Axiomatizing $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-General Concept Inclusions in the Presence of Untrusted Individuals

Daniel Borchmann*

TU Dresden

Abstract. To extract terminological knowledge from data, Baader and Distel have proposed an effective method that allows for the extraction of a base of all valid general concept inclusions of a given finite interpretation. In previous works, to be able to handle small amounts of errors in our data, we have extended this approach to also extract general concept inclusions which are "almost valid" in the interpretation. This has been done by demanding that general concept inclusions which are "almost valid" are those having only an allowed percentage of counterexamples in the interpretation. In this work, we shall further extend our previous work to allow the interpretation to contain both *trusted* and *untrusted individuals*, i. e. individuals from which we know and do not know that they are correct, respectively. The problem we then want to solve is to find a compact representation of all terminological knowledge that is valid for all trusted individuals and is almost valid for all others.

1 Introduction

Constructing description logic knowledge bases is an expensive, time-consuming and often cumbersome task. The main reason for this is that it almost always has to be conducted by human experts, since they provide the means to (more or less) reliably transform informally stated knowledge into a formal reformulation. Thus, methods to assist human experts in constructing these ontologies would be highly helpful. For example, one could provide a first, rough approximation of the desired knowledge base, i. e. a sketch of the ontology, that the expert then could build upon.

The extraction of such first knowledge bases is a widely studied topic. Interesting approaches in this direction have been proposed by Völker and Niepert [14] with their method of *Statistical Schema Induction*, which relies on methods from data mining to extract certain terminological axioms from data given as a set of RDF triples. Another approach has been discussed by Baader and Distel [3], which proposes a method to extract a *finite base* of all valid $\mathcal{EL}^{\perp}$-general concept inclusions (GCIs) of a given finite interpretation, relying on methodology from the mathematical field of *formal concept analysis*. Such a finite base could then be used as a starting point for the terminological part of a future knowledge base.

The latter approach has a particular appeal: since bases are *complete* in the sense that all valid GCIs expressible in $\mathcal{EL}^{\perp}$ already follow from it, such a base provides

* Supported by DFG Graduiertenkolleg 1763 (QuantLA)

all the knowledge which is present in the data. However, this approach also has a drawback: since only valid GCIs are considered, even very rare counterexamples could invalidate otherwise correct (and useful) GCIs, leading the algorithm to come up with a lot of special cases to circumvent such errors.

Thus, based upon the approach of Baader and Distel, the author has developed an extension that allows for finding finite bases of "almost valid" GCIs of a given finite interpretation [8, 6, 7], an extension which turns out to be somehow similar to Statistical Schema Induction. To define the notion of "almost valid", we use the notion of *confidence* from data mining [1]: GCIs are "almost valid" if and only if their confidence is high enough. In other words, the amount of counterexamples for such a GCI has to be rather small compared to all individuals to which this GCIs applies.

However, within this extension, *all* individuals in a given interpretation are suspected to be erroneous, an assumption which may not be correct in general. We thus extend these results in the present work to include the possibility that we "trust" certain individuals, i. e. that we state a-priori that they are correct as they are. This implies that all GCIs which are invalidated by such trusted individuals are false in our original domain of interest, and should thus not be included in an according knowledge base.

The target description logic we would like to consider is $\mathcal{EL}^{\bot}$, i. e. we would like to extract bases of $\mathcal{EL}^{\bot}$-GCIs which are "almost valid" for untrusted individuals and valid for trusted individuals. However, due to technical reasons, we have to take a detour and need to also consider $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$, an extension of $\mathcal{EL}^{\bot}$ by means of cyclic concept descriptions. More precisely, we shall show in this work how to find bases of $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-GCIs in the presence of untrusted individuals. To then obtain from this base of $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-GCIs a base of $\mathcal{EL}^{\bot}$-GCIs we can use results from [6, 10] showing theses $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-bases can effectively be turned into equivalent $\mathcal{EL}^{\bot}$-bases. However, due to space restrictions, we shall not consider this transformation here.

The paper is structured as follows. Firstly, we shall introduce the necessary notions from the field of description logics as they are needed in this work. More precisely, we shall introduce the logics $\mathcal{EL}^{\bot}$ and $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ as well as the notion of *model-based most-specific concept descriptions* from [10]. Then, in Section 3, we shall give a formalization of "almost valid" GCIs and the notion of "trusted" and "untrusted" individuals. Based upon this, we shall give a first finite base of all such GCIs in Section 4.3, and extend this result in Section 4.4 to allow us to perform this computation by means of formal concept analysis. Finally, we close this work by some conclusions and outlook on future work.

2 Preliminaries

We introduce the necessary definitions from the field of description logics that are necessary for our further considerations. More precisely, we shall introduce the description logics $\mathcal{EL}^{\bot}$ and $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ in Section 2.1 as well as the notion of general concept inclusions in Section 2.2. In Section 2.3 we shall discuss the notion of *model-based most-specific concept descriptions.*

2.1 The Description Logics $\mathcal{EL}^{\bot}$ and $\mathcal{EL}^{\bot}_{\mathbf{gfp}}$

Let N_C and N_R be two disjoint sets. An $\mathcal{EL}$-concept description C is formed according to the rule

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

where $A \in N_C$ and $r \in N_R$. An $\mathcal{EL}^{\bot}$-concept description is either $\bot$ or an $\mathcal{EL}$-concept description.

The semantics of $\mathcal{EL}^{\bot}$-concept descriptions are defined through the notion of *interpretations*. An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a set $\Delta^{\mathcal{I}}$ of *individuals* and an *interpretation function* $\cdot^{\mathcal{I}}$ which maps concept names $A \in N_C$ to subsets $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and role name $r \in N_R$ to subsets $r^{\mathcal{I}}$ of $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

The interpretation function $\cdot^{\mathcal{I}}$ can naturally be extended to all $\mathcal{EL}^{\bot}$-concept descriptions C, D in the usual way, i. e.

$$\begin{gathered}\bot^{\mathcal{I}} = \varnothing, \qquad \top^{\mathcal{I}} = \Delta^{\mathcal{I}}, \qquad (C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}, \\ (\exists r.C)^{\mathcal{I}} = \{\, d \in \Delta^{\mathcal{I}} \mid \exists e \in \Delta^{\mathcal{I}} \colon (d, e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}} \,\},\end{gathered}$$

where $r \in N_R$. We shall call the set $C^{\mathcal{I}}$ the *extension* of C in $\mathcal{I}$. For each $x \in \Delta^{\mathcal{I}}$ we shall say that x *satisfies* C if and only if $x \in C^{\mathcal{I}}$.

