I N F S Y S

R E S E A R C H

R E P O R T

INSTITUT FÜR INFORMATIONSSYSTEME

ARBEITSBEREICH WISSENSBASIERTE SYSTEME

QUERY ANSWERING IS UNDECIDABLE IN DLS WITH REGULAR EXPRESSIONS, INVERSES, NOMINALS, AND COUNTING

Magdalena Ortiz Sebastian Rudolph Mantas Šimkus

INFSYS RESEARCH REPORT 1843-10-03

APRIL 2010

Institut für Informationssysteme
AB Wissensbasierte Systeme
Technische Universität Wien
Favoritenstrassße 9-11
A-1040 Wien, Austria
Tel: +43-1-58801-18405
Fax: +43-1-58801-18493
sek@kr.tuwien.ac.at
www.kr.tuwien.ac.at

TU WIEN

INFSYS RESEARCH REPORT
INFSYS RESEARCH REPORT 1843-10-03, APRIL 2010

QUERY ANSWERING IS UNDECIDABLE IN DLS WITH REGULAR EXPRESSIONS, INVERSES, NOMINALS, AND COUNTING

Magdalena Ortiz,[1] Sebastian Rudolph[2] Mantas Šimkus[3]

Abstract. We show that the entailment of unions of conjunctive queries is undecidable in every description logic that extends $\mathcal{ALC}$ with nominals, inverse roles, functionality, and transitive closure of roles.

[1]Institute of Information Systems, Vienna University of Technology, Austria. E-mail: ortiz@kr.tuwien.ac.at.
[2]Karlsruhe Institute of Technology, Germany. E-mail: rudolph@kit.edu.
[3]Institute of Information Systems, Vienna University of Technology, Austria. E-mail: simkus@kr.tuwien.ac.at.

Acknowledgements: This work has been partially supported by the Austrian Science Fund (FWF) grant P20840, the Mexican National Council for Science and Technology (CONACYT) grant 187697, and the EU project OntoRule (IST-2009-231875).

1 Introduction and Preliminaries

In this technical note we show that in every extension of the DL $\mathcal{ALC}$ that supports nominals, inverse roles, counting constraints such as role functionality, and transitive closure of roles, the entailment problem for unions of conjunctive queries is undecidable (provided that transitive closure is also allowed to occur in the queries). It has been established that if only two among nominals, inverse roles and counting constraints are allowed, then query answering is decidable even in more expressive logics such as $\mathcal{ZIQ}$, $\mathcal{ZOQ}$, and $\mathcal{ZOI}$, and for more expressive query languages such as *2-way positive regular path queries* [4]. The combination of all three, however, is known to be challenging, and the algorithm in [5] is the only positive result for query answering in such a logic. For logics that additionally support transitive closure, such as the DL $\mathcal{ALCOIF}^*$ that we consider here, even the decidability of standard reasoning tasks (like knowledge base satisfiability) remains an open problem. It is known, however, that adding instead of transitive closure, one adds the full power of least and greatest fixed-point operators, results makes standard reasoning undecidable [2].

We start by introducing the DL $\mathcal{ALCOIF}^*$ and the corresponding query entailment problem.

1.1 The Description Logics $\mathcal{ALCOIF}^*$

We deal with the logic $\mathcal{ALCOIF}^*$ that extends the basic DL $\mathcal{ALC}$ with nominals ($\mathcal{O}$), inverse roles ($\mathcal{I}$), role functionality ($\mathcal{F}$) and transitive closure of roles (*). We briefly recall its syntax and semantics.

Syntax. Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be fixed, countably infinite sets of *concept*, *role*, and *individual names*, respectively. We assume that $\mathsf{N_C}$ contains $\top$ and $\bot$, denoting respectively the universal and the empty concept. *Concepts* C and *roles* r obey the following EBNF grammar, where $a \in \mathsf{N_I}$, $A \in \mathsf{N_C}$ and $p \in \mathsf{N_R}$:

$$C ::= A \mid \{a\} \mid C \sqcap C \mid C \sqcup C \mid \forall r.C \mid \exists r.C$$
$$r ::= p \mid p^- \mid r^*$$

A *knowledge base* (KB) is is a set of *concept inclusion axioms (CIAs)* of the form $C \sqsubseteq C'$ where C and C' are concepts, and *functionality assertions* $\mathsf{func}(p)$ and $\mathsf{func}(p^-)$ where p is a role name.[1]

