Complexity of the Description Logic $ALCM$

Mónica Martinez and Edelweis Rohrer
Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay

Paula Severi
Department of Computer Science, University of Leicester, England

Abstract
In this paper we show that the problem of deciding the consistency of a knowledge base in the Description Logic $ALCM$ is ExpTime-complete. The $M$ stands for meta-modelling as defined by Motz, Rohrer and Severi. To show our main result, we define an ExpTime Tableau algorithm as an extension of an algorithm for $ALC$ by Nguyen and Szalas.

1 Introduction
The main motivation of the present work is to study the complexity of meta-modelling as defined in (Motz, Rohrer, and Severi 2014; 2015). No study of complexity has been done so far for this approach and we would like to analyse if it increases the complexity of a given description logic. The standard tableau algorithm for $ALC$ which builds completion trees, e.g. see (Baader et al. 2003), can be extended with the expansion rules for meta-modelling of (Motz, Rohrer, and Severi 2015). However, it has a high (worse case) complexity, namely $NExpTime$, and cannot be used to prove that the consistency problem for $ALCM$ is ExpTime-complete. In our approach we define a flexible syntax to equate an individual to a concept, with a strong semantics which ensures that the interpretation of the individual coincides with that of the concept. So, the domain of an interpretation can no longer consist of only basic objects, but it has to be a well-founded set. Unlike other approaches in the literature, in our approach, general translations to the consistency of a DL without metamodelling does not work because our reasoner has to check inside the algorithm that the canonical model is well-founded.

It is well-known that consistency of a (general) knowledge base in $ALC$ is ExpTime-complete (Schild 1991; De Giacomo and Lenzerini 1996). The main contribution of this paper is to show that the consistency problem for $ALCM$ is ExpTime-complete too. So, complexity does not change when moving from $ALC$ to $ALCM$. For proving our result, we define a tableau algorithm for checking consistency as an extension of an algorithm for $ALC$ by (Nguyen and Szalas 2009), and prove that it is ExpTime. Hardness follows trivially from the fact that $ALCM$ is an extension of $ALC$ since any algorithm that decides consistency of a knowledge base in $ALCM$ can be used for a knowledge base in $ALC$.

Details of our ExpTime algorithm for $ALCM$ along with proofs of correctness and the complexity result can be found in (Martinez, Rohrer, and Severi 2015).

2 A Flexible Meta-modelling Approach
A knowledge base in $ALCM$ contains an Mbox besides of a Tbox and an Abox. An Mbox is a set of equalities of the form $a=m A$ where $a$ is an individual and $A$ is a concept (Motz, Rohrer, and Severi 2015). Figure 1 shows an example of two ontologies separated by a horizontal line, where concepts are denoted by large ovals and individuals by bullets. The two ontologies conceptualize the same entities at different levels of granularity. In the ontology above the horizontal line, rivers and lakes are formalized as individuals while in the one below the line they are concepts. If we want to integrate these ontologies into a single ontology it is necessary to interpret the individual river and the concept River as the same real object. Similarly for lake and Lake. The Mbox for this example is:

\[
\text{river} =_m \text{River} \quad \text{lake} =_m \text{Lake}
\]

These equalities are called meta-modelling axioms and in this case, we say that the ontologies are related through meta-modelling. In Figure 1, meta-modelling axioms are represented by dashed edges. After adding the meta-modelling axioms, the concept HydrographicObject is now also a meta-concept because it is a concept that contains an individual which is also a concept. This kind of meta-modelling can be expressed in the undecidable logic of OWL Full (Motik 2007) but it cannot be expressed in OWL DL.

OWL 2 DL has a very restricted form of meta-modelling called punning where the same identifier can be used as an individual and as a concept (Hitzler, Krötzsch, and Rudolph 2009). We next illustrate two examples where OWL would not detect inconsistencies because the identifiers, though they look syntactically equal, are treated as different objects.

