TECHNISCHE UNIVERSITÄT DRESDEN

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

LTCS–Report

On Confident GCIs of Finite Interpretations

Daniel Borchmann

LTCS-Report 12-06

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

On Confident GCIs of Finite Interpretations

Daniel Borchmann

January 24, 2013

Abstract

In the work of Baader and Distel, a method has been proposed to axiomatize all general concept inclusions (GCIs) expressible in the description logic $\mathcal{EL}^\bot$ and valid in a given interpretation $\mathcal{I}$. This provides us with an effective method to learn $\mathcal{EL}^\bot$-ontologies from interpretations, which itself can be seen as a different representation of *linked data*. In another report, we have extended this approach to handle *errors* in the data. This has been done by not only considering *valid* GCIs but also those whose *confidence* is above a certain threshold c. In the present work, we shall extend the results by describing another way to compute bases of confident GCIs. We furthermore provide experimental evidence that this approach can be useful for practical applications. We finally show that the technique of unravelling can also be used to effectively turn confident $\mathcal{EL}^\bot_{\mathrm{gfp}}$-bases into $\mathcal{EL}^\bot$-bases.

1 Introduction

Description logic ontologies provide a practical yet formally well-defined way of representing large amounts of knowledge. They have been applied especially successfully in the area of medical and biological knowledge, examples being the widely used ontologies SNOMED CT [16], GALEN [17] and the Gene Ontology [2].

A part of description logic ontologies, the so called *TBox*, contains the *terminological knowledge* of the ontology. Terminological knowledge constitutes connections between *concept descriptions* and is represented by *general concept inclusions* (GCIs). For example, we could fix in an ontology the fact that everything that has a child is actually a person. Using the description logic $\mathcal{EL}^\bot$, this could be written as

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

Here, $\exists \mathsf{child}.\top$ and Person are examples of concept descriptions, and the $\sqsubseteq$ sign can be read as "implies." General concept inclusions are, on this intuitive level, therefore quite similar to implications.

The construction of TBoxes of ontologies, which are supposed to represent the knowledge of a certain domain of interest, is normally conducted by human experts. Although this guarantees a high level of quality of the resulting ontology, the process itself is long and expensive. Automating this process would both decrease the time and cost for creating ontologies and would therefore foster the use of formal ontologies in other applications. However, one cannot expect to entirely replace human experts in the process of creating domain-specific ontologies, as these experts are the original source of this knowledge. Hence constructing ontologies completely automatically does not seem reasonable.

A compromise for this would be to devise a *semi-automatic* way of constructing ontologies, for example by *learning* relevant parts of the ontology from a set of *typical examples* of the domain of interest. The resulting ontologies could be used by ontology engineers as a starting point for further refinement and development.

This approach has been taken by Baader and Distel [5, 6, 11] for constructing $\mathcal{EL}^{\perp}$-ontologies from *finite interpretations*. The reason why this approach is restricted to $\mathcal{EL}^{\perp}$ is manifold. Foremost, the approach exploits a tight connection between the description logic $\mathcal{EL}^{\perp}$ and *formal concept analysis* [12], and such a connection has not been worked out for other description logics. Moreover, the description logic $\mathcal{EL}^{\perp}$ can be sufficient for practical applications, as, for example, SNOMED CT is formulated in a variant of $\mathcal{EL}^{\perp}$. Lastly, $\mathcal{EL}^{\perp}$ is computationally much less complex as other description logics, say $\mathcal{ALC}$ or even $\mathcal{FL}_0$.

In their approach, Baader and Distel are able to effectively construct a *base* of all valid GCIs of a given interpretation, where this interpretation can be understood as the collection of typical examples of our domain of interest. This base therefore constitutes the complete terminological knowledge that is valid in this interpretation. Moreover, these interpretations can be seen as a different way to represent *linked data* [7], the data format used by the semantic web community to store its data. Hence, this approach allows us to construct ontologies from parts of the linked data cloud, providing us with a vast amount of real-world data for experiments and practical applications.

In [10], a sample construction has been conducted on a small part of the DBpedia data set [8], which is part of the linked open data cloud. As it turned out, the approach is effective. However, one result of these experiments was a different observation: in the data set extracted from DBpedia, a small set of errors were present. These errors, although very few, greatly influenced the result of the construction in the way these errors invalidated certain GCIs, and hence these GCIs were not extracted by the algorithm anymore. Then, instead of these general GCIs, more special GCIs were extracted that "circumvent" these errors by being more specific. This not only lead to more extracted GCIs, but also to GCIs which may be hard to comprehend.

As the original approach by Baader and Distel considers only valid GCIs, even a single error may invalidate a certain, otherwise valid GCI. Since we cannot assume from real-world data that it does not contain any errors, this approach is quite limited for practical applications. Therefore, we want to present in this work a generalization to the approach of Baader and Distel which does not only consider valid GCIs but also those which are "almost valid." The rationale behind this is that these GCIs should be much less sensitive to a small amount of errors than valid GCIs. To decide whether a GCIs is "almost valid," we shall use its *confidence* in the given interpretation. We then consider the set of all GCIs of a finite interpretation whose confidence is above a certain threshold $c \in [0, 1]$, and try to find a base for them. This base can then be seen as the terminological part of an ontology learned from the data set.

This report sets out to extend the results found in [9]. In this report, first results have been given on how to construct bases of confident GCIs of finite interpretations. We augment these results by another construction that allows us to directly obtain a confident base from a set of implications of a suitable formal contexts. Furthermore, we shall provide experimental results using the DBpedia data set. With these results we want to show that our approach of considering confident GCI may provide useful information in practical applications. Lastly, we answer an open question raised in [9] and show that confident $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-bases can effectively turned into confident $\mathcal{EL}^{\perp}$-bases. For this, we shall use the techniques of unravelling that have also been used in [11] to show a similar result for bases of valid GCIs.

This report is structured as follows. In the following two section we shall introduce the necessary notions from the field of formal concept analysis and description logics needed for this paper. We shall then discuss a construction of a confident base from a suitable formal context. Afterwards, we apply our results to the same interpretation as it has been used in [10], where we not only consider particular confident GCIs and discuss their validity, but where we also examine the

	1	2	3	4	5	6	7	8	9	10
2	.	×	.	×	.	×	.	×	.	×
3	.	.	×	.	.	×	.	.	×	.
5	.	.	.	.	×	.	.	.	.	×
7	.	.	.	.	.	.	×	.	.	.

Figure 1: A formal context depicted as cross table

number of confident GCIs for varying thresholds c an ontology engineer would have to examine. Then, in the subsequent section, we show that unravelling applied to confident bases of finite interpretations can effectively be used to obtain $\mathcal{EL}^{\perp}$-bases from $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-bases. We finish this report with some conclusions and outlook on future work.

2 Formal Concept Analysis

In this section we want to introduce the necessary definitions from formal concept analysis [12] needed in this work.

2.1 Formal Contexts and Contextual Derivation Operators

Formal concept analysis originated as an attempt to unify modern lattice theory with philosophical ideas about *concepts* as hierarchies [12]. The fundamental definition of formal concept analysis is the one of a *formal context*.

2.1 Definition Let G, M be two sets and let $I \subseteq G \times M$. Then the triple $\mathbb{K} = (G, M, I)$ is called a *formal context*, whereas the set G is denoted as the set of *objects* of $\mathbb{K}$ and the set M is denoted as the set of *attributes* of $\mathbb{K}$. For $g \in G, m \in M$ we read $(g, m) \in I$ as "object g has attribute m" and write $g\ I\ m$ in this case. ♢

If a formal context $\mathbb{K} = (G, M, I)$ is finite, i. e. if the sets G and M are finite, it is sometimes convenient to depict $\mathbb{K}$ as a *cross table*, as shown in the following example.

2.2 Example Let $G = \{\, 2, 3, 5, 7 \,\}, M = \{\, 1, \dots, 10 \,\}$ and

$$I = \{\, (g, m) \in G \times M \mid g \text{ divides } m \,\}.$$

Then $\mathbb{K} = (G, M, I)$ is a formal context, which is depicted as a cross table in Figure 1. Here, we have a table where the rows are labeled with elements from G and the rows are labeled with elements from M. In a cell corresponding to a pair $(g, m) \in G \times M$ we write a cross "×" if and only if $(g, m) \in I$. Otherwise, we leave this cell blank or write a single dot "." in it. ♢

Given a formal context $\mathbb{K} = (G, M, I)$ and some set $A \subseteq G$ of objects one can ask what the largest set of attributes is that all objects in A share. Likewise, one can ask for a set $B \subseteq M$ of attributes what the largest set of objects is that have all attributes in B. To answer this question we introduce the *derivation operators* for a formal context $\mathbb{K}$.

2.3 Definition Let $\mathbb{K} = (G, M, I)$ and $A \subseteq G, B \subseteq M$. Then we define the *derivations in the formal context* $\mathbb{K}$ as

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

The set A is called an *extent* of $\mathbb{K}$ if and only if $A = (A')'$. The set B is called an *intent* of $\mathbb{K}$ if and only if $B = (B')'$. $\diamondsuit$

For convenience, we shall drop the extra parentheses and write shorter $(A')' = A''$ and $(B')' = B''$.

As a first observation on the derivation operators let us note that the functions

$$\begin{aligned} \cdot' &\colon \mathfrak{P}(G) \to \mathfrak{P}(M), \\ \cdot' &\colon \mathfrak{P}(M) \to \mathfrak{P}(G) \end{aligned}$$

form a so called *Galois connection*. For this let us recall that for a set P an *order relation* $\leqslant_P$ is just a set $\leqslant_P \subseteq P \times P$ such that $\leqslant_P$ is *reflexive*, *antisymmetric* and *transitive*.

2.4 Definition Let P, Q be two sets and let $\leqslant_P$ and $\leqslant_Q$ be order relations on P and Q, respectively. Then the two mappings

$$\begin{aligned} \varphi &\colon P \to Q, \\ \psi &\colon Q \to P \end{aligned}$$

form an *antitone Galois connection* between $(P, \leqslant_P)$ and $(Q, \leqslant_Q)$ if and only if for all $x \in P, y \in Q$ holds

$$x \leqslant_P \psi(y) \iff y \leqslant_Q \varphi(x). \qquad \diamondsuit$$

We can now see the Galois connection of the derivation operators between the ordered sets $(\mathfrak{P}(G), \subseteq)$ and $(\mathfrak{P}(M), \subseteq)$. We collect this fact, among other, immediate consequences, in the following proposition.

2.5 Proposition *Let $\mathbb{K} = (G, M, I)$ be a formal context, $A_1, A_2 \subseteq G, B_1, B_2 \subseteq M$. Then the following conditions hold:*

- $A_1 \subseteq A_2 \implies A_1' \supseteq A_2'$,
- $B_1 \subseteq B_2 \implies B_1' \supseteq B_2'$,
- $A_1 \subseteq A_1''$,
- $B_1 \subseteq B_1''$,
- $A_1' = A_1'''$,
- $B_1' = B_1'''$,
- $A_1' \subseteq B_1 \iff A_1 \supseteq B_1'$.

Another easy observation regarding derivation operators is the following: If $A \subseteq M$ and $(B_i \mid i \in I)$ is a family of subsets of A such that $\bigcup_{i \in I} B_i = A$, then

$$A' = \bigcap \{\, B_i' \mid i \in I \,\}. \tag{2.1}$$

In particular, for $\mathcal{A} \subseteq \mathfrak{P}(M)$ it is true that

$$\bigcap_{A \in \mathcal{A}} A' = (\bigcup_{A \in \mathcal{A}} A)'. \tag{2.2}$$

We shall make use of these observations in our further discussions.

2.2 Implications

If we have given a formal context $\mathbb{K} = (G, M, I)$, it may very well be that all objects that have certain attributes $A \subseteq M$ always have the attributes $B \subseteq M$ in addition. In this case, we say may that the attributes from A *imply* the attributes from B in the formal context $\mathbb{K}$.

2.6 Definition Let M be a set. An *implication* $A \to B$ on M is a pair (A, B) where $A, B \subseteq M$. In this case, A is called the *premise* and B is called the *conclusion* of the implication $A \to B$. We shall denote the set of all implications on M by $\mathrm{Imp}(M)$.

Let $\mathbb{K} = (G, M, I)$ be a formal context. An *implication* $A \to B$ of $\mathbb{K}$ is an implication on M. The set of all implications of $\mathbb{K}$ is denoted by $\mathrm{Imp}(\mathbb{K})$, i. e.

$$\mathrm{Imp}(\mathbb{K}) = \mathrm{Imp}(M).$$

The implication $A \to B$ *holds in* $\mathbb{K}$ (or is *valid* in $\mathbb{K}$) if $B \subseteq A''$. We then write $\mathbb{K} \models (A \to B)$. If $\mathcal{J}$ is a set of implications of $\mathbb{K}$ such that each implication in $\mathcal{J}$ holds in $\mathbb{K}$, then we may denote this with $\mathbb{K} \models \mathcal{J}$. The set of all implications of $\mathbb{K}$ that hold in $\mathbb{K}$ is denoted by $\mathrm{Th}(\mathbb{K})$.$\diamondsuit$

Note that the condition $B \subseteq A''$ is equivalent to $A' \subseteq B'$ by Proposition 2.5, i. e. an implication $A \to B$ holds in $\mathbb{K} = (G, M, I)$ if and only if every object $g \in G$ that has all attributes in A also has all attribute in B.

2.7 Definition Let M be a set and let $\mathcal{J} \subseteq \mathrm{Imp}(M)$ be a set of implications. Then an implication $A \to B$ *is entailed by* $\mathcal{J}$ if for every context $\mathbb{K}$ with attribute set M in which all implications from $\mathcal{J}$ hold, the implication $A \to B$ holds as well. In this case, we write $\mathcal{J} \models (A \to B)$. The set of all implications in $\mathrm{Imp}(M)$ entailed by $\mathcal{J}$ shall be denoted by $\mathrm{Cn}(\mathcal{J})$. $\diamondsuit$

Implications on a set M give rise to a certain class of mappings on the powerset lattices $(\mathfrak{P}(M), \subseteq)$, namely *closure operators on* M. Abstractly, a closure operator is a mapping

$$c\colon \mathfrak{P}(M) \to \mathfrak{P}(M)$$

such that

- $A \subseteq c(A)$, i. e. c is *extensive*,
- $A \subseteq B \Rightarrow c(A) \subseteq c(B)$, i. e. c is *monotone*, and
- $c(c(A)) = c(A)$, i. e. c is *idempotent*,

is true for all sets $A, B \subseteq M$. A set $A \subseteq M$ is said to be *closed under* c if and only if $c(A) = A$.

Now, implications give rise to closure operators on M, as described in the following definition. Additionally, it is not hard to see that every closure operator on M is equal to a closure operator induced by implications.

2.8 Definition Let M be a set and $\mathcal{L} \subseteq \mathrm{Imp}(M)$. Then define for $A \subseteq M$

$$\begin{aligned} \mathcal{L}^1(A) &:= \bigcup \{\, Y \mid (X \to Y) \in \mathcal{L}, X \subseteq A \,\}, \\ \mathcal{L}^{i+1}(A) &:= \mathcal{L}(\mathcal{L}^i(A)) \quad (i \in \mathbb{N}_{>0}), \\ \mathcal{L}(A) &:= \bigcup_{i \in \mathbb{N}_{>0}} \mathcal{L}^i(A). \end{aligned}$$

The mapping $\mathcal{L} : \mathfrak{P}(M) \to \mathfrak{P}(M)$ with $A \mapsto \mathcal{L}(A)$ is then called the *closure operator induced by* $\mathcal{L}$. A set $A \subseteq M$ is said to be *closed under* $\mathcal{L}$ if and only if $\mathcal{L}(A) = A$. $\diamondsuit$

It is easy to see that every closure operator induced by a set of implications on a set M is indeed a closure operator on M in the sense of the aforementioned definition.

An interesting observation now is that entailment for implications can be rephrased in terms of the induced closure operators. See [9, 12] for more details on this.

2.9 Lemma *Let M be a set and let $\mathcal{L} \subseteq \operatorname{Imp}(M)$, $(A \to B) \in \operatorname{Imp}(M)$. Then*

$$\mathcal{L} \models (A \to B) \iff B \subseteq \mathcal{L}(A).$$

2.3 Bases of Implications

Implications can be understood as logical objects for which we can decide validity in formal contexts. This automatically yields the following definition of *implicational bases*, which results in a way to represent all valid implications of a formal context in a compact way.

2.10 Definition Let $\mathbb{K}$ be a formal context. A set $\mathcal{J}$ of implications of $\mathbb{K}$ is an *implicational base* (or just a *base*) of $\mathbb{K}$ if the following conditions hold:

1) $\mathcal{J}$ is *sound* for $\mathbb{K}$, i. e. every implication in $\mathcal{J}$ holds in $\mathbb{K}$,
2) $\mathcal{J}$ is *complete* for $\mathbb{K}$, i. e. every implication holding in $\mathbb{K}$ follows from $\mathcal{J}$.

Moreover, a base $\mathcal{J}$ of $\mathbb{K}$ is said to be *non-redundant* if each proper subset of $\mathcal{J}$ is not a base of $\mathbb{K}$. $\diamondsuit$

An obvious base is the following.

2.11 Theorem *Let $\mathbb{K}$ be a formal context. Then the set*

$$\mathcal{L} := \{\, A \to A'' \mid A \subseteq M_{\mathbb{K}} \,\}$$

is a base of $\mathbb{K}$.

Checking completeness of a set $\mathcal{L}$ of implications may be a tedious task, as, naively, one may have to consider all valid implications of $\mathbb{K}$. However, completeness of $\mathcal{L}$ can also be verified by considering the intents of $\mathbb{K}$, as the following lemma shows.

2.12 Lemma *Let $\mathbb{K} = (G, M, I)$ be a formal context and let $\mathcal{L} \subseteq \operatorname{Imp}(M)$. Then $\mathcal{L}$ is complete for $\mathbb{K}$ if and only if*

$$\forall U \subseteq M \colon \mathcal{L}(U) = U \implies U = U'',$$

i. e. the closed sets of $\mathcal{L}$ are intents of $\mathbb{K}$.

It is easy to see that if we reverse the direction of the implication in the previous lemma, that we then obtain a characterization for $\mathcal{L}$ to be sound for $\mathbb{K}$.

The base that is described in Theorem 2.11 is not very practical, as it always contains exponentially many implications measured in the size of M. Luckily, we can explicitly describe a base that always has *minimal cardinality* among all bases of a formal context. Unfortunately, even this base may exponentially many elements in the size of M [13].

2.13 Definition ($\mathcal{K}$-pseudo-intent) Let $\mathbb{K}$ be a finite formal context and let $\mathcal{K} \subseteq \operatorname{Imp}(M)$. A set $P \subseteq M$ is said to be a *$\mathcal{K}$-pseudo-intent* of $\mathbb{K}$ if and only if

i. $P \neq P''$,

ii. $\mathcal{K}(P) = P$ and

iii. for all $\mathcal{K}$-pseudo-intents $Q \subsetneq P$ it holds that $Q'' \subseteq P$.

If $\mathcal{K} = \varnothing$, then P is also called a *pseudo-intent* of $\mathbb{K}$. $\diamondsuit$

Let us define for a formal context $\mathbb{K}$ and $\mathcal{K} \subseteq \mathrm{Th}(M)$ the *canonical base of* $\mathbb{K}$ *with background knowledge* $\mathcal{K}$ to be the set

$$\mathrm{Can}(\mathbb{K}, \mathcal{K}) := \{\, P \to P'' \mid P \text{ a } \mathcal{K}\text{-pseudo-intent of } \mathbb{K} \,\}.$$

We may write $\mathrm{Can}(\mathbb{K})$ if $\mathcal{K} = \varnothing$ and just call it the *canonical base* of $\mathbb{K}$.

We can consider the canonical base of $\mathbb{K}$ with background knowledge $\mathcal{K}$ as a smallest set of valid implications of $\mathbb{K}$ such that $\mathrm{Can}(\mathbb{K}, \mathcal{K}) \cup \mathcal{K}$ is a base for $\mathbb{K}$. Intuitively, if we assume that we already know the implications of $\mathcal{K}$ but want to learn all valid implications of $\mathbb{K}$, then $\mathrm{Can}(\mathbb{K}, \mathcal{K})$ is a smallest set of valid implications that we need to add.

