A Finite Basis for the Set of $\mathcal{EL}$-Implications Holding in a Finite Model

Franz Baader* and Felix Distel**

Theoretical Computer Science, TU Dresden, Germany
{baader,felix}@tcs.inf.tu-dresden.de

Abstract. Formal Concept Analysis (FCA) can be used to analyze data given in the form of a formal context. In particular, FCA provides efficient algorithms for computing a minimal basis of the implications holding in the context. In this paper, we extend classical FCA by considering data that are represented by relational structures rather than formal contexts, and by replacing atomic attributes by complex formulae defined in some logic. After generalizing some of the FCA theory to this more general form of contexts, we instantiate the general framework with attributes defined in the Description Logic (DL) $\mathcal{EL}$, and with relational structures over a signature of unary and binary predicates, i.e., models for $\mathcal{EL}$. In this setting, an implication corresponds to a so-called general concept inclusion axiom (GCI) in $\mathcal{EL}$. The main technical result of this paper is that, in $\mathcal{EL}$, for any finite model there is a *finite* set of implications (GCIs) holding in this model from which all implications (GCIs) holding in the model follow.

1 Introduction

Classical Formal Concept Analysis [12] assumes that data from an application are given by a formal context, i.e., by a set of objects G, a set of attributes M, and an incidence relation I that states whether or not an object satisfies a certain attribute. To analyze the data given by such a context, FCA provides tools for computing a minimal basis for the implications between sets of attributes holding in the context [11,8]. An implication $A \to B$ between sets of attributes A, B holds in a given context if all objects satisfying every attribute in A also satisfy every attribute in B. A classical result by Duquenne and Guigues [13] says that such a unique minimal basis always exists. If the set of attributes is finite, which is usually assumed, this basis is trivially finite as well.

From a model-theoretic or (first-order predicate) logical point of view, a formal context is a very simple relational structure where all predicates (the attributes) are unary. In many applications, however, data are given by more complex relational structures where objects can be linked by relations of arities greater than 1. In order to take these more complex relationships between objects into account

* Partially supported by NICTA, Canberra Research Lab.
** Supported by the Cusanuswerk.

R. Medina and S. Obiedkov (Eds.): ICFCA 2008, LNAI 4933, pp. 46–61, 2008.

when analyzing the data, we consider concepts defined in a certain logic rather than simply sets of atomic attributes (i.e., conjunctions of unary predicates). Intuitively, a concept is a formula with one free variable, and thus determines a subset of the domain (the extension of the concept) for any model of the logic used to construct these formulae. We show that, under certain conditions on this logic, many of the basic results from FCA can be extended to this more general framework. Basically, this requirement is that a finite set of objects (i.e., elements of the domain of a given model) always has a most specific concept describing these objects. The operator that goes from a finite set of objects to its most specific concept corresponds to the prime operator in classical FCA, which goes from a set of objects A to the set of attributes A' that all objects from the set have in common. The classical prime operator in the other direction, which goes from a set of attributes B to the set of objects B' satisfying all these attributes, has as its corresponding operator the one that goes from a concept to its extension.

We instantiate this general framework with concepts defined in the Description Logic $\mathcal{EL}$ [2,3], i.e., formal contexts are replaced by finite models of this DL and attributes are $\mathcal{EL}$-concepts. Though being quite inexpressive, $\mathcal{EL}$ has turned out to be very useful for representing biomedical ontologies such as SNOMED [22] and the Gene Ontology [23]. A major advantage of using an inexpressive DL like $\mathcal{EL}$ is that it allows for efficient reasoning procedures [3,5]. Actually, it turns out that $\mathcal{EL}$ itself does not satisfy the requirements on the logic needed to transfer results from FCA since objects need not have a most specific concept. However, if we extend $\mathcal{EL}$ to $\mathcal{EL}_{\mathrm{gfp}}$ by allowing for cyclic concept definitions interpreted with greatest fixpoint semantics, then the resulting logic satisfies all the necessary requirements. Implications in this setting correspond to so-called general concept inclusion axioms (GCIs), which are available in modern ontology languages such as OWL [14] and are supported by most DL systems [15].

The main technical result of this paper is that, in $\mathcal{EL}$ and in $\mathcal{EL}_{\mathrm{gfp}}$, the set of GCIs holding in a finite model always has a finite basis, i.e., although there are in general infinitely many such GCIs, we can always find a finite subset from which the rest follows. We construct such a finite basis first for $\mathcal{EL}_{\mathrm{gfp}}$, and then show how this basis can be modified to yield one for $\mathcal{EL}$. Due to the space limitation, we cannot give complete proofs of these results. They can be found in [4].

Related work. There have been previous approaches for dealing with more complex contexts involving relations between objects. So-called power context families [24] allow for the representation of relational structures by using a separate (classical) context for each arity, where the objects of the context for arity n are n-tuples. As such, power context families are just an FCA-style way of representing relational structures. In order to make use of the more complex relational structure given by power context families, Prediger [16,18,17] and Priss [19] allow the knowledge engineer to define new attributes, and provide means for handling the dependencies between the newly defined attributes and existing attributes by means of formal concept analysis. However, rather than considering all complex attributes definable by the logical language, as our approach does, they restrict the attention to finitely many attributes explicitly defined by the knowledge engineer.

Similar to our general framework, Ferré [6,7] considers complex attributes definable by some logical language. The equivalent of a formal context, called logical context in [6,7], associates a formula (i.e., a complex attribute) with each object. Since the authors assume that formulae form a join-semilattice, the formula associated with a set of objects is obtained as the join of the formulae associated with the elements of the set. Our general framework can be seen as an instance of the one defined in [6,7], where the association of formulae to (sets of) objects is defined using the semantics of the logic in question. However, Ferré's work does not consider implications, which is the main focus of the present paper (see [4] for a more detailed comparison of our approach with the one in [6,7]). An approach similar to the one of [6,7] was developed in [10] motivated by an application in biochemistry.

The work whose objectives are closest to ours is the one by Rudolph [20,21], who considers attributes defined in the DL $\mathcal{FLE}$, which is more expressive than $\mathcal{EL}$. However, instead of using one generalized context with infinitely many complex attributes, he considers an infinite family of contexts, each with finitely many attributes, obtained by restricting the so-called role depth of the concepts. He then applies attribute exploration [9] to the classical contexts obtained this way, in each step increasing the role depths until a certain termination condition applies. Rudolph shows that, for a finite model, this condition will always be satisfied eventually, and that the implication bases of the contexts considered up to that step contain enough information to decide, for any GCI between $\mathcal{FLE}$-concepts, whether this GCI holds in the given model or not. However, these implication bases do not appear to yield a basis for all the GCIs holding in the given finite model, though it might be possible to modify Rudolph's approach such that it produces a basis in our sense. The main problem with this approach is, however, that the number of attributes grows very fast when the role depth grows (this number increases at least by one exponential in each step).

2 The General Framework

In classical FCA, a formal context (G, M, I) consists of a set of objects G, a set of attributes M, and an incidence relation $I \subseteq G \times M$. Such a formal context induces two operators (both usually denoted by $\cdot'$), one mapping each set of objects A to the set of attributes A' these objects have in common, and the other mapping each set of attributes B to the set of objects satisfying these attributes. A formal concept is a pair (A, B) such that $G \supseteq A = B'$ and $M \supseteq B = A'$. The set A is the extensional description of the concept whereas B is its intensional description. The two $\cdot'$ operators form a Galois connection, and if applied twice yield closure operators $\cdot''$ on the set of objects and the set of attributes, respectively.

