Definitorially Complete Description Logics

Balder ten Cate
ISLA, Informatics Institute
Universiteit van Amsterdam
The Netherlands
balder.tencate@uva.nl

Willem Conradie
Department of Mathematics
University of Johannesburg
South Africa
wec@rau.ac.za

Maarten Marx
ISLA, Informatics Institute
Universiteit van Amsterdam
The Netherlands
marx@science.uva.nl

Yde Venema
ILLC
Universiteit van Amsterdam
The Netherlands
yde@science.uva.nl

Abstract

The Terminology Box (TBox) of a Description Logic (DL) knowledge base is used to *define* new concepts in terms of primitive concepts and relations. The topic of this paper is the effect of the available operations in a DL on the length and the syntactic shape of definitions in a Terminology Box.

Defining new concepts can be done in two ways: (1) in an explicit syntactical manner as in $\mathsf{NewConcept} \equiv \mathsf{C}$, with C an expression in which $\mathsf{NewConcept}$ does not occur. Acyclic TBoxes only contain such axioms. (2) implicitly, by writing a set of general inclusion axioms T with the property that in any model of T, the interpretation of $\mathsf{NewConcept}$ is uniquely determined by the interpretation of the primitive concepts and relations. The explicit manner is preferred because its syntactic simplicity makes it immediately clear that $\mathsf{NewConcept}$ is nothing but a defined concept, and leads to algorithms with a lower worst case complexity. The focus of this paper is on the following property of DL's: every new concept defined in the implicit way can also be defined in the explicit manner. DL's with this property are called *Definitorially Complete*.

It is known that $\mathcal{ALC}$ is definitorially complete. We provide n concrete algorithm for computing the equivalent explicit definitions. We also investigate definitorial completeness for a number of extensions of $\mathcal{ALC}$. We show that definitorial completeness is preserved when $\mathcal{ALC}$ is extended with qualified number restrictions ($\mathcal{ALCQ}$), but is lost when nominals are added ($\mathcal{ALCO}$). On the other hand, definitorial completeness is regained when $\mathcal{ALCO}$ is further extended with the @-operator. We also show that all extensions of $\mathcal{ALC}$ and $\mathcal{ALCO}@$ with transitive roles, role inclusions, inverse roles, role intersection, and/or functionality restrictions, are definitorially complete.

1 Introduction

The Terminology Box (TBox) of a Description Logic (DL) knowledge base is used to assign names to complex concept descriptions. More formally, to *define* new concepts in terms of primitive concepts and relations. The topic of this paper is the effect of the available operations in a DL on the length and the syntactic shape of definitions in a Terminology Box.

We concentrate on the description logic $\mathcal{ALC}$ and extensions with qualified number restrictions ($\mathcal{ALCQ}$) and the one-of operator and nominals ($\mathcal{ALCO}$). This set of operations has gained particular importance because, together with inverse roles and role hierarchies (cf. Section 6), it forms the basis of the Semantic Web description logic OWL-DL (Horrocks, Patel-Schneider, & van Harmelen 2003).

Defining new concepts can be done in two ways:

1. In an explicit syntactical manner as in $\mathsf{NewConcept} \equiv \mathsf{C}$, with C an expression in which $\mathsf{NewConcept}$ does not occur. Acyclic TBoxes only contain such axioms.
2. Implicitly, by writing a set of general inclusion axioms T with the property that in any model of T, the interpretation of $\mathsf{NewConcept}$ is uniquely determined by the interpretation of the primitive concepts and relations.

The explicit manner is preferred because of two reasons: (1) Its syntactic simplicity makes it immediately clear that $\mathsf{NewConcept}$ is nothing but a defined concept. (2) It yields *Acyclic TBoxes*, for which reasoning algorithms exist with a lower worst case complexity than those for general TBoxes (e.g., the satisfiability problem is in PSPACE versus in EXPTIME for $\mathcal{ALC}$ (Baader & Sattler 2000; Donini & Massacci 2000; Donini 2003)).

We call a description logic *Definitorially Complete* if every new concept defined in the implicit way can also be defined in the explicit manner. This is Beth's Definability property for First Order Logic, but then stated for description logic. Baader & Nutt (2003) showed how Beth's property for modal logic yields the result that every definitorial but potentially cyclic $\mathcal{ALC}$ TBox is equivalent to an acyclic $\mathcal{ALC}$ TBox.

In this paper we expand this result in three directions: we consider richer languages, allow for additional axioms in the TBox like role hierarchies, and we consider the size of the acyclic TBox compared to the cyclic TBox. Our main results are

1. An algorithm for turning any definitorial $\mathcal{ALC}$ TBox into an acyclic one together with upper and lower bounds on the size of the obtained acyclic TBox. (Section 2).
2. A case study of extensions of $\mathcal{ALC}$ with qualified number restrictions and nominals. (Sections 3 and 4). Concretely, we show that all of the following description logics are definitorially complete: $\mathcal{ALC}$, $\mathcal{ALCQ}$ and $\mathcal{ALCO}@$, as well as any extension of $\mathcal{ALC}$ or $\mathcal{ALCO}@$ with transitive

roles, role inclusion, inverse roles, role intersection and/or functionality restrictions.

3. An analysis of *length conservativity*, that is, the question whether adding operations to a language L may lead to shorter definitions of concepts definable in L (Sections 2, 3, 4).

One can define new concepts from old concepts in a number of ways, of which we consider only the "traditional" one. We end this introduction with a brief survey of these different ways.

Different types of definitions

Definitorial completeness is intimately related to *Beth's definability property* for modal logics, and notions of *definition* play an important role in what is to follow. We briefly consider two other such notions.

Sometimes concepts may be defined by the use of additional, unrelated, concept symbols. Suppose that we do not have the counting apparatus of $\mathcal{ALCQ}$ (see section 4) at our disposal, but want to define the concept BiMom, of mothers that have at least two children. We can do this as follows: Let $\mathsf{BiMom} \doteq \mathsf{Woman} \sqcap \exists \mathsf{hasChild}.(Q) \sqcap \exists \mathsf{hasChild}.(\neg Q)$, where Q is a new concept symbol. The definitional power that may be attributed to this construction hinges on the fact that the desired property will hold if and only if Q *can be interpreted* to satisfy the righthand side of the equation. These are so called *projective definitions*, and will not be considered further.

Cyclic concept definitions sometimes have a natural interpretation in terms of *fixed points*. Consider this example, due to Baader & Nutt (2003): let $\mathsf{Momd} \doteq \mathsf{Man} \sqcap \forall \mathsf{hasChild}.\mathsf{Momd}$, i.e. a Momd is a man who has only male descendants. Or, in other words, Momds are those men from whom no non-males are accessible via the transitive closure of the $\mathsf{hasChild}$-relation. Here semantics with greatest fixpoints is required to capture the intended meaning (and to make the definition really definitorial). We will restrict attention to definitions that are already definitorial under the ordinary *descriptive* semantics, and these types of definitions will not be further considered.

2 Preliminaries

In this section we recall the definitions of the description logic $\mathcal{ALC}$, some welll known notions like satisfiability, subsumption and acyclicTBoxes, and we define definitorial completeness.

Given disjoint sets of atomic concept symbols $\mathcal{C}$ and roles $\mathcal{R}$, the concepts of $\mathcal{ALC}$ are given by the following recursive definition:

$$C ::= \top \mid \bot \mid A \mid C \sqcap C \mid C \sqcup C \mid \neg C \mid \exists R.C \mid \forall R.C$$

where $A \in \mathcal{C}$ and $R \in \mathcal{R}$. An $\mathcal{ALC}$-TBox (short for *Terminology Box*) is a finite set of inclusion axioms $C \sqsubseteq D$ and/or concept definitions $A \doteq C$, where A is an atomic concept symbol and C and D are $\mathcal{ALC}$-concepts.

The *length* of a concept C (notation: $|C|$) is the number of subconcepts of C (which is in the same order of size as the number of symbols in C). For a TBox T, $|T|$ denotes the sum of all $|C|$, for C the left or right hand side of an inclusion axiom or a concept definition in T.

The semantics of description logics such as $\mathcal{ALC}$ is given in terms of *interpretations*. Formally, a $(\mathcal{C},\mathcal{R})$-interpretation $\mathcal{I}$ consists of a set $\Delta^{\mathcal{I}}$ (called the *domain* of $\mathcal{I}$), and a function $(\cdot)^{\mathcal{I}}$ that assigns to each atomic concept symbol $A \in \mathcal{C}$ a subset of $\Delta^{\mathcal{I}}$ and to each role $R \in \mathcal{R}$ a binary relation over $\Delta^{\mathcal{I}}$. This interpretation function naturally extends to complex $\mathcal{ALC}$ concepts, as indicated in Table 1.

The following notions will also be used for DL's other than $\mathcal{ALC}$. A concept C is *satisfiable* if $C^{\mathcal{I}} \neq \emptyset$ for some interpretation $\mathcal{I}$. Given an interpretation $\mathcal{I}$ and a TBox A, we say that an $\mathcal{I}$ *satisfies* T (notation: $\mathcal{I} \models T$, if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all inclusion axioms $(C \sqsubseteq D) \in T$, and $A^{\mathcal{I}} = C^{\mathcal{I}}$ for all concept definitions $(A \doteq C) \in T$. In this case, we will also say that $\mathcal{I}$ *is an model of* T. Two TBoxes are said to be *equivalent* if they have the same models. A concept C is *subsumed* by a concept D (notation: $\models C \sqsubseteq D$) if for all interpretations $\mathcal{I}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. C is subsumed by D given a TBox T (notation: $T \models C \sqsubseteq D$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds for all $\mathcal{I}$ with $\mathcal{I} \models T$.