Example 1 If we introduce an axiom expressing that HydrographicObject is a subclass of River, then OWL’s reasoner will not detect that the interpretation of River is not a well founded set (it is a set that belongs to itself).
In this section, we extend the description logic ALC with meta-modelling (Motz, Rohrer, and Severi 2014; 2015). In our approach it is specially important to include expressions of the form \( a = b \) and \( a \neq b \) in the Abox. Individuals with meta-modelling represent now concepts. Since we can express equality and difference between concepts, we also need to be able to express equality and difference between the corresponding individuals. If we have an equality \( \forall b \in B \) between concepts then \( a = b \) should be equal. So, without equalities, the language lacks expressibility for doing inferences of the form \( K \models a = b \). Similarly, if we have that \( A \) and \( B \) are different, (i.e. there exists an element in \( A \) that is not in \( B \), since inequalities cannot be expressed by axioms in the Tbox), then \( a \) and \( b \) should be different. In our semantics this “correspondence” is in both directions (from individuals to concepts and viceversa), it is what we call Equality Transference (Motz, Rohrer, and Severi 2015).

Definition 1 (Model of a Knowledge Base in ALCM)

An interpretation \( I \) is a model of a knowledge base \( K = (T, A, M) \) in ALCM if the following holds:

1. the domain \( \Delta \) of the interpretation is a subset of some \( S_n \) where \( S_n \) is defined by starting from an arbitrary set \( S_0 \) of atomic objects and by giving as inductive step \( S_{n+1} = S_n \cup \mathcal{P}(S_n) \).
2. \( I \) is a model of \((T, A)\) in ALC.
3. \( a^I = A^I \) for all \( a =_m A \in M \).

In the first part of Definition 1 we restrict the domain of an interpretation in ALCM to be a subset of \( S_n \), which can now contain sets since the set \( S_n \) is defined recursively. It is easy to prove that \( S_n \) is well-founded for all \( n \in \mathbb{N} \).

The second part of Definition 1 refers to the ALC-knowledge base without the Mbox axioms.

The third part of the definition restricts the interpretation of an individual that has a corresponding concept through meta-modelling to be equal to the concept interpretation. Figure 3 shows a model for the knowledge base of Figure 2.

Definition 2 (Consistency in ALCM)

We say that a knowledge base \( K = (T, A, M) \) is consistent (satisfiable) if there exists a model of \( K \).
An algorithm for checking consistency gives only one model amongst many that, depending on the choices, e.g. the application of the or rule, may or may not be well-founded. A (general) reduction from the consistency of a DL with meta-modelling to the consistency of a DL without meta-modelling does not work. So, checking for circularities has to be done inside the algorithm. To prove that the consistency problem for \( ALC_M \) is \( \text{ExpTime} \)-complete (not greater than \( ALC \)), we define an algorithm for \( ALC_M \) that is \( \text{ExpTime} \).

4 A Tableau Calculus for \( ALC_M \)

We extend the Tableau Calculus given by (Nguyen and Szalas 2009) to handle meta-modelling axioms. This calculus uses structures called \( \text{and-or graphs} \) where both branches of a non-deterministic choice introduced by disjunction are explicitly represented. Satisfiability of the branches is propagated bottom-up and if it reaches an initial node, we can be sure that a model exists. A global catching of nodes and a proper rule-application strategy is used to guarantee the exponential bound on the size of the graph.

We make several changes to this algorithm to accommodate meta-modelling. Basically, The key feature of our extension is given by four new rules (Figure 4) for the expansion of the nodes of the graph, which are the mechanism to handle meta-modelling. Some of these rules adds new axioms to the TBox and others modify the MBox, so we add TBox and MBox axioms to the labels of the nodes in the and-or graph. Besides checking for contradictions, it is necessary to check for circularities w.r.t. the membership relation. This is done by means of the predicate \text{circular}(A,M)\) in \((\bot_3)\) which guarantees that the canonical model is well-founded. We say that \text{circular}(A,M)\) holds if there is a sequence of meta-modelling axioms \( a_1 =_m A_1, a_2 =_m A_2, \ldots, a_n =_m A_n\) all in \( M \) such that \( A_1(a_2), A_2(a_3), \ldots, A_n(a_1)\) are in \( A \). In that case we say that \( A \) has a circularity w.r.t. \( M \).