In our general framework, we assume that intensional descriptions of sets of objects are given by concept descriptions. A *concept description language* is a pair $(\mathcal{L}, \mathcal{I})$, where $\mathcal{L}$ is a set, whose elements are called *concept descriptions*, and

$\mathcal{I}$ is a set of tuples $i = (\Delta_i, \cdot^i)$, called *models*, consisting of a non-empty set Δ_i (of objects) and a mapping $\cdot^i : \mathcal{L} \to \mathfrak{P}(\Delta_i) : f \mapsto f^i$ that assigns an *extension* $f^i \subseteq \Delta_i$ to each concept description $f \in \mathcal{L}$.

Intuitively, models correspond to formal contexts, and the operator $\cdot^i$ corresponds to the $\cdot'$ operator that assigns an extension B' to each set of attributes B. In order to define an analogon to the $\cdot'$ operator in the other direction, we introduce the subsumption preorder on concept descriptions: $f_1 \in \mathcal{L}$ *is subsumed by* $f_2 \in \mathcal{L}$ (written $f_1 \sqsubseteq f_2$) if $f_1^i \subseteq f_2^i$ for all models $i \in \mathcal{I}$. If $f_1 \sqsubseteq f_2$ and $f_2 \sqsubseteq f_1$, then we say that f_1 and f_2 are *equivalent* ($f_1 \equiv f_2$).

Given a set of objects A in a formal context, its intensional description A' is the largest set of attributes B such that $A \subseteq B'$. Since $B_1' \subseteq B_2'$ if $B_1 \supseteq B_2$, such a largest set should correspond to the least one w.r.t. subsumption. This motivates the following definition.

Definition 1 (Most specific concept). *Let* $(\mathcal{L}, \mathcal{I})$ *be a concept description language,* $i \in \mathcal{I}$ *be a model, and* $X \subseteq \Delta_i$*. Then* $f \in \mathcal{L}$ *is a* most specific concept *for* X *in* i *if*

$$X \subseteq f^i, \tag{1}$$

and f *is a least concept description with this property, i.e., for all* $g \in \mathcal{L}$ *with* $X \subseteq g^i$ *we have* $f \sqsubseteq g$.

The most specific concept of a set $X \subseteq \Delta_i$ need not exist, but if it exists then it is unique up to equivalence. In case X has a most specific concept in i, we denote it (or, more precisely, an arbitrary element of its equivalence class) by X^i. The concept description X^i is called the *intensional description* of the set of objects X. An example of a concept description language for which X^i always exists is $\mathcal{EL}_{\mathrm{gfp}}$, which will be introduced in Section 3 below.

The following lemma shows that the mappings

$$\cdot^i : \mathfrak{P}(\Delta_i) \to \mathcal{L} \quad \text{and} \quad \cdot^i : \mathcal{L} \to \mathfrak{P}(\Delta_i)$$

do indeed form a Galois-connection with properties similar to the $\cdot'$ operators in classical FCA. Because of these similarities to FCA we will sometimes use the term *description context* for a model $i \in \mathcal{I}$.

Lemma 2. *Let* $(\mathcal{L}, \mathcal{I})$ *be a concept description language such that* X^i *exists for every* $i \in \mathcal{I}$ *and every* $X \subseteq \Delta_i$*. Let* $i \in \mathcal{I}$ *be a model,* $X, X_1, X_2 \subseteq \Delta_i$ *sets of objects, and* $f, f_1, f_2 \in \mathcal{L}$ *concept descriptions. Then the following holds:*

(a) $X_1 \subseteq X_2 \Rightarrow X_1^i \sqsubseteq X_2^i$,
(b) $f_1 \sqsubseteq f_2 \Rightarrow f_1^i \subseteq f_2^i$,
(c) $X \subseteq X^{ii}$,
(d) $f^{ii} \sqsubseteq f$,
(e) $X^i \equiv X^{iii}$,
(f) $f^i = f^{iii}$,
(g) $X \subseteq f^i \Leftrightarrow X^i \sqsubseteq f$.

Proofs of these facts can be obtained by adapting the proofs from classical FCA. They can be found in [4], but also in [6,7] since the framework introduced above can be seen as an instance of the framework defined in [6,7].

In the remainder of this section, we assume that $(\mathcal{L},\mathcal{I})$ is an arbitrary, but fixed, concept description language. All definitions given below are implicitly parameterized with this language. Our goal is to characterize the subsumption relations that are valid in a given description context of this language by determining a minimal basis of implications comparable to the Duquenne-Guigues basis in classical FCA. We start by defining the notion of an implication and by showing some general results that hold for arbitrary concept description languages. Later on, we will look at the concept description language $\mathcal{EL}_{\text{gfp}}$ in more detail.

Definition 3 (Implication). *An* implication *is a pair (f_1, f_2) of concept descriptions $(f_1, f_2) \in \mathcal{L} \times \mathcal{L}$, which we will usually denote as $f_1 \rightarrow f_2$. We say that the implication $f_1 \rightarrow f_2$* holds *in the description context $\iota = (\Delta_\iota, \iota)$ if $f_1^\iota \subseteq f_2^\iota$.*

Obviously, we have $f_1 \sqsubseteq f_2$ iff $f_1 \rightarrow f_2$ holds in every description context $\iota \in \mathcal{I}$. However, as said above, we are now interested in the implications that hold in a fixed description context rather than in all of them.

In order to define the notion of a basis of the implications holding in a description context, we must first define a consequence operator on implications. Let $\mathcal{B} \subseteq \mathcal{L} \times \mathcal{L}$ be a set of implications and $f_1 \rightarrow f_2$ an implication. If $f_1 \rightarrow f_2$ holds in all description contexts $i \in \mathcal{I}$ in which all implications from $\mathcal{B}$ hold, then we say that $f_1 \rightarrow f_2$ *follows* from $\mathcal{B}$. It is not hard to see that the relation *follows* is

- *reflexive*, i. e. every implication $f_1 \rightarrow f_2 \in \mathcal{B}$ follows from $\mathcal{B}$, and
- *transitive*, i. e. if $f_1 \rightarrow f_2$ follows from $\mathcal{B}_2$, and every implication in $\mathcal{B}_2$ follows from $\mathcal{B}_1$, then $f_1 \rightarrow f_2$ follows from $\mathcal{B}_1$.

Definition 4 (Basis). *For a given description context ι we say that $\mathcal{B} \subseteq \mathcal{L} \times \mathcal{L}$ is a* basis *for the implications holding in ι if $\mathcal{B}$ is*

- sound *for ι, i.e., it contains only implications holding in ι;*
- complete *for ι, i.e., any implication that holds in ι follows from $\mathcal{B}$; and*
- minimal *for ι, i.e., no strict subset of $\mathcal{B}$ is complete for ι.*

Since the above definitions use only the $\cdot^\iota$ operator that assigns an extension to every concept description, but not the one in the other direction, they also make sense for concept description languages where the most specific concept of a set of objects need not always exist. An example of such a language is $\mathcal{EL}$, i.e., the sublanguage of $\mathcal{EL}_{\text{gfp}}$ that does not allow for cyclic concept definitions (see Section 3 below).