Often, the set of concept symbols $\mathcal{C}$ can be partitioned into two disjoint sets: the *P*rimary concept symbols $\mathcal{C}_P$ and the *D*efined concept symbols $\mathcal{C}_D$. The idea is that the interpretation of the symbols in $\mathcal{C}_D$ is defined by the TBox in terms of that of the symbols in $\mathcal{C}_P$, whereas the interpretation of the latter comes directly from the application domain. In order to test if the TBox indeed defines the symbols of $\mathcal{C}_D$ in terms of the symbols in $\mathcal{C}_P$, one can employ the following notion.

Definition 1 Let $\mathcal{C}_P$ and $\mathcal{C}_D$ be disjoint sets of atomic concept symbols. A TBox T is $(\mathcal{C}_P,\mathcal{C}_D)$-**definitorial** if every $(\mathcal{C}_P,\mathcal{R})$-interpretation $\mathcal{I}$ can be expanded in at most one way to a $(\mathcal{C}_P \cup \mathcal{C}_D,\mathcal{R})$-interpretation $\mathcal{I}'$ satisfying T.

By expansion we mean that $\mathcal{I}$ and $\mathcal{I}'$ restricted to $\mathcal{C}_P$ are the same, but $\mathcal{I}'$ also interprets the atomic symbols in $\mathcal{C}_D$. The notion of definitoriality captures the fact that, in interpretations satisfying the TBox, the denotation of the concepts in $\mathcal{C}_D$ is fully determined by that of the concepts in $\mathcal{C}_P$.

Let T be a TBox and T' the TBox obtained form it by uniformly replacing every occurrence of each concept symbol $C \in \mathcal{C}_D$ with a new concept symbol C'. It is not difficult to see that T is $(\mathcal{C}_P,\mathcal{C}_D)$-definitorial if and only if $T \cup T' \models \bigsqcap_{C \in \mathcal{C}_D} C \doteq C'$. It is therefore possible to determine whether a Tbox is $(\mathcal{C}_P,\mathcal{C}_D)$-definitorial, or whether a concept is implicitly defined by the TBox, by simply performing a suitable subsumption check.

In some cases definitoriality follows from the syntactic shape of a TBox. For instance:

Definition 2 An $\mathcal{ALC}$-TBox T is $(\mathcal{C}_P,\mathcal{C}_D)$-**acyclic** if it satisfies the following two properties:

1. T consists of exactly one concept definition $A \doteq C$ for each $A \in \mathcal{C}_D$, plus a number of inclusion axioms of the form $C_1 \sqsubseteq C_2$, where C_1 and C_2 are concepts without any atomic concept symbols from $\mathcal{C}_D$.

Table 1: Semantics of the $\mathcal{ALC}$ connectives

$$\begin{array}{lcl}
\top^{\mathcal{I}} & = & \Delta^{\mathcal{I}} \\
\bot^{\mathcal{I}} & = & \emptyset \\
(C \sqcap D)^{\mathcal{I}} & = & C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(C \sqcup D)^{\mathcal{I}} & = & C^{\mathcal{I}} \cup D^{\mathcal{I}} \\
(\neg C)^{\mathcal{I}} & = & \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}} \\
(\exists R.C)^{\mathcal{I}} & = & \{d \in D \mid \text{there exists } e \in D \text{ with } (d,e) \in R^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\} \\
(\forall R.C)^{\mathcal{I}} & = & \{d \in D \mid \text{for all } e \in D, \text{ if } (d,e) \in R^{\mathcal{I}} \text{ then } e \in C^{\mathcal{I}}\}
\end{array}$$

2. There is no cycle in the concept definitions in T, neither directly, as in $A \doteq (\cdots A \cdots)$, nor indirectly, by transitivity.

Clearly, every $(\{A,B\},\{C\})$-acyclic TBox is $(\{A,B\},\{C\})$-definitorial, just in virtue of its syntactic shape. Although sufficient, acyclicity is not necessary in order to be definitorial. For instance, the TBox $\{C \doteq A \sqcup (B \sqcap (C \sqcup \neg A))\}$ is equivalent to $\{C \doteq A \sqcup B\}$, and hence is $(\{A,B\},\{C\})$-definitorial without being $(\{A,B\},\{C\})$-acyclic.

Whether each definitorial TBox is equivalent to an acyclic TBox depends on the description logic being studied. If this holds for a DL, we call it definitorially complete.

Definition 3 A description logic $\mathcal{L}$ is called **definitorially complete** if each $(\mathcal{C}_P, \mathcal{C}_D)$-definitorial $\mathcal{L}$-TBox T is equivalent to a $(\mathcal{C}_P, \mathcal{C}_D)$-acyclic $\mathcal{L}$-TBox T'.

3 The basic description logic $\mathcal{ALC}$

The following result is shown in (Baader & Nutt 2003).

Theorem 4 *$\mathcal{ALC}$ is definitorially complete.*

Theorem 4 is of limited practical use, unless one has both a concrete algorithm for obtaining the equivalent acyclic TBox T' and an upperbound on its size. In the remainder of this section, we will provide a partial solution to this problem. We will describe an explicit algorithm for computing T', that involves at most a triple exponential blowup. We do not know at present whether this result can be improved, although a single exponential blowup is unavoidable:

Theorem 5 *There is an $(\mathcal{C}_P, \mathcal{C}_D)$-definitorial $\mathcal{ALC}$ TBox T, such that the smallest equivalent $(\mathcal{C}_P, \mathcal{C}_D)$-acyclic TBox T' is exponentially longer than T.*

Proof. Let $A_1, \ldots, A_n$ be atomic concept symbols and R a role, and let T be the TBox consisting of the inclusion axioms

$$\neg A_1 \sqcap \cdots \sqcap \neg A_n \sqsubseteq \forall R.\bot$$

and

$$A_1 \sqcup \cdots \sqcup A_n \sqsubseteq \exists R.\top$$

and

$$\begin{aligned}\neg A_1 \sqcap \cdots \sqcap \neg A_{k-1} \sqcap A_k \;\sqsubseteq\; & \forall R.(A_1 \sqcap \cdots \sqcap A_{k-1} \sqcap \neg A_k) \sqcap \\ & \bigsqcap_{k < \ell \leq n} ((A_\ell \sqcap \forall R.A_\ell) \sqcup (\neg A_\ell \sqcap \forall R.\neg A_\ell))\end{aligned}$$

for $k = 1 \ldots n$. Note that the size of T is polynomial in n. T expresses that $A_1, \ldots, A_n$ form an n-bit counter (with A_1 the least significant bit), counting the lengths of the maximal $R^{\mathcal{I}}$-paths starting from elements in an models, $\mathcal{I}$, of T. To be more precise, any such model, $\mathcal{I}$, will contain no $R^{\mathcal{I}}$-path of length greater than $2^n - 1$, and every $(A_i)^{\mathcal{I}}$ can be defined explicitly as follows: $x \in (A_i)^{\mathcal{I}}$ iff the n-bit binary encoding of the length of the longest $R^{\mathcal{I}}$-path starting from x has a 1 as ith least significant bit. (Note that the explicit definition is definable as an $\mathcal{ALC}$ concept without occurrances of atomic concepts). It follows that T is $(\emptyset, \{A_1, \ldots, A_n\})$-definitorial. Now, every $(\emptyset, \{A_1, \ldots, A_n\})$-acyclic TBox equivalent to T must define at least one of $A_1, \ldots, A_n$ without reference to any atomic concept symbols. Now, the depth of nesting of $\exists R$ and/or $\forall R$-constructors in any such explicit definition of A_i will have to be at least 2^n, for any shallower concept will not be able to distinguish all the domain elements distinguished by A_i. Hence, the length of any such definition will be exponential in n. Consequently every $(\emptyset, \{A_1, \ldots, A_n\})$-acyclic TBox equivalent to T must be exponentially longer than T. QED

We will now proceed to give an algorithm for turning a definitorial $\mathcal{ALC}$ TBox into an acyclic one. The algorithm is based on a special normal form for $\mathcal{ALC}$ concepts which was introduced by Janin & Walukiewicz (1995), in the setting of the modal μ-calculus. By a *literal* we mean an atomic concept or its negation.

Definition 6 For any role R and finite set of $\mathcal{ALC}$ concepts Φ, let $\nabla R.\Phi$ be shorthand for

$$\bigsqcap_{C \in \Phi} \exists R.C \sqcap \forall R. \bigsqcup_{C \in \Phi} C$$

expressing that each $C \in \Phi$ is satisfied by some R-successor, and each R-successor satisfies some $C \in \Phi$. An $\mathcal{ALC}$ concept is said to be in **disjunctive form** if it is generated by the following recursive definition:

$$C ::= \top \mid \bot \mid \pi \sqcap \nabla R_1.\Phi_1 \sqcap \cdots \sqcap \nabla R_n.\Phi_n \mid C \sqcup C$$

where π is a consistent conjunction of literals, $R_1, \ldots R_n$ are distinct roles, and $\Phi_1, \ldots, \Phi_n$ are finite sets of concepts in disjunctive form. For $\Phi = \emptyset$, $\nabla R.\Phi$ is shorthand for $\forall R.\bot$.

