Information Processing Letters 120 (2017) 30–39

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Containment of acyclic conjunctive queries with negated atoms or arithmetic comparisons

Evgeny Sherkhonov [a,*,1], Maarten Marx [b]

[a] University of Oxford, Parks Road, OX1 3QD, Oxford, UK
[b] University of Amsterdam, Science Park 904, 1098XH, Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form 2 November 2016
Accepted 16 December 2016
Available online 23 December 2016
Communicated by Jef Wijsen

Keywords:
Databases
Query containment
Conjunctive query

ABSTRACT

We study the containment problem for conjunctive queries (CQ*s*) expanded with negated atoms or arithmetic comparisons. It is known that the problem is Π_2^P-complete [14,16]. The aim of this article is to find restrictions on CQ*s* that allow for tractable containment. In particular, we consider acyclic conjunctive queries. Even with the most restrictive form of acyclicity (Berge-acyclicity), containment is coNP-hard. But for a particular fragment of Berge-acyclic CQ*s* with negated atoms or arithmetic comparisons —child-only tree patterns— containment is solvable in PTIME.

1. Introduction

We revisit the containment problem for conjunctive queries, one of the classical fundamental problems in database theory. Conjunctive queries (CQ*s*) correspond to select-from-where SQL queries, a class of most frequent queries used in practice. The containment problem is to decide, given two conjunctive queries Q_1 and Q_2, whether, over every database, the answers of Q_1 are contained in the answers of Q_2. A well-known result of Chandra and Merlin is NP-completeness of the containment problem for CQ*s* [4]. Because of relevance to practice, there have been a number of papers dedicated to finding syntactic restrictions on CQ*s* allowing polynomial-time algorithms for containment. *Acyclic* conjunctive queries have been studied as one of the restrictions [17,8].

Conjunctive queries expanded with negated atoms or arithmetic comparisons are used in practice as well. The containment problem is harder for these classes than for CQ*s* – Π_2^P-complete [14,10,15]. There has been little work on finding fragments of CQ*s* with negated atoms or comparisons that have tractable query containment. Even the restriction of acyclicity for CQ*s* has not been considered in presence of negated atoms or arithmetic comparisons. Indeed, acyclicity is a restriction on CQ*s* that allows polynomial-time containment and, furthermore, the known Π_2^P-lower bounds proofs (both in presence of negated atoms and comparisons) involve cyclic queries.

In this article we show that in some cases acyclicity does make containment easier. We show a coNP upper bound for containment of acyclic conjunctive queries with negated atoms of bounded arity. Moreover, we show that containment for acyclic conjunctive queries with arithmetic comparisons of the form $x \text{ op } c$, where x is a variable, c a constant and op a comparison operator from $\{=, \neq, <, >, \leq, \geq\}$, is also solvable in coNP. We obtain several coNP-hardness results for containment of acyclic CQ*s* with negated atoms or comparisons. These lower bounds indicate that the usual notions of acyclicity are not sufficient to obtain tractability, even with the most restrictive form of acyclicity – Berge acyclicity [7]. On a positive side we show that containment for a particular fragment

* Corresponding author at: Wolfson Building, Parks Road, OX1 3QD, Oxford, UK.
E-mail addresses: evgeny.sherkhonov@cs.ox.ac.uk (E. Sherkhonov), maartenmarx@uva.nl (M. Marx).

[1] This work was done while the author was at University of Amsterdam.

http://dx.doi.org/10.1016/j.ipl.2016.12.005

Table 1
Complexity of the containment problem: known results and the results of this article. Here ¬ denotes presence of negated atoms and ACQ*s* denotes α-acyclic CQ*s*.

Class	Complexity
CQ*s* w. ¬	Π_2^P-c [14,16]
CQ*s* w. comparisons	Π_2^P-c [10,15]
ACQ*s* w. ¬, ACQ*s* w. comparisons	coNP-c (Theorem 2, Corollary 1)
Child-only tree patterns w. ¬	PTime (Corollary 2)
Child-only tree patterns w. comparisons	PTime (Corollary 3)

of Berge-acyclic conjunctive queries with negated atoms, namely child-only tree patterns, is decidable in PTime. We extend this PTime result to the case with arithmetic comparisons. These results are based on the characterization of containment in terms of existence of a homomorphism. The latter can be checked by reducing to the known efficient algorithms for positive acyclic queries [8].