The main distinction between $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ and $\mathcal{EL}^{\bot}$ is that the former allows for *cyclic concept descriptions*. More formally, let N_D be a set disjoint to both N_C and N_R. We call this set the set of *defined concept names*. A *concept definition* is then an expression of the form $A \equiv C$, where $A \in N_D$ and C is an $\mathcal{EL}^{\bot}$-concept description which can use in the place of concept names from N_C also concept names from N_D. Let $\mathcal{T}$ be a finite set of concept definitions where each defined concept names appears exactly once on the left-hand side of a concept definition in $\mathcal{T}$. Then $\mathcal{T}$ is called a *cyclic TBox*. We define $N_D(\mathcal{T})$ as the set of all concept names from N_D that appear in some concept definition in $\mathcal{T}$.

Then, an $\mathcal{EL}_{\mathrm{gfp}}$-concept description is defined as a pair $(A, \mathcal{T})$, where $\mathcal{T}$ is a cyclic TBox and $A \in N_D$ appears on the left-hand side of a concept definition $(A \equiv C) \in \mathcal{T}$. An $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description is either of the form $\bot$ or is an $\mathcal{EL}_{\mathrm{gfp}}$-concept description.

As an example of an $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description we can consider the concept description E, where

$$\begin{aligned} E = (A, \{\, & A \equiv \mathsf{Cat} \sqcap \exists \mathsf{hunts}.B \\ & B \equiv \mathsf{Mouse} \sqcap \exists \mathsf{hunts}.A \,\}).\end{aligned}$$

Intuitively, E represents all cats hunting a mouse which again hunts a cat — a common situation in the old cartoon series "Tom and Jerry." In other words, given an interpretation $\mathcal{I}$, the semantics of E can be understood as follows: let $A^{\mathcal{I}}$ and $B^{\mathcal{I}}$ be the $\subseteq$-maximal subsets of $\Delta^{\mathcal{I}}$ such that

- all individuals in $A^{\mathcal{I}}$ satisfy Cat and have a hunts-successor in $B^{\mathcal{I}}$, and
- all individuals in $B^{\mathcal{I}}$ satisfy Mouse and have a hunts-successor in $A^{\mathcal{I}}$.

It can be shown that those maximal sets always exist. We then define $E^{\mathcal{I}} := A^{\mathcal{I}}$.

To define the semantics for $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ formally and in general, it is necessary to resolve the cyclic dependencies within $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-concept descriptions. This is done using *greatest fixpoint semantics* [13, 2]. Since the exact definition of the semantics of $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$ exceeds the space available for this publication, we leave out the details and refer the reader to the corresponding publications.

2.2 General Concept Inclusions

Terminological knowledge is represented in the form of *general concept inclusions* (GCIs). These are expressions of the form $C \sqsubseteq D$, where C and D are concept descriptions. We speak of $\mathcal{EL}^{\perp}$-GCIs if both C and D are $\mathcal{EL}^{\perp}$-concept descriptions, and likewise for other description logics.

The semantics of GCIs is again defined via interpretations. We say that an interpretation $\mathcal{I}$ is a *model* of a GCI $C \sqsubseteq D$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. If $\mathcal{I}$ is a model of $C \sqsubseteq D$, then we shall also say that $C \sqsubseteq D$ is *valid* in $\mathcal{I}$. If $C \sqsubseteq D$ is valid in every possible interpretation we shall say that C is *subsumed* by D. This fact is commonly also denoted by $C \sqsubseteq D$ (as a statement, not an expression).

If $\mathcal{L}$ is a set of GCIs and $C \sqsubseteq D$ is another GCI, then we say that $\mathcal{L}$ *entails* $C \sqsubseteq D$ and write $\mathcal{L} \models (C \sqsubseteq D)$, if and only if for every interpretation $\mathcal{J}$ which is a model of all GCIs in $\mathcal{L}$, $\mathcal{J}$ is also a model of $C \sqsubseteq D$.

Since we are going to extract terminological knowledge from interpretations, we can ask for the set of all $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-GCIs for which $\mathcal{I}$ is a model. We shall denote this set $\mathrm{Th}(\mathcal{I})$ and call it the *theory* of $\mathcal{I}$. A *base* of $\mathcal{I}$ is a set $\mathcal{B} \subseteq \mathrm{Th}(\mathcal{I})$ such that every GCI from $\mathrm{Th}(\mathcal{I})$ is already entailed by $\mathcal{B}$.

2.3 Model-Based Most-Specific Concept Descriptions

Let us fix a finite interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. A first attempt to extract terminological knowledge from $\mathcal{I}$ is to consider the set $\mathrm{Th}(\mathcal{I})$ of all valid $\mathcal{EL}^{\perp}$-GCIs of $\mathcal{I}$. However, it is quite easy to see that the number of valid $\mathcal{EL}^{\perp}$-GCIs of $\mathcal{I}$ is infinite in general, for if $C \sqsubseteq D$ is such a GCI, then $\exists r.C \sqsubseteq \exists r.D$ for $r \in N_R$ is a valid $\mathcal{EL}^{\perp}$-GCI as well. Therefore, we cannot simply use the set $\mathrm{Th}(\mathcal{I})$ as a TBox for an ontology. Instead, we try to find a *finite* base $\mathcal{B}$ of $\mathrm{Th}(\mathcal{I})$. Such a base would contain the same information as $\mathrm{Th}(\mathcal{I})$, and since it is finite it could be used as a TBox for an ontology. One of the main results from [10] is to prove that such bases always exist, and also to give an effective method to compute them. These results have been achieved using formal concept analysis [11].