2.14 Theorem (Theorem 3.8 from [11]) *Let $\mathbb{K}$ be a finite formal context and $\mathcal{K} \subseteq \mathrm{Th}(M)$. Then the set $\mathrm{Can}(\mathbb{K}, \mathcal{K}) \cup \mathcal{K}$ is base of $\mathbb{K}$ having the least number of elements among all bases of $\mathbb{K}$ containing $\mathcal{K}$.*

This theorem assumes the background knowledge $\mathcal{K}$ to contain only valid implications of $\mathbb{K}$. However, this is not necessary, as the following theorem shows.

2.15 Theorem (Theorem 2.17 from [9]) *Let $\mathbb{K} = (G, M, I)$ be a formal context and let $\mathcal{K} \subseteq \mathrm{Imp}(M)$. Then $\mathrm{Can}(\mathbb{K}, \mathcal{K})$ is the set of valid implications with minimal cardinality such that $\mathrm{Can}(\mathbb{K}, \mathcal{K}) \cup \mathcal{K}$ is complete for $\mathbb{K}$.*

2.4 Canonical Bases of Sets of Implications

We have discussed the canonical base $\mathrm{Can}(\mathbb{K})$ of a formal context $\mathbb{K}$. We can understand $\mathrm{Can}(\mathbb{K})$ as a smallest set of implications $\mathcal{L}$ such that $\mathrm{Cn}(\mathcal{L}) = \mathrm{Th}(\mathbb{K})$. Indeed, instead of only considering the set $\mathrm{Th}(\mathbb{K})$, we can consider *any* set of implications $\mathcal{K}$ and ask for a smallest set $\mathcal{L}$ such that

$$\mathrm{Cn}(\mathcal{L}) = \mathrm{Cn}(\mathcal{K}).$$

We shall give such sets $\mathcal{L}$ a special name.

2.16 Definition Let M be a finite set and let $\mathcal{K} \subseteq \mathrm{Imp}(M)$. A set $\mathcal{L} \subseteq \mathrm{Imp}(M)$ is called a *base* of $\mathcal{K}$ if and only if $\mathrm{Cn}(\mathcal{L}) = \mathrm{Cn}(\mathcal{K})$. $\diamondsuit$

In [18], Rudolph describes a method to effectively convert the set $\mathcal{K}$ into a base $\mathrm{Can}(\mathcal{K})$ of $\mathcal{K}$ of least cardinality. We shall call this set the *canonical base of* $\mathcal{K}$, since this construction yields $\mathrm{Can}(\mathrm{Th}(\mathbb{K})) = \mathrm{Can}(\mathbb{K})$. It is the purpose of this section to repeat these results, as we shall make use of them later on.

We shall first introduce the notion of pseudo-closed sets of $\mathcal{K}$.

2.17 Definition Let M be a finite set and let $\mathcal{K} \subseteq \mathrm{Imp}(M)$. A set $P \subseteq M$ is called a *pseudo-closed set* of $\mathcal{K}$ if and only if the following conditions hold:

i. $P \neq \mathcal{K}(P)$,

ii. for all $Q \subsetneq P$, it is true that $\mathcal{K}(Q) \subseteq P$. ♢

Now we expect that the set

$$\operatorname{Can}(\mathcal{K}) := \{\, P \to \mathcal{K}(P) \mid P \text{ pseudo-closed set of } \mathcal{K} \,\}$$

is a base of $\mathcal{K}$ of minimal cardinality. The correctness of this intuition is guaranteed by the following result. Before we are going to prove, let us note that if $\mathcal{K}_1$ and $\mathcal{K}_2$ are two sets of implications on a finite set M such that $\operatorname{Cn}(\mathcal{K}_1) = \operatorname{Cn}(\mathcal{K}_2)$, that then $\operatorname{Can}(\mathcal{K}_1) = \operatorname{Can}(\mathcal{K}_2)$ is true. This follows immediately from the definition of pseudo-closed sets, as $\operatorname{Cn}(\mathcal{K}_1) = \operatorname{Cn}(\mathcal{K}_2)$ implies $\mathcal{K}_1(A) = \mathcal{K}_2(A)$ for all $A \subseteq M$.

2.18 Theorem *Let M be a finite set and let $\mathcal{K} \subseteq \operatorname{Imp}(M)$. Then the set $\operatorname{Can}(\mathcal{K})$ is a base of $\mathcal{K}$ of minimal cardinality.*

Proof We can find a formal context $\mathbb{K}$ with attribute set M such that

$$A'' = \mathcal{L}(A)$$

is true for each $A \subseteq M$. From this, we can immediately infer that $\operatorname{Cn}(\mathcal{L}) = \operatorname{Cn}(\operatorname{Th}(\mathbb{K})) = \operatorname{Th}(\mathbb{K})$, because for $A, B \subseteq M$ it is true by Lemma 2.9.

$$\begin{aligned} \mathcal{L} \models (A \to B) &\iff B \subseteq \mathcal{L}(A) \\ &\iff B \subseteq A'' \\ &\iff \mathbb{K} \models (A \to B) \end{aligned}$$

It is now easy to see that $\operatorname{Can}(\mathcal{K}) = \operatorname{Can}(\operatorname{Th}(\mathbb{K})) = \operatorname{Can}(\mathbb{K})$. By Theorem 2.15 (with empty background knowledge) it is true that $\operatorname{Can}(\mathbb{K})$ is a base of $\operatorname{Th}(\mathbb{K})$ with minimal cardinality. As $\operatorname{Th}(\mathbb{K}) = \operatorname{Cn}(\mathcal{K})$ and $\operatorname{Can}(\mathbb{K}) = \operatorname{Can}(\mathcal{K})$, it follows that $\operatorname{Can}(\mathcal{K})$ is a base of $\mathcal{K}$ of minimal cardinality. □

2.19 Algorithm (Computing the Canonical Base for a Given Set of Implications)

define `canonical-base/implications`$(\mathcal{K})$
  $\mathcal{C} := \emptyset$
  $\mathcal{K}' := \{\, A \to \mathcal{K}(A \cup B) \mid (A \to B) \in \mathcal{K} \,\}$
  while $(\mathcal{K}' \neq \emptyset)$
    $(A \to B) :=$ random element of $\mathcal{K}'$
    $\mathcal{K}' := \mathcal{K}' \setminus \{\, A \to B \,\}$
    if $(\mathcal{K}' \cup \mathcal{C})(A) \neq B$ **then**
      $\mathcal{C} := \mathcal{C} \cup \{\, (\mathcal{K}' \cup \mathcal{C})(A) \to B \,\}$
    end if
  end while
  return $\mathcal{C}$
end define

Obtaining the canonical base of the set $\mathcal{K}$ can be done effectively. As shown in [18], Algorithm 2.19 computes for the set $\mathcal{K}$ of implications on M its canonical base $\operatorname{Can}(\mathcal{K})$. Note that the expression $(\mathcal{K}' \cup \mathcal{C})(A)$ just denotes the application to the set A of the closure operator induced by $\mathcal{K}' \cup \mathcal{C}$.

3 The Description Logics $\mathcal{EL}^{\perp}$ and $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$

Description logics are part of the field of knowledge representation, a branch of artificial intelligence. Its main focus lies in the representation of knowledge using well-defined semantics.

For this, description logics provide the notion of *ontologies*. These ontologies can be understood as a collection of axioms. More specifically, description logic ontologies consist of *assertional axioms* and *terminological axioms*. Examples for an assertional axioms are “Tom is a cat” and “Jerry is a mouse”, written in description logic syntax as

$$\mathsf{Cat}(\mathsf{Tom}) \quad \text{and} \quad \mathsf{Mouse}(\mathsf{Jerry}).$$

An example for terminological axiom would be to say that “every cat hunts a mouse”, written as

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

The use of the existential quantifier may be a bit surprising here, but it can be explained as follows. Consider the reformulation of “every cat hunts a mouse” to “whenever there is a cat, there exists a mouse it hunts.” The above statement should be read with this reformulation in mind.

Another example would be to say that “nothing is both a cat and a mouse”, written as

$$\mathsf{Cat} \sqcap \mathsf{Mouse} \sqsubseteq \bot.$$

Again, a reformulation may clarify the used syntax. The phrase “nothing is both a cat and a mouse” can be understood as “whenever there is something that is both a cat and a mouse, we have a contradiction.” The bottom sign $\bot$ denotes this contradiction.

These examples are formulated in the description logic $\mathcal{EL}^\bot$, the logic we shall mainly use in this work. The constructors used in $\mathcal{EL}^\bot$ are *conjunction* $\sqcap$, *existential restriction* $\exists$ and the *bottom concept* $\bot$.

During the course of our considerations, however, it shall turn out that $\mathcal{EL}^\bot$ does not suffice for all our purposes. We shall therefore latter on introduce another description logic called $\mathcal{EL}^\bot_{\mathrm{gfp}}$ that can be understood as an extension of $\mathcal{EL}^\bot$ that allows for cyclic concept descriptions. The main motivation to consider this description logic shall become clear when we introduce *model-based most-specific concept descriptions*, which allow us to reformulate notions from formal concept analysis in the language of description logics.

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

We are now going to introduce the syntax and semantics of the description logic $\mathcal{EL}^\bot$. For this, let us fix three disjoint sets N_C, N_R and N_I. We think of these sets as the sets of *concept names*, *role names* and *individual names*, respectively. We may sometimes refer to the triple (N_C, N_R, N_I) as the current *signature*.

3.1 Definition The set $\mathcal{C}$ of *$\mathcal{EL}$-concept description* is defined as follows:

i. If $A \in N_C$, then $A \in \mathcal{C}$.

ii. If $C, D \in \mathcal{C}$, then $C \sqcap D \in \mathcal{C}$.

iii. If $C \in \mathcal{C}$ and $r \in N_R$, then $\exists r.C \in \mathcal{C}$.

iv. $\top \in \mathcal{C}$.

v. $\mathcal{C}$ is minimal with these properties.

An *$\mathcal{EL}^\bot$-concept description* is either $\bot$ or an $\mathcal{EL}$-concept description. $\diamondsuit$

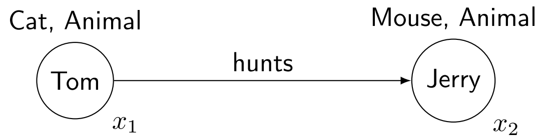

Figure 2: An example interpretation

We may simply talk about *concept descriptions* if it is clear from the context that we refer to $\mathcal{EL}^\perp$-concept descriptions.

We have already seen some examples for $\mathcal{EL}^\perp$-concept descriptions, but let us consider one more example, this time a bit more formally.

3.2 Example Let us consider the sets

$$\begin{aligned} N_C &= \{ \mathsf{Cat}, \mathsf{Mouse}, \mathsf{Animal} \}, \\ N_R &= \{ \mathsf{hunts} \}. \end{aligned}$$

Then

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

is a valid $\mathcal{EL}^\perp$-concept description. Informally, it can be understood as the set of all cats that are (at this very moment) hunting a mouse. ♢

Intuitively associating a meaning with an $\mathcal{EL}^\perp$-concept description is not sufficient for a knowledge representation formalism. Therefore, description logics define the semantics of concept descriptions in terms of *interpretations*. An interpretation can be understood as a directed graph where the vertices are labeled with concept names from N_C and edges are labeled with role names from N_R. Additionally, some of the vertices are explicitly named with elements from N_I and no vertex has more than one name.

3.3 Definition An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists of a set $\Delta^\mathcal{I}$ and an *interpretation function* $\cdot^\mathcal{I}$ such that

$$\begin{aligned} A^\mathcal{I} &\subseteq \Delta^\mathcal{I} \quad \text{for all } A \in N_C, \\ r^\mathcal{I} &\subseteq \Delta^\mathcal{I} \quad \text{for all } r \in N_R, \\ a^\mathcal{I} &\in \Delta^\mathcal{I} \quad \text{for all } a \in N_I. \end{aligned}$$

In addition, the *unique name assumption* holds: If $a, b \in N_I, a \neq b$, then $a^\mathcal{I} \neq b^\mathcal{I}$. ♢

3.4 Example Let us choose again $N_C = \{ \mathsf{Cat}, \mathsf{Mouse}, \mathsf{Animal} \}$, $N_R = \{ \mathsf{hunts} \}$ and in addition $N_I = \{ \mathsf{Tom}, \mathsf{Jerry} \}$. An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ would then be given by

$$\begin{aligned} \Delta^\mathcal{I} &= \{ x_1, x_2 \}, \\ \cdot^\mathcal{I} &= \{ (\mathsf{Cat}, \{ \mathsf{x}_1 \}), (\mathsf{Mouse}, \{ \mathsf{x}_2 \}), (\mathsf{Animal}, \{ \mathsf{x}_1, \mathsf{x}_2 \}) \}, \\ \mathsf{Tom}^\mathcal{I} &= x_1, \\ \mathsf{Jerry}^\mathcal{I} &= x_2, \end{aligned}$$

where we have specified the interpretation function $\cdot^\mathcal{I}$ through its graph. Figure 2 shows the interpretation $\mathcal{I}$ as a directed and labeled graph. ♢

Given an interpretation $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, we can extend the interpretation function $\cdot^\mathcal{I}$ to the set of all $\mathcal{EL}^\perp$-concept descriptions as follows. Let C be an $\mathcal{EL}^\perp$-concept description.

- If $C = \top$, then $C^{\mathcal{I}} = \Delta^{\mathcal{I}}$.
- If $C = \bot$, then $C^{\mathcal{I}} = \emptyset$.
- If $C = C_1 \sqcap C_2$, then $C^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$.
- If $C = \exists r.C_1$ with $r \in N_R$, then

$$C^{\mathcal{I}} = \{\, x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} \colon (x,y) \in r^{\mathcal{I}} \text{ and } y \in C_1^{\mathcal{I}} \,\}.$$

3.5 Definition If C is an $\mathcal{EL}^{\bot}$-concept description and $\mathcal{I}$ is an interpretation, then $C^{\mathcal{I}}$ is said the be the *extension of C in $\mathcal{I}$*. The elements of $C^{\mathcal{I}}$ are said to *satisfy* the concept description C and the elements of $\Delta^{\mathcal{I}} \backslash C^{\mathcal{I}}$ are said to *not satisfy* the concept description C. ◊

The notion of interpretations also allows us to speak of concept descriptions that are *more specific* than other concept descriptions.

3.6 Definition Let C, D be two $\mathcal{EL}^{\bot}$-concept descriptions. Then C is said to be *more specific* then D (or C is *subsumed by D*), written as $C \sqsubseteq D$, if and only if for all interpretations $\mathcal{I}$ it is true that

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

Two $\mathcal{EL}^{\bot}$-concept descriptions C and D are *equivalent*, written as $C \equiv D$, if and only if C is more specific than D and D is more specific than C, i. e.

$$C \equiv D \iff (C \sqsubseteq D) \text{ and } (D \sqsubseteq C).$$ ◊

We shall now introduce the notions of *terminological axioms* and *TBoxes*.

3.7 Definition An *terminological axiom* is of the form

$$C \sqsubseteq D \quad \text{or} \quad A \equiv D,$$

where $A \in N_C$ and C, D are $\mathcal{EL}^{\bot}$-concept descriptions. Terminological axioms of the form $C \sqsubseteq D$ are called *general concepts inclusions (GCIs)*, axioms of the form $A \equiv D$ are called *concept definitions*. If $C \sqsubseteq D$ is a GCI, then C is called the *subsumee* and D is called the *subsumer* of $C \sqsubseteq D$.

Let $\mathcal{I}$ be an interpretation. Then a general concept inclusion $C \sqsubseteq D$ *holds* in $\mathcal{I}$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. A concept definition $A \equiv C$ *holds* in $\mathcal{I}$ if and only if $A^{\mathcal{I}} = C^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a *model* of a set $\mathcal{T}$ of terminological axioms if and only if all axioms in $\mathcal{T}$ hold in $\mathcal{I}$. ◊

3.8 Example We can define the notion of a *hunting cat* by the concept definition

$$\mathsf{HuntingCat} \equiv \mathsf{Cat} \sqcap \exists \mathsf{hunts}.\top.$$

A general concept inclusions which expresses that every Cat is also an Animal would be

$$\mathsf{Cat} \sqsubseteq \mathsf{Animal}.$$ ◊

A word of caution is appropriate here. We have introduced the symbol $\sqsubseteq$ for denoting both subsumption and general concept inclusions. This may cause some confusions, but is an established convention in the field of description logics. It may even sometimes be that both meanings of this sign occur together. In those situations we have to exercise some extra care on clearly distinguishing both meanings of $\sqsubseteq$.

Collections of terminological axioms are called *TBoxes* (for *terminological boxes*). We shall define two types of TBoxes, namely *cyclic TBoxes* and *general TBoxes.* For this, let us fix another set N_D, begin pairwise disjoint to all N_C, N_R, N_I, which we shall call the set of *defined concept names.*

3.9 Definition Let $\mathcal{T}$ be a set of concept definitions and define

$$N_D(\mathcal{T}) := \{\, A \mid \exists C\colon (A \equiv C) \in \mathcal{T} \,\}.$$

Then $\mathcal{T}$ is called a *cyclic TBox*, if every concept definition $(A \equiv C) \in \mathcal{T}$ is such that A is a defined concept name, C is an $\mathcal{EL}$-concept description with concept names from N_C and $N_D(\mathcal{T})$, and each $A \in N_D(\mathcal{T})$ appears at most once on the left-hand side of a concept definition of $\mathcal{T}$.

The set $N_D(\mathcal{T})$ is then called the set of *defined concept names* of the cyclic TBox $\mathcal{T}$. The set $N_P(\mathcal{T})$ of concept names that appear in concept descriptions in $\mathcal{T}$ but are not defined concept names is called the set of *primitive concept names.* $\diamondsuit$

3.10 Example In the case of Tom and Jerry, it is often not really clear who hunts whom. We can therefore define

$$\begin{aligned} \mathsf{HuntingCat} &\equiv \mathsf{Cat} \sqcap \exists \mathsf{hunts}.\mathsf{HuntingMouse}, \\ \mathsf{HuntingMouse} &\equiv \mathsf{Mouse} \sqcap \exists \mathsf{hunts}.\mathsf{HuntingCat}. \end{aligned}$$

The set containing these two concept definitions is a cyclic TBox. Its defined concept names are $\{\, \mathsf{HuntingMouse}, \mathsf{HuntingCat} \,\}$, its primitive concept names are $\{\, \mathsf{Cat}, \mathsf{Mouse} \,\}$. $\diamondsuit$

Concept definitions are not really necessary if we can use general concept inclusions. To see this, let us recall the definition of a concept definition to hold in an interpretation $\mathcal{I}$. A concept definition $A \equiv C$ holds in $\mathcal{I}$ if and only if $A^{\mathcal{I}} = C^{\mathcal{I}}$. But this is the case if and only if $A^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ and $A^{\mathcal{I}} \supseteq C^{\mathcal{I}}$. Hence $A \equiv C$ holds in $\mathcal{I}$ if and only if $A \sqsubseteq C$ and $C \sqsubseteq A$ both hold in $\mathcal{I}$. Therefore, general concept inclusions can express concept definitions. Thus, if we are given a cyclic TBox $\mathcal{T}_1$ that contains concept definitions, we can always transform it into a set $\mathcal{T}_2$ containing only general concept inclusions such that the models of $\mathcal{T}_1$ are precisely the models of $\mathcal{T}_2$. In this respect, sets containing only general concept inclusions are a generalization of cyclic TBoxes. We shall call such sets *general TBoxes.*

3.11 Definition A *general TBox* is a set of general concept inclusions $C \sqsubseteq D$, where C, D are $\mathcal{EL}^{\bot}$-concept descriptions. $\diamondsuit$

To make our argumentation easier to read, we may simply refer to $\mathcal{T}$ as a *TBox* whenever $\mathcal{T}$ is a cyclic or general TBox.

We have just defined the semantics of both cyclic and general TBoxes. If $\mathcal{T}$ is such a TBox, an interpretation $\mathcal{I}$ is a model of $\mathcal{T}$ if and only if all definitions in $\mathcal{T}$ hold in $\mathcal{I}$. For this we need that the interpretation mapping $\cdot^{\mathcal{I}}$ of $\mathcal{I}$ has been extended to the set $N_D(\mathcal{T})$ of defined concept names of $\mathcal{T}$. This semantics then is called *descriptive semantics.* As we shall see later, there are also other kinds of semantics for TBoxes. As a particular example, we shall introduce *greatest fixpoint semantics* when we discuss the description logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$.

