept descriptions and C being another $\mathcal{EL}^\bot$-concept description, let us define the *projection* $\mathrm{pr}_M(C)$ of C on M as

$$\mathrm{pr}_M(C) \coloneqq \{\, D \in M \mid C \sqsubseteq D \,\}.$$

Then the following statement holds.

4.26 Proposition (Lemma 4.11 of [15])**.** *Let M be a set of $\mathcal{EL}^\bot$-concept descriptions, and let C be an $\mathcal{EL}^\bot$-concept description that is expressible in terms of M. Then*

$$C^{\mathcal{I}} = (\mathrm{pr}_M(C))'_{\mathbb{K}_{M,\mathcal{I}}}.$$

So far we have only considered one of the derivation operators in the induced context $\mathbb{K}_{M,\mathcal{I}}$. The following assertion is concerned with the other derivation operator, and it should come with no surprise that it relates this operator to model-based most-specific concept descriptions.

4.27 Proposition. *Let M be a set of $\mathcal{EL}^\bot$-concept descriptions of role-depth at most d. Then every set $O \subseteq \Delta^{\mathcal{I}}$ satisfies*

$$\mathrm{pr}_M(O^{\mathcal{I}^d}) = O'_{\mathbb{K}_{M,\mathcal{I}}}.$$

Proof. Let $\mathbb{K}_{M,\mathcal{I}} = (\Delta^{\mathcal{I}}, M, \nabla)$, and consider $D \in M$. Then it is true that

$$\begin{aligned} D \in O' &\iff \forall x \in O \colon x \nabla D \\ &\iff \forall x \in O \colon x \in D^{\mathcal{I}} \\ &\iff O \subseteq D^{\mathcal{I}}. \end{aligned}$$

By Lemma 4.2 we obtain

$$\begin{aligned} O \subseteq D^{\mathcal{I}} &\iff O^{\mathcal{I}^d} \sqsubseteq D \\ &\iff D \in \mathrm{pr}_M(O^{\mathcal{I}_d}). \end{aligned}$$

Thus $O' = \mathrm{pr}_M(O^{\mathcal{I}^d})$ as it has been claimed. □

Putting the previous statements together we immediately obtain the following observation.

4.28 Proposition. *Let M be a set of $\mathcal{EL}^\bot$-concept descriptions with role-depth at most d. Then for every set $U \subseteq M$ it is true that*

$$\mathrm{pr}_M\big((\textstyle\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big) = U''_{\mathbb{K}_{M,\mathcal{I}}}.$$

Proof. Proposition 4.25 yields $(\bigsqcap U)^{\mathcal{I}} = U'$, and thus

$$\mathrm{pr}_M\Big(\big((\textstyle\bigsqcap U)^{\mathcal{I}}\big)^{\mathcal{I}^d}\Big) = \mathrm{pr}_M\big((U')^{\mathcal{I}^d}\big).$$

Then $U' \subseteq \Delta^{\mathcal{I}}$, and thus Proposition 4.27 yields

$$\mathrm{pr}_M\big((U')^{\mathcal{I}^d}\big) = U'',$$

i.e., $\mathrm{pr}_M\big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big) = U''$ as required. □

With the technical apparatus ready we can now come back to our original problem of reducing the size of the base from Theorem 4.10. The main idea behind our further considerations is the following. To obtain "small" bases of the finite interpretation $\mathcal{I}$ we first consider the induced context $\mathbb{K}_{\mathcal{I},d} := \mathbb{K}_{M_{\mathcal{I},d},\mathcal{I}}$, and from this context its canonical base $\mathrm{Can}(\mathbb{K}_{M_{\mathcal{I},d},\mathcal{I}})$. As this base is of minimal cardinality we can hope that by transferring it to a base of GCIs that it will at least be a "small" base of $\mathcal{I}$. Indeed, with some further adaption we shall even be able to show that we can obtain a base of *minimal cardinality* of $\mathcal{I}$.

The adaption we have to make is concerned with the following problem. When we consider the set $M_{\mathcal{I},d}$ of concept descriptions as the set of attributes of $\mathbb{K}_{\mathcal{I},d}$, we lose the ability to automatically detect subsumption relationships between concept descriptions in $M_{\mathcal{I},d}$. More precisely, if $C, D \in M_{\mathcal{I},d}$ such that $C \sqsubseteq D$, then the GCI $C \sqsubseteq D$ is trivial, but the implication $\{\, C \,\} \to \{\, D \,\}$, which is valid in $\mathbb{K}_{\mathcal{I},d}$, is *not necessarily* trivial. Therefore, if we compute the canonical base of $\mathbb{K}_{\mathcal{I},d}$, we will certainly obtain some implications in $\mathrm{Can}(\mathbb{K}_{\mathcal{I},d})$ that correspond to trivial GCIs. Those trivial GCIs will increase the size of our desired base unnecessarily.

To remedy this we shall make use of bases of $\mathbb{K}_{\mathcal{I},d}$ with *background knowledge*. More precisely, let us define

$$\mathcal{S}_{\mathcal{I},d} := \big\{\, \{\, C \,\} \to \{\, D \,\} \;\big|\; C, D \in M_{\mathcal{I},d}, C \sqsubseteq D \,\big\}.$$

Then $\mathcal{S}_{\mathcal{I},d}$ contains all implications which correspond to trivial GCIs as mentioned above. Using $\mathcal{S}_{\mathcal{I},d}$ as background knowledge when computing bases of $\mathbb{K}_{\mathcal{I},d}$ will then eliminate these redundancies. As we shall see shortly, this even allows us to retain the property of the canonical base of being of minimal cardinality.

The proof of the following theorem is a straight-forward adaption of the proof of [15, Theorem 5.12].

4.29 Theorem. *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a finite interpretation, and let $d \in \mathbb{N}$. Let $\mathcal{L}$ be a base of $\mathbb{K}_{\mathcal{I},d}$ with background knowledge $\mathcal{S}_{\mathcal{I},d}$. Then*

$$\mathcal{B}_3 := \Big\{ \bigsqcap U \sqsubseteq \big(\bigsqcap U\big)^{\mathcal{I}\mathcal{I}^d} \;\Big|\; (U \to V) \in \mathcal{L} \Big\}$$

is a finite base of all valid GCIs of $\mathcal{I}$ with role-depth at most d.

Proof. Clearly $\mathcal{B}_3$ is a finite set, and all GCIs are valid in $\mathcal{I}$. Thus we only need to show that $\mathcal{B}_3$ is complete for $\mathrm{Th}^d(\mathcal{I})$. For this, we assume without loss of generality that $\mathcal{L}$ only contains implications of the form $U \to U''$ for some $U \subseteq M_{\mathcal{I},d}$.

Let $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ be a model of $\mathcal{B}_3$. Recall that $\mathbb{K}_{M_{\mathcal{I},d},\mathcal{J}}$ denotes the induced formal context of $M_{\mathcal{I},d}$ and $\mathcal{J}$. Let us write $\mathbb{K}_{\mathcal{J}} := \mathbb{K}_{M_{\mathcal{I},d},\mathcal{J}}$ and $\mathbb{K}_{\mathcal{I}} := \mathbb{K}_{\mathcal{I},d}$. We shall show the following subclaims:

(i) all implications from $\mathcal{L} \cup \mathcal{S}_{\mathcal{I},d}$ are valid in $\mathbb{K}_{\mathcal{J}}$,
(ii) all implications $V \to V''_{\mathbb{K}_{\mathcal{I}}}$ are valid in $\mathbb{K}_{\mathcal{J}}$, for $V \subseteq M_{\mathcal{I},d}$,
(iii) all GCIs $\bigsqcap V \sqsubseteq (\bigsqcap V)^{\mathcal{I}\mathcal{I}^d}$ are valid in $\mathcal{J}$, for $V \subseteq M_{\mathcal{I},d}$.

The last claim states that $\mathcal{B}_3$ entails $\mathcal{B}_2$, and thus shows by Theorem 4.10 that $\mathcal{B}_3$ is complete for $\mathrm{Th}^d(\mathcal{I})$.