The central notion that has been introduced in [10] for bringing together the description logic $\mathcal{EL}^{\perp}$ and formal concept analysis is the one of *model-based most-specific concept descriptions*. Roughly, for a set $X \subseteq \Delta^{\mathcal{I}}$ we are looking for a concept description that *describes* the individuals in X in the best way possible. More formally, we call an $\mathcal{EL}^{\perp}$-concept description C a *model-based most-specific concept description* for X (in $\mathcal{EL}^{\perp}$) if and only if

- $X \subseteq C^{\mathcal{I}}$ and

– for all $\mathcal{EL}^{\bot}$-concept descriptions D such that $X \subseteq D^{\mathcal{I}}$, it is true that $C \sqsubseteq D$.

It is clear that, if a model-based most-specific concept description for X exists, it is unique up to equivalence. In this case, we shall denote it with $X^{\mathcal{I}}$, since computing model-based most-specific concept descriptions is somehow "dual" to computing extensions $C^{\mathcal{I}}$ of concept descriptions C. If C is a concept description, we shall write $C^{\mathcal{I}\mathcal{I}}$ instead of $(C^{\mathcal{I}})^{\mathcal{I}}$, and likewise for $X^{\mathcal{I}\mathcal{I}} = (X^{\mathcal{I}})^{\mathcal{I}}$ for $X \subseteq \Delta^{\mathcal{I}}$.

However, it may happen that model-based most-specific concept descriptions in $\mathcal{EL}^{\bot}$ may not exists. To see this, consider the example interpretation

$$N_C = \emptyset,\ N_R = \{\, r \,\},\ \Delta^{\mathcal{I}} = \{\, x \,\},\ r^{\mathcal{I}} = \{\, (x,x) \,\}.$$

Then all $\mathcal{EL}^{\bot}$-concept descriptions $\exists r.\exists r.\ldots\exists r.\top$ have the set $X = \{\, x \,\}$ in their extension, but there does not exist a most specific one. On the other hand, it can be seen quite easily that the $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description $(A, \{\, A \equiv \exists r.A \,\})$ is a model-based most-specific concept description of X, if we consider $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$ instead of $\mathcal{EL}^{\bot}$ in the above definition. Indeed, this not a coincidence, as the following result shows.

Theorem 1 (Lemma 4.5 of [10]). *Let $\mathcal{I}$ be a finite interpretation and $X \subseteq \Delta^{\mathcal{I}}$. Then there exists a model-based most-specific concept description of X in $\mathcal{EL}^{\bot}_{gfp}$.*

Because of this result we shall implicitly assume from now on that we are talking about model-based most-specific concept descriptions in $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$.

Before we continue, let us note two facts about model-base most-specific concept descriptions. Firstly, if C is an $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description, then $C \sqsubseteq C^{\mathcal{I}\mathcal{I}}$ is always a valid GCI of $\mathcal{I}$. Furthermore, $C^{\mathcal{I}\mathcal{I}}$ is subsumed by C, again for each $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept description C. Establishing these two facts is not difficult, see [10].

3 Confidence and Trusted Individuals

Recall that we have fixed a finite interpretation $\mathcal{I}$. A first attempt to learn terminological knowledge from $\mathcal{I}$ was to consider $\mathrm{Th}(\mathcal{I})$. However, this approach is not really applicable if we regard this interpretation as somehow *faulty*, in the sense that for certain individuals we are not quite sure whether their concept names or role successors are correct.

Example 2. In [9] the approach by Baader and Distel has been applied to an interpretation $\mathcal{I}_{\mathrm{DBpedia}}$ that has been extracted from RDF Triples from the DBpedia data set by considering the child-relation only. This interpretation had 5262 individuals and $\mathrm{Th}(\mathcal{I}_{\mathrm{DBpedia}})$ can be axiomatized by 1252 GCIs.

However, the GCI $\exists\mathsf{child}.\top \sqsubseteq \mathsf{Person}$ was found not to be valid in $\mathcal{I}_{\mathrm{DBpedia}}$, because of the presence of 4 erroneous counterexamples. However, such a GCI would be considered correct, and it would be preferable if it could be learned, too. Furthermore, 2547 individuals in $\mathcal{I}_{\mathrm{DBpedia}}$ were positive examples for this GCI, i. e. they satisfied both $\exists\mathsf{child}.\top$ and Person. One could then argue that the 4 counterexamples are "not enough" to invalidate the GCI $\exists\mathsf{child}.\top \sqsubseteq \mathsf{Person}$, i. e. this GCI should have been learned as well.

However, it might be the case that for certain individuals we are indeed sure that the are correct. These individuals we shall call *trusted individuals*, whereas the other ones are called *untrusted individuals*.

Example 3. We consider a classical example here. Suppose that we want to learn terminological knowledge about birds, and we consider all GCIs were the number of counterexamples is "small" compared to the number of positive examples (we shall give a formalization for this shortly). Then, the GCI $\mathsf{Birds} \sqsubseteq \mathsf{Flies}$ could be extracted from the data set, as counterexamples to this are quite rare (penguins, ostriches and the like). However, these counterexamples are proper counterexamples, i. e. they are correct.

We can understand the set of all untrusted individuals as a subinterpretation of $\mathcal{I}$.

Definition 4. *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a finite interpretation. A* subinterpretation *of $\mathcal{I}$ is an interpretation $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ such that*

i. $\Delta^{\mathcal{J}} \subseteq \Delta^{\mathcal{I}}$,
ii. $\{ A \in N_C \mid x \in A^{\mathcal{J}} \} = \{ A \in N_C \mid x \in A^{\mathcal{I}} \}$ *for all* $x \in \Delta^{\mathcal{J}}$, *and*
iii. $\{ y \in \Delta^{\mathcal{J}} \mid (x, y) \in r^{\mathcal{J}} \} \subseteq \{ y \in \Delta^{\mathcal{I}} \mid (x, y) \in r^{\mathcal{I}} \}$ *for all* $x \in \Delta^{\mathcal{J}}$ *and* $r \in N_R$.

As already stated in the introduction, we are interested in extracting a compact representation of all GCIs of $\mathcal{I}$ which hold for all trusted individuals and are "almost valid" for all untrusted ones. To formalize the notion of "almost valid", we shall make use of the notion of *confidence* as follows.