3.2 The Description Logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$

In the work of Distel [11], various parallels between the fields of formal concept analysis and description logics are noted. In particular, in both areas certain elements can be *described.* Let

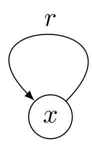

Figure 3: An interpretation where $\{ x \}$ has no model-based most-specific concept description in $\mathcal{EL}^{\bot}$.

$\mathbb{K} = (G, M, I)$ be a formal context. Then an object $g \in G$ can be *described* by a set $A \subseteq M$ of attributes if $x \in A'$. The same is true for an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. An element $x \in \Delta^{\mathcal{I}}$ is *described* by a concept description C if $x \in C^{\mathcal{I}}$. Furthermore, in both $\mathbb{K}$ and $\mathcal{I}$ we can obtain for a description A and C the set of objects A' and elements $C^{\mathcal{I}}$ described by it.

However, in $\mathbb{K}$ we can associate for g a *most-specific description* $B := \{ g \}'$. By Proposition 2.5, $g \in B'$, i. e. B describes g. If then $g \in A'$, then $\{ g \} \subseteq A'$, i. e. $\{ g \}'' \subseteq A''' = A'$. But then $B' \subseteq A'$, and hence B describes the fewest objects of all sets $A \subseteq M$ that describe g. In other words, B describes g in the most specific way.

An analogous notion of a *most-specific concept-description* with respect to an interpretation $\mathcal{I}$ has been introduced in [11] as *model-based most-specific concept description.*

3.12 Definition Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a interpretation and let $X \subseteq \Delta^{\mathcal{I}}$. Then a *model-based most-specific concept description* for X over $\mathcal{I}$ is a concept description C such that

- $X \subseteq C^{\mathcal{I}}$ and
- for all concept descriptions D with $X \subseteq D^{\mathcal{I}}$ it is true that $C \sqsubseteq D$. $\diamondsuit$

Intuitively speaking, a model-based most-specific concept description for $X \subseteq \Delta^{\mathcal{I}}$ is a most-specific concept description that describes all elements in X.

Model-based most-specific concept descriptions may not exist. We shall see in the next example an interpretation $\mathcal{I}$ where some sets of elements do not have model-based most-specific concept descriptions in $\mathcal{EL}^{\bot}$. To compensate for this we shall consider the description logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ that allows for cyclic concept descriptions. In this logic, model-based most-specific concept descriptions always exist.

The following example is a minor variation of one given in [11].

3.13 Example Let $N_C = \emptyset$ and $N_R = \{ r \}$. We consider the interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ with $\Delta^{\mathcal{I}} = \{ x \}$ and $r^{\mathcal{I}} = \{ (x, x) \}$. The interpretation depicted as a graph is shown in Figure 3.

Now suppose that C is an $\mathcal{EL}^{\bot}$-concept description that is at the same time a model-based most-specific concept description for $X = \{ x \}$ over $\mathcal{I}$. Because $N_C = \emptyset$ and $N_R = \{ r \}$, C is equivalent to one of the concept descriptions

$$\top, \exists r.\top, \exists r.\exists r.\top, \dots,$$

i. e.

$$C \equiv \underbrace{\exists r. \dots \exists r.}_{n \text{ times}} \top$$

for some $n \in \mathbb{N}$. Then define

$$D := \underbrace{\exists r. \dots \exists r.}_{n+1 \text{ times}} \top.$$

Then $D^{\mathcal{I}} = \{ x \}$ and $D \sqsubseteq C, D \not\equiv C$, contradicting the fact that C is a model-based most-specific concept description of X over $\mathcal{I}$. $\diamondsuit$

On the other hand, if model-based most-specific concept descriptions exist, they are necessarily unique up to equivalence. Therefore, if X is a set of elements of an interpretation $\mathcal{I}$, we can denote the model-based most-specific concept description of X over $\mathcal{I}$ by the special name $X^{\mathcal{I}}$. This notation has been used to stress the similarity to the derivation operators from formal concept analysis.

In the remained of this section, we shall introduce the description logic $\mathcal{EL}^{\perp}_{\text{gfp}}$ to overcome the deficiency of $\mathcal{EL}^{\perp}$ that there may not always exist model-based most-specific concept descriptions. We start this introduction by definition the syntax of $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept descriptions.

3.14 Definition Let $\mathcal{T}$ be a cyclic TBox. A concept definition $(A \equiv C) \in \mathcal{T}$ is said to be *normalized*, if C is of the form

$$C = B_1 \sqcap \ldots \sqcap B_m \sqcap \exists r_1.A_1 \sqcap \ldots \sqcap \exists r_n.A_n$$

where $m, n \in \mathbb{N}$, $B_1, \ldots, B_m \in N_P(\mathcal{T})$ and $A_1, \ldots, A_n \in N_D(\mathcal{T})$. If $n = m = 0$, then $C = \top$. We call $\mathcal{T}$ normalized if and only if it contains only normalized concept definitions.

An $\mathcal{EL}_{gfp}$*-concept description* now is of the form $C = (A, \mathcal{T})$ where $\mathcal{T}$ is a normalized TBox and A is a defined concept name of $\mathcal{T}$. An $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept description is either $\perp$ or an $\mathcal{EL}_{\text{gfp}}$-concept description. $\diamondsuit$

3.15 Example Let us reconsider the TBox from Example 3.10, i. e.

$$\begin{aligned}\mathcal{T} := \{\, &\mathsf{HuntingCat} \equiv \mathsf{Cat} \sqcap \exists \mathsf{hunts}.\mathsf{HuntingMouse}, \\ &\mathsf{HuntingMouse} \equiv \mathsf{Mouse} \sqcap \exists \mathsf{hunts}.\mathsf{HuntingCat}\,\}.\end{aligned}$$

Then $\mathcal{T}$ is a normalized cyclic TBox and the pair

$$(\mathsf{HuntingMouse}, \mathcal{T})$$

is a valid $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept description. $\diamondsuit$

We have already defined the notion of $\mathcal{EL}^{\perp}$-GCIs. Of course, this definition can be easily modified to yield the notion of $\mathcal{EL}^{\perp}_{\text{gfp}}$-GCIs: these are just expressions of the form $C \sqsubseteq D$, where C and D are $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept descriptions.

We shall sometimes omit the logic and call an $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept description just a concept description and likewise shall call an $\mathcal{EL}^{\perp}_{\text{gfp}}$-GCIs just a GCI, if the description logic used is clear from the context.

As we have defined the syntax of $\mathcal{EL}^{\perp}_{\text{gfp}}$, the natural next step is to define the semantics of $\mathcal{EL}^{\perp}_{\text{gfp}}$. This, however, is not as straight forward as in the case of $\mathcal{EL}^{\perp}$, as we have to deal with circular concept descriptions. As we shall see shortly, semantics can be defined using *fixpoint semantics*. This has been done in [3, 15].

Let C be an $\mathcal{EL}^{\perp}_{\text{gfp}}$-concept description and let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation. If $C = \perp$, then certainly $C^{\mathcal{I}} = \varnothing$. Hence let $C = (A, \mathcal{T})$. Then $A \in N_D(\mathcal{T})$. The idea to define the extension of C in $\mathcal{I}$ is now to *extend* the interpretation mapping $\cdot^{\mathcal{I}}$ such that

$$B^{\mathcal{I}} = D^{\mathcal{I}}$$

is true for all $(B \equiv D) \in \mathcal{T}$. If we have given this, we could simply define

$$C^{\mathcal{I}} := A^{\mathcal{I}}.$$

To make this approach into an actual definition, we have to resolve two issues. Firstly, it is not clear if such an extension of $\cdot^{\mathcal{I}}$ to $N_D(\mathcal{T})$ always exists. Secondly, if such an extensions exists, it may not necessarily be unique, so we have to make an explicit choice. As it turns out, we can describe the extensions of $\cdot^{\mathcal{I}}$ we are looking for as *fixpoints* of a particular mapping and can thus prove the existence of such extensions. Furthermore, it turns out that these fixpoint are naturally ordered, and we can just choose the largest one. See also [4, 15] for more details and motivation.

We are now going to work out this approach in more detail. For this, we start by formally defining the notion of an *extension* of $\cdot^{\mathcal{I}}$.

3.16 Definition Let $\mathcal{I}$ be an interpretation and let $\mathcal{T}$ be a TBox. Then an interpretation $\mathcal{J}$ is an *extension* of the interpretation $\mathcal{I}$ with respect to $\mathcal{T}$ if and only if $\Delta^{\mathcal{I}} = \Delta^{\mathcal{J}}$, $A^{\mathcal{J}}$ is defined for all $A \in N_D(\mathcal{T})$ and

- $\forall A \in N_C \colon A^{\mathcal{I}} = A^{\mathcal{J}}$,
- $\forall r \in N_R \colon r^{\mathcal{I}} = r^{\mathcal{J}}$ and
- $\forall a \in N_I \colon a^{\mathcal{I}} = a^{\mathcal{J}}$.

We shall denote with $\mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$ the set of all extensions of $\mathcal{I}$ with respect to $\mathcal{T}$. $\diamondsuit$

We can define an order relation $\preceq$ on $\mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$ by

$$\mathcal{I}_1 \preceq \mathcal{I}_2 \iff A^{\mathcal{I}_1} \subseteq A^{\mathcal{I}_2} \quad \text{for all } A \in N_D(\mathcal{T})$$

for $\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$. It is clear that $(\mathrm{Ext}_{\mathcal{T}}(\mathcal{I}), \preceq)$ is an ordered set.

3.17 Proposition *For each interpretation $\mathcal{I}$ and TBox $\mathcal{T}$, the ordered set $(\mathrm{Ext}_{\mathcal{T}}(\mathcal{I}), \preceq)$ is a complete lattice.*

Indeed, it is easy to see that

$$\mathrm{Ext}_{\mathcal{T}}(\mathcal{I}) \simeq \prod_{A \in N_D(\mathcal{T})} (\mathfrak{P}(\Delta^{\mathcal{I}}), \subseteq),$$

and the latter is, as a product of complete lattices, again a complete lattice.

As already noted, we are interested only in those extensions of $\mathcal{I}$ such that

$$A^{\mathcal{J}} = C^{\mathcal{J}}$$

is true for all $(A \equiv C) \in \mathcal{T}$. In other words, we are only interested in extensions $\mathcal{J}$ of $\mathcal{I}$ that are models of $\mathcal{T}$.

This fact can also be seen from another perspective: let us define a mapping $f \colon \mathrm{Ext}_{\mathcal{T}}(\mathcal{I}) \to \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$ by

$$A^{f(\mathcal{J})} := C^{\mathcal{J}}$$

for all $(A \equiv C) \in \mathcal{T}$ and $\mathcal{J} \in \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$. Since for each $A \in N_D(\mathcal{T})$, there is exactly one concept definition $(A \equiv C) \in \mathcal{T}$, the function f is well-defined. Furthermore, it is sufficient to define $f(\mathcal{J})$ only on defined concept names, as the value of $f(\mathcal{J})$ is already fixed for concept and role names, since $f(\mathcal{J}) \in \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$. Moreover, this mapping is monotone, i. e.

$$\mathcal{I}_1 \preceq \mathcal{I}_2 \implies f(\mathcal{I}_1) \preceq f(\mathcal{I}_2)$$

for all $\mathcal{I}_1, \mathcal{I}_2 \in \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$. This is easy to see if one recalls that the concept description C is normalized, i. e.

$$C = B_1 \sqcap \ldots \sqcap B_m \sqcap \exists r_1.A_1 \sqcap \ldots \sqcap \exists r_n.A_n$$

where $B_1, \ldots, B_m \in N_C$ and $A_1, \ldots, A_n \in N_D(\mathcal{T})$.

We can now see that the extensions of $\mathcal{I}$ that are models of $\mathcal{T}$ are actually *fixpoints* of f. This is because $\mathcal{J} \in \mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$ is a model of $\mathcal{T}$ if and only if

$$A^{\mathcal{J}} = C^{\mathcal{J}} \quad \text{for all } (A \equiv C) \in \mathcal{T}.$$

But this means that

$$A^{f(\mathcal{J})} = C^{\mathcal{J}} = A^{\mathcal{J}},$$

i. e. $f(\mathcal{J}) = \mathcal{J}$. Hence to show that there exist extensions of $\mathcal{I}$ that are models of $\mathcal{T}$ it is sufficient to show that f has fixpoints. To do this, we use the fact that f is monotone and the following, well-known theorem by Tarski [19].

3.18 Theorem *Let $(L, \leqslant)$ be a complete lattice and let $h \colon L \to L$ be a monotone mapping on $(L, \leqslant)$, i. e.*

$$x \leqslant y \implies h(x) \leqslant h(y)$$

holds for all $x, y \leqslant L$. Then the set

$$F := \{\, z \in L \mid h(z) = z \,\}$$

is such that $(F, \leqslant)$ is a complete sublattice of $(L, \leqslant)$. In particular, $F \neq \varnothing$ and there exists a least and greatest fixpoint of h.

As a corollary, we obtain the fact that the mapping f has fixpoints in $\mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$ and that there exists a greatest fixpoint of f in $\mathrm{Ext}_{\mathcal{T}}(\mathcal{I})$. We call this fixpoint the *greatest fixpoint model (gfp-model)* of $\mathcal{T}$ in $\mathcal{I}$. Having this, we are finally able to define the extension of the concept description C.

3.19 Definition Let C be an $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description and let $\mathcal{I}$ be an interpretation. Then

$$C^{\mathcal{I}} := \begin{cases} \varnothing & \text{if } C = \perp \\ A^{\mathcal{J}} & \text{if } C = (A, \mathcal{T}) \text{ and } \mathcal{J} \text{ is the gfp-model of } \mathcal{T} \text{ in } \mathcal{I}. \end{cases} \qquad \diamondsuit$$

The main result for our considerations about $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ is now the following theorem from [5, 11].

3.20 Theorem (Theorem 4.7 of [11]) *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation and $X \subseteq \Delta^{\mathcal{I}}$. Then there exists a model-based most-specific $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description of X over $\mathcal{I}$.*

Now that we can guarantee the existence of model-based most-specific concept descriptions we can consider some first properties. The following result can also be found in [5].

3.21 Lemma (Lemma 4.1 of [11]) *Let $\mathcal{I}$ be a finite interpretation. Then for each $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description D and every $X \subseteq \Delta^{\mathcal{I}}$, it holds*

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

Proof Suppose $X \subseteq D^{\mathcal{I}}$. Then $X^{\mathcal{I}} \sqsubseteq D$ holds by the definition of model-based most-specific concept descriptions (Definition 3.12). This shows the direction from left to right.

Suppose conversely that $X^{\mathcal{I}} \sqsubseteq D$. Then $X^{\mathcal{I}}$ is a concept description that is satisfied by all elements of X, therefore

$$X \subseteq (X^{\mathcal{I}})^{\mathcal{I}} \subseteq D^{\mathcal{I}},$$

as $X^{\mathcal{I}} \sqsubseteq D$ implies $(X^{\mathcal{I}})^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. This shows the converse direction. □

This lemma may remind one of the definition of a Galois connection, however the relation $\sqsubseteq$ is not an order relation on the set of all model-based most-specific concept descriptions. This is because model-based most-specific concept descriptions are only unique up to equivalence. Yet, most of the properties of a Galois connection are still valid. More precisely, if $\mathcal{I}$ is a finite interpretation, C, D are concept descriptions and $X, Y \subseteq \Delta^{\mathcal{I}}$, then the following statements are true.

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

They can be proven in the same way as for any Galois connection. We shall write $X^{\mathcal{I}\mathcal{I}}$ instead of $(X^{\mathcal{I}})^{\mathcal{I}}$.

Another property that was already claimed is that $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ can be considered as an extension of the description logic $\mathcal{EL}^{\perp}$. This may not be obvious at a first glance, since the definition of $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions is quite different from the one of $\mathcal{EL}^{\perp}$-concept descriptions. Still, $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ can be understood as an extension of $\mathcal{EL}^{\perp}$. To see this we shall first define conjunction and existential restriction for $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions.

Let C, D be two $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions. If $C = \perp$, then $C \sqcap D := \perp$ and $\exists r.C := \perp$. Likewise for $D = \perp$. Hence we may assume that both C, D are not the $\perp$ concept description. Then $C = (A_C, \mathcal{T}_C), D = (A_D, \mathcal{T}_D)$ and we can assume that the defined concept names of $\mathcal{T}_C$ and $\mathcal{T}_D$ are disjoint. Then let us define

$$C \sqcap D := (A, \mathcal{T}_C \cup \mathcal{T}_D \cup \{ A \equiv A_C \sqcap A_D \}),$$

where A is a fresh defined concept name. Furthermore, if $r \in N_R$, then

$$\exists r.C := (A, \mathcal{T}_C \cup \{ A \equiv \exists r.A_C \})$$

where again A is a fresh defined concept name. These definitions preserve the semantics, i. e. for each interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ it holds

$$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}},$$
$$(\exists r.C)^{\mathcal{I}} = \{ x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} \colon (x, y) \in r^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}} \}.$$

We can use these definitions to see that $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ can indeed be regarded as an extension of $\mathcal{EL}^{\perp}$. For this we assign for the $\mathcal{EL}^{\perp}$-concept description $\top$ the $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description $(A, \{ A \equiv \top \})$.

Furthermore, if B is a concept name, then it is equivalent to the $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description $(A, \{ A \equiv B \})$. Using the definitions for conjunction and existential restriction for $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions, we can inductively assign for each $\mathcal{EL}^{\bot}$-concept description an equivalent $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description. As these constructors preserve the semantics, $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ can be seen as an extension of $\mathcal{EL}^{\bot}$.

3.3 Bases for GCIs of Interpretations

In the case of formal contexts, we were able to extract bases of implications form them. As we view GCIs as the description logic analogue of implications, we want to do the same for GCIs and finite interpretations.

In [11], the algorithm for computing the canonical base has been generalized to the description logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$. This generalized algorithm is then able to compute *bases of valid GCIs* of a finite interpretation $\mathcal{I}$. In this short subsection we want to introduce the notion of a base and some related definitions.

3.22 Definition Let $\mathcal{I}$ be a finite interpretation. The set of valid GCIs of $\mathcal{I}$ that consist of $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions is denoted by $\mathrm{Th}(\mathcal{I})$. $\diamondsuit$

One of the main results of [11] was to find a finite set of valid GCIs of $\mathcal{I}$ such that every valid GCI of $\mathcal{I}$ was already entail by this finite set. These finite sets are then called *bases* of $\mathcal{I}$. But we can also introduce this notion in a more general setting, namely for arbitrary sets of GCIs.

3.23 Definition Let $\mathcal{C}$ be a set of GCIs. Let $\mathcal{D}$ be a set of GCIs.

i. $\mathcal{D}$ is said to be *sound for* $\mathcal{C}$ if and only if $\mathcal{C} \models \mathcal{D}$, i. e. every GCI in $\mathcal{D}$ is entailed by $\mathcal{C}$;

ii. $\mathcal{D}$ is said to be *complete for* $\mathcal{C}$ if and only if $\mathcal{D} \models \mathcal{C}$, i. e. every GCI in $\mathcal{C}$ is entailed by $\mathcal{D}$;

iii. $\mathcal{D}$ is said to be a *base for* $\mathcal{C}$ if and only if $\mathcal{D}$ is both sound and complete for $\mathcal{C}$.

If $\mathcal{D}$ is a base of $\mathcal{C}$, then $\mathcal{D}$ is said to be a *non-redundant base* of $\mathcal{C}$ if and only if no proper subset of $\mathcal{D}$ is a base of $\mathcal{C}$. $\diamondsuit$

3.24 Definition Let $\mathcal{I}$ be a finite interpretation. Then a set $\mathcal{B}$ of GCIs is said to be a *base for* $\mathcal{I}$ if and only if $\mathcal{B}$ is a base for $\mathrm{Th}(\mathcal{I})$. $\diamondsuit$

Equivalently, $\mathcal{B}$ is a base for $\mathcal{I}$ if and only if it contains only valid GCIs of $\mathcal{I}$ and every valid GCI of $\mathcal{I}$ is already entailed from $\mathcal{B}$.

One of the main results of Baader and Distel is now to give explicit descriptions of some finite bases for $\mathcal{I}$. We shall discuss their results in detail in Section 4.2.