For the first subclaim we first observe that all GCIs $C \sqsubseteq D$ with $C, D \in M_{\mathcal{I},d}$ hold in every interpretation, and in particular in $\mathcal{J}$. Thus, all implications $(\{ C \} \to \{ D \}) \in \mathcal{S}_{\mathcal{I},d}$ hold in $\mathbb{K}_{\mathcal{J}}$, since by Proposition 4.25

$$\{ C \}'_{\mathbb{K}_{\mathcal{J}}} = C^{\mathcal{J}} \subseteq D^{\mathcal{J}} = \{ D \}'_{\mathbb{K}_{\mathcal{J}}}.$$

Thus let $(U \to U''_{\mathbb{K}_{\mathcal{I}}}) \in \mathcal{L}$. We need to show that

$$U'_{\mathbb{K}_{\mathcal{J}}} \subseteq (U''_{\mathbb{K}_{\mathcal{I}}})'_{\mathbb{K}_{\mathcal{J}}}.$$

For this we first observe that

$$U'_{\mathbb{K}_{\mathcal{J}}} = (\bigsqcap U)^{\mathcal{J}}$$

by Proposition 4.25. Since $(\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}$ is expressible in terms of $M_{\mathcal{I},d}$, Proposition 4.26 yields

$$\big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big)^{\mathcal{J}} = \Big(\mathrm{pr}_{M_{\mathcal{I},d}}\big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big)\Big)'_{\mathbb{K}_{\mathcal{J}}}.$$

Proposition 4.28 yields $\mathrm{pr}_{M_{\mathcal{I},d}}\big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big) = U''_{\mathbb{K}_{\mathcal{I}}}$, and thus

$$\big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big)^{\mathcal{J}} = (U''_{\mathbb{K}_{\mathcal{I}}})'_{\mathbb{K}_{\mathcal{J}}}.$$

Then

$$U'_{\mathbb{K}_{\mathcal{J}}} = (\bigsqcap U)^{\mathcal{J}} \subseteq \big((\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}\big)^{\mathcal{J}} = (U''_{\mathbb{K}_{\mathcal{I}}})'_{\mathbb{K}_{\mathcal{J}}},$$

which shows the first subclaim.

For the second subclaim let $V \subseteq M_{\mathcal{I},d}$. Then $V \to V''_{\mathbb{K}_{\mathcal{I}}}$ is valid in $\mathbb{K}_{\mathcal{I}}$. Since $\mathcal{L} \cup \mathcal{S}_{\mathcal{I},d}$ is a base of $\mathbb{K}_{\mathcal{I}}$, it follows that $V \to V''_{\mathbb{K}_{\mathcal{I}}}$ is entailed by $\mathcal{L} \cup \mathcal{S}_{\mathcal{I},d}$. Since $\mathcal{L} \cup \mathcal{S}_{\mathcal{I},d}$ is sound for $\mathbb{K}_{\mathcal{J}}$, the implication $V \to V''_{\mathbb{K}_{\mathcal{I}}}$ is also valid in $\mathbb{K}_{\mathcal{J}}$. This finishes the second subclaim.

For the final subclaim let again $V \subseteq M_{\mathcal{I},d}$. Since $V \to V''_{\mathbb{K}_{\mathcal{J}}}$ is valid in $\mathbb{K}_{\mathcal{J}}$ by the second subclaim, it is true that

$$V'_{\mathbb{K}_{\mathcal{J}}} \subseteq (V''_{\mathbb{K}_{\mathcal{I}}})'_{\mathbb{K}_{\mathcal{J}}},$$

and a similar argumentation as above shows that

$$(\bigsqcap V)^{\mathcal{J}} \subseteq ((\bigsqcap V)^{\mathcal{I}\mathcal{I}^d})^{\mathcal{J}},$$

i.e., $\bigsqcap V \sqsubseteq \bigsqcap V^{\mathcal{I}\mathcal{I}^d}$ holds in $\mathcal{J}$, as it was claimed. □

The previous theorem allows us to consider small bases of $\mathbb{K}_{\mathcal{I},d}$ and transform them into bases of $\mathcal{I}$. This is useful on its own, but does not directly help us in finding "small" bases of $\mathcal{I}$, as it may happen that "small" bases of $\mathbb{K}_{\mathcal{I},d}$ do not give rise to "small" bases of $\mathcal{I}$, for some suitable notion of "small". We shall remedy this by showing in the remainder of this section that the canonical base of $\mathbb{K}_{\mathcal{I},d}$ with background knowledge $\mathcal{S}_{\mathcal{I},d}$ gives rise to a minimal base of $\mathcal{I}$.

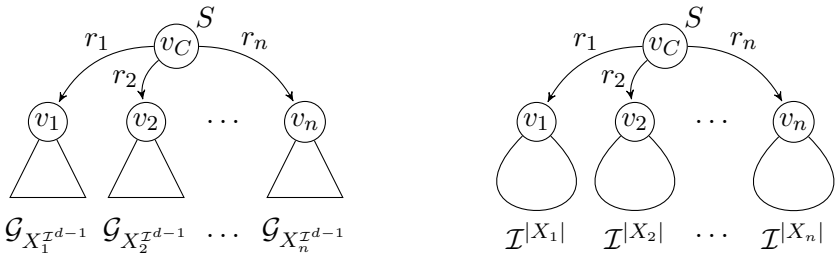


Figure 1. Left: the description graph $\mathcal{G}_C$; Right: the description graph $\mathcal{H}_C$

The main line of argumentation is the same as in [15]. Let

$$\mathcal{B}_{\mathrm{Can}}(\mathcal{I}, d) \coloneqq \Big\{ \bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}^d} \;\Big|\; (U \to U'') \in \mathrm{Can}(\mathbb{K}_{\mathcal{I},d}, \mathcal{S}_{\mathcal{I},d}) \Big\}.$$

To show that we indeed obtain a base of minimal cardinality this way we consider some arbitrary base $\mathcal{B}$ of $\mathcal{I}$. To this base we then associate a base $\mathcal{L}_{\mathcal{B}}$ of implications of $\mathbb{K}_{M_{\mathcal{I},d},\mathcal{I}}$ such that $|\mathcal{L}_{\mathcal{B}}| \leq |\mathcal{B}|$. Then because $\mathrm{Can}(\mathbb{K}_{M_{\mathcal{I},d},\mathcal{I}}, \mathcal{S}_{\mathcal{I},d})$ has minimal cardinality we obtain $|\mathrm{Can}(\mathbb{K}_{M_{\mathcal{I},d},\mathcal{I}}, \mathcal{S}_{\mathcal{I},d})| \leq |\mathcal{L}_{\mathcal{B}}|$, and from this $|\mathcal{B}| \leq |\mathcal{B}_{\mathrm{Can}}|$ as required.

We start the proof by providing two auxiliary statements that we shall make use of in the proof of the following theorem. For the first statement we shall in turn make use of the following fact: for each interpretation $\mathcal{I}$ and each $n \in \mathbb{N} \setminus \{0\}$ we can consider the interpretation $\mathcal{I}^n$ that arises from the product of the description graph of $\mathcal{I}$ with itself n times. Then if $C \sqsubseteq D$ is a GCI that is valid in $\mathcal{I}$, this GCI is also valid in $\mathcal{I}^n$. This is true intuitively, and has been proven formally in [15, Lemma 5.15].

The following lemma is a variation of [15, Lemma 5.16]. Its proof is an extension of the argumentation given there.

4.30 Lemma. *Let $\mathcal{I}$ be a finite interpretation, let $d \in \mathbb{N}$, and let $\mathcal{B}$ be a set of valid GCIs of $\mathcal{I}$ where $\mathcal{B}$ only contains concept descriptions from $\mathcal{EL}^{\bot}(N_C, N_R)_d$. Let C be a concept description such that $C \equiv \bigsqcap U$ for some $U \subseteq M_{\mathcal{I},d}$. Let D be some $\mathcal{EL}^{\bot}$-concept description such that $C \not\sqsubseteq D$ and $\mathrm{rd}(D) \leq d$. If $C \sqsubseteq D$ follows from $\mathcal{B}$, then there is some $(E \sqsubseteq F) \in \mathcal{B}$ such that $C \sqsubseteq E$ and $C \not\sqsubseteq F$.*

Proof. Clearly $C \neq \bot$, as otherwise $C \sqsubseteq D$. Thus there exists $S \subseteq N_C$ and $\Pi \subseteq N_R \times \mathfrak{P}(\Delta^{\mathcal{I}})$ such that

$$C \equiv \bigsqcap S \sqcap \bigsqcap_{(r,X)\in\Pi} \exists r.X^{\mathcal{I}^{d-1}}.$$

Let $\mathcal{G}_C = (V_C, E_C, L_C, v_C)$ be the description graph of C, and let $\mathcal{G}_D$ be the description graph of D. Denote with $\mathcal{I}_{\mathcal{G}_C}$ the interpretation that corresponds to the description graph $\mathcal{G}_C$. Then $v_C \in C^{\mathcal{I}_{\mathcal{G}_C}}$ by Corollary 4.16. On the other hand, as $C \not\sqsubseteq D$, Proposition 4.17 yields that there does not exist a homomorphism from $\mathcal{G}_D$ to $\mathcal{G}_C$, and hence $v_C \notin D^{\mathcal{I}_{\mathcal{G}_C}}$ by Corollary 4.16. Therefore, $\mathcal{I}_{\mathcal{G}_C} \not\models (C \sqsubseteq D)$.

The description graph $\mathcal{G}_C$ is a tree with root v_C. If we denote the children of v_C by $v_1, \ldots, v_n$, then each such v_i is the root of the description tree $\mathcal{G}_{X_i^{\mathcal{I}^{d-1}}}$ of $X_i^{\mathcal{I}^{d-1}}$. Let $X_i = \{x_i^1, \ldots, x_i^k\}$. By Corollary 4.23, the tree $\mathcal{G}_{X_i^{\mathcal{I}^{d-1}}}$ is a product of trees which arise from the description graph of $\mathcal{I}$ by unravelling at the elements x_i^j up to depth $d-1$. A graphical representation of $\mathcal{G}_C$ is shown on the left of Figure 1.