Lemma 7 *Every $\mathcal{ALC}$ concept is equivalent to an $\mathcal{ALC}$ concept in disjunctive form whose length is at most singly exponential in the length of the original concept.*

Proof. D'Agostino & Lenzi (2002) already showed (in the equivalent context of the modal logic **K**) that every $\mathcal{ALC}$ concept is equivalent to one in disjunctive form, but their argument gives a non-elementary upperbound. The proof we give here improves on this: it involves only a single exponential blow-up.

Let C be any $\mathcal{ALC}$ concept. We may assume without loss of generality that C is in negation normal form. For convenience, we will here consider conjunction as an operator that can take any finite set of concepts as its arguments. Thus, C is built up from literals and $\top, \bot$ using $\exists R$, $\forall R$, $\sqcup$ and $\bigsqcap$. We inductively translate C into a formula ϕ^* in disjunctive form. Most clauses of the translation are straightforward:

$$\begin{array}{lcl}
\top^* & = & \top \\
\bot^* & = & \bot \\
A^* & = & A \quad \text{(for } A \text{ a literal)} \\
(\exists R.D)^* & = & \nabla R.\{D^*, \top\} \\
(\forall R.D)^* & = & \nabla R.\emptyset \ \sqcup \ \nabla R.\{D^*\} \\
(D \sqcup E)^* & = & D^* \sqcup E^*
\end{array}$$

The only difficult case is when the concept under consideration is a conjunction, i.e., when C is of the form $\bigsqcap \Phi$. We can consider several subcases. If one of the elements of Φ is of the form $\top$, $\bot$ or $\psi \sqcup \chi$ or $\bigsqcap \Psi$, then we apply one of the following rules:

$$\begin{array}{lcl}
\bigsqcap(\Phi \cup \{\top\})^* & = & (\bigsqcap \Phi)^* \\
\bigsqcap(\Phi \cup \{\bot\})^* & = & \bot \\
\bigsqcap(\Phi \cup \{D \sqcup E\})^* & = & \bigsqcap(\Phi \cup \{D\})^* \ \sqcup \ \bigsqcap(\Phi \cup \{E\})^* \\
\bigsqcap(\Phi \cup \{\bigsqcap \Psi\})^* & = & \bigsqcap(\Phi \cup \Psi)^*
\end{array}$$

Finally, suppose that our concept C is of the form $\bigsqcap \Phi$, such that no element of Φ is of the form $\top$, $\bot$, $D \sqcup E$ or $\bigsqcap \Phi$. Then each element of Φ must be either a literal or a concept of the form $\exists R.D$ or $\forall R.D$, for some role R. We can split Φ accordingly into disjoint subsets Φ_{lit}, $\Phi_{\exists R_1}$, $\Phi_{\forall R_1}$, $\ldots$, $\Phi_{\exists R_n}$, $\Phi_{\forall R_n}$. Let $\Psi_i = \{C : \exists R_i.C \in \Phi_{\exists R_i}\}$, and $\Gamma_i = \{C : \forall R_i.C \in \Phi_{\forall R_i}\}$. If Φ_{lit} contains some atomic concept symbol and its negation, then, clearly, C is inconsistent, and we may define C^* as $\bot$. Otherwise, let $C^* = \bigsqcap \Phi_{lit} \sqcap \bigsqcap_i \phi_{R_i}$, where

$$\phi_{R_i} = \begin{cases} \nabla R_i.\{(D \sqcap \bigsqcap \Gamma_i)^* \mid D \in \Psi_i \cup \{\top\}\} & \text{if } \Phi_{\exists R_i} \neq \emptyset \\ \nabla R_i.\{(\bigsqcap \Gamma_i)^*\} \ \sqcup \ \nabla R_i.\, \emptyset & \text{otherwise} \end{cases}$$

It can be shown by induction on C that C^* is equivalent to C and that the length of C^* is singly exponential in the length of C, more precisely is $2^{O(|C| \cdot \log |C|)}$, even taking into account that $\nabla R.\Phi$ is shorthand for $\bigsqcap_{D \in \Phi} \exists R.D \sqcap \forall R. \bigsqcup_{D \in \Phi} D$. QED

The next result, but without reference to the size of the interpolant, was first proved by (Visser 1996; Ghilardi 1995). The idea of bisimulation quantifiers used in the proof may be traced via (D'Agostino & Lenzi 2002) back to (Pitts 1992).

Theorem 8 (Uniform interpolation) *For each $\mathcal{ALC}$ concept C and set of atomic concept symbols Φ, there is an $\mathcal{ALC}$ concept, which we will denote by $\tilde{\exists}\Phi.C$, satisfying the following conditions:*

1. *$\tilde{\exists}\Phi.C$ contains only atomic concept symbols that occur in C and are not in Φ.*
2. $\models C \sqsubseteq \tilde{\exists}\Phi.C$
3. *For all $\mathcal{ALC}$ concepts D not containing any atomic concepts from Φ, $\models C \sqsubseteq D$ iff $\models \tilde{\exists}\Phi.C \sqsubseteq D$.*
4. *The length of $\tilde{\exists}\Phi.C$ is singly exponential in the length of C.*

Proof. First, apply Lemma 7 to turn C into disjunctive form. This might involve a single exponential blowup. Then, define $\tilde{\exists}\Phi.\phi$ inductively as follows:

$$\begin{array}{lcl}
\tilde{\exists}\Phi.\top & = & \top \\
\tilde{\exists}\Phi.\bot & = & \bot \\
\tilde{\exists}\Phi.(\pi \sqcap \bigsqcap_i \nabla R_i.\Psi_i) & = & \pi' \sqcap \bigsqcap_i \nabla R_i \{\tilde{\exists}\Phi.C \mid C \in \Psi_i\} \\
\tilde{\exists}\Phi.(D \sqcup E) & = & \tilde{\exists}\Phi.D \sqcup \tilde{\exists}\Phi.E
\end{array}$$

where π' is obtained from the consistent conjunction of literals π by removing all (positive and negative) occurrences of atomic concept symbols in Φ. A straightforward inductive argument establishes the following fact:

> For all interpretations $\mathcal{I}$ and elements $d \in \Delta^{\mathcal{I}}$, $d \in (\tilde{\exists}\Phi.C)^{\mathcal{I}}$ iff there is an interpretation $\mathcal{J}$ and an element $e \in C^{\mathcal{J}}$ such that d and e are bisimilar with respect to all atomic concept symbols except possibly those in Φ.

In other words, $\tilde{\exists}$ is a 'bisimulation quantifier'. It follows that $\tilde{\exists}\Phi.C$ satisfies the requirements (2) and (3). That $\tilde{\exists}\Phi.C$ satisfies the requirements (1) and (4) follows directly from its definition. QED

Lemma 9 *For all $\mathcal{ALC}$ concepts C_1, C_2 and TBoxes T consisting only of inclusion axioms, the following are equivalent:*

(1) $T \models \quad C_1 \sqsubseteq C_2$,

(2) $$\models \quad \Big(C_1 \sqcap \bigsqcap_{\substack{(C \sqsubseteq D) \in T \\ R_1 \ldots R_n \in \mathcal{R},\ n \leq 2^{(|T| + |C_1| + |C_2|)}}} \forall R_1 \cdots \forall R_n (\neg C \sqcup D)\Big) \sqsubseteq C_2.$$

where $\mathcal{R}$ is the set of roles occuring in C_1, C_2 and T.

(The assumption that T consists only of inclusion axioms is not an essential restriction: this can always be ensured at the cost of at most a doubling of the size of T.)

Proof. Trivially (2) implies (1). For the other direction, suppose there is an interpretation $\mathcal{I}$ and an element $d \in \Delta^{\mathcal{I}}$, such that

$$d \in \Big(\big(C_1 \sqcap \bigsqcap_{\substack{(C \sqsubseteq D) \in T \\ R_1 \ldots R_n \in \mathcal{R},\ n \le 2^{(|T|+|C_1|+|C_2|)}}} \forall R_1 \cdots \forall R_n(\neg C \sqcup D)\big)\Big)^{\mathcal{I}}$$

and $d \notin (C_2)^{\mathcal{I}}$. We may assume without loss of generality that $\mathcal{I}$ is generated by d, meaning that every $e \in \Delta^{\mathcal{I}}$ is reachable from d in finitely many steps along the union of all relations $R^{\mathcal{I}}$ for $R \in \mathcal{R}$. We will construct a new interpretation $\mathcal{J}$ with an element $e \in \Delta^{\mathcal{J}}$, such that $\mathcal{J} \models T$ and $e \in (C_1 \sqcap \neg C_2)^{\mathcal{J}}$.

First, we need to introduce some terminology. Let Σ be the set of subconcepts of concepts in $T \cup \{C_1, C_2\}$. Let a *type* be any subset $\tau \subseteq \Sigma$. There are precisely $2^{(|T|+|C_1|+|C_2|)}$ many such types. We say that an element $e \in \Delta^{\mathcal{I}}$ has a type τ (or, that e realizes τ), if, of all subconcepts in Σ, e satisfies precisely those that are in τ. We write $e \sim e'$ if the e and e' have the same type.