3.4 Unravelling $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions

The base of described by Baader and Distel makes use of model-based most-specific concept descriptions, and therefore in general contains $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions. This may be undesired, as $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions may be very hard to understand due to their cyclic nature. To overcome this issue, Distel [11] present a method to convert bases of finite interpretations into equivalent set of GCIs which only contains $\mathcal{EL}^{\bot}$-concept descriptions. We shall generalize this technique to special kinds of confident bases in Section 6. For this, it is necessary to introduce the notion of *unravelling $\mathcal{EL}_{gfp}$-concept descriptions up to a certain depth.* This is the purpose of this section.

Of course, the concept description $\bot$ is not interesting for this problem, and we therefore restrict our attention to unravelling $\mathcal{EL}_{\mathrm{gfp}}$-concept descriptions C. The idea of doing this is very natural: we can view C as a graph (with cycles allowed), which we then just "unravel" into an possibly infinite tree. Then to unravel C *to a certain depth* $d \in \mathbb{N}$ just means describes the concept description that corresponds to the unravelling of C cut at depth d.

To make this intuition into a formal definition, we shall first define the notion of *$\mathcal{EL}$-description graphs* of $\mathcal{EL}_{\mathrm{gfp}}$-concept descriptions, which goes back to [4]. We then give a formal definition as in [11] of the *unravelling* of such a description graph, possibly only up to a certain depth d.

3.25 Definition Let $C = (A, \mathcal{T})$ be an $\mathcal{EL}_{\mathrm{gfp}}$-concept description. Then its *$\mathcal{EL}$-description graph* $G := (V, E, L)$ is defined as follows.

Recall that every concept definition in $\mathcal{T}$ is of the form $B \equiv D$, where

$$D = P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.B_1 \sqcap \ldots \exists r_m.B_m,$$

where $P_1, \ldots, P_n \in N_C$, $r_1, \ldots, r_m \in N_R$ and $B_1, \ldots, B_m \in N_D(\mathcal{T})$. We set

$$\begin{aligned} \mathit{names}(B) &:= \{ P_1, \ldots, P_n \}, \\ \mathit{succ}_r(B) &:= \{ B_i \mid 1 \leqslant i \leqslant m, r_i = r \}. \end{aligned}$$

Then define $V := N_D(\mathcal{T}), L := \mathit{names}$ and $E := \{ (B_1, r, B_2) \mid B_2 \in \mathit{succ}_r(B_1) \}$. The vertex $A \in V$ is called the *root* of the $\mathcal{EL}$-description graph of C.

We shall call V the set of *vertices*, E is the set of *edges* and L is the *labeling function* of the $\mathcal{EL}$-description graph of C. $\diamondsuit$

It is easy to see that every description graph can easily be turned back into an $\mathcal{EL}_{\mathrm{gfp}}$-concept description and that the concept description of the $\mathcal{EL}$-description graph of a concept description C is equivalent to C.

In accordance to the definition of unravelling as given in [11], we shall introduce the notion of a *directed path* in an $\mathcal{EL}$-description graph $G = (V, E, L)$ as a word $w = A_1 r_1 A_2 r_2 \ldots r_n A_{n+1}$, where $A_1, \ldots A_{n+1} \in V$ and for each $i \in \{ 1 \ldots n \}$ it is true that $(A_i, r_i, A_{i+1}) \in E$. We shall say that the path w *starts at* $A \in V$ if and only if $A = A_1$, and that w *ends at* $B \in V$ if and only if $A_{n+1} = B$. We shall also write $A_{n+1} =: \delta(w)$ and call it its *destination.* Finally, we shall say that the *length* $\mathit{len}(w)$ of w is n.

3.26 Definition Let $C = (A, \mathcal{T})$ be an $\mathcal{EL}_{\mathrm{gfp}}$-concept description and let $G = (V, E, L)$ its $\mathcal{EL}$-description graph.

The *unravelling* of G is defined as the triple $G^\infty = (V^\infty, E^\infty, L^\infty)$, where

i. V^∞ is the set of all directed paths of G starting at A;

ii. $E^\infty := \{ (w, r, wrB) \mid w, wrB \in V^\infty \}$;

iii. $L^\infty(w) := L(\delta(w))$.

Let $d \in \mathbb{N}$. The *unravelling up to depth* d of G is defined as the description graph $G^d = (V^d, E^d, L^d)$, where

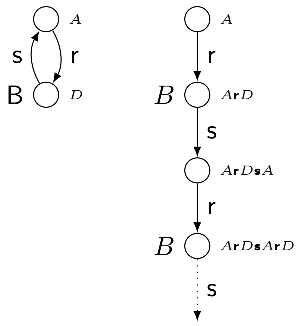

Figure 4: Description graphs of the concept description C (left) and its insinuated unravelling

i. $V^d := \{ w \in V^\infty \mid len(w) \leqslant d \}$;

ii. $E^d := \{ (A, r, B) \in E^\infty \mid A, B \in V^d \}$;

iii. $L^d(w) := L^\infty(w)$, for each $w \in V^d$.

We shall denote with C_d the concept description corresponding to G^d. Then C_d is called the *unravelling up to depth* d of C. $\diamondsuit$

It is easy to see that C is equivalent to an $\mathcal{EL}$-concept description if and only if its $\mathcal{EL}$-description graph does not contain cycles. Consequently, for each $d \in \mathbb{N}$, C_d is equivalent to an $\mathcal{EL}$-concept description.

3.27 Example As an example to illustrate these definitions, let us consider the concept description

$$C = (A, \{ A \equiv \exists \mathsf{r}.D, D \equiv \mathsf{B} \sqcap \exists \mathsf{s}.A \}),$$

where B is a concept name. In Figure 4 the description graph of C and its unravelling are depicted.

Let us compute the concept description C_3, the unravelling of C up to depth 3. For this, we use the unravelling of the description graph of C as shown in Figure 4, and cut it at depth 3. We obtain

$$C_3 = \exists \mathsf{r}.(\mathsf{B} \sqcap \exists \mathsf{s}.\exists \mathsf{r}.\mathsf{B}).$$ $\diamondsuit$

Now, the results we need for our further considerations are the following.

3.28 Lemma (Lemma 5.3 of [11]) *Let C be an $\mathcal{EL}^\bot_{\mathrm{gfp}}$-concept description and $d \in \mathbb{N}$. Then $C \sqsubseteq C_d$.*

3.29 Lemma (Lemma 5.5 of [11]) *Let $\mathcal{I}$ be a finite interpretation. Then there exists a $d \in \mathbb{N}$ such that $C_d^\mathcal{I} = C^\mathcal{I}$ is true for each $\mathcal{EL}^\bot_{\mathrm{gfp}}$-concept description C.*

Lemma 5.5 of [11] also gives a formula to compute the number d. However, we are not interested in this formula here and shall not go into further detail here.

4 A Base for Confident GCIs

The goal of this section is to present a way to effectively obtain bases of confident GCIs of finite interpretations. For this, we shall briefly introduce the notion of *confidence* in Section 4.1

and use it to define *confident GCIs* of finite interpretations $\mathcal{I}$ as those GCIs whose confidence in $\mathcal{I}$ is above a certain, user-defined threshold $c \in [0,1]$. Then, to obtain a base of all those confident GCIs, we shall make use of methods of formal concept analysis. We introduce some necessary machinery in Section 4.2, which allows us to describe a close relationship between formal concept analysis and the description logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$. We then make use of this machinery in Section 4.3 to obtain bases of confident GCIs of $\mathcal{I}$ from bases of certain implications of $\mathbb{K}_{\mathcal{I}}$.

4.1 Confident GCIs of Finite Interpretations

The notion of *confidence* has been introduced in [1] as a measure of "interest" for *association rules*. Translated into the language of formal concept analysis, one can regard association rules simply as implications. Then the notion of confidence of an implication $A \to B$ just is the empirical probability that an object that has all attributes from A also has all attributes from B. See also [20].

This idea of considering this empirical probability fits very well in our plan of considering GCIs which are "almost true." Furthermore, the notion of confidence admits a straight-forward generalization to our setting.

4.1 Definition Let $\mathbb{K}$ be a finite formal context and let $(X \to Y) \in \mathrm{Imp}(M)$. Then its *confidence* $\mathrm{conf}_{\mathbb{K}}(X \to Y)$ is defined as

$$\mathrm{conf}_{\mathbb{K}}(X \to Y) := \begin{cases} 1 & \text{if } X' = \emptyset \\ \frac{|(X \cup Y)'|}{|X'|} & \text{otherwise.} \end{cases}$$

Let $\mathcal{I}$ be a finite interpretation and let C, D be $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions. Then the *confidence* $\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D)$ is defined as

$$\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) := \begin{cases} 1 & \text{if } C^{\mathcal{I}} = \emptyset, \\ \frac{|(C \sqcap D)^{\mathcal{I}}|}{|C^{\mathcal{I}}|} & \text{otherwise.} \end{cases}$$

Let $c \in [0,1]$. We shall denote with $\mathrm{Th}_c(\mathcal{I})$ the set of all implications of $\mathbb{K}$ whose confidence is at least c, and with $\mathrm{Th}_c(\mathcal{I})$ we shall denote the set of all GCI whose confidence is at least c, i. e.

$$\mathrm{Th}_c(\mathcal{I}) := \{\, C \sqsubseteq D \mid C, D \text{ some } \mathcal{EL}^{\bot}_{\mathrm{gfp}}\text{-concept descriptions}, \mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) \geqslant c \,\}. \qquad \diamondsuit$$

Note that $\mathrm{Th}(\mathcal{I}) \subseteq \mathrm{Th}_c(\mathcal{I})$, and that $\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) = 1$ if and only if $C \sqsubseteq D$ holds in $\mathcal{I}$. Also note that contrary to the case of $\mathrm{Th}(\mathcal{I})$, the set $\mathrm{Th}_c(\mathcal{I})$ is not necessarily closed under entailment.

The idea is now to consider the set $\mathrm{Th}_c(\mathcal{I})$ of GCIs instead of $\mathrm{Th}(\mathcal{I})$ for our construction of terminological axioms from $\mathcal{I}$. To make this approach reasonable, we need a finite representation of $\mathrm{Th}_c(\mathcal{I})$, i. e. a base. In this particular case, it may also be interesting to look for special bases where all GCIs have confidence at least c. This is because those GCIs may be of most interest to the ontology engineer.

4.2 Definition Let $c \in [0,1]$. Let $\mathbb{K}$ be a finite formal context. A set $\mathcal{L} \subseteq \mathrm{Imp}(M)$ is called a *confident base* of $\mathrm{Th}_c(\mathbb{K})$ if and only if $\mathcal{L}$ is a base of $\mathrm{Th}_c(\mathbb{K})$ and $\mathcal{L} \subseteq \mathrm{Th}_c(\mathbb{K})$.

Let $\mathcal{I}$ be a finite interpretation. Then a set $\mathcal{B}$ of GCIs is called a *confident base* of $\mathrm{Th}_c(\mathcal{I})$ if and only if $\mathcal{B}$ is a base of $\mathrm{Th}_c(\mathcal{I})$ and $\mathcal{B} \subseteq \mathrm{Th}_c(\mathcal{I})$. $\diamondsuit$

Note that in the case of $c = 1$, bases of $\mathrm{Th}(\mathcal{I}) = \mathrm{Th}_1(\mathcal{I})$ are always confident bases of $\mathrm{Th}(\mathcal{I})$ as well.

4.2 Projections and Induced Contexts

The main purpose of this section is to introduce the notions of *projections* and *induced contexts*. These notions are important for our further discussions because it forms the basis of connection the description logic $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ and formal concept analysis.

Projections have been introduced in [5, 6, 11]. The main idea behind its definition is the following: given a finite interpretation $\mathcal{I}$, we are mainly interested in its model-based most-specific concept description. To make methods from formal concept analysis applicable, we shall construct a special formal context $\mathbb{K}_{\mathcal{I}}$, whose set of attributes will be a set of certain concept descriptions. Then, if we have given another concept description C, we would like to "approximate" this concept description in terms of attributes of $\mathbb{K}$. By approximation we mean that we want to find a set $N \subseteq M$ such that $C \sqsubseteq \bigsqcap_{V \in N} V$ is true "as good as possible." Of course, such a set can readily be defined by

$$N := \{ V \in M \mid C \sqsubseteq V \}.$$

This is exactly the definition of *projections*.

4.3 Definition Let M be a set of concept descriptions and let C be another concept description. Then the *projection* $\mathrm{pr}_M(C)$ *of* C *onto* M is defined as

$$\mathrm{pr}_M(C) := \{ D \in M \mid C \sqsubseteq D \}. \qquad \diamondsuit$$

As projections allow us to approximate C in terms of M, conjunction $U \mapsto \bigsqcap_{V \in U} V$ allows us to go the way back, i. e. from sets of concept descriptions to concept descriptions. For brevity, let us define for $U \subseteq M$

$$\bigsqcap U := \begin{cases} \top & \text{if } U = \varnothing \\ \bigsqcap_{V \in U} V & \text{otherwise} \end{cases}$$

We also want to lift this definition to sets of implications. Let us define for a set $\mathcal{L}$ of implications the set of GCIs

$$\bigsqcap \mathcal{L} := \{ \bigsqcap X \sqsubseteq \bigsqcap Y \mid (X \to Y) \in \mathcal{L} \}.$$

It now turns out that the mappings $C \mapsto \mathrm{pr}_M(C)$ and $U \mapsto \bigsqcap U$ satisfy the main condition of a Galois connection. But note again that since $\sqsubseteq$ does not constitute an order relation on the set of all $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions, the aforementioned mappings actually cannot be a Galois connection.

4.4 Lemma *Let M be a set of concept descriptions. Then for each $U \subseteq M$ and for each concept description C it is true that*

$$C \sqsubseteq \bigsqcap U \iff U \subseteq \mathrm{pr}_M(C).$$

Proof Let us first show the direction from left to right. From $C \sqsubseteq \bigsqcap U$ we can conclude $\mathrm{pr}_M(\bigsqcap U) \subseteq \mathrm{pr}_M(C)$, since every concept description $D \in M$ satisfying $\bigsqcap U \sqsubseteq D$ also satisfies $C \sqsubseteq D$. Furthermore, for each $F \in U$ we have $\bigsqcap U \sqsubseteq F$, therefore $U \subseteq \mathrm{pr}_M(\bigsqcap U)$ and hence

$$U \subseteq \mathrm{pr}_M(\bigsqcap U) \subseteq \mathrm{pr}_M(C)$$

as desired.

For the other direction let us suppose that $U \subseteq \mathrm{pr}_M(C)$. Then $\bigsqcap U \sqsupseteq \bigsqcap \mathrm{pr}_M(C)$. Now $C \sqsubseteq \mathrm{pr}_M(C)$ is true as well, as we have already argued.Therefore,

$$C \sqsubseteq \bigsqcap \mathrm{pr}_M(C) \sqsubseteq \bigsqcap U$$

as desired. □

We have introduced $\mathrm{pr}_M(C)$ as an approximation of the concept description C in terms of M. Occasionally, it may happen that this approximation is as good as possible, i. e. that the approximation $\mathrm{pr}_M(C)$ indeed describes C completely. We shall capture this situation in the following definition.

4.5 Definition Let M be a set of concept descriptions and let C be another concept description. We say that C is *expressible in terms of* M if and only if there exists a subset $N \subseteq M$ such that $C \equiv \bigsqcap N$. ♢

Unsurprisingly, expressibility in terms of M can be characterized easily using projections, as the following result shows.

4.6 Proposition *Let M be a set of $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions and let C be an $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description. Then C is expressible in terms of M if and only if*

$$C \equiv \bigsqcap \mathrm{pr}_M(C).$$

Proof If $C \equiv \bigsqcap \mathrm{pr}_M(C)$, then clearly C is expressible in terms of M. Conversely, let $N \subseteq M$ such that $C \equiv \bigsqcap N$. Then $C \sqsubseteq D$ for each $D \in N$ and hence

$$N \subseteq \mathrm{pr}_M(C),$$

which implies $C \sqsupseteq \bigsqcap \mathrm{pr}_M(C)$. On the other hand, $C \sqsubseteq \bigsqcap \mathrm{pr}_M(C)$ by Lemma 4.4 and hence $C \equiv \bigsqcap \mathrm{pr}_M(C)$ follows as required. □

One of the crucial observations of [11] is that we can explicitly describe a set $M_{\mathcal{I}}$ that is able to express all model-based most-specific concept descriptions of $\mathcal{I}$:

$$M_{\mathcal{I}} := \{ \perp \} \cup N_C \cup \{ \exists r.X^{\mathcal{I}} \mid X \subseteq \Delta^{\mathcal{I}}, X \neq \varnothing \}.$$

4.7 Theorem (Lemma 5.9 from [11]) *Let $\mathcal{I}$ be a finite interpretation and let C be a concept description. Then $C^{\mathcal{I}\mathcal{I}}$ is expressible in terms of $M_{\mathcal{I}}$.*

The definitions and results given so far allow us to formulate one of the main results of [11], which is an explicit description of a finite base of $\mathcal{I}$.

4.8 Theorem (Theorem 5.10 of [11]) *Let $\mathcal{I}$ be a finite interpretation. Then the set*

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

is a finite base for $\mathcal{I}$.

We now turn out attention to the notion of *induced contexts*, as they are define in [11]. Using a special induced context $\mathbb{K}_{\mathcal{I}}$ for the interpretation $\mathcal{I}$, we shall be able to derive a close relationship of the model-based most-specific concept descriptions of $\mathcal{I}$ and the intents of $\mathbb{K}_{\mathcal{I}}$.

4.9 Definition Let $\mathcal{I}$ be a finite interpretation and let M be a set of concept descriptions. Define the formal context $\mathbb{K}_{\mathcal{I},M} := (\Delta^{\mathcal{I}}, M, \nabla)$, where

$$x \nabla C \iff x \in C^{\mathcal{I}}$$

for all $x \in \Delta^{\mathcal{I}}$ and $C \in M$. The formal context $\mathbb{K}_{\mathcal{I},M}$ is the *induced formal context* of $\mathcal{I}$ and M. If $M = M_{\mathcal{I}}$, we write $\mathbb{K}_{\mathcal{I}}$ instead of $\mathbb{K}_{\mathcal{I},M}$ and call this *induced formal context* of $\mathcal{I}$. ♢

Induced context play a crucial role in combining formal concept analysis and $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$. In particular, they allow us to reduce the size of the base described in Theorem 4.8 as much as possible.

4.10 Theorem (Corollary 5.13 and Theorem 5.18 of [11]) *Let $\mathcal{I}$ be a finite interpretation and define*

$$S_{\mathcal{I}} := \{\, \{\, C \,\} \to \{\, D \,\} \mid C, D \in M_{\mathcal{I}}, C \sqsubseteq D \,\}.$$

Then the set $\mathcal{B}_{\mathrm{Can}}$ defined by

$$\mathcal{B}_{\mathrm{Can}} := \{\, \bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}} \mid (U \to U'') \in \mathrm{Can}(\mathbb{K}_{\mathcal{I}}, S_{\mathcal{I}}) \,\}$$

is a base of $\mathcal{I}$ of minimal cardinality.

Projections and induced formal context allow to express very close relationships between operations in $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ and in $\mathbb{K}_{\mathcal{I},M}$ for suitable choices of M. In particular, we can express the extensions of concept descriptions C in $\mathcal{I}$, which are expressible in terms of M, as $\mathrm{pr}_M(C)'$ in $\mathbb{K}_{\mathcal{I},M}$ for some set $N \subseteq M$. In addition, if we can express for $X \subseteq \Delta^{\mathcal{I}}$ its model-based most-specific concept descriptions $X^{\mathcal{I}}$ in terms of M, we are also able to represent these concept description $X^{\mathcal{I}}$ as $\bigsqcap X'$.

We formulate these relationships in the following two propositions. They already appear in [11].