The description language $(\mathcal{L}',\mathcal{I}')$ is a *sublanguage* of the description language $(\mathcal{L},\mathcal{I})$ if $\mathcal{L}' \subseteq \mathcal{L}$ and $\mathcal{I}' = \left\{ i|_{\mathcal{L}'} \;\middle|\; i \in \mathcal{I} \right\}$, where $i|_{\mathcal{L}'}$ is the restriction of i to $\mathcal{L}'$, i.e., $\Delta_i = \Delta_{i|_{\mathcal{L}'}}$ and $\cdot^{i|_{\mathcal{L}'}}$ is the restriction of the mapping $\cdot^i$ to $\mathcal{L}'$.

Proposition 5. *Assume that $(\mathcal{L}',\mathcal{I}')$ is a sublanguage of $(\mathcal{L},\mathcal{I})$, that $f_1 \rightarrow f_2 \in \mathcal{L}' \times \mathcal{L}'$, and that $\mathcal{B} \subseteq \mathcal{L}' \times \mathcal{L}'$. Then $f_1 \rightarrow f_2$ follows from $\mathcal{B}$ in $(\mathcal{L},\mathcal{I})$ iff $f_1 \rightarrow f_2$ follows from $\mathcal{B}$ in $(\mathcal{L}',\mathcal{I}')$.*

This proposition will be used later on to transfer results from $\mathcal{EL}_{\rm gfp}$ to $\mathcal{EL}$.

In the remainder of this section, we will characterize complete subsets of the set of all implications holding in a given description context ι. Whenever we use the $\cdot^\iota$ operator from sets of objects to concept descriptions, we implicitly assume that it is defined.

Analogously to the situation in classical FCA, we can restrict the attention to implications whose right-hand sides are closed under the operator $\cdot^{\iota\iota}$.

Lemma 6. *If the implication $f_1 \to f_2$ holds in ι, then it follows from $\{f_1 \to f_1^{\iota\iota}\}$, and the set $\{f_1 \to f_1^{\iota\iota}\}$ is sound for ι.*

Proof. By Lemma 2(f), all implications of the form $f \to f^{\iota\iota}$ hold in ι, which yields soundness of $\{f_1 \to f_1^{\iota\iota}\}$.

Let $f_1 \to f_2$ be any implication that holds in ι, i.e., $f_1^\iota \subseteq f_2^\iota$. By Lemma 2(g), this is equivalent to

$$f_1^{\iota\iota} \sqsubseteq f_2. \tag{2}$$

Let i be some model in which $f_1 \to f_1^{\iota\iota}$ holds. By definition this means that $f_1^i \subseteq (f_1^{\iota\iota})^i$. Using Lemma 2(g) again we obtain $f_1^{ii} \sqsubseteq f_1^{\iota\iota}$. Together with (2) and transitivity of $\sqsubseteq$, this yields $f_1^{ii} \sqsubseteq f_2$, and hence $f_1^i \subseteq f_2^i$. Thus, we have shown that $f_1 \to f_2$ holds in any model i in which $f_1 \to f_1^{\iota\iota}$ holds. $\square$

Corollary 7. *The set of implications $\{f \to f^{\iota\iota} \mid f \in \mathcal{L}\}$ is complete for ι.*

Having reduced the number of right-hand sides that need to be considered, our goal is now to restrict the left-hand sides. This is possible if we can find a so-called dominating set of concept descriptions.

Definition 8 (Dominating set). *The set $\mathcal{D} \subseteq \mathcal{L}$* dominates *the description context ι if, for every $f \in \mathcal{L}$, there is some $g \in \mathcal{D}$ such that $f \sqsubseteq g$ and $f^\iota = g^\iota$.*

It is sufficient to consider implications whose left-hand sides belong to a dominating set.

Lemma 9. *If $\mathcal{D} \subseteq \mathcal{L}$ dominates ι, then $\mathcal{B} := \{f \to f^{\iota\iota} \mid f \in \mathcal{D}\}$ is sound and complete for ι.*

Proof. Soundness has already been shown. To show completeness, let $f_1 \to f_2$ be an implication that holds in ι. By Lemma 6, $f_1 \to f_2$ follows from $f_1 \to f_1^{\iota\iota}$. Hence it is sufficient to show that $f_1 \to f_1^{\iota\iota}$ follows from $\mathcal{B}$. Since $\mathcal{D}$ dominates ι, there exists $g \in \mathcal{D}$ such that $f_1 \sqsubseteq g$ and $g^\iota = f_1^\iota$.

Let i be a model in which all implications of $\mathcal{B}$ hold. From $f_1 \sqsubseteq g$ and Lemma 2(b) it follows that $f_1^i \subseteq g^i$. Since $g \to g^{\iota\iota} \in \mathcal{B}$ holds in i, we also have $g^i \subseteq (g^{\iota\iota})^i$, and thus $f_1^i \subseteq (g^{\iota\iota})^i$. In addition, $g^\iota = f_1^\iota$ yields $g^{\iota\iota} = f_1^{\iota\iota}$. Thus, $f_1^i \subseteq (f_1^{\iota\iota})^i$, which shows that $f_1 \to f_1^{\iota\iota}$ holds in i. $\square$

The sound and complete set of implications $\mathcal{B}$ induced by a dominating set $\mathcal{D}$ need not be a basis since it need not be minimal. However, if $\mathcal{D}$ is finite, then $\mathcal{B}$ is finite as well. Thus, a subset of $\mathcal{B}$ that is a basis can be obtained by removing redundant elements.

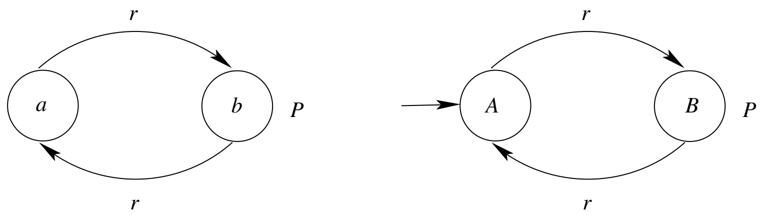

Fig. 1. A model (left) and a description graph (right)

3 $\mathcal{EL}_{\mathrm{gfp}}$ as an Instance of the General Framework

We start by defining $\mathcal{EL}$, and then show how it can be extended to $\mathcal{EL}_{\mathrm{gfp}}$. Concept descriptions of $\mathcal{EL}$ are built from a set $\mathcal{N}_c$ of concept names and a set $\mathcal{N}_r$ of role names, using the constructors top concept, conjunction, and existential restriction:

- concept names and the top concept $\top$ are $\mathcal{EL}$-concept descriptions;
- if C, D are $\mathcal{EL}$-concept descriptions and r is a role name, then $C \sqcap D$ and $\exists r.C$ are $\mathcal{EL}$-concept descriptions.

In the following, we assume that the sets $\mathcal{N}_c$ and $\mathcal{N}_r$ of concept and role names are finite. This assumption is reasonable since in practice data are usually represented over a finite signature.