For each type $\tau \subseteq \Sigma$ that is realized in $\mathcal{I}$, pick a witness $d_\tau \in \Delta^{\mathcal{I}}$ at minimal distance from d. We will now create a new interpretation $\mathcal{J}$, whose domain is the set of these witnesses. In particular, d itself belongs to the domain of $\mathcal{J}$. The interpretation of the atomic concepts in $\mathcal{J}$ is the same as in $\mathcal{I}$ but restricted to the new domain (i.e., $A^{\mathcal{J}} = A^{\mathcal{I}} \cap \Delta^{\mathcal{J}}$). For each role R, we let $R^{\mathcal{J}}$ be the set of pairs $(d_\tau, d_{\tau'})$ such that, in $\mathcal{I}$, d_τ has an R-successor of type τ'. Finally, let $\mathcal{J}_d$ be the submodel of $\mathcal{J}$ generated by d.

A straightforward inductive argument shows that the truth value of concepts in Σ is preserved:

Fact 1: For each $e \in \Delta^{\mathcal{J}_d}$ and $C \in \Sigma$, $e \in (C)^{\mathcal{J}_d}$ iff $e \in (C)^{\mathcal{I}}$.

Thus, since $d \in \Delta^{\mathcal{J}_d}$ and $d \in (C_i \sqcap \neg C_2)^{\mathcal{I}}$, also $d \in (C_i \sqcap \neg C_2)^{\mathcal{J}_d}$. Now we show that $\mathcal{J}_d$ is also a model of the TBox T. It follows, by the construction of $\mathcal{J}_d$, that

Fact 2: No two distinct $e, e' \in \Delta^{\mathcal{J}_d}$ have the same type in $\mathcal{J}_d$.

By a straightforward induction on the length of the shortest path from d to e in $\mathcal{J}_d$, we have that

Fact 3: The shortest path from d to a world e in $\mathcal{J}_d$ is no shorter than the shortest path from d to e in $\mathcal{I}$.

To see that $\mathcal{J}_d \models T$, consider any $e \in \mathcal{J}_d$. Since $\mathcal{J}_d$ is generated by d, there must be a path from d to e along the union of all relations $R^{\mathcal{J}_d}$ for $R \in \mathcal{R}$. Consider any shortest such path:

$$d = d_{\tau_0}\ (R_1)^{\mathcal{J}_d}\ d_{\tau_1}\ (R_2)^{\mathcal{J}_d}\ \ldots\ (R_n)^{\mathcal{J}_d}\ d_{\tau_n} = e$$

Since this is a shortest path and $\mathcal{J}_d$ contains only one representative of each type (cf. Fact 2), no two distinct worlds on the path can have the same type. It follows that $n \le 2^{(|T|+|C_1|+|C_2|)}$. Hence, by Fact 3, e is reachable from d in at most $2^{(|T|+|C_1|+|C_2|)}$ steps along the union of all relations $R^{\mathcal{I}}$ for $R \in \mathcal{R}$, which, by our initial assumption, implies that $e \in (\neg C \sqcup D)^{\mathcal{I}}$ for all $(C \sqsubseteq D) \in T$, and hence, by Fact 1, $e \in (\neg C \sqcup D)^{\mathcal{J}_d}$ for all $(C \sqsubseteq D) \in T$. In other words, $\mathcal{J}_d \models T$. QED

Theorem 10 *Every (C_P, C_D)-definitorial $\mathcal{ALC}$ TBox T is equivalent to an (C_P, C_D)-acyclic $\mathcal{ALC}$ TBox T^*, the size of which is at most triply exponential in the size of T.*

Proof. Let T be a (C_P, C_D)-definitorial $\mathcal{ALC}$ TBox. Introduce for each concept symbol $C \in C_D$ a distinct concept symbol C', and let $C'_D = \{C' \mid C \in C_D\}$. Let T' be obtained by replacing in T each concept $C \in C_D$ by C'. As noted before, the (C_P, C_D)-definitoriality of T implies that

$$T \cup T' \models C \sqsubset C'$$

for each $C \in C_D$. With $\Box^{\le n}$ shorthand for $\bigsqcap_{R_1 \ldots R_i \in \mathcal{R}, i \le n} \forall R_1 \cdots \forall R_i$, it follows by Lemma 9 that there is an $n \in O(2^{|T|})$ such that, for each $C \in C_D$,

$$\models \Box^{\le n}(\bigsqcap(T \cup T')) \sqcap C \sqsubseteq C'.$$

and hence

$$\models (\Box^{\le n}(\bigsqcap T) \sqcap C) \sqsubseteq (\neg\Box^{\le n}(\bigsqcap T') \sqcup C').$$

By Theorem 8 we find

$$\models (\Box^{\le n}(\bigsqcap T) \sqcap C) \sqsubseteq \tilde{\exists} C_D.(\Box^{\le n}(\bigsqcap T) \sqcap C)$$

and

$$\models \tilde{\exists} C_D.(\Box^{\le n}(\bigsqcap T) \sqcap C) \sqsubseteq (\Box^{\le n}(\bigsqcap T') \sqsubseteq C').$$

Hence

$$T \models C \sqsubseteq \tilde{\exists} C_D.(\Box^{\le n}(\bigsqcap T) \sqcap C)$$

and

$$T' \models \tilde{\exists} C_D.(\Box^{\le n}(\bigsqcap T) \sqcap C) \sqsubseteq C'$$

Substituting C for C' for every $C \in C_D$ in the latter entailment, we can combine the two to obtain

$$T \models C \doteq \tilde{\exists} C_D.(\Box^{\le n}(\bigsqcap T) \sqcap C).$$

For each $C \in C_D$, denote the right-hand-side of this equality by Δ_C. Finally, let T^* be the TBox obtained from T by replacing each occurrence of a $C \in C_D$ by Δ_C, and adding the relevant equations ($C \doteq \Delta_C$). Then T^* is (C_P, C_D)-acyclic and equivalent to T. Finally, the length of T^* is easily seen to be at most triply exponential in the length of T. QED

4 Adding qualified number restrictions

One limitation of the expressive power of $\mathcal{ALC}$ is the inability to count objects. The description logic $\mathcal{ALCQ}$ removes this limitation, by extending $\mathcal{ALC}$ with qualified number restrictions. Formally, for all natural numbers n and concepts C, $(\le n\ R\ C)$ is admitted a concept. The semantics of this new operator is as follows:

$$(\le n\ R\ C)^{\mathcal{I}} = \{a \in \Delta^{\mathcal{I}} \mid \text{there are at most } n \text{ elements } b \in C^{\mathcal{I}} \text{ with } (a,b) \in R^{\mathcal{I}}\}$$

We use $(\ge n\ R\ C)$ as shorthand for $\neg(\le (n-1)\ R\ C)$.

Clearly, there are $\mathcal{ALCQ}$ concepts that cannot be defined in $\mathcal{ALC}$. However, $\mathcal{ALCQ}$ extends $\mathcal{ALC}$ in a length-conservative manner: no $\mathcal{ALC}$-definable concept can be defined in $\mathcal{ALCQ}$ by a shorter formula. This is what we will mean when we will say that $\mathcal{ALCQ}$ is a *length-conservative extension* of $\mathcal{ALC}$.

Theorem 11 *$\mathcal{ALCQ}$ is a length-conservative extension of $\mathcal{ALC}$.*

Proof. Consider any complex $\mathcal{ALCQ}$ concept C, and suppose it is equivalent to an $\mathcal{ALC}$ concept. Let C' be the $\mathcal{ALC}$-concept obtained from C by replacing every subformula of the form $(\geq\ k\ R\ C)$ by $\exists R.C$ if $k \geq 1$ and $\top$ otherwise. Then C' is equivalent to C. To see this, consider any interpretation $\mathcal{I}$. Let $\mathcal{J}$ be the interpretation with $\Delta^{\mathcal{J}} = \mathbb{N} \times \Delta^{\mathcal{I}}$, $A^{\mathcal{J}} = \{(n, a) \mid n \in \mathbb{N}, a \in A^{\mathcal{I}}\}$ for each atomic concept A, and $R^{\mathcal{J}} = \{((n, a), (m, b)) \mid (a, b) \in R^{\mathcal{I}}\}$ for each role R. A straightforward bisimulation argument, using the fact that C is equivalent to a $\mathcal{ALC}$-concept and hence preserved under bisimulation, shows that for each $(n, d) \in \Delta^{\mathcal{J}}$, $(n, d) \in C^{\mathcal{J}}$ iff $d \in C^{\mathcal{I}}$, and $(n, d) \in (C')^{\mathcal{J}}$ iff $d \in (C')^{\mathcal{I}}$. Furthermore, by construction, C and C' have the same interpretation in $\mathcal{J}$. It follows that C and C' also have the same interpretation in $\mathcal{I}$. Since $\mathcal{I}$ was chosen arbitrarily, we have thus showed that C and C' are equivalent in all interpretations. This proves the result, since C' is clearly no longer than C. QED

Theorem 12 (Definitorial completeness of $\mathcal{ALCQ}$) *Every $(\mathcal{C}_P, \mathcal{C}_D)$-definitorial $\mathcal{ALCQ}$-TBox is equivalent to a $(\mathcal{C}_P, \mathcal{C}_D)$-acyclic $\mathcal{ALCQ}$-TBox.*

Proof. Let T be a $(\mathcal{C}_P, \mathcal{C}_D)$-definitorial $\mathcal{ALCQ}$-TBox. We will show that each concept symbol $A \in \mathcal{C}_D$ has an *explicit definition*, by which we mean a complex concept C involving only concept symbols from $\mathcal{C}_P$, such that $T \models A \doteq C$. The result then follows, since replacing each $A \in \mathcal{C}_D$ by the relevant concept C and adding these concept definitions will make the TBox $(\mathcal{C}_P, \mathcal{C}_D)$-acyclic. Suppose, for the sake of contradiction that some $A \in \mathcal{C}_D$ has no such explicit definition.