4.11 Proposition (Lemma 4.11 and Lemma 4.12 from [11]) *Let $\mathcal{I}$ be a finite interpretation and M a set of concept descriptions. Let C be a concept description expressible in terms of M. Then*

$$C^{\mathcal{I}} = \mathrm{pr}_M(C)'$$

where the derivation are computed within the induced context of $\mathcal{I}$ and M. Furthermore, every set $O \subseteq \Delta^{\mathcal{I}}$ satisfies

$$O' = \mathrm{pr}_M(O^{\mathcal{I}}).$$

Proof Since C is expressible in terms of M, $C \equiv \bigsqcap \mathrm{pr}_M(C)$ by Proposition 4.6. Therefore

$$\begin{aligned} x \in C^{\mathcal{I}} &\iff x \in (\bigsqcap \mathrm{pr}_M(C))^{\mathcal{I}} \\ &\iff \forall D \in \mathrm{pr}_M(C) : x \in D^{\mathcal{I}} \\ &\iff x \in \mathrm{pr}_M(C)' \end{aligned}$$

as $\mathrm{pr}_M(C)' = \{\, x \in \Delta^{\mathcal{I}} \mid \forall D \in \mathrm{pr}_M(C) : x \in D^{\mathcal{I}} \,\}$.

For the second claim we observe

$$\begin{aligned} D \in O' &\iff \forall g \in O \colon g \in D^{\mathcal{I}} \\ &\iff O \subseteq D^{\mathcal{I}} \\ &\iff O^{\mathcal{I}} \sqsubseteq D \\ &\iff D \in \mathrm{pr}_M(O^{\mathcal{I}}), \end{aligned}$$

where $O \subseteq D^{\mathcal{I}} \iff O^{\mathcal{I}} \sqsubseteq D$ holds due to Lemma 3.21. □

4.12 Proposition (Lemma 4.10 and 4.11 from [11]) *Let $\mathcal{I}$ be a finite interpretation and let M be a set of concept descriptions. Then each $B \subseteq M$ satisfies*

$$B' = (\bigsqcap B)^{\mathcal{I}},$$

where the derivations are computed in $\mathbb{K}_{\mathcal{I},M}$.

Let $A \subseteq \Delta^{\mathcal{I}}$. *If* $A^{\mathcal{I}}$ *is expressible in terms of* M, *then*

$$\bigsqcap A' \equiv A^{\mathcal{I}}.$$

Proof Remember that an object $x \in \Delta^{\mathcal{I}}$ has an attribute $D \in M$ if and only if $x \in D^{\mathcal{I}}$. Hence

$$x \in B' \iff \forall D \in B : x \in D^{\mathcal{I}} \iff x \in (\bigsqcap B)^{\mathcal{I}}.$$

Let $A \subseteq \Delta^{\mathcal{I}}$ such that $A^{\mathcal{I}}$ is expressible in terms of M. By Proposition 4.6,

$$A^{\mathcal{I}} \equiv \bigsqcap \operatorname{pr}_M(A^{\mathcal{I}}).$$

By Proposition 4.11, $\operatorname{pr}_M(A^{\mathcal{I}}) = A'$ and hence the claim follows. □

In Theorem 4.14 we shall precisely formulate a connection between the model-based most-specific concept descriptions of $\mathcal{I}$ and the intents of $\mathbb{K}_{\mathcal{I}}$. To prove this connection, we shall make use of the following proposition.

4.13 Proposition *Let* $\mathcal{I}$ *be a finite interpretation and let* $X \subseteq M_{\mathcal{I}}$. *Then*

$$X \subseteq \operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap X) \subseteq X'',$$

where the derivation is computed in $\mathbb{K}_{\mathcal{I}}$.

Proof By Lemma 4.4, $X \subseteq \operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap X)$ holds. Now

$$\begin{aligned} D \in \operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap X) &\iff \bigsqcap X \sqsubseteq D \\ &\implies (\bigsqcap X)^{\mathcal{I}} \subseteq D^{\mathcal{I}} \\ &\iff X' \subseteq \{ D \}' \\ &\iff X'' \supseteq \{ D \}'' \ni D \\ &\implies D \in X'' \end{aligned}$$

as required. □

Having defined the formal context $\mathbb{K}_{\mathcal{I}}$, we are now going to show that this formal context indeed allows us to view model-based most-specific concept descriptions as intents of a formal context. We have already seen that all model-based most-specific concept descriptions are expressible in terms of $M_{\mathcal{I}}$, the set of attributes of $\mathbb{K}_{\mathcal{I}}$. It is therefore not surprising that the lattice of intents of $\mathbb{K}_{\mathcal{I}}$ and the equivalence classes of model-based most-specific concept descriptions ordered by $\sqsupseteq$ are order-isomorphic.

Before we prove the following theorem, we have to deal with a technical detail. This is because model-based most-specific concept descriptions are only unique up to equivalence. In particular, $\sqsubseteq$ is in general not an order relation on the set of all model-based most-specific concept descriptions. To overcome this we use the standard approach of considering classes of equivalent concept descriptions instead.

Let M be a set of concept descriptions. Then let us define

$$M/_{\equiv} := \{ [X] \mid X \in M \}$$

where

$$[X] := \{ Y \in M \mid X \equiv Y \}.$$

Furthermore, for $X, Y \in M$ we set

$$[X] \sqsubseteq [Y] \iff X \sqsubseteq Y.$$

Note that this is well-defined because if $\hat{X} \in [X], \hat{Y} \in [Y]$, then $\hat{X} \equiv X, \hat{Y} \equiv Y$ and hence $X \sqsubseteq Y \iff \hat{X} \sqsubseteq \hat{Y}$. With this definition it is easy to see that $({}^{M}\!/\!_{\equiv}, \sqsubseteq)$ is an ordered set.

4.14 Theorem *Let $\mathcal{I}$ be a finite interpretation and let $\mathcal{M}$ be the set of all model-based most-specific concept descriptions of $\mathcal{I}$. Then the mappings*

$$\bigsqcap : \mathfrak{P}(M_{\mathcal{I}}) \to \mathcal{M} \qquad \text{and} \qquad \mathrm{pr}_{M_{\mathcal{I}}} : \mathcal{M} \to \mathfrak{P}(M_{\mathcal{I}})$$

describe an order-isomorphism between the ordered sets $(\mathfrak{P}(M_{\mathcal{I}}), \subseteq)$ and $({}^{\mathcal{M}}\!/\!_{\equiv}, \sqsupseteq)$ via

$$\begin{array}{rccc} \varphi : & \mathfrak{P}(M_{\mathcal{I}}) & \to & {}^{\mathcal{M}}\!/\!_{\equiv} \\ & N & \mapsto & [\bigsqcap N] \end{array}$$

and $\varphi^{-1}([X]) = \mathrm{pr}_{M_{\mathcal{I}}}(X)$. More precisely, the following statements hold:

i. $\bigsqcap U \in \mathcal{M}$ *for each* $U \in \mathrm{Int}(\mathbb{K}_{\mathcal{I}})$.

ii. $\mathrm{pr}_{M_{\mathcal{I}}}(C) \in \mathrm{Int}(\mathbb{K}_{\mathcal{I}})$ *for each* $C \in \mathcal{M}$.

iii. $U \subseteq V$ *implies* $\mathrm{pr}_{M_{\mathcal{I}}}(U) \sqsupseteq \mathrm{pr}_{M_{\mathcal{I}}}(V)$ *for all* $U, V \subseteq M_{\mathcal{I}}$.

iv. $C \sqsubseteq D$ *implies* $\bigsqcap C \supseteq \bigsqcap D$ *for all* $C, D \in \mathcal{M}$.

v. $\mathrm{pr}_{M_{\mathcal{I}}}(\bigsqcap U) = U$ *for each* $U \in \mathrm{Int}(\mathbb{K}_{\mathcal{I}})$.

vi. $\bigsqcap \mathrm{pr}_{M_I}(C) \equiv C$ *for each* $C \in \mathcal{M}$.

Additionally, $U'' = \mathrm{pr}_{M_{\mathcal{I}}}((\bigsqcap U)^{\mathcal{I}\mathcal{I}})$ and $C^{\mathcal{I}\mathcal{I}} \equiv \bigsqcap(\mathrm{pr}_{M_{\mathcal{I}}}(C))''$ for each set $U \subseteq M_{\mathcal{I}}$ and each concept description C expressible in terms of $M_{\mathcal{I}}$, where the derivations are computed in $\mathbb{K}_{\mathcal{I}}$.

Proof We show each claim step by step.

For i, let $U \in \mathrm{Int}(\mathbb{K}_{\mathcal{I}})$, i. e. $U = U''$. Then

$$\bigsqcap U = \bigsqcap U'' \equiv (U')^{\mathcal{I}} = (\bigsqcap U)^{\mathcal{I}\mathcal{I}}$$

by Proposition 4.12. Hence $\bigsqcap U \equiv (\bigsqcap U)^{\mathcal{I}\mathcal{I}}$ and therefore $\bigsqcap U \in \mathcal{M}$.

For ii, let $C \in \mathcal{M}$, i. e. $C \equiv C^{\mathcal{I}\mathcal{I}}$. By Theorem 4.7, C is expressible in terms of $M_{\mathcal{I}}$ and hence by Proposition 4.11

$$\begin{aligned} \mathrm{pr}_{M_{\mathcal{I}}}(C) &= \mathrm{pr}_{M_{\mathcal{I}}}(C^{\mathcal{I}\mathcal{I}}) \\ &= (C^{\mathcal{I}})' \\ &= (\mathrm{pr}_{M_{\mathcal{I}}}(C))'', \end{aligned}$$

thus $\mathrm{pr}_{M_{\mathcal{I}}}(C) \in \mathrm{Int}(\mathbb{K}_{\mathcal{I}})$.

Claims iii and iv are already contained in Lemma 4.4.

For v we need to show that

$$\mathrm{pr}_{M_{\mathcal{I}}}(\bigsqcap U) = U$$

for $U \in \operatorname{Int}(\mathbb{K}_{\mathcal{I}})$. By Proposition 4.13, $U \subseteq \operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap U) \subseteq U''$, and since $U = U''$, equality follows. Claim vi follows from Proposition 4.6, as $C \in \mathcal{M}$ is expressible in terms of $M_{\mathcal{I}}$ by Theorem 4.7.

Finally for $U \subseteq M_{\mathcal{I}}$

$$\begin{aligned} \operatorname{pr}_{M_{\mathcal{I}}}((\bigsqcap U)^{\mathcal{I}\mathcal{I}}) &= \operatorname{pr}_{M_{\mathcal{I}}}((U')^{\mathcal{I}}) \\ &= U'' \end{aligned}$$

by Proposition 4.12 and Proposition 4.11, and

$$\begin{aligned} \bigsqcap(\operatorname{pr}_{M_{\mathcal{I}}}(C))'' &\equiv (\operatorname{pr}_{M_{\mathcal{I}}}(C)')^{\mathcal{I}} \\ &= C^{\mathcal{I}\mathcal{I}} \end{aligned}$$

for every $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description C, again by Proposition 4.11 and Proposition 4.12. □

An immediate consequence of this theorem is the following.

4.15 Corollary *Let $\mathcal{I}$ be a finite interpretation. Then for each $A \subseteq M_{\mathcal{I}}$ it is true that*

$$(\bigsqcap A)^{\mathcal{I}\mathcal{I}} = \bigsqcap A'',$$

where the derivations are done in $\mathbb{K}_{\mathcal{I}}$.

Proof It is true that $\bigsqcap A$ is expressible in terms of $M_{\mathcal{I}}$, therefore Theorem 4.14 yields

$$(\bigsqcap A)^{\mathcal{I}\mathcal{I}} = \bigsqcap(\operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap A))''.$$

Now by Proposition 4.11 we have $\operatorname{pr}_{M_{\mathcal{I}}}(\bigsqcap A)' = (\bigsqcap A)^{\mathcal{I}}$. Therefore

$$\begin{aligned} (\bigsqcap A)^{\mathcal{I}\mathcal{I}} &= \bigsqcap((\bigsqcap A)^{\mathcal{I}})' \\ &= \bigsqcap A'', \end{aligned}$$

since $(\bigsqcap A)^{\mathcal{I}} = A'$ by Proposition 4.12. □

4.3 Computing Confident Bases of Finite Interpretations

Building upon the results of the previous sections, we are now able to describe a first confident base of $\operatorname{Th}_c(\mathcal{I})$ for arbitrary choices of $c \in [0,1]$. For this, we shall make use of results of [9], which itself uses ideas from Luxenburger [14]. As [9] already gives a thorough introduction and motivation of Luxenburger's results, we shall not repeat it here. Instead, we shall extend the results obtained in [9] by the result of Theorem 4.20.

Roughly speaking, the ideas by Luxenburger applied to our setting of confident GCIs can be formulated as follows. We consider the partition $\operatorname{Th}_c(\mathcal{I}) = \operatorname{Th}(\mathcal{I}) \cup (\operatorname{Th}_c(\mathcal{I}) \setminus \operatorname{Th}(\mathcal{I}))$ and try to separately find a base for $\operatorname{Th}(\mathcal{I})$ and a confident base for $\operatorname{Th}_c(\mathcal{I}) \setminus \operatorname{Th}(\mathcal{I})$. Of course, a base $\mathcal{B}$ of $\operatorname{Th}(\mathcal{I})$ has already been given by Distel [11], so it remains to find a confident base of $\operatorname{Th}_c(\mathcal{I}) \setminus \operatorname{Th}(\mathcal{I})$.

To achieve this we use the following observation from Luxenburger, translated to the language of description logics: if $(C \sqsubseteq D) \in \operatorname{Th}_c(\mathcal{I}) \setminus \operatorname{Th}(\mathcal{I})$, it is true that

$$\mathcal{B} \cup \{\, C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}} \,\} \models (C \sqsubseteq D),$$

because $\mathcal{B} \models (C \sqsubseteq C^{\mathcal{I}\mathcal{I}})$ and $\varnothing \models (D^{\mathcal{I}\mathcal{I}} \sqsubseteq D)$ (note that $C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$ always holds in $\mathcal{I}$.) Therefore, it suffices to consider only GCIs of the form $C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}$.

Let us define

$$\mathrm{Conf}(\mathcal{I}, c) := \{\, X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}} \mid Y, X \subseteq \Delta^{\mathcal{I}} \text{ and } \mathrm{conf}_{\mathcal{I}}(X^{\mathcal{I}} \sqsubseteq Y^{\mathcal{I}}) \in [c, 1) \,\}.$$

Then we can formulate the following result.

4.16 Theorem *Let $\mathcal{I}$ be a finite interpretation, let $c \in [0,1]$ and let $\mathcal{B}$ be a base of $\mathcal{I}$. Then $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c)$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$.*

Proof Clearly $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c) \subseteq \mathrm{Th}_c(\mathcal{I})$ and it only remains to be shown that $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c)$ entails all GCIs with confidence at least c in $\mathcal{I}$.

Let $C \sqsubseteq D$ be a GCI with $\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) \geqslant c$. We have to show that $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c) \models C \sqsubseteq D$. If $C \sqsubseteq D$ is already valid in $\mathcal{I}$, then $\mathcal{B} \models C \sqsubseteq D$ and nothing remains to be shown. We therefore assume that $\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) \neq 1$.

As $C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$ is valid in $\mathcal{I}$, $\mathcal{B} \models C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$. Furthermore, $\mathrm{conf}_I(C \sqsubseteq D) = \mathrm{conf}_{\mathcal{I}}(C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}})$ and hence $(C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}) \in \mathrm{Conf}(\mathcal{I}, c)$. Finally, $\varnothing \models D^{\mathcal{I}\mathcal{I}} \sqsubseteq D$. We therefore obtain

$$\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c) \models C \sqsubseteq C^{\mathcal{I}\mathcal{I}}, C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}, D^{\mathcal{I}\mathcal{I}} \sqsubseteq D$$

and hence $\mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c) \models C \sqsubseteq D$ as required. □

It is not hard to see that the prerequisites of the previous theorem can be weakened in the following way: instead of considering the whole set $\mathrm{Conf}(\mathcal{I}, c)$, it is sufficient to choose a base $\mathcal{C} \subseteq \mathrm{Conf}(\mathcal{I}, c)$ of $\mathrm{Conf}(\mathcal{I}, c)$, since then

$$\mathcal{B} \cup \mathcal{C} \models \mathcal{B} \cup \mathrm{Conf}(\mathcal{I}, c).$$

Furthermore, it is not necessary for $\mathcal{B}$ to be a base of $\mathcal{I}$. Instead, one can choose a set $\hat{\mathcal{B}}$ of valid GCIs such that $\hat{\mathcal{B}} \cup \mathcal{C}$ is complete for $\mathcal{I}$, because then

$$\hat{\mathcal{B}} \cup \mathcal{C} \models \mathcal{B} \cup \mathcal{C}.$$

4.17 Corollary *Let $\mathcal{I}$ be a finite interpretation, $c \in [0,1]$. Let $\mathcal{C} \subseteq \mathrm{Conf}(\mathcal{I}, c)$ be a base of $\mathrm{Conf}(\mathcal{I}, c)$ and let $\mathcal{B} \subseteq \mathrm{Th}(\mathcal{I})$ such that $\mathcal{B} \cup \mathcal{C}$ is complete for $\mathcal{I}$. Then $\mathcal{B} \cup \mathcal{C}$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$.*

Now, this results allows us to describe confident bases of $\mathrm{Th}_c(\mathcal{I})$ in a very simple way: if $\mathcal{L}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$, then the set

$$\{ \textstyle\bigsqcap X \sqsubseteq \bigsqcap Y \mid (X \to Y) \in \mathcal{L} \}$$

is a confident base of $\mathrm{Th}_c(\mathcal{I})$. This is the content of Theorem 4.20, which we shall prepare with the following lemmas.

4.18 Lemma *Let M be a set of concept descriptions and let $\mathcal{L} \subseteq \mathrm{Imp}(M)$ and $(X \to Y) \in \mathrm{Imp}(M)$. Then $\mathcal{L} \models (X \to Y)$ implies $\bigsqcap \mathcal{L} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$.*

Proof Let $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ be an interpretation such that $\mathcal{J} \models \bigsqcap \mathcal{L}$. Recall that we denote with $\mathbb{K}_{\mathcal{J},M}$ the formal context induced by M and $\mathcal{J}$. We shall show that $\mathbb{K}_{\mathcal{J},M} \models \mathcal{L}$. This then implies $\mathbb{K}_{\mathcal{J},M} \models (X \to Y)$ and from this we shall infer that $\mathcal{J} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$, as required.

Let $(E \to F) \in \mathcal{L}$. Then $(\bigsqcap E)^{\mathcal{J}} \subseteq (\bigsqcap F)^{\mathcal{J}}$ since $\mathcal{J} \models \bigsqcap \mathcal{L}$. By Proposition 4.12, $(\bigsqcap E)^{\mathcal{J}} = E'$, where the derivation is done in $\mathbb{K}_{\mathcal{J},M}$. Therefore, $E' \subseteq F'$ is true in $\mathbb{K}_{\mathcal{J},M}$ and hence $\mathbb{K}_{\mathcal{J},M} \models \mathcal{L}$. Since $\mathcal{L} \models (X \to Y)$, we obtain $X' \subseteq Y'$. Again by Proposition 4.12 we obtain $(\bigsqcap X)^{\mathcal{J}} \subseteq (\bigsqcap Y)^{\mathcal{J}}$ and therefore $\mathcal{J} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$. ▫

Note that the converse direction is not true in general, i. e. $\bigsqcap \mathcal{L} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$ does in general not imply $\mathcal{L} \models (X \to Y)$. This fact is illustrated by the following example.

4.19 Example Let $N_C := \{\, \mathsf{A}, \mathsf{B} \,\}, N_R := \{\, \mathsf{r} \,\}$ and $M = \{\, \mathsf{A}, \mathsf{B}, \exists \mathsf{r}.\mathsf{A}, \exists \mathsf{r}.\mathsf{B} \,\}$. Define

$$\begin{aligned}\mathcal{L} &:= \{\, \{\, \mathsf{A} \,\} \to \{\, \mathsf{B} \,\} \,\},\\ X &:= \{\, \exists \mathsf{r}.\mathsf{A} \,\},\\ Y &:= \{\, \exists \mathsf{r}.\mathsf{B} \,\}.\end{aligned}$$

Then clearly $\mathcal{L} \not\models (X \to Y)$, but $\bigsqcap \mathcal{L} = \{\, \mathsf{A} \sqsubseteq \mathsf{B} \,\}$, $(\bigsqcap X \sqsubseteq \bigsqcap Y) = (\exists \mathsf{r}.\mathsf{A} \sqsubseteq \exists \mathsf{r}.\mathsf{B})$ and therefore $\bigsqcap \mathcal{L} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$. ♢

We now prove the main result of this section.