Models of this language are pairs $(\Delta_I, \cdot^I)$ where Δ_I is a finite,[1] non-empty set, and $\cdot^I$ maps role names r to binary relations $r^I \subseteq \Delta_I \times \Delta_I$ and $\mathcal{EL}$-concept descriptions to subsets of Δ_I such that

$$\top^I = \Delta_I, \qquad (C \sqcap D)^I = C^I \cap D^I, \quad \text{and}$$
$$(\exists r.C)^I \quad = \{d \in \Delta_i \mid \exists e \in C^I \text{ such that } (d,e) \in r^I\}.$$

Subsumption and equivalence between $\mathcal{EL}$-concept descriptions is defined as in our general framework, i.e., $C \sqsubseteq D$ iff $C^I \sqsubseteq D^I$ for all models I, and $C \equiv D$ iff $C \sqsubseteq D$ and $D \sqsubseteq C$.

Unfortunately, $\mathcal{EL}$ itself cannot be used to instantiate our framework since in general a set of objects need not have a most specific concept in $\mathcal{EL}$. This is illustrated by the following simple example. Assume that $\mathcal{N}_c = \{P\}$, $\mathcal{N}_r = \{r\}$, and consider the model I with $\Delta_I = \{a, b\}$, $r^I = \{(a,b),(b,a)\}$, and $P^I = \{b\}$ (see the left-hand side of Fig. 1 for a graphical representation of this model). To see that the set $\{a\}$ does not have a most specific concept, consider the $\mathcal{EL}$-concept descriptions

$$C_k := \underbrace{\exists r.\exists r \ldots \exists r.}_{k \text{ times}} \top.$$

[1] Usually, the semantics given for description logics allows for models of arbitrary cardinality. However, in the case of $\mathcal{EL}$ the restriction to finite models is without loss of generality since it has the finite model property, i.e., a subsumption relationship holds w.r.t. all models iff it holds w.r.t. all finite models.

We have $\{a\} \subseteq C_k^I = \{a, b\}$ for all k, and thus a most specific concept C for $\{a\}$ would need to satisfy $C \sqsubseteq C_k$ for all $k \geq 0$. However, it is easy to see that $C \sqsubseteq C_k$ can only be true if the role depth of C, i.e., the maximal nesting of existential restrictions, is at least k. Since any $\mathcal{EL}$-concept description has a finite role depth, this shows that such a most specific concept C cannot exist.

However, most specific concepts always exist in $\mathcal{EL}_{\text{gfp}}$, the extension of $\mathcal{EL}$ by cyclic concept definitions interpreted with greatest fixpoint (gfp) semantics.[2] In $\mathcal{EL}_{\text{gfp}}$, we assume that the set of concept names is partitioned into the set $\mathcal{N}_{\text{prim}}$ of primitive concepts and the set $\mathcal{N}_{\text{def}}$ of defined concept. A *concept definition* is of the form

$$B_0 \equiv P_1 \sqcap \ldots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \sqcap \exists r_n.B_n$$

where $B_0, B_1, \ldots, B_n \in \mathcal{N}_{\text{def}}$, $P_1, \ldots, P_m \in \mathcal{N}_{\text{prim}}$, and $r_1, \ldots, r_n \in \mathcal{N}_r$. The empty conjunction (i.e., $m = 0 = n$) stands for $\top$. A *TBox* is a finite set of concept definitions such that every defined concept occurs at most once as a left-hand side of a concept definition.

Definition 10 ($\mathcal{EL}_{\text{gfp}}$-concept description). *An* $\mathcal{EL}_{\text{gfp}}$-concept description *is a tuple $(A, \mathcal{T})$ where $\mathcal{T}$ is a TBox and A is a defined concept occurring on the left-hand side of a definition in $\mathcal{T}$.*

For example, $(A, \mathcal{T})$ with $\mathcal{T} := \{A \equiv \exists r.B, B \equiv P \sqcap \exists r.A\}$ is an $\mathcal{EL}_{\text{gfp}}$-concept description. Any $\mathcal{EL}_{\text{gfp}}$-concept description $(A, \mathcal{T})$ can be represented by a directed, rooted, edge- and node-labeled graph: the nodes of this graph are the defined concepts in $\mathcal{T}$, with A being the root; the edge label of node B_0 is the set of primitive concepts occurring in the definition of B_0; and every conjunct $\exists r_i.B_i$ in the definition of B_0 gives rise to an edge from B_0 to B_i with label r_i. In the following, we call such graphs *description graphs*. The description graph associated with the $\mathcal{EL}_{\text{gfp}}$-concept description from our example is shown on the right-hand side of Fig. 1, where A is the root.

Models of $\mathcal{EL}_{\text{gfp}}$ are of the form $I = (\Delta_I, \cdot^I)$ where Δ_I is a finite, non-empty set, and $\cdot^I$ maps role names r to binary relations $r^I \subseteq \Delta_I \times \Delta_I$ and primitive concepts to subsets of Δ_I. The mapping $\cdot^I$ is extended to $\mathcal{EL}_{\text{gfp}}$-concept descriptions $(A, \mathcal{T})$ by interpreting the TBox $\mathcal{T}$ with gfp-semantics: consider all extensions of I to the defined concepts that satisfy the concept definitions in $\mathcal{T}$, i.e., assign the same extension to the left-hand side and the right-hand side of each definition. Among these extensions of I, the *gfp-model of $\mathcal{T}$ based on I* is the one that assigns the largest sets to the defined concepts (see [1] for a more detailed definition of gfp-semantics). The *extension $(A, \mathcal{T})^I$ of $(A, \mathcal{T})$* in I is the set assigned to A by the gfp-model of $\mathcal{T}$ based on I.

Again, subsumption and equivalence of $\mathcal{EL}_{\text{gfp}}$-concept descriptions is defined as in the general framework. The following theorem shows that the description language $\mathcal{EL}_{\text{gfp}}$ we have just defined is indeed an instance of the framework introduced in Section 2.

[2] Because of the space restriction, we can only give a very compact introduction of this DL. See [1,4] for more details.

Theorem 11. *In $\mathcal{EL}_{\text{gfp}}$, the most specific concept of a set of objects always exists.*

The proof of this theorem given in [4] is based on the methods and results from [2]. It proceeds in two steps. First, it is shown how to construct the most specific concept of a singleton set $\{a\}$. The main idea is that the graph representing the model can also be viewed as the description graph of an $\mathcal{EL}_{\text{gfp}}$-concept description, where the root is the node corresponding to a. In the example (see Fig. 1), we have simply renamed the lower case individual names into upper case concept names. The $\mathcal{EL}_{\text{gfp}}$-concept description $(A, \mathcal{T})$ represented by the description graph on the right-hand side of Fig. 1 is the most specific concept of $\{a\}$ in the model represented by the graph on the left-hand side of Fig. 1. The most specific concept of a set of objects $\{a_1, \ldots, a_n\}$ is the least common subsumer (lcs) of the most specific concepts of the singleton sets $\{a_i\}$. In [2] it is shown that the lcs in $\mathcal{EL}_{\text{gfp}}$ always exists and how to compute it.