Let us now consider the description graph $\mathcal{H}_C$ that we obtain from $\mathcal{G}_C$ by replacing all unravellings of $\mathcal{I}$ up to depth $d-1$ by the full description graph of $\mathcal{I}$. By this each subtree

$\mathcal{G}_{X_i^{\mathcal{I}^{d-1}}}$ in $\mathcal{G}_C$ is transformed into a graph isomorphic to $\mathcal{I}^{|X_i|}$. This transformation is sketched on the right of Figure 1.

We shall now show that in the interpretation $\mathcal{I}_{\mathcal{H}_C}$ that correspond to $\mathcal{H}_C$ the GCI $C \sqsubseteq D$ does not hold as well. To this end we observe that there exists a homomorphism from $\mathcal{G}_C$ to $\mathcal{H}_C$, showing that $v_C \in C^{\mathcal{I}_{\mathcal{H}_C}}$. On the other hand, a homomorphism from $\mathcal{G}_D$ to $\mathcal{H}_C$ could easily be transformed into a homomorphism from $\mathcal{G}_D$ to $\mathcal{G}_C$, as $\mathrm{rd}(D) \leq d$. As such a homomorphism does not exist we obtain $v_C \notin D^{\mathcal{I}_{\mathcal{H}_C}}$. Therefore, $\mathcal{I}_{\mathcal{H}_C}$ is not a model of $C \sqsubseteq D$.

As $\mathcal{B}$ entails $C \sqsubseteq D$ there must exist a GCI $(E \sqsubseteq F) \in \mathcal{B}$ that is not valid in $\mathcal{I}_{\mathcal{H}_C}$. As $E \sqsubseteq F$ is valid in $\mathcal{I}$, by the remark immediately preceding this proof it is valid in all interpretations $\mathcal{I}^{|X_1|}, \ldots, \mathcal{I}^{|X_n|}$ as well. But then the only element $E \sqsubseteq F$ can fail for in $\mathcal{I}_{\mathcal{H}_C}$ is v_C, and thus $v_C \in E^{\mathcal{I}_{\mathcal{H}_C}}$, $v_C \notin F^{\mathcal{I}_{\mathcal{H}_C}}$. As $\mathrm{rd}(E), \mathrm{rd}(F) \leq d$ we obtain $v_C \in E^{\mathcal{I}_{\mathcal{G}_C}}$, $v_C \notin F^{\mathcal{I}_{\mathcal{G}_C}}$.

Now by Corollary 4.16 there exists a homomorphism from the description graph $\mathcal{G}_E$ to $\mathcal{G}_C$, and there does not exist a homomorphism from $\mathcal{G}_F$ to $\mathcal{G}_C$. This is because the description graph of $\mathcal{I}_{\mathcal{G}_C}$ is isomorphic to $\mathcal{G}_C$. But then Proposition 4.17 yields $C \sqsubseteq E$ and $C \not\sqsubseteq F$, as required. □

The following proposition is a variation of [15, Lemma 5.17].

4.31 Proposition. *Let $C \in \mathcal{EL}^{\bot}(N_C, N_R)_d$, $U \subseteq M_{\mathcal{I},d}$. Then $\bigsqcap U \sqsubseteq C$ implies $\bigsqcap U \sqsubseteq \mathrm{approx}_{\mathcal{I},d}(C)$.*

Proof. We can write C as

$$C = \bigsqcap S \sqcap \bigsqcap_{(r,D)\in\Pi} \exists r.D$$

for some $S \subseteq N_C$ and some $\Pi \subseteq N_R \times \mathcal{EL}^{\bot}(N_C, N_R)_{d-1}$. Then

$$\mathrm{approx}_{\mathcal{I},d}(C) = \bigsqcap S \sqcap \bigsqcap_{(r,D)\in\Pi} \exists r.D^{\mathcal{I}\mathcal{I}^{d-1}}.$$

Since $\bigsqcap U \sqsubseteq C$, for each $A \in S$ we also have $A \in U$. Furthermore, for each $(r, D) \in \Pi$ there must exist some $(\exists r.X^{\mathcal{I}^{d-1}}) \in U$ such that $\exists r.X^{\mathcal{I}^{d-1}} \sqsubseteq \exists r.D$. Since $X^{\mathcal{I}^{d-1}} \equiv X^{\mathcal{I}^{d-1}\mathcal{I}\mathcal{I}^{d-1}} \sqsubseteq D^{\mathcal{I}\mathcal{I}^{d-1}}$ we obtain $\exists r.X^{\mathcal{I}^{d-1}} \sqsubseteq \exists r.D^{\mathcal{I}\mathcal{I}^{d-1}}$. But then

$$\bigsqcap U \sqsubseteq \bigsqcap S \sqcap \bigsqcap_{(r,D)\in\Pi} \exists r.D^{\mathcal{I}\mathcal{I}^{d-1}} \sqsubseteq \mathrm{approx}_{\mathcal{I},d}(C)$$

as required. □

4.32 Theorem. *Let $\mathcal{I}$ be a finite interpretation over N_C and N_R, and let $d \in \mathbb{N}$. Then $\mathcal{B}_{\mathrm{Can}}(\mathcal{I}, d)$ is a base of all valid GCIs of $\mathcal{I}$ with role-depth at most d. Furthermore, $\mathcal{B}_{\mathrm{Can}}(\mathcal{I}, d)$ has minimal cardinality among all bases of all valid GCIs of $\mathcal{I}$ with role-depth at most d.*

Proof. Let $\mathcal{B}$ be a base of $\mathrm{Th}^d(\mathcal{I})$. Without loss of generality we can assume that all GCIs in $\mathcal{B}$ are of the form $E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d}$ for some $\mathcal{EL}^{\bot}$-concept description E with $\mathrm{rd}(E) \leq d$.

We know that $|\mathcal{B}_{\mathrm{Can}}(\mathcal{I}, d)| \leq |\mathrm{Can}(\mathbb{K}_{\mathcal{I},d}, \mathcal{S}_{\mathcal{I},d})|$. The idea of this proof is to define a set $\mathcal{L}_{\mathcal{B}}$ of implications such that

$$|\mathrm{Can}(\mathbb{K}_{\mathcal{I},d}, \mathcal{S}_{\mathcal{I},d})| \leq |\mathcal{L}_{\mathcal{B}}| \leq |\mathcal{B}|.$$

If we succeed in this, then we clearly have shown the claim of the theorem. Thus let us define

$$\mathcal{L}_{\mathcal{B}} \coloneqq \left\{ \operatorname{pr}_{M_{\mathcal{I},d}}(\operatorname{approx}_{\mathcal{I},d}(E)) \to \operatorname{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d}) \;\middle|\; (E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d}) \in \mathcal{B} \right\}.$$

Then clearly $|\mathcal{L}_{\mathcal{B}}| \leq |\mathcal{B}|$. To show $|\operatorname{Can}(\mathbb{K}_{\mathcal{I},d}, \mathcal{S}_{\mathcal{I},d})| \leq |\mathcal{L}_{\mathcal{B}}|$, we show that $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$ is a base of $\mathbb{K}_{\mathcal{I},d}$. Then by the minimality of the canonical base we know that $\mathcal{L}_{\mathcal{B}}$ has at least as many elements as $\operatorname{Can}(\mathbb{K}_{\mathcal{I},d}, \mathcal{S}_{\mathcal{I},d})$.

In the remainder of this proof we shall therefore show that $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$ is a base of $\mathbb{K}_{\mathcal{I},d}$. For this we shall show that $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$ is sound and complete for $\mathbb{K}_{\mathcal{I},d}$.

Let us first show soundness of $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$. Clearly, $\mathcal{S}_{\mathcal{I},d}$ holds in any induced context with attribute set $M_{\mathcal{I},d}$, and thus in particular in $\mathbb{K}_{\mathcal{I},d}$. To see that $\mathcal{L}_{\mathcal{B}}$ is also sound for $\mathbb{K}_{\mathcal{I},d}$, consider some implication

$$\left(\operatorname{pr}_{M_{\mathcal{I},d}}(\operatorname{approx}_{\mathcal{I},d}(E)) \to \operatorname{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d})\right) \in \mathcal{L}_{\mathcal{B}}.$$

Since $\operatorname{approx}_{\mathcal{I},d}(E)$ is expressible in terms of $M_{\mathcal{I},d}$ by definition, Proposition 4.26 implies

$$\left(\operatorname{pr}_{M_{\mathcal{I},d}}(\operatorname{approx}_{\mathcal{I},d}(E))\right)'_{\mathbb{K}_{\mathcal{I},d}} = \operatorname{approx}_{\mathcal{I},d}(E)^{\mathcal{I}}.$$

Since $\operatorname{rd}(E) \leq d$, Proposition 4.8 yields

$$(\operatorname{approx}_{\mathcal{I},d}(E))^{\mathcal{I}} \subseteq E^{\mathcal{I}} \equiv (E^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}}$$

Finally, since $E^{\mathcal{I}\mathcal{I}^d}$ is expressible in terms of $M_{\mathcal{I},d}$ by Lemma 4.9, Proposition 4.26 applies again and yields

$$(E^{\mathcal{I}\mathcal{I}^d})^{\mathcal{I}} = \left(\operatorname{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d})\right)'_{\mathbb{K}_{\mathcal{I},d}}.$$

Thus we have shown that the implication

$$\left(\operatorname{pr}_{M_{\mathcal{I},d}}(\operatorname{approx}_{\mathcal{I},d}(E)) \to \operatorname{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d})\right) \in \mathcal{L}_{\mathcal{B}}.$$

holds in $\mathbb{K}_{\mathcal{I},d}$. Since the choice of this implication was arbitrary, we have shown that $\mathcal{L}_{\mathcal{B}}$ is sound for $\mathbb{K}_{\mathcal{I},d}$.