Claim 1. *There is an model $\mathcal{I}$ of T and $a \in (\neg A)^{\mathcal{I}}$ and $b \in A^{\mathcal{I}}$ such that a and b agree on all concepts involving only concept symbols from $\mathcal{C}_P$.*

Proof of claim. We use a standard argument, involving a double application of the compactness theorem. Consider $Cons_{\mathcal{C}_P}(A) = \{C \mid T \models A \sqsubseteq C$ and C only involves concept symbols from $\mathcal{C}_P\}$. By compactness, there is an model $\mathcal{I}_1$ of T and an element $a \in (\neg A)^{\mathcal{I}_1}$ such that $a \in C^{\mathcal{I}_1}$ for each $C \in Cons_{\mathcal{C}_P}(A)$. For, otherwise, there would be $C_1, \ldots, C_n \in Cons_{\mathcal{C}_P}(A)$ such that $T \models C_1 \sqcap \cdots \sqcap C_n \sqsubseteq A$, which would imply that $C_1 \sqcap \cdots \sqcap C_n$ is an explicit definition of A. Next, consider $Th_{\mathcal{C}_P}(a) = \{C \mid a \in C^{\mathcal{I}_1}$ and C only involves concept symbols from $\mathcal{C}_P\}$. Again by compactness, there is an model $\mathcal{I}_2$ of T and an element $b \in A^{\mathcal{I}_2}$ such that $b \in C^{\mathcal{I}_2}$ for each $C \in Th_{\mathcal{C}_P}(a)$. For, otherwise, there would be $C_1, \ldots, C_n \in Th_{\mathcal{C}_P}(a)$ such that $T \models A \sqsubseteq \neg(C_1 \sqcap \cdots \sqcap C_n)$, which would imply that $\neg(C_1 \sqcap \cdots \sqcap C_n) \in Cons_{\mathcal{C}_P}(A)$, thus contradicting the fact that $a \in C^{\mathcal{I}_1}$ for each $C \in Cons_{\mathcal{C}_P}(A)$. By construction, a and b agree on all complex concepts only involving concept symbols from $\mathcal{C}_P$. Finally, let $\mathcal{I}$ be the disjoint union $\mathcal{I}_1 \uplus \mathcal{I}_2$. Then $\mathcal{I} \models T$, $a \in (\neg A)^{\mathcal{I}}$, $b \in A^{\mathcal{I}}$ and a and b agree on all concepts involving only concept symbols from $\mathcal{C}_P$. **End of proof of claim**.

We may assume without loss of generality that $\mathcal{I}$ is countable and recursively saturated. (An interpretation is *recursively saturated* if it realizes every recursively enumerable type that is consistent with its theory. Every interpretation is elementarily equivalent to a countable recursively saturated interpretation. For details, see for instance (Doets 1996).)

Claim 2. *For each $d, e \in \Delta^{\mathcal{I}}$, agreeing on all complex $\mathcal{ALCQ}[C_P]$-concepts, and for all roles R, there is a bijection between the R-successors of d and the R-successors of e preserving all complex $\mathcal{ALCQ}[C_P]$-concepts.*

Proof of claim: Let $\{d_1, d_2, \ldots\}$ be the R-successors of d, and $\{e_1, e_2, \ldots\}$ those of e (note that there are at most countably many). We obtain the desired bijection between $\{d_1, d_2, \ldots\}$ and $\{e_1, e_2, \ldots\}$ as the limit of a sequence of finite partial bijections. Roughly speaking, we will alternate between finding images for elements of $\{d_1, d_2, \ldots\}$ and finding pre-images for elements of $\{e_1, e_2, \ldots\}$, so ensuring that, in the limit, the constructed function is indeed total and surjective (the injectivity will be apparent from the individual steps of the construction.)

Let $f_0 = \emptyset$, which is trivially a finite partial bijection. Next, we will now show how to construct f_{n+1} on the basis of f_n. Let $dom(f_n)$ and $rng(f_n)$ denote the domain and the range of f_n, respectively. Depending on the parity of n, proceed as follows:

n **even:** Let i be the least natural number such that $d_i \notin dom(f_n)$. For each of the finitely many $d' \in rng(f_n) \cup \{e, d_i\}$, introduce a new constant $c_{d'}$ denoting d'. Consider the type

$$\tau = \{R(c_e, y), y \neq c_{e'}\} \cup$$
$$\{(ST_{c_{d_i}}(\phi) \rightarrow ST_y(\phi)) \mid e' \in rng(f_n), \phi \in \mathcal{ALCQ}[C_P]\}$$

This is clearly a recursive type (i.e., membership of τ can be decided by a Turing machine). Furthermore, every finite subtype of τ is realized in $\mathcal{I}$. Indeed, consider any finite subtype $\tau' \subseteq \tau$, and let $\phi_1, \ldots, \phi_m$ be the (finitely many) $\mathcal{ALCQ}[C_P]$-concepts occurring in τ' that are true at d_i. Note that there may be $d' \in \mathrm{dom}(f_n)$ satisfying $(\phi_1 \sqcap \cdots \sqcap \phi_m)$. Suppose there are k many such d' (where $0 \leq k \leq n$). Then d satisfies $(\geq (k+1)R.(\phi_1 \sqcap \cdots \sqcap \phi_m))$, and hence also e (recall that d and e agree on all $\mathcal{ALCQ}[C_P]$-concepts). By the inductive hypothesis, elements linked by f_n satisfy the same $\mathcal{ALCQ}[C_P]$-concepts, and hence exactly k elements $e' \in rng(f_n)$ satisfy $(\phi_1 \sqcap \cdots \sqcap \phi_m)$. It follows that there must be an R-successor e^* of e distinct from all elements of $rng(f_n)$, satisfying $(\phi_1 \sqcap \cdots \sqcap \phi_m)$. Thus, τ' is realized by e^*.

We now appeal to the recursive saturation of $\mathcal{I}$, and conclude that τ is realized in $\mathcal{I}$, i.e. that there is an element e^* that realizes τ, hence belongs to $\{e_1, e_2, \ldots\}$, is distinct from all elements of $rng(f_n)$, and agrees with d_i on all complex $\mathcal{ALCQ}[C_P]$-concepts. We set $f_{n+1} = f_n \cup \{(d_i, e^*)\}$.

n **odd:** We consider the least natural number i such that $e_i \notin rng(f_n)$, and proceed symmetrically to the previous case. **End of proof of claim**

Next, we take the tree-unravelings of $\mathcal{I}$ around a and b, respectively. More precisely, let $\widehat{\mathcal{I}}[a]$ be the interpretation whose domain consists of all finite sequences $\langle a_1, R_1, a_2, R_2, \ldots, R_{n-1}, a_n \rangle$ with $a_1 = a$ and $(a_k, a_k + 1) \in R_k^{\mathcal{I}}$ for $k = 1, \ldots, n-1$, and such that for any role R, $R^{\widehat{I}[a]}$ consists of all pairs $(\langle a_1, R_1, a_2, \ldots, a_n \rangle, \langle a_1, R_1, a_2, \ldots, a_n, R, a_{n+1} \rangle)$ where $(a_n, a_{n+1}) \in R^{\mathcal{I}}$. We regard the *length of a sequence* $\langle a_1, R_1, a_2, R_2, \ldots, a_n \rangle$ to be n, i.e. as the number of domain elements occurring in it. An inductive argument shows that for all complex $\mathcal{ALCQ}$-concepts C and sequences $\sigma = \langle a_1, R_1, a_2, R_2, \ldots, R_{n-1}, a_n \rangle$, $\sigma \in C^{\widehat{\mathcal{I}}[a]}$ iff $a_n \in C^{\mathcal{I}}$. In particular, $\widehat{\mathcal{I}}[a]$ is an model of T and $\langle a \rangle \in (\neg A)^{\widehat{\mathcal{I}}[a]}$. Define $\widehat{\mathcal{I}}[b]$ analogously, starting from the node b. Observe that $\widehat{\mathcal{I}}[b]$ is an model of T and that $\langle b \rangle \in A^{\widehat{\mathcal{I}}[b]}$ and that $\langle a \rangle$ and $\langle b \rangle$ still agree on all complex $\mathcal{ALCQ}[\mathcal{C}_P]$-concepts.