4.20 Theorem *Let $\mathcal{I}$ be a finite interpretation and let $c \in [0, 1]$. Let $\mathcal{L}$ be a confident base of* $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$. *Then the set*

$$\mathcal{B} := \bigsqcap \mathcal{L} = \{ \bigsqcap X \sqsubseteq \bigsqcap Y \mid (X \to Y) \in \mathcal{L} \}$$

is a confident base of $\mathrm{Th}_c(\mathcal{I})$.

Proof We have to show that $\mathcal{B}$ is sound and complete for $\mathrm{Th}_c(\mathcal{I})$, i. e. $\mathcal{B} \subseteq \mathrm{Th}_c(\mathcal{I})$ and for each $(C \sqsubseteq D) \in \mathrm{Th}_c(\mathcal{I})$ follows $\mathcal{B} \models (C \sqsubseteq D)$.

To see that $\mathcal{B}$ is sound let $(\bigsqcap X \to \bigsqcap Y) \in \mathcal{B}$. We have to show that $\mathrm{conf}_{\mathcal{I}}(\bigsqcap X \to \bigsqcap Y) \geqslant c$. To do this, we shall verify

$$\mathrm{conf}_{\mathcal{I}}(\bigsqcap X \sqsubseteq \bigsqcap Y) = \mathrm{conf}_{\mathbb{K}_{\mathcal{I}}}(X \to Y).$$

Let $(\bigsqcap X)^{\mathcal{I}} \neq \varnothing$. Then by Proposition 4.12, $X' \neq \varnothing$ and hence

$$\begin{aligned}\mathrm{conf}_{\mathcal{I}}(\bigsqcap X \sqsubseteq \bigsqcap Y) &= \frac{|(\bigsqcap X \sqcap \bigsqcap Y)^{\mathcal{I}}|}{|(\bigsqcap X)^{\mathcal{I}}|}\\ &= \frac{|(\bigsqcap (X \cup Y))^{\mathcal{I}}|}{|(\bigsqcap X)^{\mathcal{I}}|}\\ &= \frac{|(X \cup Y)'|}{|X'|}\\ &= \mathrm{conf}_{\mathbb{K}_{\mathcal{I}}}(X \to Y).\end{aligned}$$

If $(\bigsqcap X)^{\mathcal{I}} = \varnothing$, then $X' = \varnothing$ and therefore

$$\mathrm{conf}_{\mathcal{I}}(\bigsqcap X \sqsubseteq \bigsqcap Y) = 1 = \mathrm{conf}_{\mathbb{K}_{\mathcal{I}}}(X \to Y).$$

This shows that $\mathcal{B}$ is sound for $\mathrm{Th}_c(\mathcal{I})$.

We shall now show that $\mathcal{B}$ is complete for $\mathrm{Th}_c(\mathcal{I})$. For this we shall show that

i. $\mathcal{B} \models (\bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}})$ for each $U \subseteq M_{\mathcal{I}}$;

ii. $\mathcal{B} \models \mathrm{Conf}(\mathcal{I}, c)$.

Recall from Theorem 4.8 that

$$\mathcal{B}_{\mathcal{I}} = \{ \bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}} \mid U \subseteq M_{\mathcal{I}} \}$$

is a finite base of $\mathcal{I}$. By Theorem 4.16, $\mathcal{B}_{\mathcal{I}} \cup \mathrm{Conf}(\mathcal{I}, c)$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$. If we show the two claims from above, we have shown $\mathcal{B} \models (\mathcal{B}_{\mathcal{I}} \cup \mathrm{Conf}(\mathcal{I}, c))$ which then implies that $\mathcal{B}$ is complete for $\mathrm{Th}_c(\mathcal{I})$ as well.

For the first case let $U \subseteq M_{\mathcal{I}}$. Since $\mathcal{L}$ is a confident base for $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$, it is complete for $\mathbb{K}_{\mathcal{I}}$ and therefore

$$\mathcal{L} \models (U \to U'').$$

Then Lemma 4.18 yields

$$\mathcal{B} \models (\bigsqcap U \sqsubseteq \bigsqcap U'')$$

and Corollary 4.15 implies $\bigsqcap U'' \equiv (\bigsqcap U)^{\mathcal{I}\mathcal{I}}$, hence

$$\mathcal{B} \models (\bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}})$$

as desired.

For the second case let $(U^{\mathcal{I}} \sqsubseteq V^{\mathcal{I}}) \in \mathrm{Conf}(\mathcal{I}, c)$, i. e. $U, V \subseteq \Delta^{\mathcal{I}}$ and $\mathrm{conf}_{\mathcal{I}}(U^{\mathcal{I}} \sqsubseteq V^{\mathcal{I}}) \in [c, 1)$. By Proposition 4.12, $U^{\mathcal{I}} \equiv \bigsqcap U'$ and $V^{\mathcal{I}} \equiv \bigsqcap V'$, since $U^{\mathcal{I}}, V^{\mathcal{I}}$ are expressible in terms of $M_{\mathcal{I}}$ by Theorem 4.7. Therefore,

$$\mathcal{B} \models (U^{\mathcal{I}} \sqsubseteq V^{\mathcal{I}}) \iff \mathcal{B} \models (\bigsqcap U' \sqsubseteq \bigsqcap V').$$

As before we see that

$$\mathrm{conf}_{\mathcal{I}}(\bigsqcap U' \sqsubseteq \bigsqcap V') = \mathrm{conf}_{\mathbb{K}_{\mathcal{I}}}(U' \to V').$$

Since $\mathcal{L}$ is a confident base for $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$ it is true that $\mathcal{L} \models U' \to V'$, hence $\mathcal{B} \models (\bigsqcap U' \sqsubseteq \bigsqcap V')$ by Lemma 4.18 and therefore $\mathcal{B} \models (U^{\mathcal{I}} \sqsubseteq V^{\mathcal{I}})$. □

Since $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$ is a confident base of itself, we immediately obtain that $\bigsqcap \mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$.

In addition to entailing all confident GCIs, a base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$ also compromises the knowledge about when two concept descriptions $C, D \in M_{\mathcal{I}}$ subsume each other. If $C \sqsubseteq D$, then the corresponding implication $\{ C \} \to \{ D \}$ always holds in $\mathbb{K}_{\mathcal{I}}$ and is therefore entailed by any base $\mathcal{L}$ of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$. However, this knowledge is not needed in the base $\bigsqcap \mathcal{L}$ of $\mathrm{Th}_c(\mathcal{I})$, and therefore may cause some redundancies in the base $\bigsqcap \mathcal{L}$. The following result shows that at least these redundancies can be avoided.

4.21 Corollary *Let $\mathcal{I}$ be a finite interpretation, $c \in [0, 1]$ and let $\mathcal{L} \subseteq \mathrm{Imp}(\mathbb{K}_{\mathcal{I}})$ be such that $\mathcal{L} \cup S_{\mathcal{I}}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}})$. Then $\bigsqcap \mathcal{L}$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$.*

Proof By Theorem 4.20, $\bigsqcap \mathcal{L} \sqcap \bigsqcap S_{\mathcal{I}} = \bigsqcap(\mathcal{L} \cup S_{\mathcal{I}})$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$. Since

$$\bigsqcap S_{\mathcal{I}} = \{ C \sqsubseteq D \mid C, D \in M_{\mathcal{I}}, C \sqsubseteq D \}$$

is valid in every interpretation, the set $\bigsqcap \mathcal{L}$ is already a base of $\mathrm{Th}_c(\mathcal{I})$. □

More generally, if a base $\mathcal{L}$ of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$ is redundant, then there exists an implication $(X \to Y) \in \mathcal{L}$ such that

$$\mathcal{L}\backslash\{ X \to Y \} \models X \to Y.$$

By Lemma 4.18 this yields

$$\bigsqcap \mathcal{L}\backslash\{ \bigsqcap X \sqsubseteq \bigsqcap Y \} \models (\bigsqcap X \sqsubseteq \bigsqcap Y),$$

i. e. redundancies in a set of implications yield redundancies in the corresponding set of GCIs. Therefore, removing these redundancies is a good starting point for reducing the size of the result set of GCIs.

One way to do this is to use the results from Section 2.4 in the following way: if $\mathcal{L}$ is a (confident) base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, then we can reduce the size of the base by considering $\mathrm{Can}(\mathcal{L})$ instead. As $\mathrm{Cn}(\mathcal{L}) = \mathrm{Cn}(\mathrm{Can}(\mathcal{L}))$, we know that $\mathrm{Can}(\mathcal{L})$ is a base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$. Also in this case does the set $\bigsqcap \mathrm{Can}(\mathcal{L})$ yield a base of $\mathrm{Th}_c(\mathcal{I})$.

4.22 Corollary *Let $\mathcal{I}$ be a finite interpretation, $c \in [0, 1]$ and let $\mathcal{K} \subseteq \mathrm{Imp}(\mathbb{K}_\mathcal{I})$ be a base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$. Then $\bigsqcap \mathcal{K}$ is a base of $\mathrm{Th}_c(\mathcal{I})$.*

Proof By the observation after Theorem 4.20 we know that $\bigsqcap \mathrm{Th}_c(\mathbb{K}_\mathcal{I})$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$. As $\mathcal{K}$ is a base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, from Lemma 4.18 we infer that $\bigsqcap \mathcal{K}$ is also a base of $\bigsqcap \mathrm{Th}_c(\mathbb{K}_\mathcal{I})$. Hence, $\bigsqcap \mathcal{K}$ entails all GCIs from $\mathrm{Th}_c(\mathcal{I})$. Therefore, $\bigsqcap \mathcal{K}$ is complete for $\mathrm{Th}_c(\mathcal{I})$.

Conversely, as $\mathcal{K}$ is a base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$ is also a base of $\mathcal{K}$. But then $\mathcal{K}$ is entailed by $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, i. e. $\bigsqcap \mathcal{K}$ is entailed by $\bigsqcap \mathrm{Th}_c(\mathbb{K}_\mathcal{I}) \subseteq \mathrm{Th}_c(\mathcal{I})$. Therefore, $\bigsqcap \mathcal{K}$ is sound for $\mathrm{Th}_c(\mathcal{I})$ and hence a base of $\mathrm{Th}_c(\mathcal{I})$. □

However, the approach of considering $\mathrm{Can}(\mathcal{L})$ instead of $\mathcal{L}$ has the drawback that we cannot guarantee anymore that $\mathrm{Can}(\mathcal{L}) \subseteq \mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, i. e. that $\bigsqcap \mathrm{Can}(\mathcal{L})$ is a confident base of $\mathrm{Th}_c(\mathcal{I})$.

Another observation is that in general we cannot transfer non-redundancy of a base $\mathcal{L} \cup S_\mathcal{I}$ of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$ to non-redundancy of $\bigsqcap \mathcal{L}$. This is illustrated by the following example.

4.23 Example We want to find an interpretation $\mathcal{I}$, a number $c \in [0, 1]$ and a non-redundant set $\mathcal{L}$ of implications of $\mathbb{K}_\mathcal{I}$ such that $\mathcal{L} \cup S_\mathcal{I}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$ but $\bigsqcap \mathcal{L}$ contains redundancies.

The main idea to obtain such an example is to construct the interpretation $\mathcal{I}$ in such a way that for two concept names $\mathsf{A}, \mathsf{B} \in N_C$, both implications $\{ \mathsf{A} \} \to \{ \mathsf{B} \}$ and $\{ \exists \mathsf{r}.\mathsf{A} \} \to \{ \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \}$ have confidence at least c in $\mathbb{K}_\mathcal{I}$. Of course, this immediately implies that both $\mathsf{A} \sqsubseteq \mathsf{B}$ and $\exists \mathsf{r}.\mathsf{A} \sqsubseteq \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})$ have confidence at least c in $\mathcal{I}$. Then, if we can include the two implications in the set $\mathcal{L}$ we are looking for, this will immediately result in $\bigsqcap \mathcal{L}$ being redundant. If, in addition, we can make sure that $\mathcal{L}$ is a non-redundant set of implications such that $\mathcal{L} \cup S_\mathcal{I}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_\mathcal{I})$, then we have obtained our desired example.

Let $N_C = \{ \mathsf{A}, \mathsf{B} \}, N_R = \{ \mathsf{r} \}$ and consider the interpretation $\mathcal{I}_1$ given in Figure 5, where every edge denotes an r-edge. In this interpretation, it is true that

$$\mathrm{conf}_{\mathcal{I}_1}(\mathsf{A} \sqsubseteq \mathsf{B}) \geqslant \frac{1}{2}$$
$$\mathrm{conf}_{\mathcal{I}_1}(\exists \mathsf{r}.\mathsf{A} \sqsubseteq \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})) \geqslant \frac{1}{2}.$$

So, let us choose $c = \frac{1}{2}$. Then we want to find a non-redundant set $\mathcal{L} \subseteq \mathrm{Imp}(\mathbb{K}_{\mathcal{I}_1})$ such that $\mathcal{L} \cup S_{\mathcal{I}_1}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_1})$ and $(\{ \mathsf{A} \} \to \{ \mathsf{B} \}), (\{ \exists \mathsf{r}.\mathsf{A} \} \to \{ \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \}) \in \mathcal{L}$.

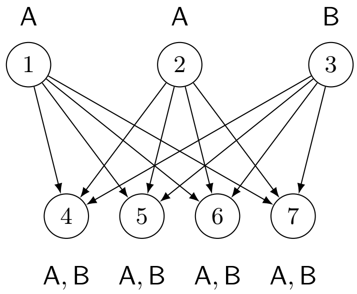

Figure 5: Example interpretation $\mathcal{I}_1$

$\mathbb{K}_{\mathcal{I}_1}$	$\bot$	A	B	$\exists \mathsf{r}.(\mathsf{A} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}))$	$\exists \mathsf{r}.(\mathsf{B} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}))$	$\exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})$	$\exists \mathsf{r}.\exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})$	$\exists \mathsf{r}.\mathsf{A}$	$\exists \mathsf{r}.\mathsf{B}$	$\exists \mathsf{r}.\top$
1	.	×	.	.	.	×	.	×	×	×
2	.	×	.	.	.	×	.	×	×	×
3	.	.	×	.	.	×	.	×	×	×
4	.	×	×	.	.	.	.	.	.	.
5	.	×	×	.	.	.	.	.	.	.
6	.	×	×	.	.	.	.	.	.	.
7	.	×	×	.	.	.	.	.	.	.

Figure 6: Induced formal context of $\mathcal{I}_1$

For this, we compute

$$M_{\mathcal{I}_1} = \{\, \bot, \mathsf{A}, \mathsf{B}, \exists \mathsf{r}.(\mathsf{A} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})),\\ \exists \mathsf{r}.(\mathsf{B} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})), \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}), \exists \mathsf{r}.\exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}), \exists \mathsf{r}.\mathsf{A}, \exists \mathsf{r}.\mathsf{B}, \exists \mathsf{r}.\top \,\}$$

and obtain the induced formal context $\mathbb{K}_{\mathcal{I}_1}$ as shown Figure 6.

We can now compute an irredundant and complete subset $\mathcal{L}_1$ of the confident base $\mathrm{Can}(\mathbb{K}_{\mathcal{I}_1}) \cup \mathrm{Conf}(\mathcal{I}_1, c)$, which, after removing redundancies from $\mathcal{L}_1$ with respect to $S_{\mathcal{I}_1}$, contains the following implications:

$$\begin{aligned}
\varnothing &\to \{\, \mathsf{A} \,\}\\
\{\, \mathsf{A} \,\} &\to \{\, \mathsf{B} \,\}\\
\{\, \exists \mathsf{r}.\mathsf{A} \,\} &\to \{\, \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \,\}\\
\{\, \exists \mathsf{r}.\mathsf{B} \,\} &\to \{\, \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \,\}\\
\{\, \exists \mathsf{r}.\top \,\} &\to \{\, \exists r.(A \sqcap B) \,\}\\
\{\, \mathsf{A}, \mathsf{B}, \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \,\} &\to \{\, \bot \,\}\\
\{\, \exists \mathsf{r}.\exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B}) \,\} &\to \{\, \bot \,\}\\
\{\, \exists \mathsf{r}.(\mathsf{B} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})) \,\} &\to \{\, \bot \,\}\\
\{\, \exists \mathsf{r}.(\mathsf{A} \sqcap \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})) \,\} &\to \{\, \bot \,\}
\end{aligned}$$

As $\mathrm{Can}(\mathcal{I}, c) \cup \mathrm{Conf}(\mathcal{I}, c)$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_1})$, the set $\mathcal{L}$ of all these implications is such that $\mathcal{L} \cup S_{\mathcal{I}_1}$ is a confident base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_1})$ as well. Furthermore, the set $\mathcal{L}$ is irredundant.

However,

$$(\mathsf{A} \sqsubseteq \mathsf{B}), (\exists \mathsf{r}.\mathsf{A} \sqsubseteq \exists \mathsf{r}.(\mathsf{A} \sqcap \mathsf{B})) \in \bigsqcap \mathcal{L},$$

and therefore $\bigsqcap \mathcal{L}$ is not irredundant. $\diamondsuit$

5 Experiments with Confident GCIs

The main motivation to consider confident GCIs is the idea that they may provide helpful information on finding errors in the data. But of course this is only a heuristic idea, and it is not clear a-priori whether this approach really is useful. Indeed, it is at least very hard to give some theoretical insight into the usefulness of this approach, as even formalizing the notion of *interpretations with errors* in accordance to practical observations is far from obvious.

Therefore, in this section we want to show the usefulness of considering confident GCIs by means of a real-world example. The data set we use stems from the DBpedia data set [8] as of March 2010, and is given as an interpretation $\mathcal{I}_{\mathrm{DBpedia}}$ that represents the child-relation in this data set. A detailed construction of this interpretation has been given in [9], and we shall not repeat it here. Instead, we can think of $\mathcal{I}_{\mathrm{DBpedia}}$ as an interpretation containing all elements that appear in the child-relation of the DBpedia data set as of March 2010. For these elements, we can collect properties such as Artist or Criminal, which then serve as concept names. As role name we just use child. Collecting these information in the interpretation $\mathcal{I}_{\mathrm{DBpedia}}$, we obtain 5626 elements and 60 concept names.

We have to mention a special peculiarity of this interpretation here, to not cause confusion when we present our experimental results. One would expect that the child-relation only relates persons to persons, i. e. only persons can be children of persons. However, DBpedia suffers from the liberal structure of Wikipedia Infoboxes, where it draws its information from. These infoboxes are not standardized in any way, and extracting information from them is really a difficult task. If in such an infobox a link to another article appears under the rubric listing children, then this link is collected as a children. However, sometimes there are some links under this rubric that link to articles somehow related to these children. For example, in $\mathcal{I}_{\mathrm{DBpedia}}$, the element `Ellen_Harper` has as a child the element `The_Carol_Burnett_Show`, which is a US american comedy show of the late 1960s and 1970s. This is because in the infobox of the Wikipedia article on Ellen Harper, one of its children is related to the Carol Burnett Show, with a link to the corresponding article.

Despite this oddity in the child-relation of DBpedia, the data set itself contains a lot of valuable information. Even better, one could argue that *because of* this peculiar child-relation, the interpretation $\mathcal{I}_{\mathrm{DBpedia}}$ is very well suited for our experiments, because this allows us to verify in how far confident GCIs are able to detect some of these errors.

In the following, we want to assume that an ontology engineer wants to use $\mathcal{I}_{\mathrm{DBpedia}}$ to construct an ontology that represents the properties of the child-relation. This ontology engineer want to consider confident GCIs to overcome some of the errors present in this interpretation. To do this, she has to extract some confident GCIs from $\mathcal{I}_{\mathrm{DBpedia}}$ and has to check them for usefulness. This she has to do manually, and therefore this can be an expensive task. To show how much of extra work this can be for our particular example of $\mathcal{I}_{\mathrm{DBpedia}}$, we propose the following three experiments:

i. We want to explicitly consider the sets $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.95)$ and $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.90)$, to see whether the GCIs thus obtained are of any use for our ontology engineer working on $\mathcal{I}_{\mathrm{DBpedia}}$.

ii. We want to consider for all $c \in \{ 0, 0.01, 0.02, \dots, 0.99 \}$ the sizes of the set $\mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$, to see how many GCIs have to be consider when varying the threshold on the minimal confidence. During this, we may also want to consider the canonical base of $\mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$, since this may give rise to a much smaller base of $\mathrm{Th}_c(\mathcal{I})$.

iii. Finally, we want to consider for all $c \in \{ 0, 0.01, 0.02, \dots, 0.99 \}$ sizes of the canonical base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$. The rationale behind this is the following: if we consider confident GCIs, we assume that we can circumvent certain errors and can extract more general patterns which are falsified by erroneous counterexamples. We therefore expect that, if we consider confident GCIs as actually valid GCIs, that the resulting theory extracted from a finite interpretation gets more succinct. With this experiment we want to examine in how far this is true for $\mathcal{I}_{\mathrm{DBpedia}}$ for varying values of c.