It remains to be shown that $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$ is complete for $\mathbb{K}_{\mathcal{I},d}$. For this we show that no set $U \subseteq M_{\mathcal{I},d}$ with $U \neq U''_{\mathbb{K}_{\mathcal{I},d}}$ is a model of $\mathcal{L}_{\mathcal{B}} \cup \mathcal{S}_{\mathcal{I},d}$.

So let $U \subseteq M_{\mathcal{I},d}$ with $U \neq U''$. Assume that U is closed under $\mathcal{S}_{\mathcal{I},d}$. Since U is not an intent, there exists some $D \in U'' \setminus U$ with $\operatorname{rd}(D) \leq d$. By definition of $\mathbb{K}_{\mathcal{I},d}$, it is true that $D^{\mathcal{I}} \subseteq U'$. Lemma 4.25 implies $U' = (\bigsqcap U)^{\mathcal{I}}$, i.e., $D^{\mathcal{I}} \subseteq (\bigsqcap U)^{\mathcal{I}}$. By Lemma 4.2 we obtain

$$(\bigsqcap U)^{\mathcal{I}\mathcal{I}} \sqsubseteq D. \tag{7}$$

Since U is closed under $\mathcal{S}_{\mathcal{I},d}$, and since $D \notin U$, we obtain $F \not\sqsubseteq D$ for all $F \in U$. This is because if $F \sqsubseteq D$ for some $F \in U$, then since U is closed under $\mathcal{S}_{\mathcal{I},d}$, we would obtain $D \in U$. Since D is either a concept name or of the form $\exists r.X^{\mathcal{I}^{d-1}}$, we obtain

$$\bigsqcap U \not\sqsubseteq D. \tag{8}$$

Equations (7) and (8) now imply that

$$\bigsqcap U \not\sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}^d}.$$

By Lemma 4.30 there exists some $(E \sqsubseteq E^{\mathcal{I}\mathcal{I}^d}) \in \mathcal{B}$ such that

$$\bigsqcap U \sqsubseteq E, \quad \bigsqcap U \not\sqsubseteq E^{\mathcal{I}\mathcal{I}^d}.$$

We now claim that U is not a model of the implication

$$\big(\mathrm{pr}_{M_{\mathcal{I},d}}(\mathrm{approx}_{\mathcal{I},d}(E)) \to \mathrm{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d})\big) \in \mathcal{L}_{\mathcal{B}}. \tag{9}$$

To see this, we first observe that by Proposition 4.31 and $\bigsqcap U \sqsubseteq E$, we see that $\bigsqcap U \sqsubseteq \mathrm{approx}_{\mathcal{I},d}(E)$. This implies that

$$\mathrm{pr}_{M_{\mathcal{I},d}}\big(\mathrm{approx}_{\mathcal{I},d}(E)\big) \subseteq U.$$

To show that U is not a model of the implication in (9) we thus need to show that $\mathrm{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d}) \not\subseteq U$. Let us assume by contradiction that this is not the case, i.e., $\mathrm{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d}) \subseteq U$. Then

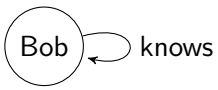
$$\bigsqcap U \sqsubseteq \bigsqcap \mathrm{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d}) \equiv E^{\mathcal{I}\mathcal{I}^d},$$

where the last equivalence is due to the fact that $E^{\mathcal{I}\mathcal{I}^d}$ is expressible in terms of $M_{\mathcal{I},d}$. But this is a contradiction to $\bigsqcap U \not\sqsubseteq E^{\mathcal{I}\mathcal{I}^d}$. We have thus shown that $\mathrm{pr}_{M_{\mathcal{I},d}}(E^{\mathcal{I}\mathcal{I}^d}) \not\subseteq U$, and the proof is finished. □

4.5 *Bases of GCIs with Unlimited Role Depth*

In the last sections we have seen how to compute bases of general concept inclusions with a role-depth bound. But it is also possible to compute bases of GCIs without a role-depth bound, which then entail *all* valid GCIs that hold in the input interpretation. The corresponding theory has been developed in [15], and it is the purpose of this section to give a brief overview of the main points of this work, and to point out differences to our argumentation. As this exposition is meant to be just an overview, we shall not give proofs of the statements of this section, and refer the reader to [15] for the details.

The main obstacle one has to overcome to learn GCIs without a role-depth bound is to generalize the notion of model-based most-specific concept descriptions to an unbounded case. The trouble here is caused by interpretations with cycles. To see what we mean by this let us consider the following example interpretation $\mathcal{I}_{\mathsf{Bob}} = (\{\mathsf{Bob}\}, \cdot^{\mathcal{I}})$ over the signature $(N_C, N_R) = (\emptyset, \{\mathsf{knows}\})$:



(10)

It can be seen easily that there does not exist a most-specific concept description that has Bob in its extension. For this we note that the only $\mathcal{EL}^{\bot}$-concept description that can be formed over (N_C, N_R) are $\bot$ and

$$C_n \coloneqq \underbrace{\exists \mathsf{knows}. \ldots \exists \mathsf{knows}.}_{n \text{ times}} \top \quad (n \in \mathbb{N}). \tag{11}$$

Then $\mathsf{Bob} \in C_n^{\mathcal{I}_{\mathsf{Bob}}}$ for all $n \in \mathbb{N}$, and since C_{n_1} is more specific than C_{n_2} for $n_1 > n_2$, it follows that no C_n can be most-specific with containing Bob in its extension. Intuitively, one would need to use an infinite chain of existential quantifiers to express such a most-specific concept description.

The resort to this dilemma is to extend the description logic $\mathcal{EL}^\bot$ to allow for expressing such infinite chains of quantifiers. In [15] this has been done using *greatest fixpoint semantics*, resulting in a logic called $\mathcal{EL}^\bot_{\text{gfp}}$.

4.5.1 An Extension of $\mathcal{EL}^\bot$ with Greatest Fixpoints

Let us fix a signature (N_R, N_C). An $\mathcal{EL}^\bot_{\text{gfp}}$*-concept description* C over this signature is a pair $C = (A_C, \mathcal{T}_C)$ consisting of a concept name $A_C \notin N_C$ and an $\mathcal{EL}^\bot$-TBox $\mathcal{T}_C$ of concept definitions. The TBox $\mathcal{T}_C$ is only allowed to contain *primitive concept definitions*, i.e., expressions of the form $A \equiv C$, where A is a new concept name taken from a set $N_D(\mathcal{T}_C)$ of *defined concept names* of $\mathcal{T}_C$ that is disjoint with N_C, and where C is an $\mathcal{EL}^\bot$-concept description over $(N_C \cup N_D(\mathcal{T}_C), N_R)$. It is furthermore required that $\mathcal{T}_C$ contains a concept definition for A_C. An example of a $\mathcal{EL}^\bot_{\text{gfp}}$-concept description is

$$(A_{\mathsf{Bob}}, \{A_{\mathsf{Bob}} \equiv \exists \mathsf{knows}.A_{\mathsf{Bob}}\}). \tag{12}$$

Let us briefly sketch how the semantics of such concept descriptions is defined. To this end, let $C = (A_C, \mathcal{T}_C)$ be an $\mathcal{EL}^\bot_{\text{gfp}}$-concept description and let $\mathcal{I}$ be an interpretation. To define $C^\mathcal{I}$, we consider *extensions* of $\mathcal{I}$: an *extension* $\mathcal{J}$ of $\mathcal{I}$ is an interpretation whose interpretation function $\cdot^\mathcal{J}$ extends $\cdot^\mathcal{I}$ to the set $N_D(\mathcal{T}_C)$ of defined concept names of $\mathcal{T}_C$. It can then be shown that the set of all extensions of $\mathcal{I}$ forms a complete lattice, and that the fixpoints of a suitable monotone function f on this lattice are exactly the models of $\mathcal{T}_C$. Let $\mathcal{I}_{\text{gfp}}$ be the greatest such fixpoint of f, which exists due to Tarski's Fixpoint Theorem [34]. Then $C^\mathcal{I}$ is defined to be the extension of A_C in this greatest fixpoint $\mathcal{I}_{\text{gfp}}$, i.e., $C^\mathcal{I} := (A_C)^{\mathcal{I}_{\text{gfp}}}$.