4 A Finite Basis for Implications in $\mathcal{EL}_{\text{gfp}}$

We show that the set of implications holding in a given model always has a finite basis in $\mathcal{EL}_{\text{gfp}}$. A first step in this direction is to show that it is enough to restrict the attention to implications with acyclic $\mathcal{EL}_{\text{gfp}}$-concept descriptions as left-hand sides. The $\mathcal{EL}_{\text{gfp}}$-concept description $(A, \mathcal{T})$ is *acyclic* if the graph associated with it is acyclic. It is easy to see that there is a 1–1-relationship between $\mathcal{EL}$-concept descriptions and acyclic $\mathcal{EL}_{\text{gfp}}$-concept descriptions. For example, $(A, \{A \equiv B \sqcap \exists r.B, B \equiv P\})$ corresponds to $P \sqcap \exists r.P$, and $\exists r.P$ corresponds to $(A, \{A \equiv \exists r.B, B \equiv P\})$. This shows that $\mathcal{EL}$ can indeed be seen as a sublanguage of $\mathcal{EL}_{\text{gfp}}$. In the following, we will not distinguish an acyclic $\mathcal{EL}_{\text{gfp}}$-concept description from its equivalent $\mathcal{EL}$-concept description.

Given an $\mathcal{EL}_{\text{gfp}}$-concept description, its *node size* is the number of nodes in the description graph corresponding to it. For an acyclic $\mathcal{EL}_{\text{gfp}}$-concept description, we define its *depth* to be the maximal length of a path starting at the root in the description graph corresponding to it. Any $\mathcal{EL}_{\text{gfp}}$-concept description $(A, \mathcal{T})$ can be approximated by acyclic $\mathcal{EL}_{\text{gfp}}$-concept descriptions $(A, \mathcal{T})_d$ of increasing depth d. To obtain $(A, \mathcal{T})_d$, the description graph associated with $(A, \mathcal{T})$ is unraveled into a (possibly infinite) tree, and then all branches are cut at depth d. It is easy to see that $(A, \mathcal{T}) \sqsubseteq (A, \mathcal{T})_d$ holds for all $d \geq 0$.

Lemma 12. *Let $\mathcal{U}$ be an $\mathcal{EL}_{\text{gfp}}$-concept description of node size m, I a model of cardinality n, and $d = m \cdot n + 1$. Then $a \in (\mathcal{U}_d)^I$ implies $a \in \mathcal{U}^I$.*

A detailed proof of this lemma can be found in [4].

Theorem 13. *In $\mathcal{EL}_{\text{gfp}}$, the set of acyclic concept descriptions dominates every description context I.*

Proof. Let $\mathcal{U}$ be an $\mathcal{EL}_{\text{gfp}}$-concept description and I a description context. We must find an acyclic $\mathcal{EL}_{\text{gfp}}$-concept description $\mathcal{V}$ such that $\mathcal{U} \sqsubseteq \mathcal{V}$ and $\mathcal{U}^I = \mathcal{V}^I$.

Let m be the node size of $\mathcal{U}$, n the cardinality of I, and $d = m \cdot n+1$. We know that $\mathcal{U} \sqsubseteq \mathcal{U}_d$, and thus also $\mathcal{U}^I \subseteq (\mathcal{U}_d)^I$. Lemma 12 shows that the inclusion in the other direction holds as well. Thus, $\mathcal{V} := \mathcal{U}_d$ does the job. □

By Lemma 9, this theorem immediately implies the following corollary.

Corollary 14. *For any description context I of $\mathcal{EL}_{\mathrm{gfp}}$, the set*

$$\{\mathcal{U} \rightarrow \mathcal{U}^{II} \mid \mathcal{U} \text{ is an acyclic } \mathcal{EL}_{\mathrm{gfp}}\text{-concept description}\}$$

is sound and complete for I.

The complete set of implications given in the corollary is, of course, infinite. Also note that, though the left-hand sides $\mathcal{U}$ of implications in this set are acyclic, the right-hand sides $\mathcal{U}^{II}$ need not be acyclic. We show next that there is also a *finite* sound and complete set of implications. As mentioned before, a finite basis can then be obtained by removing redundant elements.

Theorem 15. *In $\mathcal{EL}_{\mathrm{gfp}}$, for any description context I, there exists a* finite *set $\mathcal{B}$ of implications that is sound and complete for I.*

Proof. By Corollary 14 it suffices to find a finite and sound set of implications from which all implications of the form $\mathcal{U} \rightarrow \mathcal{U}^{II}$, where $\mathcal{U}$ is an acyclic $\mathcal{EL}_{\mathrm{gfp}}$-concept description, follow. To this purpose, consider the set $\mathcal{E} := \{\mathcal{U}^I \mid \mathcal{U}$ is an $\mathcal{EL}_{\mathrm{gfp}}$-concept description$\}$, and let $\mathcal{C}$ be a set of $\mathcal{EL}_{\mathrm{gfp}}$-concept descriptions that contains, for each set $X \in \mathcal{E}$, exactly one element $\mathcal{V}$ with $\mathcal{V}^I = X$. Because of Theorem 13, we can assume without loss of generality that $\mathcal{C}$ contains only acyclic descriptions. Since Δ_I is finite, the sets $\mathcal{E}$ and $\mathcal{C}$ are also finite.

Consider the following finite set of implications, which is obviously sound:

$$\begin{aligned}\mathcal{B} := &\{P \rightarrow P^{II} \mid P \in \mathcal{N}_{\mathrm{prim}} \cup \{\top\}\}\\ &\cup \{\exists r.C \rightarrow (\exists r.C)^{II} \mid r \in \mathcal{N}_r, C \in \mathcal{C}\}\\ &\cup \{C_1 \sqcap C_2 \rightarrow (C_1 \sqcap C_2)^{II} \mid C_1, C_2 \in \mathcal{C}\}.\end{aligned}$$

We show that, for any acyclic $\mathcal{EL}_{\mathrm{gfp}}$-concept description $\mathcal{U}$, the implication $\mathcal{U} \rightarrow \mathcal{U}^{II}$ follows from $\mathcal{B}$. Since $\mathcal{U}$ is acyclic, we can view it as an $\mathcal{EL}$-concept description. The proof is by induction on the structure of this description.

Base case: $\mathcal{U} = P \in \mathcal{N}_{\mathrm{prim}} \cup \{\top\}$. Then $P \rightarrow P^{II}$ is in $\mathcal{B}$ by definition. Thus, it also follows from $\mathcal{B}$.