5.1 Confident GCIs of $\mathcal{I}_{\mathrm{DBpedia}}$ for $c = 0.95$ and $c = 0.90$

In this section we want to show how our ontology engineer would examine confident GCIs extracted for two particular choices of c. For this, we shall examine the sets $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.95)$ and $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.90)$ and discuss whether the GCIs contained in these sets are "reasonable." Thereby we decide whether a GCI $C \sqsubseteq D$ is reasonable by considering the counterexamples to $C \sqsubseteq D$, for which we can decide whether they are valid counterexamples or not.

5.1.1 The Case of Minimal Confidence 0.95

We can compute the set $\mathcal{C} := \mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.95)$ to be[1]

$$\begin{aligned}\mathcal{C} = \{ & \mathsf{Place} \sqsubseteq \mathsf{PopulatedPlace},\\ & \exists \mathsf{child}.\top \sqsubseteq \mathsf{Person},\\ & \exists \mathsf{child}.\exists \mathsf{child}.\top \sqcap \exists \mathsf{child}.\mathsf{OfficeHolder}\\ & \quad \sqsubseteq \exists \mathsf{child}.(\mathsf{OfficeHolder} \sqcap \exists \mathsf{child}.\top) \}\end{aligned}$$

It is quite surprising that the set $\mathcal{C}$ turns out to have only three elements. Let us now consider every GCIs in more detail.

The set $\mathcal{C}$ contains the GCI $\exists \mathsf{child}.\top \sqsubseteq \mathsf{Person}$, which indeed looks very natural. However, $\mathcal{I}_{\mathrm{DBpedia}}$ contains four counterexamples, namely `Teresa_Carpio`, `Charles_Heung`, `Adam_Cheng` and `Lydia_Shum`. However, all these elements name individuals which are artists from Hong Kong, and therefore certainly are persons. In other words, these counterexamples are erroneous and the corresponding GCIs is valid.

It is also convincing that the GCI $\mathsf{Place} \sqsubseteq \mathsf{PopulatedPlace}$ is reasonable as well (places named in DBpedia appear because people have been born or lived there), and the only counterexample to this GCI is Greenwich_Village, denoting a district of New York which certainly is populated.

The last GCI which remains to be considered is

$$\exists \mathsf{child}.\exists \mathsf{child}.\top \sqcap \exists \mathsf{child}.\mathsf{OfficeHolder} \sqsubseteq \exists \mathsf{child}.(\mathsf{OfficeHolder} \sqcap \exists \mathsf{child}.\top)$$

Subjectively, this GCI appears to be too specific to be considered as a valid (or useful) GCI. The only counterexample to this GCI is Pierre_Samuel_du_Pont_de_Nemours, denoting the french government official Pierre Samuel du Pont de Nemours, who had two sons, namely Victor Marie

[1] We have removed some redundancies in the concept descriptions to make them more readable. The GCIs extracted by the algorithm are actually much longer, but equivalent to those shown here.

du Pont and Eleuthère Irénée du Pont. The first got a french diplomat and is therefore listed in $\mathcal{I}_{\text{DBpedia}}$ as an instance of $\mathsf{OfficeHolder}$. Although he had four children, none of them got famous enough to be named in the Wikipedia infobox of the corresponding Wikipedia article[2]. On the other hand, his brother Eleuthère Irénée du Pont became a famous american industrial and had a lot of famous children, which are listed in the Wikipedia infobox and therefore appear in $\mathcal{I}_{\text{DBpedia}}$.

From the point of view of the DBpedia data set, $\mathsf{Pierre\_Samuel\_du\_Pont\_de\_Nemours}$ may be considered a valid counterexample, when one considers the child relation in $\mathcal{I}_{\text{DBpedia}}$ as denoting only *famous* children (noteworthy by name in the Wikipedia infobox.) If one, however, considers the child relation simply as having children, the counterexample is not correct (as the Wikipedia article is not correct.) Deciding which of the choices to take is now up to the ontology engineer, and depends on the actual domain the ontology is to represent.

A legitimate question now is what happens if we consider the GCIs $\exists\mathsf{child}.\top \sqsubseteq \mathsf{Person}$ and $\mathsf{Place} \sqsubseteq \mathsf{PopulatedPlace}$ as valid GCIs, i. e. how much does the base of $\mathcal{I}_{\text{DBpedia}}$ change if we include those GCIs as background knowledge? Let

$$\mathcal{F} := \{\, \{\, \exists\mathsf{child}.\top \,\} \to \{\, \mathsf{Person} \,\}, \{\, \mathsf{Place} \,\} \to \{\, \mathsf{PopulatedPlace} \,\} \,\}.$$

One way to find a complete set of $\mathrm{Th}(\mathcal{I}_{\text{DBpedia}})$ such that the mentioned GCIs are valid as well is just to compute the canonical base of $\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}$ with the corresponding background knowledge, i. e. we compute

$$\mathcal{L} := \mathrm{Can}(\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}, S_{\mathcal{I}_{\text{DBpedia}}} \cup \mathcal{F}).$$

The set $\mathcal{L} \cup S_{\mathcal{I}_{\text{DBpedia}}} \cup \mathcal{F}$ is then complete for $\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}$, therefore

$$\mathcal{L} \cup \mathcal{F} \cup S_{\mathcal{I}_{\text{DBpedia}}} \models \mathrm{Can}(\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}, S_{\mathcal{I}_{\text{DBpedia}}}).$$

Since $\bigsqcap \mathrm{Can}(\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}, S_{\mathcal{I}_{\text{DBpedia}}})$ is complete for $\mathrm{Th}(\mathcal{I}_{\text{DBpedia}})$, the set $\bigsqcap(\mathcal{L} \cup \mathcal{F})$ is complete for $\mathrm{Th}(\mathcal{I}_{\text{DBpedia}})$ as well. Therefore $\bigsqcap(\mathcal{L} \cup \mathcal{F})$ is a base of

$$\mathrm{Th}(\mathcal{I}_{\text{DBpedia}}) \cup \{\, \exists\mathsf{child}.\top \sqsubseteq \mathsf{Person}, \mathsf{Place} \sqsubseteq \mathsf{PopulatedPlace} \,\}. \tag{5.1}$$

If we now compute the set $\mathcal{L}$, we obtain a set of 1245 implications, therefore $\bigsqcap(\mathcal{L} \cup \mathcal{F})$ is a base of (5.1) of size 1247. Compared to the 1252 implications needed to axiomatize $\mathrm{Th}(\mathcal{I})$, we can indeed observe a decrease in the size of the base, although this may not be very impressive.

Note, however, that another consequence of including the set $\mathcal{F}$ into a base is of course, that the size of the concept descriptions in the resulting GCIs will become smaller and more readable.

5.1.2 The Case of Minimal Confidence 0.90

Of course, it is true that $\mathrm{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.90) \supseteq \mathrm{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95)$ and hence we shall only discuss the GCIs in

$$\mathcal{D} := \mathrm{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.90) \setminus \mathrm{Conf}(\mathcal{I}_{\text{DBpedia}}, 0.95).$$

[2] as of 13. November 2012

We can compute

$$
\begin{aligned}
\mathcal{D} = \{ & \mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.\top)))) \\
& \qquad \sqsubseteq \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.\mathsf{Person})))) \\
& \exists \mathsf{child}.(\mathsf{Artist} \sqcap \mathsf{Writer} \sqcap \exists \mathsf{child}.\top) \\
& \qquad \sqsubseteq \mathsf{Writer} \sqcap \mathsf{Artist}, \\
& \exists \mathsf{child}.\exists \mathsf{child}.\top \sqcap \exists \mathsf{child}.\mathsf{President} \\
& \qquad \sqsubseteq \exists \mathsf{child}.(\mathsf{President} \sqcap \exists \mathsf{child}.\top), \\
& \mathsf{Monarch} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.\top) \\
& \qquad \sqsubseteq \exists \mathsf{child}.(\mathsf{Monarch} \sqcap \exists \mathsf{child}.\top), \\
& \exists \mathsf{child}.\mathsf{MemberOfParliament} \sqcap \mathsf{Politician} \\
& \qquad \sqsubseteq \mathsf{MemberOfParliament}, \\
& \exists \mathsf{child}.\mathsf{OfficeHolder} \sqcap \exists \mathsf{child}.\exists \mathsf{child}.\mathsf{Person} \\
& \qquad \sqsubseteq \exists \mathsf{child}.(\mathsf{OfficeHolder} \sqcap \exists \mathsf{child}.\mathsf{Person}) \}.
\end{aligned}
$$

These GCIs are all quite specific and it is doubtful whether they may be of any use for an ontology designed who tries to extract GCIs from $\mathcal{I}_{\text{DBpedia}}$. But let us still have a look at the counterexample for the given GCIs.

We shall start with the first GCI listed above, i. e.

$$
\begin{aligned}
& \mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.\top)))) \\
& \sqsubseteq \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.(\mathsf{Person} \sqcap \exists \mathsf{child}.\mathsf{Person})))). \qquad (5.2)
\end{aligned}
$$

This GCI seems to be rather complicated, and one may assume a much more general GCI to be true, namely

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

which is the $\mathcal{EL}$-approximation of the fact that all children should be persons. However, as already discussed, this GCI is not true in $\mathcal{I}_{\text{DBpedia}}$ (and has confidence only around 0.53.) Now this GCI states that if you have generations of instances of Person of at least 5 generations, then the element at the fifth generation can be chosen to be a Person. The only counterexample to this GCI is `Mayer_Amschel_Rothschild`, naming the founder of the Rothschild dynasty. The only two fifth-generation descendants not being instances of Person in $\mathcal{I}_{\text{DBpedia}}$ are `Edouard_Etienne_de_Rothschild` and `David_René_de_Rothschild`, which are certainly persons. Therefore, this counterexample is invalid and this GCI is valid.

Let us now consider the remaining GCIs. In the order of appearance above, the following list gives all the counterexamples in $\mathcal{I}_{\text{DBpedia}}$ of the corresponding GCIs:

i. `John_McManners`

ii. `Alois_Hitler`

iii. `Dejan_Dragaš`

iv. `Marion_Dewar`, `Ranasinghe_Premadasa`

v. `Pierre_Samuel_du_Pont_de_Nemours`

The last counterexample has already been discussed in the previous case, so we shall focus our discussions on the first four only.

i. The individual `John_McManners` denotes an British clergyman and historian who had a son, `Hugh_McManners`, a musician and writer, who itself has a son. However, `John_McManners`, though being a famous writer, was not an artist. Therefore, this GCI is not correct.

ii. The individual `Alois_Hitler` names the father of Adolf Hitler, who was the only of the children of Alois Hitler to rule a country. As he had no children on its own, the individual serves as a correct counterexample to the given GCI, which is therefore incorrect.

iii. The individual `Dejan_Dragaš` denotes a 14th-century Serbian noblemen and despot of Kumanovo. He had two sons, `Constantine_Dragaš`, who had children and was ruler of parts of Serbia, but not a monarch, and `Jovan_Dragaš`, who was despot of Kumanovo, but had no children. Again, this counterexample is correct and the GCI invalid.

iv. The individual `Marion_Dewar` is not a correct counterexample, as Marion Dewar was member of the Canadian House of Commons from 1987 to 1988.

 The other individual, `Ranasinghe_Premadasa`, denotes a former Prime Minister and later President of Sri Lanka. It is, however, quite hard to tell whether this means that he has ever been member of the Parliament of Sri Lanka. Hence, from the point of view of DBpedia extracting available knowledge from the Wikipedia pages, this counterexamples can be assumed correct, although further investigations by a human expert may be necessary.

5.1.3 Discussion

By considering $\operatorname{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.95)$ and $\operatorname{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.90)$ we have illustrated in which way an ontology engineer can make use of confident GCIs. As a first observation, we have seen that this may include non-trivial research for the ontology engineer. In particular, deciding whether a counterexample present in the data is correct always involves the question whether the counterexample is relevant for the particular domain the resulting ontology is to represent. It may therefore happen that an otherwise correct counterexample is rejected since it does not appear in the domain of discourse. With respect to this observation, one could also say that confident GCIs may help to model domains from data that does not fully describe these domains, but are merely an approximation of them.

5.2 Sizes of Bases of $\operatorname{Th}_c(\mathcal{I}_{\mathbf{DBpedia}})$

In this section we shall conduct the remaining two experiments presented in the introduction, i. e. we shall examine the behavior of the size of $\operatorname{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, c)$, $\operatorname{Can}(\operatorname{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, c))$ and of $\operatorname{Can}(\operatorname{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}))$ for varying values of c. For the first experiment, we shall obtain an impression on how many extra GCIs an ontology engineer has to consider. For the second experiment, we shall obtain an intuition on how many GCIs a resulting TBox will contain.

5.2.1 The Size of $\operatorname{Conf}(\mathcal{I}_{\mathbf{DBpedia}}, c)$ and $\operatorname{Can}(\operatorname{Conf}(\mathcal{I}_{\mathbf{DBpedia}}, c))$

For the first experiment, we consider as the set of values for the parameter c the set

$$V = \{ 0, 0.01, 0.02, \dots, 0.99 \}$$

For each $c \in V$, we compute $|\operatorname{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)|$. The result is shown in Figure 7, where the y-axis is scaled logarithmically.

The results given in this picture show that the number of confident GCIs the ontology engineer has to check manually declines exponentially as the minimal confidence grows. Even for $c = 0.86$,

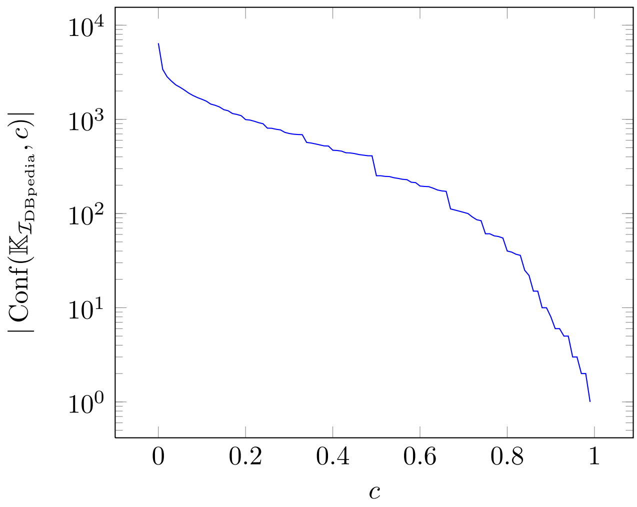

Figure 7: Size of $\mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$ for all $c \in V$

there are only 15 extra GCIs to investigate. Given the fact that a base of $\mathcal{I}_{\mathrm{DBpedia}}$ has 1252 elements, this extra effort is negligible.

Of course, it is not clear whether this behavior is typical or just particular to our data set. However, it indicates that considering confident GCIs for data, where the quality is good enough (i. e. where only few errors have been made), is not a noteworthy overhead.

A drawback for this experiment is that we ignore the fact that $\mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$ does not need to be irredundant. Of course, if our ontology engineer confirms a certain subset $\mathcal{L} \subseteq \mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$, then all implications already entailed by $\mathcal{L}$ do not need to be checked on their own. The same is true if we consider the GCIs entailed by an already confirmed subset of $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, c)$.

Therefore, we show in Figure 8 the size behavior of the canonical base of $\mathrm{Conf}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$ for all $c \in V$. Note that this canonical base is irredundant, hence the aforementioned issue does not arise anymore. And indeed, as we can see from the picture, the number of GCIs decreases significantly, especially for small values of c. But we can also observe that for larger values of c, say $c \geqslant 0.8$, the overall number of GCIs to be considered does not decrease that much. Indeed, $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.8)$ contains 40 elements, whereas $\mathrm{Can}(\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, 0.8))$ contains 32. One could argue that this does not really help, but one also has to consider that checking every GCI by hand may be such an expensive task that every GCI saved pays off.

5.2.2 The Size of $\mathrm{Can}(\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}))$

We now turn our attention to the last experiment. There, we consider the size of the sets $\mathrm{Can}(\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c))$ for all $c \in V$. The results of the experiment are shown in Figure 9.

From this data plot, we can make a couple of observations. For high values of c, i. e. $c \in [0.8, 1]$, the overall size of $\mathrm{Can}(\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c))$ does not decrease significantly. We can see this as a sign that the overall theory $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$ does not change significantly, i. e. that the errors we can handle with the aforementioned values of c do not significantly influence $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}, c)$. This is actually what we want to achieve with our method, namely correcting small errors while preserving as much of the original theory as possible. Of course, this data plot is only an indication that our methods achieves this in the particular example of $\mathcal{I}_{\mathrm{DBpedia}}$. But as $\mathcal{I}_{\mathrm{DBpedia}}$

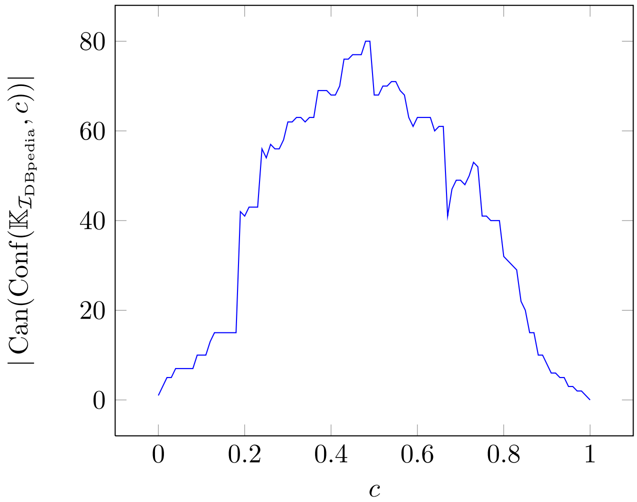

Figure 8: Size of $\operatorname{Can}(\operatorname{Conf}(\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}, c))$ for all $c \in V$

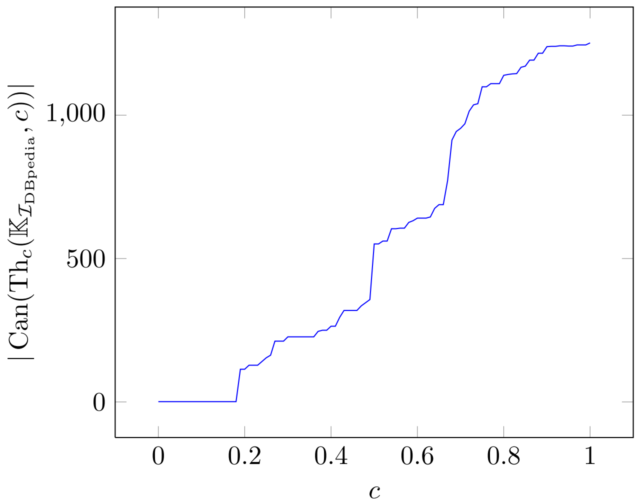

Figure 9: Size of $\operatorname{Can}(\operatorname{Th}_c(\mathbb{K}_{\mathcal{I}_{\text{DBpedia}}}))$ for all $c \in V$

arises from real-world data, we can be certain that this behavior is not accidental.

Of course, the more we decrease c, the more we depart from the original set $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$ of implications. There are three points were this becomes especially apparent, namely around $c = 0.66, c = 0.54$ and $c = 0.18$, where the changes in the size of $\mathrm{Can}(\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}}))$ is more significant then for other values of c. Incidentally, these are three special values for c, as

i. for $c \leqslant 0.18$ the implication $\varnothing \to M_{\mathcal{I}_{\mathrm{DBpedia}}}$ is entailed by $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$, resulting in a singleton canonical base;

ii. for $c \leqslant 0.54$, the implication $\{\, \exists \mathsf{child}.\top \,\} \to \{\, \exists \mathsf{child}.\mathsf{Person} \,\}$ is contained in $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$, eliminating a large number of special cases;

iii. for $c \leqslant 0.66$, the implication $\varnothing \to \{\, \mathsf{Person} \,\}$ is contained in $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$, also eliminating a variety of special cases.