The contributions of this article are summarized in Table 1. In particular,

- We identify a fragment of CQ*s* with negated atoms for which containment is coNP-complete: α-acyclic conjunctive queries with negated atoms of bounded arity. We derive the same bound for α-acyclic CQ*s* with arithmetic comparisons.
- Consider the following three conditions on a conjunctive query $\mathcal{Q}$ with negated atoms (resp. with arithmetic comparisons).
  - (i) $\mathcal{Q}$ contains an atom with a constant as an argument,
  - (ii) $\mathcal{Q}$ is connected,
  - (iii) $\mathcal{Q}$ is Berge-acyclic.

  For every class of CQ*s* with negated atoms (arithmetic comparisons) satisfying at most two of the conditions (i)–(iii), containment is coNP-hard.
- Although we could not show that CQ*s* with negated atoms or comparisons satisfying all of (i)–(iii) have a PTime containment problem, we could do that for an even further restricted case: CQ*s* corresponding to XML tree patterns with multiple labels on the nodes [11]. If these tree patterns only contain either child or the descendant edges, their expansions with negated labels or arithmetic comparisons have a PTime containment problem.

Related work. For (positive) conjunctive queries, containment and evaluation problems are equivalent. The PTime result for evaluation of α-acyclic CQ*s* from [17] implies PTime result for containment. Gottlob et al. [8] proved that in fact evaluation (and thus containment) is complete for LOGCFL, the class of problems that are logspace reducible to a context-free language. This class of problems allows for efficient parallelizable algorithms. Since then there have been a number of papers on generalizing the acyclicity condition while keeping the evaluation and containment problems tractable. Chekuri and Rajaraman [5] introduced the notion of query width and proved that containment for CQ*s* with bounded query width is in PTime. The class of α-acyclic queries is exactly the class of queries with a query width of 1. Later, Gottlob et al. [9] introduced the notion of hypertree-width. They showed that CQ*s* of bounded hypertree-width can also be evaluated efficiently, and, moreover, this class strictly generalizes the class of queries with bounded query width.

Containment for CQ*s* expanded with negated atoms and arithmetic comparisons has been considered in [14, 16] and [10,15] respectively. In both cases, containment is Π_2^P-complete. In either of the expansions, the lower bound proofs involved cyclic CQ*s*. There has been little work studying restrictions of CQ*s* (in particular, acyclicity) with negated atoms or arithmetic comparisons that lower the complexity of containment. In [15], van der Meyden considered *monadic* CQ*s* with arithmetic comparisons, which trivially are a fragment of acyclic CQ*s* with comparisons, and argued that containment for this class is solvable in PTime.

Tree pattern containment over trees has received considerable attention as well. Child-only tree patterns are acyclic queries and thus containment is in PTime. In fact, any two-combinations of the child, descendant and the wildcard (empty node label) axes allow PTime containment [2,11]. When all the three axes are allowed, the problem becomes coNP-complete [11]. In case label negation is added to tree patterns, containment becomes PSpace-complete [6].

Containment for tree patterns expanded with attribute value comparisons has also been studied in the past. Attribute value comparisons are specific to XML documents (trees), where each node can have a number of associated attribute values. In [1] it has been shown that containment for this fragment is Π_2^P-complete. Notably, the lower bound used a reduction from containment of CQ*s* with arithmetic comparisons, and used the construct $@_a X = @_b Y$ that allows to compare attributes of two distinct nodes. In [13] it has been shown that if only constructs of the form $@_a \mathrm{op}\ c$ ($\mathrm{op} \in \{=, \neq, <, >, \leq, \geq\}$), i.e., comparison with a *constant* only, are allowed, then containment remains in coNP.

Overview. Section 2 recalls the needed concepts and notation. Section 3 is about our coNP completeness results. Section 4 contains the PTime results for the expanded tree patterns. We end with conclusions, open problems and future work.

2. Preliminaries

A relational schema **S** is a set of relational names with associated arities. We assume countably infinite disjoint sets of variables and constants **Var** and **Const**. A *term* is an element from **Var** ∪ **Const**. We also assume a dense linear order $<$ on **Const**. For tuples of terms $\bar{x}$ and $\bar{y}$, by $\bar{x} \subseteq \bar{y}$ we denote the fact that every element of $\bar{x}$ is an element of $\bar{y}$. An *instance* I over **S** is a set of *facts* of the form $R(a_1, \ldots, a_n)$, where $R \in \mathbf{S}$ is a relational name of arity n and each $a_i \in \mathbf{Const}$. By $dom(I)$ we denote the domain of I, i.e., the constants appearing in I. A *positive atom* (or just an *atom*) and a *negated atom* are expressions of the form $R(x_1, \ldots, x_n)$ and $\neg R(x_1, \ldots, x_n)$ respectively, where $R \in \mathbf{S}$ is a relational name of arity n and each x_i is a term.