4.5.2 Model-Based Most-Specific Concept Descriptions

In analogy to Definition 4.1 we now define model-based most-specific concept descriptions without role-depth limits. Note that the only difference to Definition 4.1 is that the modified definition not only considers $\mathcal{EL}^\bot$-concept descriptions up to a certain role-depth, but all $\mathcal{EL}^\bot_{\text{gfp}}$-concept descriptions.

Let $\mathcal{I}$ be an interpretation and let $X \subseteq \Delta^\mathcal{I}$. Then an $\mathcal{EL}^\bot_{\text{gfp}}$-concept description C is called a *model-based most-specific concept description* (*mmsc* for short) of X in $\mathcal{I}$, if

(i) $X \subseteq C^\mathcal{I}$, and
(ii) $C \sqsubseteq D$ for every $\mathcal{EL}^\bot_{\text{gfp}}$-concept description D over (N_C, N_R) that satisfies $X \subseteq D^\mathcal{I}$.

As in the bounded case, model-based most-specific concept descriptions are unique up to equivalence. Thus, if one exists, we can speak of *the* mmsc, and shall denote it by $X^\mathcal{I}$. Of course, the corresponding mapping $X \mapsto X^\mathcal{I}$ is well-defined only up to equivalence, but this fact does not impose any major problem on our argumentation.

It can be shown that model-based most-specific concept descriptions always exist in $\mathcal{EL}^\bot_{\text{gfp}}$, and they can be computed efficiently. Let us briefly sketch how this can be done. Analogously to Lemma 4.2 and Lemma 4.3, for all interpretations $\mathcal{I}$ the mappings $\cdot^\mathcal{I} \colon \mathcal{EL}^\bot_{\text{gfp}}(N_C, N_R) \to \mathfrak{P}(\Delta^\mathcal{I})$ and $\cdot^\mathcal{I} \colon \mathfrak{P}(\Delta^\mathcal{I}) \to \mathcal{EL}^\bot_{\text{gfp}}(N_C, N_R)$ as defined above satisfy the main condition of an isotone Galois connection. In addition, as shown in Section 4.3 for the case of $\mathcal{EL}^\bot$, it can be shown that the semantics of $\mathcal{EL}^\bot_{\text{gfp}}$ can be characterized by means of description graphs and simulations.

Let $C = (A_C, \mathcal{T}_C)$ be an $\mathcal{EL}^\bot_{\text{gfp}}$-concept description. Then we say that $\mathcal{T}_C$ is in *normal*

form if all $(A \equiv D) \in \mathcal{T}_C$ are of the form

$$D = P_1 \sqcap \cdots \sqcap P_n \sqcap \exists r_1.B_1 \sqcap \ldots \exists r_k.B_k$$

for some $P_1, \ldots, P_n \in N_C$ and $B_1, \ldots, B_k \in N_D(\mathcal{T}_C)$. It can be shown that every $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description can be put into normal form efficiently [2]. In the case that $\mathcal{T}_C$ is in normal form, the *description graph* $\mathcal{G}_C := (V_C, E_C, L_C, A_C)$ of C consists of the following components. The vertex set $V_C := N_D(\mathcal{T}_C)$ consists of all defined concept names of the $\mathcal{T}_C$, the edge set E_C contains all labeled edges (A, r, B) such that $A \equiv D$ and $\exists r.B$ appears in D. Finally, $L_C(A) := \{ P_1, \ldots, P_n \}$.

Conversely, every description graph $\mathcal{G} = (V, E, L, v_0)$ canonically corresponds to an $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description $C_{\mathcal{G}} = (v_0, \mathcal{T}_{\mathcal{G}})$, where

$$\mathcal{T}_{\mathcal{G}} := \Big\{ v \equiv \prod L(v) \sqcap \prod_{(v,r,w) \in E} \exists r.w \;\Big|\; v \in V \Big\}.$$

It is readily verified that both constructions are inverse to each other, i.e. $C \equiv C_{\mathcal{G}_C}$ and $\mathcal{G} = \mathcal{G}_{C_{\mathcal{G}}}$.

Now, in analogy to Proposition 4.13, we can decide $x \in C^{\mathcal{I}}$ by checking the existence of a simulation from the description graph of C to the description graph of $\mathcal{I}$, i.e., $x \in C^{\mathcal{I}}$ if and only if there is a simulation from $\mathcal{G}_C$ to $\mathcal{G}^x_{\mathcal{I}}$. Furthermore, and similar to Proposition 4.17, $C \sqsubseteq D$ for two $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions C and D, if and only if a simulation from $\mathcal{G}_D$ to $\mathcal{G}_C$ exists. Thus we can conclude that the model-based most-specific concept description of singleton sets $\{ x \}$ always exists and is given by the $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description $C_{\mathcal{G}^x_{\mathcal{I}}}$; see also Theorem 4.19.

To compute model-based most-specific concept descriptions of arbitrary sets $X \subseteq \Delta^{\mathcal{I}}$ we again utilize the notion of *least common subsumers*: an $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description D is a *least common subsumer* of $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions $C_1, \ldots, C_n$ if and only if D subsumes $C_1, \ldots, C_n$, and D is most specific with this property. As in the role-depth-bounded case, the least common subsumer can be computed by means of the *product* of the description graphs of the concept descriptions $C_1, \ldots, C_n$ [2]. In particular, the least common subsumer of two $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions C and D can be found as the concept description $C_{\mathcal{G}_C \times \mathcal{G}_D}$ that is induced by the product of their induced description graphs. Then, as in Corollary 4.23, the model-based most-specific concept description of X is the least common subsumer of the concept descriptions $\{x\}^{\mathcal{I}}$. Finally, note that the mmsc of $\emptyset$ is always $\bot$.

4.5.3 Bases of GCIs

An expression $C \sqsubseteq D$ is called an $\mathcal{EL}^{\bot}_{\text{gfp}}$-*GCI* if C and D are $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions. Then similarly to Definition 4.5 we define an $\mathcal{EL}^{\bot}_{\text{gfp}}$-*base* of GCIs for an interpretation $\mathcal{I}$ to be a set $\mathcal{B}$ of $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCIs that is *sound* and *complete*, i.e., $\mathcal{I}$ is a model of all GCIs in $\mathcal{B}$, and every $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCI valid in $\mathcal{I}$ is entailed by $\mathcal{B}$.

If $C \sqsubseteq D$ is a valid $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCI of $\mathcal{I}$, then it can be shown that also the $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCI $C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$ is valid in $\mathcal{I}$, and that $C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$ entails $C \sqsubseteq D$ (see Lemma 4.6). As an immediate consequence we infer that the set

$$\mathcal{B}_0 := \big\{ C \sqsubseteq C^{\mathcal{I}\mathcal{I}} \;\big|\; C \in \mathcal{EL}^{\bot}_{\text{gfp}}(N_C, N_R) \big\}$$

is a base of $\mathcal{I}$. However, it is obvious that this does not yield a *finite* base in general, in contrast to the case of $\mathcal{EL}^{\bot}$ with role-depth bound, as there are infinitely many concept

descriptions for $\mathcal{EL}^{\bot}_{\text{gfp}}$. Hence $\mathcal{B}_0$ is not useful for practical purposes.

In case of a finite interpretation $\mathcal{I}$ it has been shown [15, Theorem 5.7] that it is sufficient to consider only acyclic $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions as left-hand-sides in $\mathcal{B}_0$ to obtain a base. Here we call an $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description C *acyclic* if the description graph of C is acyclic. As it can be seen easily, such an acyclic concept description is then equivalent to an $\mathcal{EL}^{\bot}$-concept description. Therefore, the set

$$\mathcal{B}_1 \coloneqq \{\, C \sqsubseteq C^{\mathcal{I}\mathcal{I}} \mid C \in \mathcal{EL}^{\bot}(N_C, N_R) \,\}$$

is also a base of $\mathcal{I}$. But while this base is "smaller" than $\mathcal{B}_0$, it is still infinite in general. It can be shown that one can obtain a finite base from $\mathcal{B}_1$ by exploiting the same connection between formal concept analysis and description logics that we have used in Section 4.4. However, in contrast to the argumentation there, considering methods from formal concept analysis is not a mere optimization anymore, but is now crucial step to obtain a finite base of $\mathcal{I}$.

4.5.4 Induced Formal Contexts

In Definition 4.24 we have introduced induced contexts, and we can apply the very same construction to a set $M \subseteq \mathcal{EL}^{\bot}_{\text{gfp}}(N_C, N_R)$ of $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions. Then all of the corresponding statements on induced contexts also hold. In analogy to the definition of the set $M_{\mathcal{I},d}$ it turns out that a suitable set of $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions is

$$M_{\mathcal{I}} \coloneqq \{\, \bot \,\} \cup N_C \cup \{\, \exists r.X^{\mathcal{I}} \mid r \in N_R, \emptyset \neq X \subseteq \Delta^{\mathcal{I}} \,\}$$

Then the *induced context* $\mathbb{K}_{\mathcal{I}}$ of $\mathcal{I}$ is defined to be the induced context of $M_{\mathcal{I}}$ and $\mathcal{I}$. More precisely, $\mathbb{K}_{\mathcal{I}} = (\Delta^{\mathcal{I}}, M_{\mathcal{I}}, \nabla)$, where $(x, C) \in \nabla$ iff $x \in C^{\mathcal{I}}$.