Definition 5. *Let $\mathcal{I}$ be a finite interpretation, and let $C \sqsubseteq D$ be an $\mathcal{EL}^{\bot}_{gfp}$-GCI. Then the* confidence *of $C \sqsubseteq D$ in $\mathcal{I}$ is defined as*

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

It can be seen quite easily that $\operatorname{conf}_{\mathcal{I}}(C \sqsubseteq D)$ is the largest number $k \leqslant 1$ such that

$$|(C \sqcap D)^{\mathcal{I}}| \geqslant k \cdot |C^{\mathcal{I}}|.$$

We shall now formally define the set of GCIs we are interested in.

Definition 6. *Let $\mathcal{I}$ be a finite interpretation, let $\mathcal{J}$ be a subinterpretation of $\mathcal{I}$ and let $c \in [0, 1]$. The set of* confident GCIs in $\mathcal{I}$ with untrusted individuals $\mathcal{J}$ *is defined to be the following set of $\mathcal{EL}^{\bot}_{gfp}$-GCIs:*

$$\operatorname{Th}_c(\mathcal{I}, \mathcal{J}) := \{ C \sqsubseteq D \mid C^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \subseteq D^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \\ \text{and } |(C \sqcap D)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |C^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \}.$$

Note that we cannot define $\operatorname{Th}_c(\mathcal{I}, \mathcal{J})$ to consist of all GCIs satisfying $\operatorname{conf}_{\mathcal{J}}(C \sqsubseteq D) \geqslant c$ instead of $|(C \sqcap D)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |C^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|$. This is because $C^{\mathcal{I}} \cap \Delta^{\mathcal{J}} = C^{\mathcal{J}}$

is not true in general. However, if one restricts one attention to *closed* subinterpretations $\mathcal{J}$, i. e. subinterpretations where no role edges exist between elements from $\Delta^{\mathcal{J}}$ and $\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}$ and vice versa, then $C^{\mathcal{I}} \cap \Delta^{\mathcal{J}} = C^{\mathcal{J}}$ is indeed true, as it has been shown in [10, Lemma 6.12].

Another observation about $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ is that this set is not necessarily closed under entailment. In other words, it is possible that $\mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \models (C \sqsubseteq D)$ but $(C \sqsubseteq D) \notin \mathrm{Th}_c(\mathcal{I}, \mathcal{J})$. However, we can consider GCIs in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ as *valid* in a certain domain of interest, and counterexamples in $\mathcal{J}$ as erroneous. Then $(C \sqsubseteq D) \notin \mathrm{Th}_c(\mathcal{I}, \mathcal{J})$, which just means that the amount of counterexamples in $\mathcal{J}$ is too high, indicates that either the data $\mathcal{I}$ is not sufficiently good enough for learning $C \sqsubseteq D$ (if it is not valid in our domain) or that some GCIs in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ are actually not valid in our domain. In both cases, further refinement is necessary, for the first case by checking *completeness with respect to the underlying domain of interest* [5], and in the second case by finding *reasons for errors* in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$, for example using *axiom pinpointing* [4].

4 Axiomatizing GCIs in the Presence of Untrusted Individuals

We shall now show how to axiomatize $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$. To this end, we show in Sections 4.3 and 4.4 how one can compute *finite bases of* $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$, i. e. finite sets $\mathcal{B} \subseteq \mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ such that all GCIs in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ are already entailed by $\mathcal{B}$. To this end, we shall make use of results obtained by Baader and Distel [10, 3], which are introduced in Section 4.2. As these results make use of formal concept analysis, we shall first introduce some notions of it first.

4.1 Formal Concept Analysis

Formal concept analysis is a subfield of mathematical order theory that is usually concerned with investigating different representations of complete lattices. However, is has also been used in different areas, as in data mining and classification.

The fundamental notion of formal concept analysis [11] is the one of a *formal context*. A formal context is a triple $\mathbb{K} = (G, M, I)$ where G and M are sets and $I \subseteq G \times M$. Intuitively, we think of the set G as the set of *objects*, the set M as the set of *attributes*, and of the set I as an *incidence relation* between objects and attributes. If $g \in G$ and $m \in M$, we say that g *has* the attribute m if and only if $(g, m) \in I$. A formal context $\mathbb{L} = (H, N, J)$ is a *subcontext* of $\mathbb{K}$ if and only if $H \subseteq G$, $N \subseteq M$, $J \subseteq I$.

For a set of objects $B \subseteq G$, we can ask for the set B' of common attributes of all objects in B, i. e.

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

The set B' is called the *derivation* of B in $\mathbb{K}$. Dually, we define for $A \subseteq M$ the set A' of all objects satisfying all attributes in A.

In the formal context $\mathbb{K}$ we can ask the question whether an object g that has all attributes from a set A_1 always also has all attributes from a set A_2, i. e. whether it is true that $g \in A_1'$ implies $g \in A_2'$. We can formalize this question as follows: we call a pair (A_1, A_2) with $A_1, A_2 \subseteq M$ an *implication*, usually written as $A_1 \to A_2$. We shall say that the implication $A_1 \to A_2$ is *valid* in $\mathbb{K}$ if and only if $A_1' \subseteq A_2'$. The set of all implications $A_1 \to A_2$ with $A_1, A_2 \subseteq M$ is denoted by $\operatorname{Imp}(M)$, and the set of all valid implications of $\mathbb{K}$ is called its *theory* and is denoted by $\operatorname{Th}(\mathbb{K})$.

A set $\mathcal{L} \subseteq \operatorname{Imp}(M)$ of implications *entails* an implication $A_1 \to A_2$ if and only if for all contexts in which $\mathcal{L}$ is true, the implication $A_1 \to A_2$ is true as well. A set $\mathcal{L} \subseteq \operatorname{Th}(\mathbb{K})$ is called a *base* of $\mathbb{K}$ if and only if all valid implications of $\mathbb{K}$ are entailed by $\mathcal{L}$.