Step case 1: $\mathcal{U} = \exists r.\mathcal{V}$ for some $r \in \mathcal{N}_r$ and some $\mathcal{EL}$-concept description $\mathcal{V}$. Let J be a description context in which all implications from $\mathcal{B}$ hold. The semantics of existential restrictions yields

$$\mathcal{U}^J = (\exists r.\mathcal{V})^J = \{x \in \Delta_J \mid \exists y \in \mathcal{V}^J : (x,y) \in r^J\}.$$

By the induction hypothesis, $\mathcal{V} \rightarrow \mathcal{V}^{II}$ follows from $\mathcal{B}$, and thus holds in J. Therefore $\mathcal{V}^J \subseteq (\mathcal{V}^{II})^J$, which yields

$$\mathcal{U}^J \subseteq \{x \in \Delta_J \mid \exists y \in (\mathcal{V}^{II})^J : (x,y) \in r^J\}.$$

Now, choose $C \in \mathcal{C}$ such that $C^I = \mathcal{V}^I$. Lemma 2(g) yields $\mathcal{V}^{II} \sqsubseteq C$, and thus

$$\mathcal{U}^J \subseteq \{x \in \Delta_J \mid \exists y \in C^J : (x,y) \in r^J\} = (\exists r.C)^J.$$

Since $\exists r.C \rightarrow (\exists r.C)^{II} \in \mathcal{B}$ holds in J by assumption, we get

$$\begin{aligned}\mathcal{U}^J \subseteq ((\exists r.C)^{II})^J &= (\{x \in \Delta_I \mid \exists y \in C^I : (x,y) \in r^I\}^I)^J = \\ &= (\{x \in \Delta_I \mid \exists y \in \mathcal{V}^I : (x,y) \in r^I\}^I)^J = ((\exists r.\mathcal{V})^{II})^J = (\mathcal{U}^{II})^J.\end{aligned}$$

Thus, we have shown that $\mathcal{U} \rightarrow \mathcal{U}^{II}$ holds in every context J in which all implications from $\mathcal{B}$ hold.

Step case 2: $\mathcal{U} = \mathcal{U}_1 \sqcap \mathcal{U}_2$ for $\mathcal{EL}$-concept descriptions $\mathcal{U}_1,\mathcal{U}_2$. Let J be a description context in which all implications from $\mathcal{B}$ hold. By the induction hypothesis, $\mathcal{U}_1^J \subseteq (\mathcal{U}_1^{II})^J$ and $\mathcal{U}_2^J \subseteq (\mathcal{U}_2^{II})^J$. Therefore

$$\mathcal{U}^J = (\mathcal{U}_1 \sqcap \mathcal{U}_2)^J = \mathcal{U}_1^J \cap \mathcal{U}_2^J \subseteq (\mathcal{U}_1^{II})^J \cap (\mathcal{U}_2^{II})^J.$$

We choose $C_1, C_2 \in \mathcal{C}$ such that $C_1^I = \mathcal{U}_1^I$ and $C_2^I = \mathcal{U}_2^I$. Then

$$\mathcal{U}^J \subseteq (C_1^{II})^J \cap (C_2^{II})^J \subseteq C_1^J \cap C_2^J = (C_1 \sqcap C_2)^J,$$

where the second inclusion holds due to Lemma 2(d). Since the implication $C_1 \sqcap C_2 \rightarrow (C_1 \sqcap C_2)^{II} \in \mathcal{B}$ holds in J, we get

$$\begin{aligned}\mathcal{U}^J \subseteq ((C_1 \sqcap C_2)^{II})^J &= ((C_1^I \cap C_2^I)^I)^J = ((\mathcal{U}_1^I \cap \mathcal{U}_2^I)^I)^J = \\ &= ((\mathcal{U}_1 \sqcap \mathcal{U}_2)^{II})^J = (\mathcal{U}^{II})^J.\end{aligned}$$

This shows that $\mathcal{U} \rightarrow \mathcal{U}^{II}$ follows from $\mathcal{B}$. □

Corollary 16. *In $\mathcal{EL}_{\mathrm{gfp}}$, for any description context I there exists a* finite *basis for the implications holding in I.*

Proof. Starting with $\mathcal{B}^* := \mathcal{B}$, where in the beginning all implications are unmarked, take an unmarked implication $\mathcal{U} \rightarrow \mathcal{V} \in \mathcal{B}^*$. If this implication follows from $\mathcal{B}^*$, then remove it, i.e., $\mathcal{B}^* := \mathcal{B}^* \setminus \{\mathcal{U} \rightarrow \mathcal{V}\}$; otherwise, mark $\mathcal{U} \rightarrow \mathcal{V}$. Continue with this until all implications in $\mathcal{B}^*$ are marked. The final set $\mathcal{B}^*$ is the desired basis. □

5 A Finite Basis for Implications in $\mathcal{EL}$

Although the sublanguage $\mathcal{EL}$ of $\mathcal{EL}_{\mathrm{gfp}}$ is not an instance of our general framework, we can nevertheless show the above corollary also for this language. Because of Proposition 5, it is sufficient to show that in $\mathcal{EL}_{\mathrm{gfp}}$ any description context I has a finite basis consisting of implications where both the left-hand *and the right-hand sides* are acyclic.

The following proposition will allow us to construct a finite set of implications *with acyclic right-hand sides* from which a given implication $\mathcal{U} \rightarrow \mathcal{U}^{II}$ (with potentially cyclic right-hand side) follows. Recall that, for any $\mathcal{EL}_{\mathrm{gfp}}$-concept description $\mathcal{U}$, we obtain the acyclic description $\mathcal{U}_d$ by unraveling the description graph and then cutting all branches at depth d.

Proposition 17. *Let k_0 be a non-negative integer, I a description context, and $\mathcal{U}$ be an $\mathcal{EL}_{\mathrm{gfp}}$-concept description. Then the implication $\mathcal{U} \to \mathcal{U}^{II}$ follows from*

$$\mathcal{B} := \{(X^I)_{k_0} \to (X^I)_{k_0+1} \mid X \subseteq \Delta_I\} \cup \{\mathcal{U} \to (\mathcal{U}^{II})_{k_0}\}.$$

Proof. The proof depends on the following technical result, whose proof can be found in [4].

(∗) For any set $X \subseteq \Delta_I$, there exist sets $\mathcal{P} \subseteq \mathcal{N}_{\mathrm{prim}}$ and $\mathcal{Y} \subseteq \mathcal{N}_r \times \mathfrak{P}(\Delta_I)$ such that

$$X^I \equiv \bigsqcap_{P \in \mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.Y^I.$$

The above equivalence is actually an abbreviation for saying that X^I is of the form $(A, \mathcal{T})$ where $\mathcal{T}$ consists of the following concept definitions:

- $A \equiv \bigsqcap_{P\in\mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.B_{r,Y}$;
- the definitions in the TBoxes $\mathcal{T}_{r,Y}$ for $(r, Y) \in \mathcal{Y}$ where $Y^I = (B_{r,Y}, \mathcal{T}_{r,Y})$.

Note that the sets of defined concepts in the TBoxes $\mathcal{T}_{r,Y}$ can be assumed to be pairwise disjoint and not to contain A.

To prove the proposition, we first show, by induction on ℓ, that the implications $(X^I)_\ell \to (X^I)_{\ell+1}$ follow from $\mathcal{B}$ for all $\ell \geq k_0$. For $\ell = k_0$ this is trivial because $(X^I)_{k_0} \to (X^I)_{k_0+1} \in \mathcal{B}$.