In analogy to Theorem 4.10, a *finite* base of $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCIs for $\mathcal{I}$ can now be obtained as

$$\mathcal{B}_2 \coloneqq \Big\{ \bigsqcap U \sqsubseteq \Big(\bigsqcap U\Big)^{\mathcal{I}\mathcal{I}} \;\Big|\; U \subseteq M_{\mathcal{I}} \Big\}.$$

This yields first finite base for an interpretation $\mathcal{I}$, which unfortunately can get quite large. This is because the set $\mathcal{B}_2$ can contain up to $2^{|M_{\mathcal{I}}|}$ general concept inclusions, and the set $M_{\mathcal{I}}$ itself might have $1 + |N_C| + |N_R| \cdot (2^{|\Delta^{\mathcal{I}}|} - 1)$ concept descriptions in the worst case. Thus the set $\mathcal{B}_2$ may contain up to $\mathcal{O}\big(2^{2^{|\Delta^{\mathcal{I}}|}}\big)$ general concept inclusions, and it is desirable from a practical point of view to reduce the number of GCIs in a finite base as much as possible. This can be achieved by means of formal concept analysis, in the same spirit of Theorem 4.32: the set

$$\mathcal{B}_3 \coloneqq \{\, \bigsqcap P \sqsubseteq \bigsqcap P''_{\mathbb{K}_{\mathcal{I}}} \mid (P \to P''_{\mathbb{K}_{\mathcal{I}}}) \in \text{Can}(\mathbb{K}_{\mathcal{I}}, \mathcal{S}_{\mathcal{I}}) \,\}$$

is a minimal base of $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCIs for the interpretation $\mathcal{I}$, where the set $\mathcal{S}_{\mathcal{I}}$ is defined as similarly to the set $\mathcal{S}_{\mathcal{I},d}$ of Theorem 4.29, i.e.,

$$\mathcal{S}_{\mathcal{I}} \coloneqq \{\, \{C\} \to \{D\} \mid C, D \in M_{\mathcal{I}}, C \sqsubseteq D \,\}.$$

4.5.5 Unravelling $\mathcal{EL}^{\bot}_{\text{gfp}}$-bases to $\mathcal{EL}^{\bot}$-bases

A drawback of using $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions instead of $\mathcal{EL}^{\bot}$-concept descriptions is the fact that $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions are notoriously hard to read because of their recursive nature. This is particularly true for domain experts that may not be trained

in logics. As a consequence, the use of $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions impairs the practical usefulness of our results discussed above.

To get around this issue we can make use of unravellings of $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions into $\mathcal{EL}^{\bot}$-concept descriptions. Put differently, we shall show that each base of $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCIs can be transformed into a base of $\mathcal{EL}^{\bot}$-GCIs by unravelling the corresponding concept descriptions.

Unravellings have already been considered in Definition 4.18 in order to construct role-depth-bounded model-based most-specific concept descriptions. We use the same technique here to transform an $\mathcal{EL}^{\bot}_{\text{gfp}}$-base into an $\mathcal{EL}^{\bot}$-base. For this we first note that, technically, unravellings have only been defined for description graphs in Definition 4.18. But since there is a one-to-one correspondence between $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept descriptions and description graphs, we can simply define the *unravelling* of an $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description up to role-depth $d \in \mathbb{N}$ as the $\mathcal{EL}^{\bot}$-concept description $C_d \coloneqq C_{\mathcal{G}_C \restriction_d}$, where $\mathcal{G}_C$ denotes the description graph of C. Note that $\mathcal{G}_C \restriction_d$ is always a description tree, and thus we can associate with it an $\mathcal{EL}^{\bot}$-concept description as we did in Section 4.3.

We shall furthermore make use of the following fact [15, Lemma 5.5 and Corollary 5.6]: for each $\mathcal{EL}^{\bot}_{\text{gfp}}$-concept description C and each finite interpretation $\mathcal{I}$ there exists a number $d_{C,\mathcal{I}}$ such that $C^{\mathcal{I}} = (C_{d_{C,\mathcal{I}}})^{\mathcal{I}}$. The idea is then to unravel all GCIs in a base $\mathcal{B}$ up to the maximum of all $d_{C,\mathcal{I}}$ for C being a concept description occurring in $\mathcal{B}$. For this we set

$$d \coloneqq \max\{\, d_{C,\mathcal{I}} \mid (C \sqsubseteq D) \in \mathcal{B} \,\}$$

and

$$\mathcal{B}_d \coloneqq \left\{\, C_d \sqsubseteq \left(C^{\mathcal{I}\mathcal{I}}\right)_d, \left(C^{\mathcal{I}\mathcal{I}}\right)_d \sqsubseteq \left(C^{\mathcal{I}\mathcal{I}}\right)_{d+1} \;\middle|\; (C \sqsubseteq D) \in \mathcal{B} \,\right\}.$$

Then the set $\mathcal{B}_d$ is indeed a base of all $\mathcal{EL}^{\bot}_{\text{gfp}}$-GCIs valid in $\mathcal{I}$ [15, Theorem 5.21] consisting only of $\mathcal{EL}^{\bot}$-concept descriptions.

5. Experimental Evaluation

With our previous argumentation we have obtained a way to extract, in some sense, all valid GCIs of a given finite interpretation. As the resulting method is an effective one we can seek to apply it to data-sets from real-world applications to evaluate the usefulness of our approach. To this end we recall our previous remark about RDF-Triple data-sets. There we observed that we can consider each such data-set as a finite interpretation, as both essentially are vertex and edge-labeled graphs. In this way our results allow to extract terminological knowledge from data-sets of the Semantic Web.

In this section we shall illustrate this method by applying it to a subset of the DBpedia data-set [10] from the release of 2014. To construct our finite interpretation $\mathcal{I}_{\text{DBpedia}}$ we shall proceed as follows. In DBpedia there are, among others, two data-sets named *mapping-based types* and *mapping-based properties* (*cleaned*). Let us call the former data-set T, and the latter P.

The data-set T consists of triples of the form

```
<http://dbpedia.org/resource/Ellson,_Minnesota>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://dbpedia.org/ontology/PopulatedPlace> .
```

```
<http://dbpedia.org/resource/Otis_Taylor_(American_football)>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://dbpedia.org/ontology/Athlete> .

<http://dbpedia.org/resource/Eddie_George>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person> .
```

i.e., it contains information about instances being of certain types. Because the predicate in all those triples is

```
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
```

we can equally consider T as a data-set of pairs instead of triples.

The second data-set P contains triples like

```
<http://dbpedia.org/resource/Jowkar,_Afghanistan>
<http://www.w3.org/2003/01/geo/wgs84_pos#lat>
"35.521388888888886"^^<http://www.w3.org/2001/XMLSchema#float> .

<http://dbpedia.org/resource/Khenaman_Rural_District>
<http://dbpedia.org/ontology/isPartOf>
<http://dbpedia.org/resource/Kerman_Province> .

<http://dbpedia.org/resource/Robert_Benchley>
<http://dbpedia.org/ontology/influenced>
<http://dbpedia.org/resource/James_Thurber> .
```

i.e., this data set contains relationships between instances as well as literal information about instances.

For readability, we shall from now on omit the prefix `http://dbpedia.org/ontology`. Where other prefixes have been used we leave them in place to avoid ambiguities.

To get a reasonably sized data-set out of T and P for our experiments we proceed as follows. First we consider all triples (s, p, o) in the mapping-based properties data-set P such that $p = \texttt{child}$. All entities s and o that occur in such a triple are collected into a set $\Delta^{\mathcal{I}_{\text{DBpedia}}}$. Then for each element x in $\Delta^{\mathcal{I}_{\text{DBpedia}}}$ we consider all pairs (x, c) in the mapping-based types data-set T and define the set N_C to be the set of all those elements c, i.e.,

$$N_C \coloneqq \{\, c \mid \exists x \in \Delta^{\mathcal{I}_{\text{DBpedia}}} \colon (x, c) \in T \,\}.$$

Then for each $A \in N_C$ we set

$$A^{\mathcal{I}_{\text{DBpedia}}} \coloneqq \{\, x \in \Delta^{\mathcal{I}_{\text{DBpedia}}} \mid (x, A) \in T \,\}.$$

Then $\mathcal{I}_{\text{DBpedia}} \coloneqq (\Delta^{\mathcal{I}_{\text{DBpedia}}}, \cdot^{\mathcal{I}_{\text{DBpedia}}})$ is an interpretation over N_C and $N_R \coloneqq \{\, \text{child} \,\}$. We have $|\Delta^{\mathcal{I}_{\text{DBpedia}}}| = 16891$ and $|N_C| = 183$.