It is obvious that we can extend each set $\mathcal{K} \subseteq \operatorname{Th}(\mathbb{K})$ to a base of $\mathbb{K}$, provided that $\mathbb{K}$ is finite. We call $\mathcal{L} \subseteq \operatorname{Th}(\mathbb{K})$ a *base with background knowledge* $\mathcal{K}$ if and only if $\mathcal{L} \cup \mathcal{K}$ is a base of $\mathbb{K}$. If $\mathcal{K} = \varnothing$, then bases with background knowledge $\mathcal{K}$ are just bases of $\mathbb{K}$.

A particularly interesting base is the so called *canonical base* $\operatorname{Can}(\mathbb{K}, \mathcal{K})$ of $\mathbb{K}$, for some given background knowledge $\mathcal{K}$. Making the definition of this base understandable is hardly possible in the given amount of space, and we refer the reader to [11] for further details. However, we still note that it is well known that $\operatorname{Can}(\mathbb{K}, \mathcal{K})$ is a base of smallest cardinality with background knowledge $\mathcal{K}$, i. e. every set of implications with less elements than $\operatorname{Can}(\mathbb{K}, \mathcal{K})$ cannot be a base of $\mathbb{K}$ with background knowledge $\mathcal{K}$.

4.2 Results by Baader and Distel

To connect description logics with formal concept analysis, Baader and Distel make use of the notion of of model-based most-specific concept descriptions, and define a formal context $\mathbb{K}_{\mathcal{I}}$ which captures all relevant information on the valid $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-GCIs of $\mathcal{I}$. For this, we define

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

The set $M_{\mathcal{I}}$ has the particular property that all model-based most-specific concept descriptions in $\mathcal{I}$ are *expressible in terms of* $M_{\mathcal{I}}$ [10]: let us denote for a set $U \subseteq M_{\mathcal{I}}$ with $\bigsqcap U$ the concept description that is either $\top$, when U is empty, or $V_1 \sqcap \ldots \sqcap V_n$, when $U = \{ V_1, \ldots, V_n \}$. Then a concept description C is *expressible in terms* of $M_{\mathcal{I}}$ if and only if there exists a set $N \subseteq M_{\mathcal{I}}$ such that $C \equiv \bigsqcap N$.

Having defined the set $M_{\mathcal{I}}$, we can introduce the notion of the *induced context* of $\mathcal{I}$. This is the formal context $\mathbb{K}_{\mathcal{I}} = (\Delta^{\mathcal{I}}, M_{\mathcal{I}}, \nabla)$, where for all $x \in \Delta^{\mathcal{I}}$ and $C \in M_{\mathcal{I}}$, it is true that $x \nabla C$ if and only if $x \in C^{\mathcal{I}}$.

The derivation operators in $\mathbb{K}_{\mathcal{I}}$, the interpretation function $\cdot^{\mathcal{I}}$ and model-based most-specific concept descriptions are closely related.

Proposition 7. *Let $A \subseteq \Delta^{\mathcal{I}}, B \subseteq M_I$. Then $A^{\mathcal{I}} \equiv \bigsqcap A'$ and $B' = (\bigsqcap B)^{\mathcal{I}}$, where the derivations are conducted in $\mathbb{K}_{\mathcal{I}}$.*

With some more technical machinery it can even be shown that $(\bigsqcap A)^{\mathcal{I}\mathcal{I}} = \bigsqcap A''$ is true for each $A \subseteq M_{\mathcal{I}}$, i. e. model-based most-specific concept descriptions and the derivation operators in the induced context of $\mathcal{I}$ are closely related. This connection also extends to the canonical base of $\mathbb{K}_{\mathcal{I}}$ and $\mathcal{EL}^{\perp}_{\mathrm{gfp}}$-bases of $\mathcal{I}$.

Theorem 8 (5.13 and 5.18 of [10]). *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

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

is a finite $\mathcal{EL}^{\bot}_{gfp}$-base of $\mathcal{I}$ of minimal cardinality.

Note that the set $S_{\mathcal{I}}$ contains knowledge which is trivially true in every interpretation, but not necessarily in every formal context. More precisely, if $C \sqsubseteq D$, then we do not need to state this GCI explicitly in a base. However, the corresponding implication $\{ C \} \to \{ D \}$ may not necessarily be valid in every formal context, and therefore bases of $\mathbb{K}$ have to contain information to entail such implications. However, we are only interested in bases of $\mathcal{I}$, which is why we add the set $S_{\mathcal{I}}$ as background knowledge.

4.3 A First Finite Base

The aim of this section is to provide a first result for axiomatizing the interpretation $\mathcal{I}$ with untrusted individuals $\mathcal{J}$. To this end we shall make use of ideas from M. Luxenburger's theory of *partial implications* in formal contexts [12]. Due to space restrictions, however, we shall not recall his ideas in the original setting here, but shall discuss them in the appropriate reformulation to our setting only.

The two ideas from Luxenburger's work we shall make use of are actually quite simple. First of all, we observe that

$$\mathrm{Th}(\mathcal{I}) \subseteq \mathrm{Th}_c(\mathcal{I}, \mathcal{J}).$$

Therefore, to finitely axiomatize the set $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$, i. e. to find a finite set $\mathcal{B} \subseteq \mathrm{Th}_c(\mathcal{I}, \mathcal{J})$ that entails all GCIs in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$, it suffices to find a finite base $\mathcal{C}$ of $\mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \setminus \mathrm{Th}(\mathcal{I})$, since then the set

$$\mathcal{B}_{\mathrm{Can}} \cup \mathcal{C}$$

will be a finite base of $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$.

We thus concentrate on finding such a finite base $\mathcal{C}$ of $\mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \setminus \mathrm{Th}(\mathcal{I})$, i. e. on finding a finite set $\mathcal{C} \subseteq \mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \setminus \mathrm{Th}(\mathcal{I})$ which already entails all GCIs in $\mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \setminus \mathrm{Th}(\mathcal{I})$. For this, we make two observations for GCIs $(C \sqsubseteq D) \in \mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \setminus \mathrm{Th}(\mathcal{I})$.