Semantics. We rely on the usual notion of *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, consisting of a *domain* $\Delta^\mathcal{I} \neq \emptyset$ and a *valuation function* $\cdot^\mathcal{I}$ such that $a^\mathcal{I} \in \Delta^\mathcal{I}$ for all $a \in \mathsf{N_I}$; $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ for all $A \in \mathsf{N_C}$; $p^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$ for all $P \in \mathsf{N_R}$; $\top^\mathcal{I} = \Delta^\mathcal{I}$, and $\bot^\mathcal{I} = \emptyset$. The function $\cdot^\mathcal{I}$ inductively extends to all roles and concepts in the standard way, and the satisfaction of CIAs and functionality assertions is also standard. If $\mathcal{I}$ satisfies all axioms and assertions in $\mathcal{K}$ we call it a *model of* $\mathcal{K}$ and write $\mathcal{I} \models \mathcal{K}$.

1.2 Query Entailment

A *conjunctive role query* (CRQ) is a formula $q = \exists \vec{z}.\varphi(\vec{z})$, where $\varphi(\vec{z})$ is a conjunction of atoms $t(z, z')$ for z, z' variables from $\vec{z}$ and t a role. Note that CRQs are a special case of ordinary *conjunctive queries* (CQs), where in addition concepts may be used in atoms. A *union of conjunctive role queries* (UCRQ) q is a disjunction of CRQs.

Given an interpretation $\mathcal{I}$, a *match* π *for* $\mathcal{I}$ *and* q is an assignment of an element $\pi(z) \in \Delta^\mathcal{I}$ to each variable z occurring in q that makes the formula true. $\mathcal{I}$ *satisfies* q, denoted $\mathcal{I} \models q$, if there is a match π for

[1] As nominal concepts $\{a\}$ allow us to express ABox assertions as CIAs, we omit the distinction between the ABox and the TBox part of KBs.

$\mathcal{I}$ and q. Given a KB $\mathcal{K}$ and a UCRQ q, *query entailment* is the problem of deciding whether $\mathcal{I} \models q$ for each model $\mathcal{I}$ of $\mathcal{K}$, denoted $\mathcal{K} \models q$.

1.3 Unbounded Tiling Problem

We show the undecidability of query entailment in $\mathcal{ALCOIF}^*$ by a reduction from the following undecidable tiling (or domino) problem [1].

Definition 1.1 [Tiling problem] A *tiling system* $\mathfrak{T}$ is a triple (T, H, V) consisting of a finite set of *tile types* $T = \{t_1, \dots, t_k\}$, $k \geq 0$, a *horizontal matching* relation $H \subseteq T \times T$ and a *vertical matching* relation $H \subseteq T \times T$. A *solution* for $\mathfrak{T}$ is a function $f : \mathbb{N} \times \mathbb{N} \rightarrow T$ such that $(f(x, y), f(x + 1, y)) \in H$ and $(f(x, y), f(x, y+1)) \in V$ for every $x, y \in \mathbb{N}$. The *tiling problem* is to decide whether a given tiling system $\mathfrak{T}$ has a solution.

2 Query Entailment is undecidable in $\mathcal{ALCOIF}^*$

We show that from a given tiling system $\mathfrak{T}$ we can obtain a $\mathcal{ALCOIF}^*$ KB $\mathcal{K}_\mathfrak{T}$ and a UCRQ q such that $\mathcal{K}_\mathfrak{T} \not\models q$ iff there is a solution for $\mathfrak{T}$. It is well known that in all extensions of $\mathcal{ALC}$ we can write a knowledge base whose models can be seen as tiled pseudo-grids, where every domain element is associated to one tile type and has a matching right and up successor. Furthermore, in the presence of functionality and inverses we can make the up and right successors unique, and enforce every node to be the right or up successor of at most one node. As we will see below, the additional presence of nominals and roles of the form r^* allows us to ensure in the models of $\mathcal{K}_\mathfrak{T}$ the existence of a unique right-up path and a unique up-right path to every node n. That is, to give n a unique horizontal coordinate x and a unique vertical coordinate y. The only reason why such an structure need not be a grid, is that a node (x, y) may have as right a successor a node different from $(x + 1, y)$, or as up successor a node different from $(x, y + 1)$. However, such 'errors' can easily be detected with the query q: we define q in such a way that it has a match in a model of $\mathcal{K}_\mathfrak{T}$ iff there is some node whose right or up successor is wrong. Hence a model of $\mathcal{K}_\mathfrak{T}$ where q has no match represents a properly tiled infinite grid, which is a solution for $\mathfrak{T}$.