By construction one could expect $\mathcal{I}_{\text{DBpedia}}$ to contain only elements that are instances of the concept `Person` as we only consider instances in the DBpedia data-set that either have or are children. But since DBpedia extracts its data from Wikipedia Infoboxes in a heuristic way, elements that are not persons are also contained in $\mathcal{I}_{\text{DBpedia}}$, example being organizations, books, and places. This is because in Wikipedia Infoboxes children are sometimes stored together with the organizations they belong to, or the places they have lived in. If this extra information points to another Wikipedia page, DBpedia may mistakenly pick up this page as the child of the current article, instead of the child itself.

Because of this, individuals that are not persons do appear in $\mathcal{I}_{\text{DBpedia}}$. For our purpose of demonstrating our approach to learn GCIs from finite interpretations, however, this fact does not play much of a role.

We now apply our approach to $\mathcal{I}_{\text{DBpedia}}$ and compute a base $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ of all $\mathcal{EL}^{\bot}$ GCIs with role-depth at most 2, using an prototypical implementation of the algorithm described before[1]. The base $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ then consists of 3880 elements. In the following we shall have a closer look on some typical elements of $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ to convey a feeling which kind of knowledge our algorithm extracts from $\mathcal{I}_{\text{DBpedia}}$.

The first GCIs computed by our algorithm only involve concept names, for example

$$\begin{aligned}
\texttt{ChemicalSubstance} &\sqsubseteq \texttt{Mineral},\\
\texttt{ChessPlayer} &\sqsubseteq \texttt{Athlete},\\
\texttt{HockeyTeam} &\sqsubseteq \texttt{SportsTeam},\\
\texttt{Saint} &\sqsubseteq \texttt{Cleric},\\
\texttt{Governor} &\sqsubseteq \texttt{Politician},\\
\texttt{IceHockeyPlayer} &\sqsubseteq \texttt{Athlete}.
\end{aligned}$$

We note that most of those GCIs are valid in the DBpedia data-set by construction. This is because Wikipedia Infoboxes, from which DBpedia extracts its knowledge, are not standardized in any way, and thus some background knowledge is necessary to produce a consistent data-set. This background knowledge is given by a manually created taxonomy of 685 classes, and the above given GCIs are all contained therein. But $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ also contains GCIs which are not represented in DBpedia's ontology, due to the data-set being too specific. Examples for these GCIs are

$$\begin{aligned}
\texttt{Coach} &\sqsubseteq \texttt{CollegeCoach},\\
\texttt{Name} &\sqsubseteq \texttt{GivenName},\\
\texttt{TimePeriod} &\sqsubseteq \texttt{Year}.
\end{aligned}$$

Furthermore, $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ contains GCIs representing disjointness constraints, as

$$\begin{aligned}
\texttt{Agent} \sqcap \texttt{ChemicalSubstance} &\sqsubseteq \bot,\\
\texttt{ChemicalSubstance} \sqcap \texttt{TimePeriod} &\sqsubseteq \bot,\\
\texttt{Agent} \sqcap \texttt{TimePeriod} &\sqsubseteq \bot,\\
\texttt{Judge} \sqcap \texttt{Politician} &\sqsubseteq \bot,\\
\texttt{Journalist} \sqcap \texttt{Judge} &\sqsubseteq \bot,\\
\texttt{HorseTrainer} \sqcap \texttt{Politician} &\sqsubseteq \bot,\\
\texttt{HorseTrainer} \sqcap \texttt{Judge} &\sqsubseteq \bot,\\
\texttt{FictionalCharacter} \sqcap \texttt{HorseTrainer} &\sqsubseteq \bot,\\
\texttt{Architect} \sqcap \texttt{Scientist} &\sqsubseteq \bot,\\
\texttt{Architect} \sqcap \texttt{Journalist} &\sqsubseteq \bot,\\
\texttt{Architect} \sqcap \texttt{FictionalCharacter} &\sqsubseteq \bot.
\end{aligned}$$

[1]The source-code for this implementation is available under `http://github.com/exot/EL-exploration`.

DBpedia's ontology implicitly contains those GCIs as well, as the subsumption hierarchy is a tree. The GCIs in $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ make these disjointness relationships explicit, using a minimal number of GCIs.

So far we have only considered GCIs that involve only concept names, but the main strength of our algorithm is to learn GCIs that contain roles. The first example of a GCI containing a role-name is

$$\texttt{Pope} \sqsubseteq \exists \texttt{child}.\texttt{Person} \sqcap \texttt{Cleric}.$$

While it is surprising that $\mathcal{I}_{\text{DBpedia}}$ contains popes, it is even more surprising that all popes contained in $\mathcal{I}_{\text{DBpedia}}$ do have children (and are not contained in $\mathcal{I}_{\text{DBpedia}}$ because they are the children of famous persons). The reason for this is that apparently popes never had famous parents, and thus appear in $\mathcal{I}_{\text{DBpedia}}$ because they have famous children. The only such popes where Alexander VI, Paul III, and Julius II, and these are the only popes contained in $\mathcal{I}_{\text{DBpedia}}$. This is why our algorithm extracts the above-mentioned GCI. Interestingly, $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ also contains the GCI

$$\exists \texttt{child}.\texttt{Person} \sqcap \texttt{Pope} \sqcap \texttt{Saint} \sqsubseteq \bot,$$

that expresses the fact that so far the Catholic Church has not canonized any popes having children.

A general pattern of GCIs involving the `child` role is to exclude that certain professions have (famous) children, like

$$\begin{aligned}
\exists \texttt{child}.\top \sqcap \texttt{Astronaut} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{Medician} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{ChessPlayer} &\sqsubseteq \bot \\
\exists \texttt{child}.\texttt{Politician} \sqcap \texttt{Engineer} &\sqsubseteq \bot \\
\exists \texttt{child}.\texttt{Mayor} \sqcap \texttt{Governor} &\sqsubseteq \bot \\
\exists \texttt{child}.\texttt{Scientist} \sqcap \texttt{Judge} &\sqsubseteq \bot \\
\exists \texttt{child}.\texttt{Journalist} \sqcap \texttt{Monarch} &\sqsubseteq \bot \\
\exists \texttt{child}.\texttt{Journalist} \sqcap \texttt{PlayboyPlaymate} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{Economist} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{BritishRoyalty} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{BeautyQueen} &\sqsubseteq \bot \\
\exists \texttt{child}.\top \sqcap \texttt{Philosopher} &\sqsubseteq \bot
\end{aligned}$$

A variation of this pattern is to express the fact that certain professions only have persons as children:

$$\begin{aligned}
\exists \texttt{child}.\top \sqcap \texttt{SoccerManager} &\sqsubseteq \exists \texttt{child}.\texttt{Person} \\
\exists \texttt{child}.\top \sqcap \texttt{BaseballPlayer} &\sqsubseteq \exists \texttt{child}.\texttt{Person} \\
\exists \texttt{child}.\top \sqcap \texttt{Saint} &\sqsubseteq \exists \texttt{child}.\texttt{Person} \\
\exists \texttt{child}.\top \sqcap \texttt{ScreenWriter} &\sqsubseteq \exists \texttt{child}.\texttt{Person}
\end{aligned}$$

Note that these GCIs are indeed interesting GCIs, as $\mathcal{I}_{\text{DBpedia}}$ contains individuals that are not persons. The above mentioned GCIs then express the fact that for certain professions DBpedia has extracted only persons as children. On the other hand, one could expect that only persons can have children (as only Wikipedia Infoboxes of persons do have entries for children), and indeed $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ contains the GCI

$$\exists \texttt{child}.\top \sqsubseteq \texttt{Person}.$$

Interestingly, one can even make use of some GCIs in $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ to find places where DBpedia has extracted children that are not persons. For example, our algorithm extracts from $\mathcal{I}_{\text{DBpedia}}$ the GCI

$$\exists \texttt{child}.\texttt{Place} \sqcap \texttt{Engineer} \sqsubseteq \exists \texttt{child}.\texttt{Http://schema.org/AdministrativeArea}$$

indicating that in at least one case a child has been extracted that is not a person, but a place, and that the person having that child is an engineer. In this case, there is only one such engineer, named Edward Snell, and his infobox lists his children together with their places of birth (among others).

Finally, the majority of GCIs contained in $\mathcal{B}_{\mathcal{I}_{\text{DBpedia}}}$ does not follow any obvious pattern, and we list some here to give an impression how they look like.

$$\begin{aligned}
\exists \texttt{child}.\texttt{Judge} \sqcap \texttt{Judge} &\sqsubseteq \exists \texttt{child}.(\texttt{Judge} \sqcap \exists \texttt{child}.\texttt{OfficeHolder}) \\
\exists \texttt{child}.\texttt{Engineer} &\sqsubseteq \texttt{Engineer} \\
\texttt{PlayboyPlaymate} &\sqsubseteq \exists \texttt{child}.\texttt{Person} \\
\exists \texttt{child}.\texttt{ComicsCreator} \sqcap \texttt{Politician} &\sqsubseteq \texttt{Congressman} \\
\exists \texttt{child}.\texttt{Criminal} \sqcap \texttt{Politician} &\sqsubseteq \texttt{MemberOfParliament}
\end{aligned}$$

6. Conclusions and Future Work

In this paper we have discussed an approach to learn valid GCIs of finite interpretations whose quantifier depth does not exceed a given bound. For this we have modified the original argumentation of [15] to allow for role-depth bounds, and we have shown that all major results are still valid. Moreover, the introduction of role-depth bounds has also simplified the logical setup in which we need to argue, as we do not need to consider logics with cyclic concept descriptions anymore. Considering those logics was necessary in the original approach of [15]. Finally, we have also demonstrated using a real-world data-set that our approach is effective, showing that we can automatically learn terminological knowledge expressible in $\mathcal{EL}^{\bot}$ from larger data-sets.