Intuitively, the algorithm works as follows:

- First of all, an and-or graph is built applying a global catching of nodes and a proper rule-application strategy.
- If the graph does not contain any node with label \( \bot \), the algorithm returns that the knowledge base is consistent.
- If not, a bottom up exploration of the graph is done, starting from the node with label \( \bot \), which is marked as inconsistent. At the end of this process, if the root node was not marked as inconsistent, the algorithm returns that the knowledge base is consistent and otherwise inconsistent.

Let \( n \) be the size of the knowledge base in \( ALC_M \). The complexity of the predicate \text{circular} \) is linear on \( n \) since it amounts to detecting cycles in a directed graph (Sedgewick and Wayne 2011). The complexity of the whole algorithm is determined by the size of the constructed graph which is exponential on \( n \). From this, we show the main new result of this paper:

**Theorem 1 (Complexity of \( ALC_M \)) Consistency of a (general) knowledge base in \( ALC_M \) is \( \text{ExpTime} \)-complete.**

\[
\begin{align*}
(\bot_1) & : (T,A,M) \text{ circular}(A,M) \\
(\bot_2) & : (T,A,M) \not\text{ circular}(A,M) \\
(\text{close}) & : (T,A,M) \not\text{ circular}(A,M) \quad \text{for } a \neq b \in A, a \neq b \\
\text{where } A' &= A[a/b], A'' = M[a/b] \\
(=) & : (T,A,U,A',M \cup \{a =_M a, a =_M b\}) \\
\text{where } T' &= T \cup ((A \cup \neg B) \cup (B \cup \neg A)) \text{ and } A' = \{(A \cup \neg B) \cap (B \cup \neg A)(d)|d \in \text{dom}(A) \cup \text{dom}(M) \cup \{a\} \}
\end{align*}
\]

**Figure 4: New Tableau Rules for Meta-modelling**

5 Related Work

ExpTime tableau algorithms for checking satisfiability w.r.t. a general TBox are shown in (De Giacomo, Donini, and Massacci 1996; Donini and Massacci 2000), which globally cache only unsatisfiable sets. An ExpTime Tableau algorithm for checking satisfiability of a concept in \( ALC \) w.r.t. a general TBox that can globally cache satisfiable and unsatisfiable sets is showed in (Goré and Nguyen 2013).

In (Nguyen and Szalas 2009), an ExpTime algorithm for checking consistency of a knowledge base, including a TBox and an Abox in \( ALC \) is presented.

In the literature of Description Logic, there are other approaches to meta-modelling (Motik 2007; Pan, Horrocks, and Schreiber 2005; Glimm, Rudolph, and Völker 2010; Jekjantuk, Gröner, and Pan 2010; Giacomo, Lenznerini, and Rosati 2011; Homola et al. 2013; 2014; Lenznerini, Lepore, and Poggi 2014). The approaches which define fixed layers or levels of meta-modelling (Pan, Horrocks, and Schreiber 2005; Jekjantuk, Gröner, and Pan 2010; Homola et al. 2013; 2014) impose a very strong limitation to the ontology engineer. The key feature in our semantics is to interpret a and A as the same object when a and A are connected through meta-modelling, i.e., if a =_M A then a^T = A^T. This allows us to detect inconsistencies in the ontologies which is not possible under the Hilog semantics (Motik 2007; Giacomo, Lenznerini, and Rosati 2011; Homola et al. 2013; 2014; Lenznerini, Lepore, and Poggi 2014; Kubincová, Kluka, and Homola 2015).

6 Conclusions and Future Work

The current algorithm was chosen only for the theoretical purpose of proving complexity. The standard tableau algorithm with meta-modelling presented in (Motz, Rohrer, and Severi 2015) is likely to work better in practice, as some
preliminary tests show (Vidal 2015).
We plan to study the complexity of more expressive logics with meta-modelling, including cardinality restrictions, role hierarchies and nominals (Tobies 2001; Nguyen and Golinska-Pilarek 2014).
We will also study the incorporation of meta-modelling to the automata approach (Calvanese, De Giacomo, and Lenzerini 1999).
Furthermore, it is also possible to show Pspace-completeness for ALCM under certain conditions of unfoldable Tboxes. The details will appear in a separate report.

7 Acknowledgements
The third author would like to acknowledge a Daphne Jackson fellowship sponsored by EPSRC and the University of Leicester. We would also like to thank Alfredo Viola for some excellent suggestions.

References