The reduction uses one concept name T_i for each tile type $t_i \in T$, as well as a nominal $\{o\}$ for the *origin* of the plane $\mathbb{N} \times \mathbb{N}$ (i.e., the point $(0, 0)$), the concepts X and Y for the horizontal and vertical axes (i.e., the points of the form $(x, 0)$ and $(0, y)$), respectively, and the roles x and y to relate each point (x, y) with its horizontal successor $(x + 1, y)$ and its vertical successor $(x, y + 1)$, respectively.

Defining the KB $\mathcal{K}_\mathfrak{T}$. We first define $\mathcal{K}_\mathfrak{T}$, which contains four functionality assertions for x, y, and their respective inverses:

$$\mathsf{func}(\mathrm{x}) \qquad \mathsf{func}(\mathrm{x}^-)$$
$$\mathsf{func}(\mathrm{y}) \qquad \mathsf{func}(\mathrm{y}^-)$$

Now we list the CIAs in $\mathcal{K}_\mathfrak{T}$. As we want each domain element to represent one point in the grid, we make sure it is associated to exactly one tile type:

$$\top \sqsubseteq \bigsqcup_{t_i \in T} (T_i \sqcap \bigsqcap_{t_j \in T, j \neq i} \neg T_j)$$

Furthermore, every element must have one x and one y successor, whose respective tile types are compatible with the H and V matching relations.

$$T_i \sqsubseteq \exists \mathrm{x}.(\bigsqcup_{(t_i,t_j) \in H} T_j) \sqcap \exists \mathrm{y}.(\bigsqcup_{(t_i,t_j) \in V} T_j)$$

Next, we define the horizontal and vertical axes X and Y as the set of all domain elements reachable from the origin by only right and up successors x and y, respectively.

$$X \equiv \exists \mathrm{x}^{-*}.\{o\}$$
$$Y \equiv \exists \mathrm{y}^{-*}.\{o\}$$

The only node that is in both axes is the origin, the nodes in the horizontal axis do not have incoming up arcs, and the nodes in the vertical axis do not have incoming right arcs.

$$X \sqcap Y \sqsubseteq \{o\}$$
$$X \sqsubseteq \forall \mathrm{y}^{-}.\bot$$
$$Y \sqsubseteq \forall \mathrm{x}^{-}.\bot$$

This already ensures that both axes are infinite, acyclic chains of objects. Finally, we make sure that every node in the grid is reachable from the horizontal axis by a sequence of y steps, and from the vertical axis by a sequence of x steps.

$$\top \sqsubseteq \exists \mathrm{y}^{-*}.X \sqcap \exists \mathrm{x}^{-*}.Y$$

Due to the functionality of x and y, we can now ensure that every domain element not in the borders is reachable from exactly one element of the horizontal border by only x steps, and from exactly one element of the vertical border by only y steps.

Hence in every model $\mathcal{I}$ of $\mathcal{K}_{\mathfrak{T}}$ we can assign to every node $n \in \Delta^{\mathcal{I}}$ a unique coordinate $c(n) = (x, y)$ as follows:

- $c(o^{\mathcal{I}}) = (0, 0)$,
- If $n \in X^{\mathcal{I}}$ and there exist $n_0, \ldots, n_x \in \Delta^{\mathcal{I}}$, $x \geq 0$, with $n_0 = o^{\mathcal{I}}$ and $n_x = n$, such that $n_i \neq n_j$ and $(n_i, n_{i+1}) \in \mathrm{x}^{\mathcal{I}}$ for every $0 \leq i < j \leq x$, then $c(n^{\mathcal{I}}) = (x, 0)$.
- If $n \in Y^{\mathcal{I}}$ and there exist $n_0, \ldots, n_y \in \Delta^{\mathcal{I}}$, $y \geq 0$, with $n_0 = o^{\mathcal{I}}$ and $n_y = n$ such that $n_i \neq n_j$ and $(n_i, n_{i+1}) \in \mathrm{y}^{\mathcal{I}}$ for every $0 \leq i < j \leq y$, then $c(n^{\mathcal{I}}) = (0, y)$.
- If $n \notin X^{\mathcal{I}} \cup Y^{\mathcal{I}}$, then $c(n^{\mathcal{I}}) = (c_x, c_y)$, where $n_x \in X^{\mathcal{I}}$ is the single such node for which $(n_x, n) \in (\mathrm{y}^*)^{\mathcal{I}}$, $c(n_x) = (c_x, 0)$ $n_y \in Y^{\mathcal{I}}$ is the single such node for which $(n_y, n) \in (\mathrm{x}^*)^{\mathcal{I}}$, and $c(n_y) = (0, c_y)$.