Claim 3. *There exists a $\mathcal{C}_P$-isomorphism between $\widehat{\mathcal{I}}[a]$ and $\widehat{\mathcal{I}}[b]$ linking $\langle a \rangle$ to $\langle b \rangle$.*

Proof of claim: We obtain the claimed isomorphism by constructing a chain of partial isomorphisms $g_1 \subseteq g_2 \subseteq \cdots$ and taking g as the union. Each g_i will have as its domain and range all sequences of length at most i in $\Delta^{\widehat{\mathcal{I}}[a]}$ and $\Delta^{\widehat{\mathcal{I}}[b]}$, respectively. Moreover, if $g_i(x) = y$, then x and y will agree on all complex $\mathcal{ALCQ}[\mathcal{C}_P]$-concepts.

Let $g_1 = \{(\langle a \rangle, \langle b \rangle)\}$. This is clearly a partial isomorphism, satisfying the above conditions. Suppose that g_n has been constructed as a partial isomorphism, satisfying the conditions. Let $\sigma = \langle a_1, R_1, a_2, \ldots, R_n, a_n \rangle \in \widehat{\mathcal{I}}[a]$, and suppose $g(\sigma) = \rho = \langle b_1, R'_1, b_2, \ldots, R'_n, b_n \rangle$. Then, by the inductive hypothesis σ and ρ agree on all complex $\mathcal{ALCQ}[\mathcal{C}_P]$-concepts, i.e. a_n and b_n agree on all such concepts. By claim 2, a bijection f, preserving all complex $\mathcal{ALCQ}[\mathcal{C}_P]$-concepts exists between $\{d' \mid a_n R^{\mathcal{I}} d'\}$ and $\{e' \mid b_n R^{\mathcal{I}} e'\}$, for any role R. Now for any sequence of the form $\sigma \circ \langle R, a_{n+1} \rangle$,[1] i.e. for any R-successor of σ, let $g^{\sigma}(\sigma \circ \langle R, a_{n+1} \rangle) = \rho \circ \langle R, f(a_{n+1}) \rangle$. Note that $\sigma \circ \langle R, a_{n+1} \rangle$ and $\rho \circ \langle R, f(a_{n+1}) \rangle$ agree on all complex $\mathcal{ALCQ}[\mathcal{C}_P]$-concepts, and that g^{σ}, so defined, constitutes a bijection between the sets of successors of σ and ρ. Let g_{n+1} be the union of g_n with all such g^{σ} for all $\sigma \in \widehat{\mathcal{I}}[a]$ which are sequences of length n. **End of proof of claim**

But we have now obtained two models of T, namely $\widehat{\mathcal{I}}[a]$ and $\widehat{\mathcal{I}}[b]$, agreeing on the interpretation of all concepts in $\mathcal{C}_P$-concepts, but differing on the interpretation of a concept in $\mathcal{C}_D$, namely A. This contradicts the assumption that T is $(\mathcal{C}_P, \mathcal{C}_D)$-definitorial. QED

5 Adding nominals

Another limitation of the expressive power of $\mathcal{ALC}$ is the inability to refer to individual objects. $\mathcal{ALCO}$ addresses this issue by extending $\mathcal{ALC}$ with *nominals*: atomic concepts that denote a unique object. Examples of $\mathcal{ALCO}$ concepts are $john$ (true of the unique individual named john), $\exists CHILD.john$ (true of all parents of the unique individual named john) and $(john \sqcup mary \sqcup jane)$ (true of individuals named john, mary and jane, and false of all other individuals).

Formally, besides the atomic concept symbols $\mathcal{C}$ and the roles $\mathcal{R}$, we assume a set of nominals $\mathcal{N} = \{i, j, \ldots\}$. Syntactically, these nominals are treated as atomic concepts, just like the atomic concept symbols in $\mathcal{C}$. Semantically, each nominal is interpreted as a singleton set. No further assumptions are made on the interpretations. In particular, we do not assume that different nominals name different objects, or that all objects are named by a nominal. Uniqueness of names can be enforced, if needed, by extending the TBox with inclusion axioms of the form $i \sqsubseteq \neg j$. Also, for N a finite set of nominals, the axiom $T \sqsubseteq \bigsqcup\{i \mid i \in N\}$ states a closed world assumption.

Again, $\mathcal{ALCO}$ extends $\mathcal{ALC}$ in a length-conservative manner:

Theorem 13 *$\mathcal{ALCO}$ is a length-conservative extension of $\mathcal{ALC}$.*

Proof. Suppose an $\mathcal{ALCO}$-concept C is definable in $\mathcal{ALC}$. Let C' be the $\mathcal{ALC}$-concept obtained from C by replacing all nominals by $\perp$. Clearly, C' is at most as long as C. Now, consider any interpretation $\mathcal{I}$ and element $d \in \Delta^{\mathcal{I}}$. Let $\mathcal{I}'$ be obtained from $\mathcal{I}$ by extending the domain with a single new point, e, and making all nominals true at e. Then it can be shown that, for all $\mathcal{ALC}$-concepts D, $d \in D^{\mathcal{I}}$ iff $d \in D^{\mathcal{I}'}$. Hence, the same holds for C, as it is equivalent to an $\mathcal{ALC}$ concept. Moreover, by construction $d \in C^{\mathcal{I}'}$ iff $d \in {C'}^{\mathcal{I}'}$. All in all, this entails that $d \in C^{\mathcal{I}}$ iff $d \in {C'}^{\mathcal{I}}$. Thus, C and C' are equivalent. QED

Theorem 14 *$\mathcal{ALCO}$ is not definitorially complete.*

Proof. Let T be the TBox consisting of the following axioms:

$$\begin{array}{l} A \sqsubseteq i \\ j \sqcap B \sqsubseteq \exists R.(i \sqcap A) \\ j \sqcap \neg B \sqsubseteq \exists R.(i \sqcap \neg A) \end{array}$$

This TBox is clearly $(\{B, i, j\}, \{A\})$-definitorial: consider any model $\mathcal{I}$ of T. If $j^{\mathcal{I}} \in B^{\mathcal{I}}$, then $A^{\mathcal{I}} = \{i^{\mathcal{I}}\}$, and otherwise, $A^{\mathcal{I}} = \emptyset$. However, T is not equivalent to any $(\{B, i, j\}, \{A\})$-acyclic TBox. To see this, let $\mathcal{I}$ be the interpretation with $\Delta^{\mathcal{I}} = \{a, b\}$, $R^{\mathcal{I}} = \{(a, b)\}$, $i^{\mathcal{I}} = b$, $j^{\mathcal{I}} = a$, $A^{\mathcal{I}} = \{b\}$, $B^{\mathcal{I}} = \{a\}$, and let $\mathcal{J}$ be defined similarly, except that $A^{\mathcal{J}} = B^{\mathcal{J}} = \emptyset$. Note that $\mathcal{I}$ and $\mathcal{J}$ both satisfy T. It's quite easy to see that an $\mathcal{ALCO}$ concept that does not contain A cannot distinguish the nodes b of the two interpretations (and, in fact, the same holds for $\mathcal{ALCQO}$ concepts). It follows that A cannot be defined in terms of the other symbols by means of an $\mathcal{ALCO}$ concept definition, and hence there can be no acyclic equivalent of T. QED

[1] where '$\circ$' denotes sequence concatenation

Adding qualified number restrictions will not help: the same argument shows that also $\mathcal{ALCQO}$ is definitorially incomplete. Fortunately, we have been able to identify a modest extension of $\mathcal{ALCO}$ which *is* definitorially complete. We extend the syntax of $\mathcal{ALCO}$ by allowing $@_i C$ as a concept, for each concept C and nominal i. The semantics of this new construct is given as follows:

$$(@_i A)^{\mathcal{I}} = \begin{cases} \Delta^{\mathcal{I}} & \text{if } i^{\mathcal{I}} \in A^{\mathcal{I}} \\ \emptyset & \text{otherwise} \end{cases}$$

With @, the concept A from the proof of Theorem 14 is explcitly definable: $T \models A \equiv i \sqcap @_j B$. The resulting description logic, which we will call $\mathcal{ALCO}@$, is closely related to the notion of a Boolean IBox (Areces *et al.* 2003).

Theorem 15 *$\mathcal{ALCO}@$ is definitorially complete.*

This was essentially proved by ten Cate, Marx, & Viana (2005) in the context of hybrid logic. (For more on the relationship between hybrid and description logics see (Areces & de Rijke 2001; Sattler, Calvanese, & Molitor 2003).) It is a special case of Theorem 18 below.

Theorem 16 *$\mathcal{ALCO}@$ is a length-conservative extension of $\mathcal{ALC}$ but is not a length-conservative extension of $\mathcal{ALCO}$.*

Proof. That $\mathcal{ALCO}@$ is a length-conservative extension of $\mathcal{ALC}$ can be shown similar as for $\mathcal{ALCO}$: suppose an $\mathcal{ALCO}@$-concept C is definable in $\mathcal{ALC}$. Let $\mathcal{C}$ and $\mathcal{N}$ be the sets of atomic concepts and nominals of the language, respectively. Let C' be the $\mathcal{ALC}$-concept obtained from C by replacing all nominals by $\bot$ and replacing subconcepts of the form $@_i\phi$ by $\top$ if $\phi \sqcap \forall R.(\bot) \sqcap \bigsqcap_{P \in \mathcal{C} \cup \mathcal{N}} P$ is satisfiable, and $\bot$ otherwise. Clearly, C' is at most as long as C. Now, consider any interpretation $\mathcal{I}$ and element $d \in \Delta^{\mathcal{I}}$. Let $\mathcal{I}'$ be obtained from $\mathcal{I}$ by extending the domain with a single new point, e, and making all nominals and atomic concepts true at e. Then it can be shown that, for all $\mathcal{ALC}$-concepts D, $d \in D^{\mathcal{I}}$ iff $d \in D^{\mathcal{I}'}$. Hence, the same holds for C. Moreover, by construction $d \in C^{\mathcal{I}'}$ iff $d \in C'^{\mathcal{I}'}$. All in all, this entails that $d \in C^{\mathcal{I}}$ iff $d \in C'^{\mathcal{I}}$. Thus, C and C' are equivalent.