Indeed, adding the implication $\varnothing \to \{\, \mathsf{Person} \,\}$ to $\mathrm{Th}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$ results in the size of the canonical base to drop from 1252 to 1210. If we additionally add the implication $\{\, \exists \mathsf{child}.\top \,\} \to \{\, \exists \mathsf{child}.\mathsf{Person} \,\}$, the resulting canonical base then even only contains 1163 implications.[3]

To understand this phenomenon, we can take the following point of view: intuitively, the lower the threshold c, the *simpler* the set $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$ gets, because we neglect more special cases in our data set. If the change in size of the respective canonical base is as significant as observed for the three values given above, we would assume that, either, a lot new implications have been accepted, or that new implications are accepted that render a lot of other implications redundant. Indeed, as we have seen above, the second case occurs (consider also Figure 7 to see that the size of $\mathrm{Conf}(\mathcal{I}_{\mathrm{DBpedia}}, c)$ does not change significantly at those points.) We can therefore consider those significant changes in the size of the canonical base of $\mathrm{Th}_c(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$ as a sign of discovery of some general implications, which may be of interest for our ontology engineer.

6 Unravelling Confident $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-Bases to Confident $\mathcal{EL}^{\bot}$-Bases

So far, we have only considered bases of $\mathrm{Th}_c(\mathcal{I})$ formulated in the description logic $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$. Although this description logic allows us to find such bases in an easy way, the resulting bases may not be suitable for practical purposes. This is mainly due to the inherit incomprehensibility of $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions: the presence of cycles in these concept descriptions make it hard, even for logicians, to find out what this concept description is supposed to mean. On the other hand, $\mathcal{EL}^{\bot}$-concept descriptions are normally easy to understand and their intention can be also be deduced by non-experts.

Therefore, we want to discuss in this section a way to obtain confident $\mathcal{EL}^{\bot}$-bases of $\mathrm{Th}_c(\mathcal{I})$ from confident $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-bases of $\mathrm{Th}_c(\mathcal{I})$. The technique we are going to use for this is based on unravelling $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions as introduced in Section 3.4. The argumentation presented here is a generalization of the argumentation of Distel [11], who showed a similar result on obtaining $\mathcal{EL}^{\bot}$-bases of $\mathcal{I}$ from $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-bases of $\mathcal{I}$.

Let $c \in [0, 1]$ and let us assume that $\mathcal{D}$ is a base of $\mathrm{Th}_c(\mathcal{I})$. We can partition $\mathcal{D} = \mathcal{B} \cup \mathcal{C}$, where $\mathcal{B} \subseteq \mathrm{Th}(\mathcal{I})$ and $\mathcal{C} \cap \mathrm{Th}(\mathcal{I}) = \varnothing$. Without loss of generality, we can also assume that $\mathcal{B}$ only contains GCIs of the form $E \sqsubseteq E^{\mathcal{I}\mathcal{I}}$.

[3] Please note that, although $\top \sqsubseteq \mathsf{Person}$ entails $\exists \mathsf{child}.\top \sqsubseteq \exists \mathsf{child}.\mathsf{Person}$, the implication $\varnothing \to \{\, \mathsf{Person} \,\}$ does not entail the implication $\{\, \exists \mathsf{child}.\top \,\} \to \{\, \mathsf{Person} \,\}$. Indeed, the implication $\{\, \exists \mathsf{child}.\top \,\} \to \{\, \exists \mathsf{child}.\mathsf{Person} \,\}$ is not even entailed by $\mathrm{Th}_{0.66}(\mathbb{K}_{\mathcal{I}_{\mathrm{DBpedia}}})$.

As a first step, we are going to define an auxiliary set $\mathcal{X}_{\mathcal{I},d}$ of $\mathcal{EL}^{\perp}$-GCIs that "capture" entailment relations between $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions. For this recall that by Lemma 3.29 there exists a $d \in \mathbb{N}$ such that

$$C^{\mathcal{I}} = (C_d)^{\mathcal{I}} \tag{6.1}$$

holds for each $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept description C, where C_d denotes the unravelling of C up to depth d.

6.1 Definition Let $\mathcal{I}$ be a finite interpretation and let $d \in \mathbb{N}$ be as in Lemma 3.29. Then define

$$\mathcal{X}_{\mathcal{I},d} := \{ (X^{\mathcal{I}})_d \sqsubseteq (X^{\mathcal{I}})_{d+1} \mid X \subseteq \Delta^{\mathcal{I}}, X \neq \varnothing \}. \qquad \diamondsuit$$

Note that $\mathcal{X}_{\mathcal{I},d}$ is a set of valid GCIs of $\mathcal{I}$, since for each $X \subseteq \Delta^{\mathcal{I}}, X \neq \varnothing$, it is true that $((X^{\mathcal{I}})_d)^{\mathcal{I}} = X^{\mathcal{I}\mathcal{I}} = ((X^{\mathcal{I}})_{d+1})^{\mathcal{I}}$.

In the course of Distel's argumentation [11, Theorem 21], the following property of $\mathcal{X}_{\mathcal{I},d}$ had been shown.

6.2 Lemma *Let $\mathcal{I}$ be a finite interpretation and let $d \in \mathbb{N}$ as in Lemma 3.29. Then for each $Y \subseteq \Delta^{\mathcal{I}}, k \in \mathbb{N}, k \geqslant d$ it is true that*

$$\mathcal{X}_{\mathcal{I},d} \models ((Y^{\mathcal{I}})_k \sqsubseteq (Y^{\mathcal{I}})_{k+1})$$

and

$$\mathcal{X}_{\mathcal{I},d} \models ((Y^{\mathcal{I}})_d \sqsubseteq Y^{\mathcal{I}}).$$

Before we are going to prove this lemma we need to formulate a preliminary result from [11]. As its prove involves notions we have not introduced in this work, we shall not repeat it here.

6.3 Lemma (Lemma 5.19 of [11]) *Let C, D be $\mathcal{EL}_{\mathrm{gfp}}$-concept descriptions. Then for each $k \in \mathbb{N}$ it is true that*

i. $(\exists r.C)_k \equiv \exists r.C_{k-1}$ *for each* $r \in N_R$ *and*

ii. $(C \sqcap D)_k \equiv C_k \sqcap D_k$.

We now prove Lemma 6.2.

Proof (Lemma 6.2) For the first claim observe that if $Y = \varnothing$, $Y^{\mathcal{I}} = \bot$ and nothing remains to be shown. We therefore assume $Y \neq \varnothing$ and shall show the claim using induction on k. The case $k = d$ is clear as $((Y^{\mathcal{I}})_d \sqsubseteq (Y^{\mathcal{I}})_{d+1}) \in \mathcal{X}_{\mathcal{I},d}$. For the step-case assume that we already now that

$$\mathcal{X}_{\mathcal{I},d} \models ((Z^{\mathcal{I}})_{k-1} \sqsubseteq (Z^{\mathcal{I}})_k)$$

is true for all $Z \subseteq \Delta^{\mathcal{I}}, Z \neq \varnothing$.

From Theorem 4.7 we know that $Y^{\mathcal{I}}$ is expressible in terms of $M_{\mathcal{I}}$. As $Y \neq \varnothing$, it is true that $Y^{\mathcal{I}} \neq \bot$ and therefore

$$Y^{\mathcal{I}} \equiv \bigsqcap U \sqcap \bigsqcap_{(r,Z)\in\Pi} \exists r.Z^{\mathcal{I}}$$

for some $U \subseteq N_C$ and $\Pi \subseteq N_R \times \mathfrak{P}(\Delta^{\mathcal{I}})$. By Lemma 6.3 we therefore obtain

$$(Y^{\mathcal{I}})_k \equiv \bigsqcap U \sqcap \bigsqcap_{(r,Z)\in\Pi} \exists r.(Z^{\mathcal{I}})_{k-1}.$$

Now the induction hypothesis yields that $(Z^{\mathcal{I}})_{k-1} \sqsubseteq (Z^{\mathcal{I}})_k$ is entailed by $\mathcal{X}_{\mathcal{I},d}$ and hence $\mathcal{X}_{\mathcal{I},d}$ also entails

$$\begin{aligned}(Y^{\mathcal{I}})_k &\sqsubseteq \bigsqcap U \sqcap \bigsqcap_{(r,Z)\in\Pi} \exists r.(Z^{\mathcal{I}})_k \\ &\equiv \Big(\bigsqcap U \sqcap \bigsqcap_{(r,Z)\in\Pi} \exists r.Z^{\mathcal{I}}\Big)_{k+1} \\ &\equiv (Y^{\mathcal{I}})_{k+1},\end{aligned}$$

as required. This completes the step-case and proves the first claim.

Let $Y \subseteq \Delta^{\mathcal{I}}$. We shall now show the second claim, i. e.

$$\mathcal{X}_{\mathcal{I},d} \models ((Y^{\mathcal{I}})_d \sqsubseteq Y^{\mathcal{I}}).$$

Again, if $Y = \varnothing$, then $Y^{\mathcal{I}} = \bot$ and the claim is trivial. Therefore, let $Y \neq \varnothing$ and let $\mathcal{J}$ be an interpretation such that $\mathcal{J} \models \mathcal{X}_{\mathcal{I},d}$. It is then true that

$$((Y^{\mathcal{I}})_k)^{\mathcal{J}} \subseteq ((Y^{\mathcal{I}})_{k+1})^{\mathcal{J}} \subseteq ((Y^{\mathcal{I}})_{k+2})^{\mathcal{J}} \subseteq \dots$$

From Lemma 3.29 we know that for some $\ell \in \mathbb{N}$ it is true that $((Y^{\mathcal{I}})_\ell)^{\mathcal{J}} = (Y^{\mathcal{I}})^{\mathcal{J}}$. Assuming without loss of generality that $\ell \geqslant k$ we immediately obtain that

$$((Y^{\mathcal{I}})_k)^{\mathcal{J}} \subseteq (Y^{\mathcal{I}})^{\mathcal{J}},$$

i. e. $\mathcal{J} \models ((Y^{\mathcal{I}})_k \sqsubseteq Y^{\mathcal{I}})$. Therefore $\mathcal{X}_{\mathcal{I},d} \models ((Y^{\mathcal{I}})_k \sqsubseteq Y^{\mathcal{I}})$, as it has been claimed. □

We now construct the $\mathcal{EL}^\perp$-base of $\mathrm{Th}_c(\mathcal{I})$ from $\mathcal{B} \cup \mathcal{C}$. For this, let us define

$$\begin{aligned}\mathcal{B}_0 &:= \{\, E_d \sqsubseteq (E^{\mathcal{II}})_d \mid (E \sqsubseteq E^{\mathcal{II}}) \in \mathcal{B} \,\} \cup \{\, C_d \sqsubseteq (C^{\mathcal{II}})_d \mid (C \sqsubseteq D) \in \mathcal{C} \,\}, \\ \mathcal{C}_0 &:= \{\, (C^{\mathcal{II}})_d \sqsubseteq (D^{\mathcal{II}})_d \mid (C \sqsubseteq D) \in \mathcal{C} \,\}.\end{aligned} \tag{6.2}$$

Note that $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d}$ only contains $\mathcal{EL}^\perp$-GCIs and that all GCIs in $\mathcal{B}_0 \cup \mathcal{X}_{\mathcal{I},d}$ are valid in $\mathcal{I}$. The claim now is that $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d}$ is a confident $\mathcal{EL}^\perp$-base of $\mathrm{Th}_c(\mathcal{I})$.

6.4 Theorem *Let $\mathcal{I}$ be a finite interpretation, $c \in [0,1]$ and let $d \in \mathbb{N}$ be as in Lemma 3.29. Let $\mathcal{B} \cup \mathcal{C}$ be a confident base of $\mathrm{Th}_c(\mathcal{I})$ such that $\mathcal{B} \subseteq \mathrm{Th}(\mathcal{I})$, $\mathcal{C} \cap \mathrm{Th}(\mathcal{I}) = \varnothing$ and $\mathcal{B}$ only contains GCIs of the form $E \sqsubseteq E^{\mathcal{II}}$. Define the sets $\mathcal{B}_0$ and $\mathcal{C}_0$ as in (6.2). Then the following statements hold:*

i. $\mathcal{C}_0 \subseteq \mathrm{Th}_c(\mathcal{I})$ *and* $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d} \models \mathcal{C}$.

ii. $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d} \models \mathcal{B}$.

In particular, the set $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d}$ is a confident $\mathcal{EL}^\perp$-base of $\mathrm{Th}_c(\mathcal{I})$.

Proof To see $\mathcal{C}_0 \subseteq \mathrm{Th}_c(\mathcal{I})$, recall that $(C_d)^{\mathcal{I}} = C^{\mathcal{I}}$ is true for each concept description C. For $(C \sqsubseteq D) \in \mathcal{C}$ with $|C^{\mathcal{I}}| \neq \varnothing$ we therefore obtain:

$$\begin{aligned}\mathrm{conf}_{\mathcal{I}}(C \sqsubseteq D) &= \mathrm{conf}_{\mathcal{I}}(C^{\mathcal{II}} \sqsubseteq D^{\mathcal{II}}) \\ &= \frac{|(C^{\mathcal{II}} \sqcap D^{\mathcal{II}})^{\mathcal{I}}|}{|C^{\mathcal{III}}|} \\ &= \frac{|((C^{\mathcal{II}} \sqcap D^{\mathcal{II}})_d)^{\mathcal{I}}|}{|((C^{\mathcal{II}})_d)^{\mathcal{I}}|} \\ &= \frac{|((C^{\mathcal{II}})_d \sqcap (D^{\mathcal{II}})_d)^{\mathcal{I}}|}{|((C^{\mathcal{II}})_d)^{\mathcal{I}}|} \\ &= \mathrm{conf}_{\mathcal{I}}((C^{\mathcal{II}})_d \sqsubseteq (D^{\mathcal{II}})_d)\end{aligned}$$

using Lemma 6.3. If $C^{\mathcal{I}} = \varnothing$, then $(C_d)^{\mathcal{I}} = C^{\mathcal{I}} = \varnothing$ and

$$\operatorname{conf}_{\mathcal{I}}(C \sqsubseteq D) = \operatorname{conf}_{\mathcal{I}}(C_d \sqsubseteq D_d)$$

is true as well. Overall, we obtain $((C^{\mathcal{II}})_d \sqsubseteq (D^{\mathcal{II}})_d) \in \operatorname{Th}_c(\mathcal{I})$ and therefore $\mathcal{C}_0 \subseteq \operatorname{Th}_c(\mathcal{I})$ as required.

We now show $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d} \models \mathcal{C}$. Let $(C \sqsubseteq D) \in \mathcal{C}$. Then the following statements are true

$$\begin{aligned}
\varnothing &\models (C \sqsubseteq C_d) \\
\mathcal{B}_0 &\models (C_d \sqsubseteq (C^{\mathcal{II}})_d) \\
\mathcal{C}_0 &\models ((C^{\mathcal{II}})_d \sqsubseteq (D^{\mathcal{II}})_d) \\
\mathcal{X}_{\mathcal{I},d} &\models ((D^{\mathcal{II}})_d \sqsubseteq D^{\mathcal{II}}) \\
\varnothing &\models (D^{\mathcal{II}} \sqsubseteq D).
\end{aligned}$$

Therefore, $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d} \models (C \sqsubseteq D)$ as required.

We now consider the second claim, i. e. that $\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_{\mathcal{I},d} \models \mathcal{B}$. Let $(E \sqsubseteq E^{\mathcal{II}}) \in \mathcal{B}$. Then it is true that

$$\begin{aligned}
\varnothing &\models (E \sqsubseteq E_d) \\
\mathcal{B}_0 &\models (E_d \sqsubseteq (E^{\mathcal{II}})_d) \\
\mathcal{X}_{\mathcal{I},d} &\models ((E^{\mathcal{II}})_d \sqsubseteq E^{\mathcal{II}})
\end{aligned}$$

again using Lemma 6.2 for the last entailment. Therefore,

$$\mathcal{B}_0 \cup \mathcal{C}_0 \cup \mathcal{X}_d \models (E \sqsubseteq E^{\mathcal{II}})$$

and the claim is proven. □

7 Conclusions and Further Work

This work extended the results obtained in [9] in various ways. Firstly, we have given another construction of a base of $\operatorname{Th}_c(\mathcal{I})$, which works by directly transforming bases of $\operatorname{Th}_c(\mathbb{K}_{\mathcal{I}})$ into confident bases of $\operatorname{Th}_c(\mathcal{I})$. Secondly, we have given experimental evidence that our approach of considering confident GCIs may be helpful during the process of construction an ontology from example data. Finally, we have shown that certain $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-bases of $\operatorname{Th}_c(\mathcal{I})$ can effectively be transformed into $\mathcal{EL}^{\perp}$-bases of $\operatorname{Th}_c(\mathcal{I})$ by generalizing the corresponding technique of [11].

From the viewpoint of both theory and practical application of confident GCIs, the most important next step is to generalize the exploration algorithm from [11] to our setting of confident GCIs. This may simplify the exploration process in the way that certain, special GCIs may not have to be considered as soon as a more general GCI, which may have some erroneous counterexamples, has already been confirmed. As exploration has as its main purpose to *complete* an ontology by missing statements, generalizing the exploration process to confident GCIs may also unify two steps of the ontology construction, namely construction from data and completing the ontology.

Certainly, another direction of research would be to clarify and formalize the vague argumentation we have given in Section 5.2.2. For this, it may also be interesting to conduct experiments directly with GCIs and not only with implications. This, however, would require a possibility to compute a smallest base of a given set of GCIs, i. e. a method to minimize the cardinality of a given TBox.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 207–216, May 1993.

[2] 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.

[3] F. Baader. Least common subsumers, most specific concepts, and role-value-maps in a description logic with existential restrictions and terminological cycles. LTCS-Report LTCS-02-07, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2002. See `http://lat.inf.tu-dresden.de/research/reports.html`.

[4] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *IJCAI*, pages 325–330. Morgan Kaufmann, 2003.

[5] Franz Baader and Felix Distel. A Finite Basis for the Set of $\mathcal{EL}$-Implications Holding in a Finite Model. In Raoul Medina and Sergei Obiedkov, editors, *Proceedings of the 6th International Conference on Formal Concept Analysis, (ICFCA 2008)*, volume 4933 of *Lecture Notes in Artificial Intelligence*, pages 46–61. Springer Verlag, 2008.

[6] Franz Baader and Felix Distel. Exploring finite models in the description logic $\mathcal{EL}_{\mathrm{gfp}}$. In Sébastien Ferré and Sebastian Rudolph, editors, *Proceedings of the 7th International Conference on Formal Concept Analysis, (ICFCA 2009)*, volume 5548 of *Lecture Notes in Artificial Intelligence*, pages 146–161. Springer Verlag, 2009.

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

[8] 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, 9 2009.

[9] Daniel Borchmann. Axiomatizing Confident $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-GCIs of Finite Interpretations. Technical Report MATH-AL-08-2012, TU Dresden, September 2012.

[10] 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, *ICDM Workshops*, pages 1083–1090. IEEE, 2011.

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

[12] Bernhard Ganter and Rudolph Wille. *Formal Concept Analysis: Mathematical Foundations.* Springer, Berlin-Heidelberg, 1999.

[13] Sergei O. Kuznetsov. On the intractability of Computing the Duquenne-Guigues Base. *J. UCS*, 10(8):927–933, 2004.

[14] M. Luxenburger. *Implikationen, Abhängigkeiten und Galois-Abbildungen.* PhD thesis, TH Darmstadt, 1993.

[15] Bernhard Nebel. Terminological Cycles: Semantics and Computational Properties. In *Principles of Semantic Networks*, pages 331–362. Morgan Kaufmann, 1991.

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

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

[18] Sebastian Rudolph. Some notes on pseudo-closed sets. In Sergei O. Kuznetsov and Stefan Schmidt, editors, *ICFCA*, volume 4390 of *Lecture Notes in Computer Science*, pages 151–165. Springer, 2007.

[19] Alfred Tarski. A Lattice-Theoretical Fixpoint Theorem and Its Applications. *Pacific Journal of Mathematics*, 1955.

[20] M. J. Zaki and M. Ogihara. Theoretical foundation of association rules. In *SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery (SIGMOD-DMKD'98), Seattle, Friday, June 5, 1998*, pages 1–8, 1998.