Firstly, it is true that

$$(C \sqsubseteq D) \in \mathrm{Th}_c(\mathcal{I}, \mathcal{J}) \iff (C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}) \in \mathrm{Th}_c(\mathcal{I}, \mathcal{J}).$$

This is mainly due to the fact that $E^{\mathcal{I}\mathcal{I}\mathcal{I}} = E^{\mathcal{I}}$ and $(E \sqcap F)^{\mathcal{I}} = (E^{\mathcal{I}\mathcal{I}} \sqcap F^{\mathcal{I}\mathcal{I}})^{\mathcal{I}}$ are true for all $\mathcal{EL}^{\bot}_{\mathrm{gfp}}$-concept descriptions E, F, as can be seen easily from the definition of model-based most-specific concept descriptions. We thus obtain

$$C^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \subseteq D^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \iff C^{\mathcal{I}\mathcal{I}\mathcal{I}} \setminus \Delta^{\mathcal{J}} \subseteq D^{\mathcal{I}\mathcal{I}\mathcal{I}} \setminus \Delta^{\mathcal{J}}$$
$$|(C \sqcap D)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |C^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \iff |(C^{\mathcal{I}\mathcal{I}} \sqcap D^{\mathcal{I}\mathcal{I}})^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |C^{\mathcal{I}\mathcal{I}\mathcal{I}} \cap \Delta^{\mathcal{J}}|$$

Secondly, if $\mathcal{B}$ is a base of $\operatorname{Th}(\mathcal{I})$, then

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

since $\mathcal{B} \models (C \sqsubseteq C^{\mathcal{I}\mathcal{I}})$, and $D^{\mathcal{I}\mathcal{I}} \sqsubseteq D$.

Thus, let us define

$$\operatorname{Conf}(\mathcal{I}, c, \mathcal{J}) := \{ C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}} \mid (C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}) \in \operatorname{Th}_c(\mathcal{I}, \mathcal{J}) \},$$

It has been shown in [10] that if $\mathcal{I}$ is finite, there exist only finitely many different model-based most-specific concept descriptions up to equivalence. Therefore, we immediately obtain the following result:

Theorem 9. *Let $\mathcal{I}$ be a finite interpretation, $\mathcal{J}$ be a subinterpretation of $\mathcal{I}$ and $c \in [0, 1]$. If $\mathcal{B}$ is a finite base of $\mathcal{I}$, then the set*

$$\mathcal{B} \cup \operatorname{Conf}(\mathcal{I}, c, \mathcal{J})$$

is a finite base of $\operatorname{Th}_c(\mathcal{I}, \mathcal{J})$.

4.4 Computing a Base by means of FCA

The result of the previous section provides us with an effective method to compute bases of confident GCIs in the presence of trusted and untrusted individuals. In this section, we shall extend these result to allow us a computation of such a base using methods from formal concept analysis. More precisely, we shall see that we can view our initial data $\mathcal{I}$ and $\mathcal{J}$ as a certain formal context, and computing a base of $\mathcal{I}$ with untrusted individuals $\mathcal{J}$ can then be understood as computing a certain set of implications from this context.

There are two reasons why we are interested in such a transformation. Firstly, the overall goal of our considerations is to adapt the algorithm of *attribute exploration* from formal concept analysis to our setting of extracting terminological knowledge from interpretations (see also Section 5). Secondly, such a reformulation might allow us to use high performance algorithms from the field of data mining to support us in our task, because the field of formal concept analysis can be understood as a theoretical foundation of data mining [15].

The actual transformation is now quite simple: as for general concept inclusions, we can define the notion of confidence for implications as well, i. e. if $\mathbb{K}$ is a formal context and $A \to B$ an implication then $\operatorname{conf}_{\mathbb{K}}(A \to B) := 1$ if $A' = \varnothing$ and $\operatorname{conf}_{\mathbb{K}}(A \to B) = |(A \cup B)'|/|A'|$ otherwise. We shall denote with $\operatorname{Imp}_c(\mathbb{K})$ the set of all implications from $\operatorname{Imp}(M)$ which have confidence at least c in $\mathbb{K}$.

The transformation then uses the induced context $\mathbb{K}_{\mathcal{I}}$ of $\mathcal{I}$ and asserts that a certain set of implications yields a base of $\operatorname{Th}_c(\mathcal{I}, \mathcal{J})$.

Theorem 10. *Let $\mathcal{I}$ be a finite interpretation, $\mathcal{J}$ be a subinterpretation of $\mathcal{I}$ and $c \in [0, 1]$. Let us denote for $X \subseteq \Delta^{\mathcal{I}}$ with $\mathbb{K}_{\mathcal{I}}|_X$ the subcontext of $\mathbb{K}_{\mathcal{I}}$ restricted to the object set X, and let us define*

$$T := \operatorname{Imp}_c(\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{J}}}) \cap \operatorname{Th}(\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}}).$$

If then $\mathcal{L} \subseteq T$ *is complete for* T*, then*

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

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

A proof of this theorem can be found in Section A of the appendix.

5 Conclusions and Further Work

In this work we have presented an extension to our previous work of axiomatizing general concept inclusions using confidence, that takes into account the existence of *trusted individuals* from which we can be sure that their properties in the given data are correct. Essentially, we not only consider general concept inclusions whose confidence is high enough, but we also require for such general concept inclusions that they are not falsified by trusted individuals. For these general concept inclusions we have discussed the existence and the computation of bases. In particular, we have shown that we can utilize methodology from the field of formal concept analysis to obtain such bases. This has the particular appeal that in the long run we might be able to utilize fast data mining algorithms to compute bases of confident general concept inclusions which respect trusted individuals.

The considerations we have presented in this work are part of a larger attempt to apply the algorithm of *attribute exploration* from formal concept analysis to our setting of confident general concept inclusions. This algorithm has been used before to complete knowledge bases [5]. Essentially, the algorithm generates logical consequences which can be drawn from the given knowledge and asks the expert for their validity. If the expert declines, then the given knowledge is incomplete and the expert has to provide additional information. This algorithm is repeated as long as new consequences can be generated.