As for the second claim, we will assume a countably infinite set of atomic concepts $\mathcal{C} = \{P_1, P_2, \ldots\}$. Consider the sequence of $\mathcal{ALCO}@$-concepts $\phi_n = i \sqcap \exists.R \bigsqcap_{k=1,\ldots,n}(P_k \leftrightarrow @_i P_k))$, with $n \in \omega$. Each ϕ_n has length linear in n, even if the bi-implication sign is treated as a defined connective. Moreover, each ϕ_n is equivalent to an $\mathcal{ALCO}$-concept. (For instance, $\phi_1 = i \sqcap \exists R.(P_1 \leftrightarrow @_i P_1) \equiv i \sqcap ((P_1 \sqcap \exists R.P_1) \sqcup (\neg P_1 \sqcap \exists r.\neg P_1)))$. Now, take any sequence $\{\psi_n\}_{n \in \omega}$ of $\mathcal{ALCO}$-concepts with the property that there is a fixed polynomial $h(n)$ such that the length of each ψ_n is less that $h(n)$. We will show that $\phi_n \not\equiv \psi_n$ for some $n \in \omega$.

For $n \in \omega$, let F_n be the set of all functions $f : \{1, \ldots, n\} \to \{0, 1\}$. For each subset $G \subseteq F_n$, define an interpretation $\mathcal{I}_G$ as follows. The domain $\Delta^{\mathcal{I}_G}$ consists of all $f \in G$, together with an extra world w. The relation $R^{\mathcal{I}_G}$ connects w to each function $f \in G$. Further, for each $f \in G$ and primitive concept P_k, $f \in P_k^{\mathcal{I}_G}$ iff $f(k) = 1$. Lastly $w \in i^{\mathcal{I}_G}$.

Now, the number of subconcepts of any ψ_n is bounded by $h(n)$. Hence, one can distinguish between at most $2^{h(n)}$ different elements in interpretations by using subconcepts of ψ_n. On the other hand, the number of subsets of F_n is doubly exponential in n, so for large enough n there must exist $G_1, G_2 \subseteq F_n$ such that $G_1 \neq G_2$ and such that $w \in C^{\mathcal{I}_{G_1}}$ iff $w \in C^{\mathcal{I}_{G_2}}$ for all subconcepts C of ψ_n. Without loss of generality, we may assume that $G_1 \setminus G_2 \neq \emptyset$. Let $g \in G_1 \setminus G_2$. As a final step, let the interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ be identical to $\mathcal{I}_{G_1}$ and $\mathcal{I}_{G_2}$, respectively, except that, for all $k \leq n$, $w \in (P_k)^{\mathcal{I}_1}$ and $w \in (P_k)^{\mathcal{I}_2}$ iff $g(k) = 1$. A simple inductive argument shows that, still, $w \in (\psi_n)^{\mathcal{I}_1}$ iff $w \in (\psi_n)^{\mathcal{I}_2}$. However, by construction $w \in (\phi_n)^{\mathcal{I}_1}$ and $w \notin (\phi_n)^{\mathcal{I}_2}$. We conclude that $\psi_n \neq \phi_n$. QED

6 Adding role axioms

Besides nominals and qualified number restrictions, DLs such as OWL often allow for certain types of role axioms. Typical examples are *transitivity* and *role inclusion*. In this section, we generalize some of our results to incorporate such role axioms. The results we obtain also apply to *role inverse* and *role intersection*.

Almost all role axioms used in description logics can be expressed by a special type of first-order formulas:

Definition 17 A PUR-formula ("Positive allowing Universal Restrictions") is a first-order formula built up from atomic formulas (of the form Rxy, $x = y$, $\top$, or $\bot$) using conjunction, disjunction, existential quantification and universal quantification, plus restricted universal quantification of the form $\forall y.(Rxy \to \cdots)$, for x, y distinct variables. A PUR *Horn condition* is a first-order sentence of the form

$$\forall x_1, \ldots, x_n.(\phi \to \psi)$$

where ϕ is a PUR formula and ψ is of the form $R(x_i, x_j)$, $x_i = x_j$, or $\bot$).

PUR Horn conditions form a generalization of universal Horn conditions. Table 2 lists examples of role axioms that can be expressed by means of PUR Horn conditions. We call a DL *definitorially complete in the presence of* PUR *Horn conditions* if, whenever a TBox is (C_P, C_D)-definitorial relative to a set of PUR Horn conditions H, then it is equivalent (relative to H) to a $C_P, C_D)$-acyclic TBox. This is clearly a strengthening of the usual notion of definitorial completeness.

Theorem 18 *$\mathcal{ALC}$ and $\mathcal{ALCO}@$ are definitorially complete in the presence of* PUR *Horn conditions.*

Proof. It was shown in (ten Cate 2005) that $\mathcal{ALC}$ and $\mathcal{ALCO}@$ have interpolation relative to any set of PUR Horn

Table 2: Role axioms that can be expressed using PUR Horn conditions

Transitivity:	$\forall xyz.(Rxy \wedge Ryz \rightarrow Rxz)$	("R is transitive")
Role inclusion:	$\forall xy.(Rxy \rightarrow Sxy)$	("$R \subseteq S$")
Role inverse:	$\forall xy.(Rxy \rightarrow Syx)$, $\forall xy.(Sxy \rightarrow Ryx)$	("$S = R^{-1}$")
Role intersection:	$\forall xy.(R_1xy \wedge R_2xy \rightarrow Sxy)$, $\forall xy(Sxy \rightarrow R_1xy)$, $\forall xy(Sxy \rightarrow R_2xy)$	("$S = R_1 \cap R_2$")
Functionality:	$\forall xy.(Sxy \rightarrow x = y)$, $\forall x(\exists y \forall z(Rxz \rightarrow z = y) \rightarrow Sxx)$, $\forall xyz(Sxx \wedge Rxy \wedge Rxz \rightarrow y = z)$	("$S = \{(x,x) \mid x$ has at most one R-successor $\}$")

conditions:[2]

Let H be any set of PUR Horn conditions, and let $\mathcal{L}$ be either of the description logics $\mathcal{ALC}$ and $\mathcal{ALCO}@$. If $H \models C \sqsubseteq D$, for some $\mathcal{L}$-concepts C, D, then there is an $\mathcal{L}$-concept E such that $H \models C \sqsubseteq E$, $H \models E \sqsubseteq D$, and all atomic concepts (but not necessarily nominals) occurring in E occur both in C and in D.

Definitorial completeness follows from this, by a similar argument as used in the proof of Theorem 10. We will only discuss in detail the argument for $\mathcal{ALC}$. The one for $\mathcal{ALCO}@$ is similar (cf. also the proof of Corollary 4.2 in (ten Cate, Marx, & Viana 2005)).

Consider any $\mathcal{ALC}$ TBox T that is (C_P, C_D)-definitorial relative to a set of univeral Horn conditions H. Introduce for each concept symbol $C \in C_D$ a distinct concept symbol C', and let $C'_D = \{C' \mid C \in C_D\}$. Let T' be obtained by replacing in T each concept $C \in C_D$ by C'. The (C_P, C_D)-definitoriality of T relative to H implies that

$$H \cup T \cup T' \models C \sqsubseteq C'$$

for each $C \in C_D$. Let $\Box^{\leq n}$ be shorthand for $\bigsqcap_{R_1 \ldots R_i \in \mathcal{R}, i \leq n} \forall R_1 \cdots \forall R_i$. It can be shown by means of a compactness argument, using the fact that PUR Horn conditions are preserved under taking generated submodels, that there is an $n \in \mathbb{N}$ such that, for each $C \in C_D$,

$$H \models \Box^{\leq n}(\bigsqcap(T \cup T')) \sqcap C \sqsubseteq C'.$$

and hence

$$H \models (\Box^{\leq n}(\bigsqcap T) \sqcap C) \sqsubseteq (\Box^{\leq n}(\bigsqcap T') \sqsubseteq C').$$

[2] More precisely, it is shown in (ten Cate 2005) that the PUR Horn conditions form precisely the fragment of first-order logic that is preserved under bisimulation products and generated subframes, and it is shown that the basic multi-modal logic, as well as the basic hybrid logic, have interpolation over proposition letters relative to any frame class closed under bisimulation products and generated subframes.