While the experiments we have conducted in Section 5 show that ontology engineers can use our approach to extract terminological knowledge from data-sets, they also illustrate a major disadvantage of our approach when applied to real-world data: when the data-set contains *errors*, then the GCIs extracted by our algorithm can be too specific, because errors can invalidate a more general pattern. This issue has been addressed in [11], where in addition to extracting valid GCIs from a finite interpretation also GCIs are considered that are "almost valid" in the given data-set. A GCI is said to be "almost valid" if it is correct in at least a certain number of cases where it is applicable. Experiments in [11] indicated that considering almost valid GCIs instead of only valid GCIs of finite interpretations increases the usefulness of the underlying approach to learn terminological knowledge from real-world data.

The work of [11] builds upon the original approach of [15], and hence does not allow for role-depth bounds. As real-world data can always be assumed to be faulty, and the approach of [11] promises to perform better on such data-sets, also introducing role-depth bounds there may increase the practicability of our ideas even more. We believe that such an extension is not difficult, albeit technical, and we leave it as future work.

The description logic $\mathcal{EL}^\bot$ we have used here is an inexpressive logic, and does not capture important forms of knowledge one may be interested in. For example, we have seen in our experiments that we extract the GCI

$$\exists \texttt{child}.\top \sqcap \texttt{SoccerManager} \sqsubseteq \exists \texttt{child}.\texttt{Person}$$

from $\mathcal{I}_{\text{DBpedia}}$. This GCI suggests, but does not exactly express, the fact that each soccer manager who has children only has persons as children. Indeed, $\mathcal{EL}^\bot$ cannot express this fact, as it is not able to talk about all successors of a given individual. However, using a more expressive logic, i.e., one that allows for the universal quantifier $\forall$, this fact can easily expressed as

$$\exists \texttt{child}.\top \sqcap \texttt{SoccerManager} \sqsubseteq \forall \texttt{child}.\texttt{Person}.$$

There is some work extending the original approach of [15] to more expressive description logics [13]. Including these extensions in our approach can provide the expressiveness to learn more interesting knowledge from given data-sets, while allowing to control the amount of the GCIs to be learned by restricting their maximal quantifier depth.

Finally, a more application-oriented line of research is to conduct a thorough study of the actual usefulness of our ideas towards learning ontologies from larger data sets. The experiment we have discussed in Section 5 only gives a first impression how our algorithm behaves, but does not reveal whether it is really able to extract knowledge suitable for inclusion in a knowledge base. For this a more sophisticated experimental setup has to be developed that formulates necessary criteria for evaluating the performance of automatic knowledge extraction algorithms. In that respect, an investigation of how our approach works together with existing approaches for learning knowledge from data would be an interesting first step.

References

[1] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, May 2000.

[2] Franz Baader. Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In Gottlob and Walsh [20], pages 319–324.

[3] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Gottlob and Walsh [20], pages 325–330.

[4] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence*, pages 364–369. Morgan Kaufmann, July–August 2005.

[5] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope further. In Kendall Clark and Peter F. Patel-Schneider, editors, *Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.

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

[7] Franz Baader, Bernhard Ganter, Barış Sertkaya, and Ulrike Sattler. Completing description logic knowledge bases using formal concept analysis. In Manuela M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 230–235, January 2007.

[8] Franz Baader, Ralf Küsters, and Ralf Molitor. Computing least common subsumers in description logics with existential restrictions. In Thomas Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pages 96–103. Morgan Kaufmann, July–August 1999.

[9] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, March 2009.

[10] Christian Bizer, Jens Lehmann, Gergi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia - A Crystallization Point of the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.

[11] Daniel Borchmann. *Learning Terminological Knowledge with High Confidence from Erroneous Data*. PhD thesis, Technische Universität Dresden, 2014.

[12] Daniel Borchmann and Felix Distel. Mining of $\mathcal{EL}$-GCIs. In Myra Spiliopoulou, Haixun Wang, Diane J. Cook, Jian Pei, Wei Wang, Osmar R. Zaïane, and Xindong Wu, editors, *Proceedings of the 11th International Conference on Data Mining, Workshops*, pages 1083–1090. IEEE, December 2011.

[13] Felix Distel. Model-based most specific concepts in description logics with value restrictions. Technical Report 08-04, Institute of Theoretical Computer Science, TU Dresden, Dresden, Germany, 2008. See http://lat.inf.tu-dresden.de/research/reports.html.

[14] Felix Distel. An Approach to Exploring Description Logic Knowledge Bases. In Kwuida and Sertkaya [24], pages 209–224.

[15] Felix Distel. *Learning Description Logic Knowledge Bases from Data Using Methods from Formal Concept Analysis*. PhD thesis, Technische Universität Dresden, 2011.

[16] Felix Distel and Yue Ma. A hybrid approach for learning concept definitions from text. In Thomas Eiter, Birte Glimm, Yevgeny Kazakov, and Markus Krötzsch, editors, *Informal Proceedings of the 26th International Workshop on Description Logics*, volume 1014 of *CEUR Workshop Proceedings*, pages 156–167. CEUR-WS.org, July 2013.

[17] Franz Baader. Terminological cycles in a description logic with existential restrictions. LTCS-Report LTCS-02-02, Chair for Automata Theory, Institute of Theoretical Computer Science, Dresden University of Technology, Germany, 2002. See http://lat.inf.tu-dresden.de/research/reports.html.

[18] Bernhard Ganter. Two Basic Algorithms in Concept Analysis. In Kwuida and Sertkaya [24], pages 312–340.

[19] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, 1999.

[20] Georg Gottlob and Toby Walsh, editors. *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, August 2003.

[21] Jean-Luc Guigues and Vincent Duquenne. Famille minimale d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences*

Humaines, 95:5–18, 1986.

[22] Monika Henzinger, Thomas Henzinger, and Peter Kopke. Computing simulations on finite and infinite graphs. In *36th Annual Symposium on Foundations of Computer Science*, pages 453–462. IEEE Computer Society, October 1995.

[23] Ian Horrocks, Peter F. Patel-Schneider, and Frank van Harmelen. From $\mathcal{SHIQ}$ and RDF to OWL: The Making of a Web Ontology Language. *Journal of Web Semantics*, 1(1), 2003.

[24] Léonard Kwuida and Barış Sertkaya, editors. *Proceedings of the 8th Interational Conference of Formal Concept Analysis*, volume 5986 of *Lecture Notes in Computer Science*. Springer, March 2010.

[25] Jens Lehmann, Nicola Fanizzi, Lorenz Bühmann, and Claudia d'Amato. *Concept learning*, pages 71–91. In Lehmann and Völker [26], 2014.

[26] Jens Lehmann and Johanna Völker, editors. *Perspectives on Ontology Learning*. Akademische Verlagsgesellschaft, Berlin, 2014.

[27] Peter F. Patel-Schneider, Patrick Hayes, and Ian Horrocks. OWL Web Ontology Language Semantics and Abstract Syntax. Technical report, W3C, 2004.

[28] Cathy Price and Kent Spackman. SNOMED Clinical Terms. *British Journal of Healtcare Computing and Information Management*, 17:27–31, 2000.

[29] Alan L. Rector, William Anthony Nowlan, and Galen Consortium. The GALEN project. *Computer Methods and Programs in Biomedicine*, 45(1-2):75–78, 1994.

[30] Sebastian Rudolph. *Relational exploration: combining description logics and formal concept analysis for knowledge specification*. PhD thesis, Technische Universität Dresden, 2006.

[31] Barış Sertkaya. A Survey on how Description Logic Ontologies Benefit from FCA. In Marzena Kryszkiewicz and Sergei Obiedkov, editors, *Proceedings of the 7th International Conference on Concept Lattices and Their Applications*, pages 2–21, 2010.

[32] Gerd Stumme. Attribute exploration with background implications and exceptions. In Hans-Hermann Bock and Wolfgang Polasek, editors, *Data Analysis and Information Systems*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 457–469. Springer, Heidelberg, 1996.

[33] Fabian Suchanek. *Information Extraction for Ontology Learning*, pages 135–151. In Lehmann and Völker [26], 2014.

[34] Alfred Tarski. A Lattice-Theoretical Fixpoint Theorem and Its Applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[35] Patricia L. Whetzel, Natalya Fridman Noy, Nigam H. Shah, Paul R. Alexander, Csongor Nyulas, Tania Tudorache, and Mark A. Musen. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*, 39(Web-Server-Issue):541–545, 2011.