An attribute exploration algorithm in the setting of confident general concept inclusions would provide an attempt to solve the issue of *rare counterexamples* which comes from the heuristic approach of confident general concept inclusions, as it has been indicated in Example 3: if we consider all general concept inclusions whose confidence is "high enough", we may neglect certain counterexamples which are appear seldom in the data but are correct otherwise. An attribute exploration algorithm could help to resolve this issue by consulting the expert whether certain confident general concept inclusions are valid or not. If the expert is then asked a general concept inclusion which has rare counterexample, the expert could add (or mark) these counterexamples as trusted individuals, thus prohibiting the algorithm from considering them as errors. Thus, the approach of confident general concept inclusions may benefit from an attribute exploration algorithm, and its design is part of future work.

Acknowledgments The author is deeply grateful for the thorough, constructive and motivating comments by the anonymous reviewers, who helped improving the quality of this work.

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.* 1993, pp. 207–216.

[2] Franz Baader. "Terminological Cycles in a Description Logic with Existential Restrictions". In: *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence.* (Acapulco, Mexico). Ed. by Georg Gottlob and Toby Walsh. Morgan Kaufmann, 2003, pp. 325–330.

[3] Franz Baader and Felix Distel. "A Finite Basis for the Set of $\mathcal{EL}$-Implications Holding in a Finite Model". In: *Proceedings of the 6th International Conference on Formal Concept Analysis.* (Montreal, Canada). Ed. by Raoul Medina and Sergei A. Obiedkov. Vol. 4933. Lecture Notes in Computer Science. Springer, 2008, pp. 46–61.

[4] Franz Baader and Rafael Peñaloza. "Axiom Pinpointing in General Tableaux". In: *TABLEAUX.* Ed. by Nicola Olivetti. Vol. 4548. Lecture Notes in Computer Science. Springer, 2007, pp. 11–27. ISBN: 978-3-540-73098-9.

[5] Franz Baader et al. "Completing Description Logic Knowledge Bases using Formal Concept Analysis". In: *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence.* AAAI Press, 2007, pp. 230–235.

[6] Daniel Borchmann. *Axiomatizing Confident $\mathcal{EL}^{\bot}_{gfp}$-GCIs of Finite Interpretations.* Report MATH-AL-08-2012. Dresden, Germany: Chair of Algebraic Structure Theory, Institute of Algebra, Technische Universität Dresden, 2012.

[7] Daniel Borchmann. *Axiomatizing $\mathcal{EL}^{\bot}$-Expressible Terminological Knowledge from Erroneous Data.* to appear in: Proceedings of the Seventh Internation Conference on Knowledge Capture, KCAP 2013.

[8] Daniel Borchmann. *On Confident GCIs of Finite Interpretations.* LTCS-Report 12-06. See `http://lat.inf.tu-dresden.de/research/reports.html`. Dresden: Institute for Theoretical Computer Science, TU Dresden, 2012.

[9] Daniel Borchmann and Felix Distel. "Mining of $\mathcal{EL}$-GCIs". In: *ICDM Workshops.* Ed. by Myra Spiliopoulou et al. IEEE, 2011, pp. 1083–1090. ISBN: 978-0-7695-4409-0.

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

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

[12] Michael Luxenburger. "Implikationen, Abhängigkeiten und Galois-Abbildungen". PhD thesis. TH Darmstadt, 1993.

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

[14] Johanna Völker and Mathias Niepert. "Statistical Schema Induction". In: *The Semantic Web: Research and Applications - 8th Extended Semantic Web Conference, Proceedings, Part I.* (Heraklion, Crete, Greece). Ed. by Grigoris

Antoniou et al. Vol. 6643. Lecture Notes in Computer Science. Springer, 2011, pp. 124–138. ISBN: 978-3-642-21033-4.
[15] Mohammed Javeed Zaki and Mitsunori Ogihara. “Theoretical foundation of association rules”. In: *SIGMOD’98 Workshop on Research Issues in Data Mining and Knowledge Discovery (SIGMOD-DMKD’98)*. 1998, pp. 1–8.

A Proofs for Section 4.4

For completeness, we provide the proofs we have omitted in Section 4.4.

The main idea of the proof is to show that set $\bigsqcap \mathcal{L}$ satisfies $\bigsqcap \mathcal{L} \models \mathcal{B}_{\text{Can}} \cup \operatorname{Conf}(\mathcal{I}, c, \mathcal{J})$, and thus Theorem 9 yields the desired result. A crucial cornerstone in this argumentation is the following lemma.

Lemma 11. *Let M be a set of concept descriptions, and let $\mathcal{L} \subseteq \operatorname{Imp}(M)$ and $(X \to Y) \in \operatorname{Imp}(M)$. Then $\mathcal{L} \models (X \to Y)$ implies $\bigsqcap \mathcal{L} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$, where*

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

Proof. Let $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ be an interpretation such that $\mathcal{J} \models \bigsqcap \mathcal{L}$. Let us define a formal context $\mathbb{K}_{\mathcal{J},M} = (\Delta^{\mathcal{J}}, M, \nabla)$ via

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

for all $x \in \Delta^{\mathcal{J}}, C \in M$.

We shall show now that $\mathbb{K}_{\mathcal{J},M} \models \mathcal{L}$. Let $(E \to F) \in \mathcal{L}$. Then $(\bigsqcap E)^{\mathcal{J}} \subseteq (\bigsqcap F)^{\mathcal{J}}$, since $\mathcal{J} \models \bigsqcap \mathcal{L}$. It is not hard to see that $(\bigsqcap E)^{\mathcal{J}} = E'$, where the derivation has been done in $\mathbb{K}_{\mathcal{J},M}$. Therefore, it is true that $E' \subseteq F'$, and thus $\mathbb{K}_{\mathcal{J},M} \models (E \to F)$.

Since $\mathcal{L} \models (X \to Y)$, it is true that $\mathbb{K}_{\mathcal{J},M} \models (X \to Y)$, i. e. $X' \subseteq Y'$. As $(\bigsqcap X)^{\mathcal{J}} = X'$, it is thus true that $(\bigsqcap X)^{\mathcal{J}} \subseteq (\bigsqcap Y)^{\mathcal{J}}$, i. e. $\bigsqcap \mathcal{L} \models (\bigsqcap X \sqsubseteq \bigsqcap Y)$.