We say that $\mathcal{I}$ has an *error* if there exist $n, n' \in \Delta^{\mathcal{I}}$ with $c(n) = (x, y)$ and $c(n') = (x', y')$ for which one of the following holds:

(HE) $(n, n') \in (\mathrm{x})^{\mathcal{I}}$ and $x' > x + 1$, or

(VE) $(n, n') \in (\mathrm{y})^{\mathcal{I}}$ and $y' > y + 1$.

A model of $\mathcal{K}_{\mathfrak{T}}$ where there are no errors is isomorphic to properly tiled $\mathbb{N} \times \mathbb{N}$ grid, and thus represents a solution of $\mathfrak{T}$. Hence we have:

Lemma 2.1 *If $\mathcal{I} \models \mathcal{K}_{\mathfrak{T}}$ and there are no errors in $\mathcal{I}$, then $\mathfrak{T}$ has a solution.*

As anticipated, we use the query q to verify whether such an $\mathcal{I}$ exists.

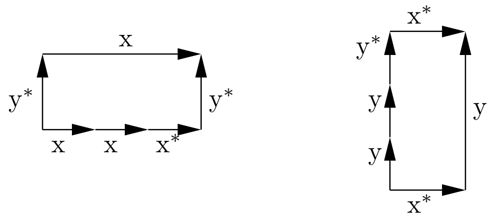

Figure 1: The queries q_H (left) and q_V (right)

Defining the query q. To conclude our reduction, we define the query q such that $\mathcal{I} \models q$ iff there is an error in $\mathcal{I}$. The query has the following form:

$$q = q_H \vee q_V$$

where q_H and q_V are CRQs such that q_H has a match in $\mathcal{I}$ if there is horizontal error (HE) in $\mathcal{I}$, and q_V has a match if there is vertical error (VE), which are defined as follows:

$$q_H = \quad \exists v_1, \ldots, v_6 \,.\, \mathrm{x}(v_1, v_6) \wedge \mathrm{y}^*(v_2, v_1) \wedge \mathrm{x}(v_2, v_3) \wedge \mathrm{x}(v_3, v_4) \wedge \mathrm{x}^*(v_4, v_5) \wedge \mathrm{y}^*(v_5, v_6)$$

$$q_V = \quad \exists v'_1, \ldots, v'_6 \,.\, \mathrm{y}(v'_1, v'_6) \wedge \mathrm{x}^*(v'_2, v'_1) \wedge \mathrm{y}(v'_2, v'_3) \wedge \mathrm{y}(v'_3, v'_4) \wedge \mathrm{y}^*(v'_4, v'_5) \wedge \mathrm{x}^*(v'_5, v'_6)$$

The queries are depicted in Figure 1. Each arrow corresponds to one atom. The variable names are omitted, but the variables are ordered counterclockwise, starting from the upper left corner for v_1 in q_H, and from the bottom left corner for v'_1 in q_V.

