Extended Abstract: Tractable Query Answering for Expressive Ontologies and Existential Rules

David Carral, Irina Dragoste, Markus Krötzsch
Center for Advancing Electronics Dresden (cfaed), TU Dresden, Germany

Answering conjunctive queries (CQs) over knowledge bases (KBs) containing disjunctive existential rules is a relevant reasoning task which can be addressed using the disjunctive chase algorithm—a sound and complete materialisation-based procedure where all relevant consequences are pre-computed, allowing queries to be directly evaluated over materialised sets of facts—and acyclicity notions [7,8]—sufficient conditions that guarantee termination of this procedure [2]. As shown in [5,6], acyclicity notions can be used to determine that the chase will indeed terminate over a large subset of real-world ontologies. Nevertheless, even if a KB is characterised as acyclic, CQ answering still remains a problem of high theoretical complexity: CQ answering over acyclic programs with disjunctive existential rules is coN2ExpTime-complete [3]. For acyclic Horn-SROIQ ontologies, it is ExpTime-complete [6]. Moreover, CQ answering becomes even harder if we consider non-deterministic ontologies.

Example 1. Let \( R_n = \{ D_{i-1}(x) \rightarrow \exists y_i. L_i(x, y_i) \land D_i(y_i), D_{i-1}(x) \rightarrow \exists z_i. R_i(x, z_i) \land D_i(z_i) \mid i = 1, \ldots, n \} \) with \( n \geq 0 \). The chase of the program \( P = \langle R_n, \{ D_0(c) \} \rangle \) is exponentially large in \( n \). Note that \( P \) is acyclic with respect to all notions described in [6] and can be expressed in most DL fragments.

We study the limits of tractable reasoning using the chase and propose a series of restrictions that, if combined, prevent the exponential blow-up highlighted in the previous example. An important concept for predicting the behaviour of the chase procedure is the dependency graph of a rule set, defined next. Our definition refers to the skolem chase, which uses functional terms to denote fresh elements.

Definition 2. Consider a rule set \( R \) where (without loss of generality) rules do not share variables. The dependency graph \( G(R) \) of \( R \) has the existentially quantified variables in \( R \) as nodes, and an edge \( y \rightarrow z \) if the skolem chase of some program \( \langle R, I \rangle \) contains terms of the form \( f_z(®t) \) and \( f_y(®s) \), where \( f_z \) and \( f_y \) are the skolem functions for \( z \) and \( y \), respectively, and \( f_y(®s) \) occurs among the terms in \( ®t \).

Intuitively, \( y \rightarrow z \) means that a domain element created for the existential variable \( y \) was involved in an application of the rule of \( z \) (to instantiate a variable that occurred in body and head). Let \( R_n \) be the rule set from Example 1. Then, \( G(R_n) = \emptyset \) if \( n \leq 1 \), and \( G(R_n) = \{ y_{i-1} \rightarrow y_i, y_{i-1} \rightarrow z_i, z_{i-1} \rightarrow y_i, z_{i-1} \rightarrow z_i \mid i = 2, \ldots, n \} \) otherwise.

The key to our tractability results is the notion of a braid, which, intuitively speaking, consists of a possibly large number of intertwined paths.

* The full version of this paper was published at ISWC 2017 [4].
**Definition 3.** Consider a directed graph $G$. A path is a sequence of nodes $\alpha_1, \ldots, \alpha_n$ with $\alpha_i \rightarrow \alpha_{i+1} \in G$ for all $i = 1, \ldots, n - 1$. The graph $G$ is acyclic if, for every path $\alpha_1, \ldots, \alpha_n$ with $n \geq 2$, $\alpha_1 \neq \alpha_n$. A simple path is a path which does not contain two occurrences of the same node. A braid is a sequence of nodes $\alpha_1, \ldots, \alpha_n$ such that, for all $i = 1, \ldots, n - 1$, there are at least two different simple paths from $\alpha_i$ to $\alpha_{i+1}$.

As our main theoretical contribution, we study the complexity of reasoning over a rule set $R$ that satisfies some combination of the following restrictions:

- (a) The graph $G(R)$ is acyclic.
- (f) The arity of all function symbols in $sk(R)$ is at most 1.
- (b) The length of the braids in $G(R)$ is bounded.
- (w) The treewidth of the rules in $R$ is bounded.
- (p) The arity of the predicates in $R$ is bounded.

We summarise our findings in Figure 1, where we assume that the rule set satisfies (a), as otherwise reasoning becomes undecidable. All the complexity results are tight.

We empirically study the generality of these restrictions using (deterministic and non-deterministic) real-world ontologies without equality from the MOWL Corp (MC) [9] and Oxford Ontology Library (OOL) [1]. To do so, we transform the DL ontologies in these corpora into KBs with disjunctive existential rules, and then check how many of them satisfy (a-p). Since rule sets resulting from the transformation of normalised DL ontologies satisfy (f), (w), and (p), we check how many satisfy (a) and (b). We found that 61.8% (974) of the ontologies from MC and 76% (171) from OOL satisfy (a). Moreover, 78.3% of the acyclic ontologies from MC contain braids of length at most 1, 90.8% of length at most 2, 95.5% at most 3, 98.8% at most 4 and 99% at most 5. In the OOL, 51.4% of the acyclic ontologies feature braids of length at most 1, 69.5% at most 2, 81.2% at most 3, 92.3% at most 4, 97.6% at most 5, and 98.2% at most 6. Our work therefore suggests a new approach to efficient CQ answering that might be applicable to many real-world ontologies.

**Acknowledgements** This work was supported by the DFG within the cfaed Cluster of Excellence, CRC 912 (HAEC), and Emmy Noether grant KR 4381/1-1.
References

1. The oxford ontology library. Available at https://www.cs.ox.ac.uk/isg/ontologies/