Now, assume that $(Y^I)_k \to (Y^I)_{k+1}$ follows from $\mathcal{B}$ for every $Y \subseteq \Delta_I$ and every $k, k_0 \leq k < \ell$. Let J be a model in which all implications from $\mathcal{B}$ hold. Then, by the induction hypothesis, we get

$$((Y^I)_k)^J \subseteq ((Y^I)_{k+1})^J \tag{3}$$

for all $k_0 \leq k < \ell$ and all $Y \subseteq \Delta_I$. By (∗), for any set $X \subseteq \Delta_I$, there exist sets $\mathcal{P} \subseteq \mathcal{N}_{\mathrm{prim}}$ and $\mathcal{Y} \subseteq \mathcal{N}_r \times \mathfrak{P}(\Delta_I)$ such that

$$X^I \equiv \bigsqcap_{P \in \mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.Y^I.$$

It is easy to see that this implies

$$(X^I)_\ell \equiv \bigsqcap_{P \in \mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.(Y^I)_{\ell-1} \tag{4}$$

and

$$(X^I)_{\ell+1} \equiv \bigsqcap_{P \in \mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.(Y^I)_\ell. \tag{5}$$

Thus, we have

$$\begin{aligned}\Big((X^I)_\ell\Big)^J &\overset{(4)}{=} \Big(\bigsqcap_{P \in \mathcal{P}} P \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \exists r.(Y^I)_{\ell-1}\Big)^J \\ &= \bigsqcap_{P \in \mathcal{P}} P^J \sqcap \bigsqcap_{(r,Y)\in\mathcal{Y}} \{x \in \Delta_J \mid \exists y \in ((Y^I)_{\ell-1})^J : (x,y) \in r^J\}.\end{aligned}$$

From (3) we obtain $((Y^I)_{\ell-1})^J \subseteq ((Y^I)_\ell)^J$, and thus

$$\begin{aligned}\Big((X^I)_\ell\Big)^J &\subseteq \prod_{P\in\mathcal{P}} P^J \sqcap \prod_{(r,Y)\in\mathcal{Y}} \{x \in \Delta_J \mid \exists y \in ((Y^I)_\ell)^J : (x,y) \in r^J\} \\ &= \Big(\prod_{P\in\mathcal{P}} P \sqcap \prod_{(r,Y)\in\mathcal{Y}} \exists r.(Y^I)_\ell\Big)^J \\ &\stackrel{(5)}{=} \Big((X^I)_{\ell+1}\Big)^J.\end{aligned}$$

Hence we have shown that $(X^I)_\ell \to (X^I)_{\ell+1}$ follows from $\mathcal{B}$, which concludes the induction proof.

Now, let J again be a model in which all implications from $\mathcal{B}$ hold, and let $x \in \mathcal{U}^J$. We must show that this implies $x \in (\mathcal{U}^{II})^J$. We have $x \in ((\mathcal{U}^{II})_{k_0})^J$ because $\mathcal{U} \to (\mathcal{U}^{II})_{k_0} \in \mathcal{B}$. Hence $x \in ((\mathcal{U}^{II})_k)^J$ for all $k \leq k_0$ since $(\mathcal{U}^{II})_{k_0} \sqsubseteq (U^{II})_k$ for all $k \leq k_0$. From what we have shown above, we know that

$$(U^{II})_k \to (\mathcal{U}^{II})_{k+1}$$

follows from $\mathcal{B}$ for all $k \geq k_0$. Thus $((\mathcal{U}^{II})_k)^J \subseteq ((\mathcal{U}^{II})_{k+1})^J$ holds in J for all $k \geq k_0$, which yields $x \in ((\mathcal{U}^{II})_k)^J$ also in this case.

Therefore $x \in ((\mathcal{U}^{II})_k)^J$ for $k = |\mathcal{G}_\mathcal{U}| \cdot |\Delta_J| + 1$, independently of whether this number is smaller or larger than k_0. It follows directly from Lemma 12 that $x \in (\mathcal{U}^{II})^J$. Thus, we have shown that

$$\mathcal{U}^J \subseteq (\mathcal{U}^{II})^J$$

if all implications from $\mathcal{B}$ hold in J. This means that $\mathcal{U} \to \mathcal{U}^{II}$ follows from $\mathcal{B}$. □

Having proved Proposition 17, we are almost finished with constructing a finite, sound and complete set of acyclic implications for the implications holding in a description context I. The idea is to replace any implication $\mathcal{U} \to \mathcal{U}^{II}$ in the finite, sound and complete set of implications constructed in the proof of Theorem 15 by the corresponding implications from Proposition 17.

The remaining problems is, however, that the set of implications obtained this way need not be sound for I. Indeed, if k_0 is too small, then the implications in $\{(X^I)_{k_0} \to (X^I)_{k_0+1} \mid X \subseteq \Delta_I\}$ need not hold in I. Therefore, we define for every $X \subseteq \Delta_I$

$$d_X := m_X \cdot n + 1,$$

where m_X is the node size of X^I and n is the cardinality of the model I. The number k_0 is the maximum of these numbers, i.e.,

$$k_0 := \max_{X\subseteq\Delta_I} d_X. \tag{6}$$

Then, because $d_X \leq k_0$ for every $X \subseteq \Delta_I$, we have

$$X^I \sqsubseteq (X^I)_{k_0+1} \sqsubseteq (X^I)_{k_0} \sqsubseteq (X^I)_{d_X}.$$

By Lemma 2(b), this implies

$$X^{II} \subseteq ((X^I)_{k_0+1})^I \subseteq ((X^I)_{k_0})^I \subseteq ((X^I)_{d_X})^I.$$

From Lemma 12 we obtain $X^{II} \supseteq ((X^I)_{d_X})^I$, and thus

$$X^{II} = ((X^I)_{k_0+1})^I = ((X^I)_{k_0})^I = ((X^I)_{d_X})^I.$$

In particular, this shows

$$((X^I)_{k_0})^I \subseteq ((X^I)_{k_0+1})^I.$$

Hence, all implications in $\{(X^I)_{k_0} \rightarrow (X^I)_{k_0+1} \mid X \subseteq \Delta_I\}$ hold in I.

Theorem 18. *In $\mathcal{EL}_{\text{gfp}}$, for any description context I, there exists a* finite *set $\mathcal{B}$ of implications that is sound and complete for I, and such that all concept descriptions occurring in $\mathcal{B}$ are* acyclic*.*

Proof. Let $\mathcal{C}$ be the set of acyclic $\mathcal{EL}_{\text{gfp}}$-concept descriptions defined in the proof of Theorem 15. We have shown in that proof that the set

$$\begin{aligned}\mathcal{B}_\star := \{&P \rightarrow P^{II} \mid P \in \mathcal{N}_{\text{prim}} \cup \{\top\}\}\\ \cup\, \{&\exists r.C \rightarrow (\exists r.C)^{II} \mid r \in \mathcal{N}_r, C \in \mathcal{C}\}\\ \cup\, \{&C_1 \sqcap C_2 \rightarrow (C_1 \sqcap C_2)^{II} \mid C_1, C_2 \in \mathcal{C}\}\end{aligned}$$

is complete for I.

Let k_0 be defined as in (6). Then, by Proposition 17, the fact that $\mathcal{B}_\star$ is complete also implies that the following set of implications is complete for I:

$$\begin{aligned}\mathcal{B} := \{&(X^I)_{k_0} \rightarrow (X^I)_{k_0+1} \mid X \subseteq \Delta_I\}\\ \cup\, \{&P \rightarrow (P^{II})_{k_0} \mid P \in \mathcal{N}_{\text{prim}} \cup \{\top\}\}\\ \cup\, \{&\exists r.C \rightarrow ((\exists r.C)^{II})_{k_0} \mid r \in \mathcal{N}_r, C \in \mathcal{C}\}\\ \cup\, \{&C_1 \sqcap C_2 \rightarrow ((C_1 \sqcap C_2)^{II})_{k_0} \mid C_1, C_2 \in \mathcal{C}\}.\end{aligned}$$

Regarding soundness, we have shown above that, due to the fact that k_0 was chosen large enough, all implications of the form $(X^I)_{k_0} \rightarrow (X^I)_{k_0+1}$ hold I. The implications $P \rightarrow (P^{II})_{k_0}$ hold because $P \rightarrow P^{II}$ holds in I, and $P^{II} \sqsubseteq (P^{II})_{k_0}$. The same arguments can be used to show that the implications of the forms $\exists r.C \rightarrow ((\exists r.C)^{II})_{k_0}$ and $C_1 \sqcap C_2 \rightarrow ((C_1 \sqcap C_2)^{II})_{k_0}$ hold in I.