Theorem 11. *Let $\mathcal{I}$ be a finite interpretation, $\mathcal{J}$ be a subinterpretation of $\mathcal{I}$ and $c \in [0,1]$. Let us denote for $X \subseteq \Delta^{\mathcal{I}}$ with $\mathbb{K}_{\mathcal{I}}|_X$ the subcontext of $\mathbb{K}_{\mathcal{I}}$ restricted to the object set X, and let us define*

$$T := \operatorname{Imp}_c(\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{J}}}) \cap \operatorname{Th}(\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}}).$$

If $\mathcal{L} \subseteq T$ is complete for T, then

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

is a base of $\operatorname{Th}_c(\mathcal{I}, \mathcal{J})$.

Proof. We need to show two claims, namely

i. $\bigsqcap \mathcal{L} \subseteq \operatorname{Th}_c(\mathcal{I}, \mathcal{J})$ and
ii. $\bigsqcap \mathcal{L}$ is complete for $\operatorname{Th}_c(\mathcal{I}, \mathcal{J})$.

For the first claim we need to show that for every GCI $(\bigsqcap X \sqsubseteq \bigsqcap Y) \in \bigsqcap \mathcal{L}$ it is true that

i. $|(\bigsqcap X \sqcap \bigsqcap Y)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |(\bigsqcap X)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|$
ii. $(\bigsqcap X)^{\mathcal{I}} \setminus \Delta^{\mathcal{J}} \subseteq (\bigsqcap Y)^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}$

For the first subclaim, we observe that $\mathrm{conf}_{\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{J}}}}(X \to Y) \geqslant c$ for all $(X \to Y) \in \mathcal{L}$, i. e.

$$|(X \cup Y)' \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |X' \cap \Delta^{\mathcal{J}}|$$

Since $(\bigsqcap X)^{\mathcal{I}} = X'$, we obtain

$$|(\bigsqcap(X \cup Y))^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |(\bigsqcap X)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|,$$

and from $\bigsqcap(X \cup Y) \equiv \bigsqcap X \sqcap \bigsqcap Y$ it follows immediately that

$$|(\bigsqcap X \sqcap \bigsqcap Y)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |(\bigsqcap X)^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|$$

as required.

For the second subclaim, we observe that $X' \setminus \Delta^{\mathcal{J}} \subseteq Y' \setminus \Delta^{\mathcal{J}}$, since $X \to Y$ is valid in the formal context $\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{I}} \setminus \Delta^{\mathcal{J}}}$. Since $X' = (\bigsqcap X)^{\mathcal{I}}$ and $Y' = (\bigsqcap Y)^{\mathcal{I}}$ the claim follows.

We have therefore shown that $\bigsqcap \mathcal{L} \subseteq \mathrm{Th}_c(\mathcal{I}, \mathcal{J})$.

We now show that $\bigsqcap \mathcal{L}$ is complete for $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$. To this end, we shall show that

i. $\bigsqcap \mathcal{L} \models (\bigsqcap U \sqsubseteq (\bigsqcap U)^{\mathcal{I}\mathcal{I}})$ for all $U \subseteq M_{\mathcal{I}}$, in particular, $\bigsqcap \mathcal{L} \models \mathcal{B}_{\mathrm{Can}}$;
ii. $\bigsqcap \mathcal{L} \models \mathrm{Conf}(\mathcal{I}, c, \mathcal{J})$.

If we can establish these claims, then by Theorem 9 we obtain from $\bigsqcap \mathcal{L} \models \mathcal{B}_{\mathrm{Can}} \cup \mathrm{Conf}(\mathcal{I}, c, \mathcal{J})$ the completeness of $\bigsqcap \mathcal{L}$ for $\mathrm{Th}_c(\mathcal{I}, \mathcal{J})$.

Let $U \subseteq M_{\mathcal{I}}$. Since $\mathcal{L}$ entails all valid implications of $\mathbb{K}_{\mathcal{I}}$, we obtain

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

By Lemma 11, it follows that $\mathcal{L} \models (\bigsqcap U \sqsubseteq (\bigsqcap U''))$. Since $(\bigsqcap U'') \equiv (\bigsqcap U)^{\mathcal{I}\mathcal{I}}$, we obtain the validity of the subclaim.

For the second subclaim, let $(C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}}) \in \mathrm{Conf}(\mathcal{I}, c, \mathcal{J})$. We define $U := C^{\mathcal{I}}, V := D^{\mathcal{I}}$. Then $U^{\mathcal{I}} \equiv \bigsqcap U'$ and $V^{\mathcal{I}} \equiv \bigsqcap V'$, so

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

We show that $\mathcal{L} \models (U' \to V')$. For this we recall that

$$|(C^{\mathcal{I}\mathcal{I}} \sqcap D^{\mathcal{I}\mathcal{I}})^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |(C^{\mathcal{I}\mathcal{I}})^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|$$

Now since $\bigsqcap U' \equiv U^{\mathcal{I}} \equiv C^{\mathcal{I}\mathcal{I}}$ and $\bigsqcap V' \equiv D^{\mathcal{I}\mathcal{I}}$, we obtain

$$|(\bigsqcap U' \sqcap \bigsqcap V')^{\mathcal{I}} \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |(\bigsqcap U')^{\mathcal{I}} \cap \Delta^{\mathcal{J}}|$$

As shown before, this implies that

$$|(U' \cup V')' \cap \Delta^{\mathcal{J}}| \geqslant c \cdot |U'' \cap \Delta^{\mathcal{J}}|,$$

where the derivations are conducted in $\mathbb{K}_{\mathcal{I}}$. In other words, it is true that

$$\mathrm{conf}_{\mathbb{K}_{\mathcal{I}}|_{\Delta^{\mathcal{J}}}}(U' \to V') \geqslant c.$$

Thus, $\mathcal{L} \models (U' \to V')$, and Lemma 11 implies $\bigsqcap \mathcal{L} \models (\bigsqcap U' \sqsubseteq \bigsqcap V')$, thus $\bigsqcap \mathcal{L} \models (U^{\mathcal{I}} \sqsubseteq V^{\mathcal{I}}) = (C^{\mathcal{I}\mathcal{I}} \sqsubseteq D^{\mathcal{I}\mathcal{I}})$, as required.