Now we show how the queries q_H and q_V detect horizontal and vertical errors, respectively. First consider q_H, and an arbitrary $\mathcal{I}$ such that $\mathcal{I} \models \mathcal{K}_{\mathfrak{T}}$. Suppose there is a horizontal error, i.e., there exist $n, n' \in \Delta^{\mathcal{I}}$ such that (HE): $c(n) = (x, y)$, $c(n') = (x', y')$, $(n, n') \in (\mathrm{x})^{\mathcal{I}}$ and $x' > x+1$. Let $n_x, m_x, n'_x \in X^{\mathcal{I}}$ be such that $c(n_x) = (x, 0)$, $c(m_x) = (x+1, 0)$, and $c(n'_x) = (x', 0)$; i.e., n_x, m_x, and n'_x are the nodes on the horizontal axis with the right-coordinates x, $x+1$ and x', respectively. Note that $(n_x, n) \in (\mathrm{y}^*)^{\mathcal{I}}$ and $(n'_x, n') \in (\mathrm{y}^*)^{\mathcal{I}}$. As $x' > x+1$, there exists some $m'_x \in X^{\mathcal{I}}$ with $c(m'_x) = (x_m, 0)$ such that $x_m > x+1$ and $x_m \geq x'$ (intuitively, m'_x can be n'_x or any node on the X axis between m_x and n'_x). Since $(n_x, m_x) \in (\mathrm{x})^{\mathcal{I}}$, $(m_x, m'_x) \in (\mathrm{x})^{\mathcal{I}}$, and $(m'_x, n'_x) \in (\mathrm{x}^*)^{\mathcal{I}}$, there is a match π for q_H that has $\pi(v_1) = n$, $\pi(v_6) = n'$, $\pi(v_2) = n_x$, $\pi(v_3) = m_x$, $\pi(v_4) = m'_x$, and $\pi(v_5) = n'_x$.

One can similarly show that if there are $n, n' \in \Delta^{\mathcal{I}}$ for which (VE) holds, then there is a match π' for q_V in $\mathcal{I}$ (with $\pi(v'_1) = n$ and $\pi(v'_6) = n'$). Hence there is a match for $q = q_H \vee q_V$ in every $\mathcal{I}$ that has an error, and we obtain the desired reduction:

Theorem 2.2 $\mathcal{K}_{\mathfrak{T}} \not\models q$ *iff there is a solution for* $\mathfrak{T}$.

Since the existence of a solution for $\mathfrak{T}$ is undecidable, we get:

Theorem 2.3 $\mathcal{K} \models q$ *for a* $\mathcal{ALCOIF}^*$ *KB* $\mathcal{K}$ *and a UCRQ* q *is undecidable.*

Finally, we note that each CRQ in the query q is a simple cycle, which can be expressed with one variable only in every query language that allows for composition of roles. In fact, if we consider conjunctive regular path queries [3], that allow full regular expressions over roles, then the query can be expressed as follows:

$$q = \quad \exists v \,.\, (\mathrm{y}^{-*} \cdot \mathrm{x} \cdot \mathrm{x} \cdot \mathrm{x}^* \cdot \mathrm{y}^* \cdot \mathrm{x}^-) \cup (\mathrm{x}^{-*} \cdot \mathrm{y} \cdot \mathrm{y} \cdot \mathrm{y}^* \cdot \mathrm{x}^* \cdot \mathrm{y}^-)\,(v, v)$$

This shows that conjunctive regular path queries are undecidable in $\mathcal{ALCOIF}^*$, even if only one atom and one variable are allowed in the queries. It also shows the undecidability of standard reasoning tasks (such as knowledge base satisfiability) in any extention of $\mathcal{ALCOIF}^*$ that allows to enforce the irreflexivity of the regular role expressions $\mathrm{y}^{-*} \cdot \mathrm{x} \cdot \mathrm{x} \cdot \mathrm{x}^* \cdot \mathrm{y}^* \cdot \mathrm{x}^-$ and $\mathrm{x}^{-*} \cdot \mathrm{y} \cdot \mathrm{y} \cdot \mathrm{y}^* \cdot \mathrm{x}^* \cdot \mathrm{y}^-$.

References

[1] R. Berger. The undecidability of the dominoe problem. *Mem. Amer. Math. Soc.*, 66:1–72, 1966.

[2] P. A. Bonatti and A. Peron. On the undecidability of logics with converse, nominals, recursion and counting. 158(1):75–96, 2004.

[3] D. Calvanese, G. De Giacomo, M. Lenzerini, and M. Y. Vardi. Containment of conjunctive regular path queries with inverse. pages 176–185, 2000.

[4] D. Calvanese, T. Eiter, and M. Ortiz. Regular path queries in expressive description logics with nominals. pages 714–720, 2009.

[5] B. Glimm and S. Rudolph. Status qio: Conjunctive query entailment is decidable. In *Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning (KR-10)*. AAAI Press, 2010.