The left-hand sides of implications in $\mathcal{B}$ are acyclic since the elements of $\mathcal{C}$ are acyclic, primitive concepts and $\top$ are acyclic, and any concept description of the form $\mathcal{U}_k$ is acyclic. This last argument also shows that the right-hand sides of implications in $\mathcal{B}$ are acyclic. □

Since $\mathcal{B}$ contains only acyclic $\mathcal{EL}_{\text{gfp}}$-concept descriptions, it can also be viewed as a set of implications in $\mathcal{EL}$. Proposition 5, together with Theorem 18, shows that $\mathcal{B}$ is also complete for the $\mathcal{EL}$-implications holding in I. As argued before, the existence of a finite, sound and complete set also implies the existence of a basis.

Corollary 19. *In $\mathcal{EL}$, for any description context I, there exists a* finite *basis for the implications holding in I.*

6 Conclusion

We have shown that any description context I (i.e., any finite relational structure over a finite signature of unary and binary predicate symbols) has a finite basis for the $\mathcal{EL}$- and $\mathcal{EL}_{\mathrm{gfp}}$-implications holding in I. Such a basis provides the knowledge engineer with interesting information on the application domain described by the context. The knowledge engineer can, for example, use these implications as starting point for building an ontology describing this domain.

In this paper, we have concentrated on showing the existence of a finite basis. Of course, if this approach is to be used in practice, we also need to find efficient algorithms for computing the basis. After that, the next step will be to generalize attribute exploration [9] to our more general setting. This would allow us to consider also relational structures that are not explicitly given, but rather "known" by a domain expert.

Finally, we will also try to show similar results for other DLs. For the DL $\mathcal{FL}_0$, which differs from $\mathcal{EL}$ in that existential restrictions are replaced by value restrictions, we are quite confident that this is possible. For more expressive DLs, like $\mathcal{ALC}$, this is less clear.

References

1. Baader, F.: Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In: Gottlob, G., Walsh, T. (eds.) Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003), Acapulco, Mexico, pp. 319–324 (2003)
2. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Gottlob, G., Walsh, T. (eds.) Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003), Acapulco, Mexico, pp. 325–330 (2003)
3. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Kaelbling, L.P., Saffiotti, A. (eds.) Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005), Edinburgh (UK), pp. 364–369 (2005)
4. Baader, F., Distel, F.: A finite basis for the set of $\mathcal{EL}$-implications holding in a finite model. LTCS-Report 07-02, Theoretical Computer Science, TU Dresden, Germany (2007), `http://lat.inf.tu-dresden.de/research/reports.html`
5. Baader, F., Lutz, C., Suntisrivaraporn, B.: CEL—a polynomial-time reasoner for life science ontologies. In: Furbach, U., Shankar, N. (eds.) IJCAR 2006. LNCS (LNAI), vol. 4130, pp. 287–291. Springer, Heidelberg (2006)
6. Ferré, S.: Systèmes d'information logiques: un paradigme logico-contextuel pour interroger, naviguer et apprendre. PhD thesis, IRISA, France (2002)
7. Ferré, S., Ridoux, O.: Introduction to logical information systems. Information Processing & Management 40(3), 383–419 (2004)
8. Ganter, B.: Algorithmen zur Formalen Begriffsanalyse. In: Ganter, B., Wille, R., Wolff, K.E. (eds.) Beiträge zur Begriffsanalyse, pp. 241–254. B.I. Wissenschaftsverlag (1987)
9. Ganter, B.: Attribute exploration with background knowledge. Theoretical Computer Science 217(2), 215–233 (1999)
10. Ganter, B., Kuznetsov, S.O.: Pattern structures and their projections. In: Delugach, H.S., Stumme, G. (eds.) ICCS 2001. LNCS (LNAI), vol. 2120, pp. 129–144. Springer, Heidelberg (2001)

11. Ganter, B., Wille, R.: Implikationen und Abhängigkeiten zwischen Merkmalen. In: Degens, P.O., Hermes, H.-J., Opitz, O. (eds.) Die Klassifikation und ihr Umfeld, Frankfurt, Indeks-Verlag (1986)
12. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, New York (1997)
13. Guigues, J.-L., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Math. Sci. Humaines 95, 5–18 (1986)
14. Horrocks, I., Patel-Schneider, P.F., van Harmelen, F.: From SHIQ and RDF to OWL: The making of a web ontology language. Journal of Web Semantics 1(1), 7–26 (2003)
15. Möller, R., Haarslev, V.: Description logic systems. In: Baader, F., et al. (eds.) The Description Logic Handbook: Theory, Implementation, and Applications, pp. 282–305. Cambridge University Press, Cambridge (2003)
16. Prediger, S.: Logical scaling in formal concept analysis. In: Delugach, H.S., et al. (eds.) ICCS 1997. LNCS, vol. 1257, pp. 332–341. Springer, Heidelberg (1997)
17. Prediger, S.: Terminologische Merkmalslogik in der Formalen Begriffsanalyse. In: Stumme, G., Wille, R. (eds.) Begriffliche Wissensverarbeitung: Methoden und Anwendungen, pp. 99–124. Springer, Heidelberg (1999)
18. Prediger, S., Wille, R.: The lattice of concept graphs of a relationally scaled context. In: Tepfenhart, W.M. (ed.) ICCS 1999. LNCS, vol. 1640, pp. 401–414. Springer, Heidelberg (1999)
19. Priss, U.: The formalization of WordNet by methods of relational concept analysis. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database and some of its applications, MIT Press, Cambridge (1998)
20. Rudolph, S.: Exploring relational structures via FLE. In: Wolff, K.E., Pfeiffer, H.D., Delugach, H.S. (eds.) ICCS 2004. LNCS (LNAI), vol. 3127, pp. 196–212. Springer, Heidelberg (2004)
21. Rudolph, S.: Relational Exploration: Combining Description Logics and Formal Concept Analysis for Knowledge Specification. PhD thesis, Technische Universität Dresden (2006)
22. Spackman, K.A., Campbell, K.E., Cote, R.A.: SNOMED RT: A reference terminology for health care. J. of the American Medical Informatics Association, Fall Symposium Supplement, 640–644 (1997)
23. The Gene Ontology Consortium. Gene Ontology: Tool for the unification of biology. Nature Genetics 25, 25–29 (2000)
24. Wille, R.: Conceptual graphs and formal concept analysis. In: Delugach, H.S., et al. (eds.) ICCS 1997. LNCS, vol. 1257, pp. 290–303. Springer, Heidelberg (1997)