By the above mentioned interpolation result, we can find an $\mathcal{ALC}$-concept E not containing any symbols from C_D, such that

$$H \models (\Box^{\leq n}(\bigsqcap T) \sqcap C) \sqsubseteq E$$

and

$$H \models E \sqsubseteq (\Box^{\leq n}(\bigsqcap T') \sqsubseteq C').$$

Hence

$$H \cup T \models C \sqsubseteq E \quad \text{and} \quad H \cup T' \models E \sqsubseteq C'.$$

Substituting C for C', we can combine the two to obtain

$$H \cup T \models C \doteq E.$$

For each $C \in C_D$, denote the concept E obtained in this way by Δ_C. Finally, let T^* be the TBox obtained from T by replacing each occurrence of a $C \in C_D$ by Δ_C, and adding the relevant equations $(C \doteq \Delta_C)$. Then T^* is (C_P, C_D)-acyclic and equivalent to T, relative to H. QED

The generality of this result comes at a price: since the result is proved model theoretically, we have no information on the size of the smallest equivalent acyclic TBox. Nevertheless, the result is quite powerful. In particular, it allows us to derive definitorial completeness of many description logics:

Corollary 19 *Let $X \subseteq \{\mathcal{S}, \mathcal{H}, \mathcal{I}, \mathcal{F}, \cap\}$. Then $\mathcal{ALC}X$ and $\mathcal{ALC}X\mathcal{O}@$ are definitorially complete.*

Proof. As shown in Table 2, transitivity axioms ($\mathcal{S}$) and role inclusions axioms ($\mathcal{H}$) can be expressed directly by means of PUR Horn conditions.

For inverse roles ($\mathcal{I}$), we apply the following trick: given a (C_P, C_D)-definitorial TBox containing inverse roles, we start by replacing all occurences of inverse roles, such as R^{-1}, by new atomic roles S, postulating by means of PUR Horn conditions that S is in fact the inverse of the R (see Table 2). We then apply Theorem 18 to obtain an (C_P, C_D)-acyclic TBox. Finally, we replace the newly introduced roles

S by the original R^{-1}. The same trick can be applied in the case of role intersection ($\cap$).

Finally, the most difficult case is that of functionality statements ($\mathcal{F}$). Recall that, in $\mathcal{ALCF}$, concepts can contain functionality statements of the from $(\leq 1\ R)$ as subconcepts. Given a (C_P, C_D)-definitorial TBox containing such functionality concepts, we replace each subconcept of the form $(\leq 1\ R)$ by $\exists S.\top$, for some new atomic role S. Next, we postulate by means of PUR Horn conditions that $S = \{(x, x) \mid x \text{ has at most one } R\text{-successor}\}$ (see Table 2). Applying Theorem 18, we obtain an equivalent (C_P, C_D)-acyclic TBox. Finally, we eliminate all occurences of S in the new TBox, by replacing $\exists S.C$ and $\forall S.C$ by $(\leq 1\ R) \sqcap C$ and $\neg(\leq 1\ R) \sqcup C$, respectively. QED

7 Conclusion

We have shown that $\mathcal{ALC}$ and $\mathcal{ALCQ}$ are milestones in the design landscape of DL's. Their sets of logical operations are so carefully balanced that whatever concept can be defined implicitly, can also be defined explicitly, in the same language. The language $\mathcal{ALCO}$ turned out to be less well behaved, but this problem could be solved by extending the language with @. In fact, we showed that $\mathcal{ALC}$ and $\mathcal{ALCO}$@ are definitorially complete even in the presence of role axioms defined by PUR Horn conditions. In particular, the extensions of $\mathcal{ALC}$ and $\mathcal{ALCO}$@ with any combination of transitive roles, role inclusions, inverse roles, role intersection, and/or functionality restrictions are all definitorially complete. These are important language extensions, because they form the basis of the Semantic Web description logic OWL-DL (Horrocks, Patel-Schneider, & van Harmelen 2003).

1 `Should we be more precise about the connection with OWL? the full language OWL-DL is more complicated due to the presence of concrete domains.`

Being definitorially complete is a fragile property. Unlike say *decidability of the satisfiability problem*, it does not behave monotonically with respect to expressive power: we saw that $\mathcal{ALC}$ has it, it is lost in $\mathcal{ALCO}$ but it is regained again in $\mathcal{ALCO}$@. While, $\mathcal{ALCQO}$, like $\mathcal{ALCO}$, is definitorially incomplete, we conjecture that $\mathcal{ALCQO}$@ is again complete. This fragile behaviour can serve as a valuable fitness test for DL's, and can help formulating clear research goals, such as:

> Find definitorially complete description logics with an EXPTIME-complete subsumption problem that are as expressive as possible.

Besides its theoretical interest, there is also a practical side to being definitorial completeness. Although larger, acyclic terminologies are often computationally more attractive than cyclic ones. For this reason, a DL designer could forbid users to make cyclic definitions. If the DL is acyclic definitorial complete, this is not a real restriction, as the user can always make an implicit definition explicit in these logics. However there might be a difference in user friendliness, because implicit definitions can be much more succinct than their equivalent explicit counterparts. Exactly how much more succinct remains an open problem, although the single exponential lower-bound and the triple exponential upper-bound we proved in Section 3 provides a partial answer in the case of $\mathcal{ALC}$. The authors are currently working on extending the algorithm from Section 3 to $\mathcal{ALCQ}$ and $\mathcal{ALCI}$.

References

[Areces & de Rijke 2001] Areces, C., and de Rijke, M. 2001. From description to hybrid logics, and back. In Wolter, F.; Wansing, H.; de Rijke, M.; and Zakharyaschev, M., eds., *Advances in Modal Logic. Volume 3*. CSLI Publications. 17–36.

[Areces *et al.* 2003] Areces, C.; Blackburn, P.; Hernandez, B. M.; and Marx, M. 2003. Handling boolean aboxes. In et al, D. C., ed., *Proceedings of the International Workshop on Description Logics (DL2003)*.

[Baader & Nutt 2003] Baader, F., and Nutt, W. 2003. Basic description logics. In Baader, F.; McGuiness, D.; Nardi, D.; and Patel-Schneider, P., eds., *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press.

[Baader & Sattler 2000] Baader, F., and Sattler, U. 2000. Tableau algorithms for description logics. In Dyckhoff, R., ed., *Proceedings of the International Conference on Automated Reasoning with Tableaux and Related Methods (Tableaux 2000)*, volume 1847 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag.

[Conradie 2002] Conradie, W. 2002. Definability and changing perspectives — the Beth property for three extensions of modal logic. Master's thesis, University of Amsterdam.

[D'Agostino & Lenzi 2002] D'Agostino, G., and Lenzi, G. 2002. On modal μ-calculus with explicit interpolants. Technical Report PP-2002-17, Institute for Logic, Language and Computaion, University of Amsterdam.

[Doets 1996] Doets, K. 1996. *Basic Model Theory*. Cambridge University Press.

[Donini & Massacci 2000] Donini, F. M., and Massacci, F. 2000. Exptime tableaux for $\mathcal{ALC}$. *Artificial Intelligence* 124(1):87–138.

[Donini 2003] Donini, F. 2003. Complexity of reasoning. In Baader, F.; McGuiness, D.; Nardi, D.; and Patel-Schneider, P., eds., *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press.

[Gabbay & Maksimova 2005] Gabbay, D., and Maksimova, L. 2005. *Definability and Interpolation: Modal and Intuisionistic Logic*. Oxford University Press.

[Ghilardi 1995] Ghilardi, S. 1995. An algebraic theory of normal forms. *Annals Pure Appl. Logic* 71(3):189–245.

[Horrocks, Patel-Schneider, & van Harmelen 2003] Horrocks, I.; Patel-Schneider, P.; and van Harmelen, F.

2003. From SHIQ and RDF to OWL: The making of a web ontology language. *Journal of Web Semantics* 1(1):7–26.

[Janin & Walukiewicz 1995] Janin, D., and Walukiewicz, I. 1995. Automata for the modal μ-calculus and related results. In *Mathematical foundations of computer science 1995 (Prague)*, volume 969 of *Lecture Notes in Comput. Sci.* Berlin: Springer. 552–562.

[Marx & Venema 1997] Marx, M., and Venema, Y. 1997. *Multi-dimensional Modal Logic*. Applied Logic Series. Kluwer Academic Publishers.

[Pitts 1992] Pitts, A. M. 1992. On an interpretation of second-order quantification in first-order intuitionistic propositional logic. *Journal of Symbolic Logic* 57(1):33–52.

[Sattler, Calvanese, & Molitor 2003] Sattler, U.; Calvanese, D.; and Molitor, R. 2003. Relationships with other formalisms. In Baader, F.; McGuiness, D.; Nardi, D.; and Patel-Schneider, P., eds., *The Description Logic Handbook: Theory, implementation and applications*. Cambridge University Press.

[ten Cate, Marx, & Viana 2005] ten Cate, B.; Marx, M.; and Viana, P. 2005. Hybrid logics with Sahlqvist axioms. *Logic Journal of the IGPL* 13(3):293–300.

[ten Cate 2005] ten Cate, B. 2005. *Model theory for extended modal languages*. Ph.D. Dissertation, University of Amsterdam. ILLC Dissertation Series DS-2005-01.

[Visser 1996] Visser, A. 1996. Uniform interpolation and layered bisimulation. In *Gödel '96 (Brno, 1996)*, volume 6 of *Lecture Notes Logic*. Berlin: Springer. 139–164.