Ontology Based Query Answering with Existential Rules

Michaël Thomazo
University of Montpellier
France
thomazo@lirmm.fr

1 Framework and objectives

Ontology-Based Query Answering (OBQA) is currently a problem that receives a lot of attention both from knowledge representation and databases communities. The aim is to answer queries that are at least as expressive as conjunctive queries while taking an ontology into account. This is important in order to improve the quality of query answering and interoperability between different sources of data.

The mainstream formalism to deal with ontologies is *description logics* (DLs) ([Baader *et al.*, 2007]). While historical DLs are very expressive, most of the OBQA research focus on recently introduced lightweight DLs ($\mathcal{EL}$ [Baader, 2003] and DL-Lite [Calvanese *et al.*, 2007]). Real-world ontologies expressed in these DLs already exists, such as the medical ontology SNOMED-CT (based on $\mathcal{EL}$).

During my Ph.D, I am considering an alternative, rule-based formalism. I use existential rules [Baget *et al.*, 2011b] (rules for short, also known as Tuple Generating Dependencies [Abiteboul *et al.*, 1994], Datalog+/- [Calì *et al.*, 2009]), which despite their simple syntactic form are very expressive. Moreover, they allow for a smooth integration to database systems since, in contrast to DLs, they allow for any predicate arity and support variable cycles. Indeed, their form is:

$$\forall x \forall y (B[\mathbf{x}, \mathbf{y}] \rightarrow \exists \mathrm{z} \mathrm{H}[\mathbf{y}, \mathbf{z}]),$$

where B (resp. H) is an arbitrary conjunction of atoms called the body (resp. the head) of the rule. The OBQA problem is thus formalized as follows: given two facts (existentially closed conjunctions of atoms) F and Q, a set of rules $\mathcal{R}$, does $F, \mathcal{R} \models Q$? (where $\models$ denotes logical entailment). However, the ability to create new terms, via existentially quantified variables, makes reasoning extremely complex: the OBQA problem is undecidable, even with a single rule with a single binary predicate ([Baget *et al.*, 2011b]). A lot of work has been done in the last years in order to define decidable classes of rules, aiming at a good tradeoff between expressivity and complexity. The interested reader can consult [Mugnier, 2011] for a survey of such classes. Some of these class cover the lightweight DLs used for OBQA[1].

Algorithms for OBQA can be split in two big categories: either they materialize, that is, they use $\mathcal{R}$ to infer new data from F, or not. When for any fact F, this materialization yields a finite fact entailing all possible consequence of F and $\mathcal{R}$, $\mathcal{R}$ is called a finite expansion set (*f.e.s*). When this fact is not finite, but has a bounded treewidth, $\mathcal{R}$ is said a bounded treewidth set (*b.t.s.*)[2]. Both conditions ensures decidability of OBQA, but none are recognizable. Non-materializing algorithms perform query rewriting: the initial query Q is rewritten into a query Q', such that Q is entailed by $F, \mathcal{R}$ iff Q' is entailed by F. When Q' is a finite union of conjunctive queries (UCQs), $\mathcal{R}$ is said a finite unification set (*f.u.s.*). Rules translating DL-Lite ontologies are both *b.t.s.* and *f.u.s.*, while rules translating $\mathcal{EL}$ are only *b.t.s.*.

The aim of my Ph.D thesis is to identify expressive decidable classes, study the complexity of reasoning for these classes, and design efficient algorithms.

2 Contributions

In my Ph.D work, I consider both materialization-based and query rewriting approaches. Natural brute-force algorithms exist for *f.e.s.* and *f.u.s.*, but no such algorithm is known for *b.t.s.*. In a joint work with J.-F. Baget, M.-L. Mugnier and S. Rudolph, we have defined an abstract class, named *greedy bounded treewidth set* (*g.b.t.s.*) and provided a worst-case optimal algorithm for it. This class is a subset of *b.t.s.* which covers most of the known recognizable b.t.s. classes of rules. Slight adaptation of this algorithm makes it also optimal for these subclasses. This work has been published in [Baget *et al.*, 2011a; Thomazo *et al.*, 2012; Thomazo, 2012].

Pure query rewriting approaches suffer from the exponential blow-up of the size of rewritings w.r.t. to the query, even with solely class or role hierarchies. UCQs are then too large to be efficiently dealt with by RDMS. In [König *et al.*, 2012], we show that this is inherent to UCQs by characterizing the smallest rewriting using that shape of formulas. I thus propose, in a paper submitted to IJCAI'13, to use *semi-conjunctive queries*, which are a more general form of positive existential formula. I present an algorithm for computing such rewritings, and experimentally evaluate the quality of these rewriting by checking the efficiency of evaluation of such queries. First results show that this approach is more efficient than using UCQs.

[1] Constraints and equality rules are also needed, but this is out of the scope of this summary.

[2] See [Baget *et al.*, 2011b] for formal definition of these classes.

3 Related work

I now briefly present some related results.

Recognizability of materialization-based approaches relies on two main criteria: guardedness and acyclicity. Guarded rules (where an atom of the body contains all variables of the body) have been generalized in several ways [Calì *et al.*, 2009; Baget *et al.*, 2011b; Krötzsch and Rudolph, 2011].

Weak-acyclicity [Fagin *et al.*, 2005] has been generalized into super-weak acyclicity [Marnette, 2009] and join acyclicity [Krötzsch and Rudolph, 2011]. An incomparable notion relies on the notion of rule dependency [Baget *et al.*, 2011b]. [Grau *et al.*, 2012] proposes a semantic condition of acyclicity that generalizes all these notions.

Last, the combined approach mixes materialization and query rewriting: it both extends the data by applying rules (independently of the query) and rewrite the queries with respect to the rules (independently of the data). This has proven to be useful for both DL-Lite and $\mathcal{EL}$ ontologies.

Query rewriting approaches are applicable in particular to linear [Calì *et al.*, 2009; Baget *et al.*, 2011b] and (join-)sticky [Gottlob *et al.*, 2011] rules, which are *f.u.s.*. Several algorithms have been implemented for linear rules or DL-Lite ontologies, rewriting either into a union of conjunctive queries (QuOnto , Requiem [Pérez-Urbina *et al.*, 2009] , Nyaya[Gottlob *et al.*, 2011], Iqaros [Venetis *et al.*, 2012], Rapid [Chortaras *et al.*, 2011]) or into a Datalog program (Presto [Rosati and Almatelli, 2010]). [König *et al.*, 2012] proposes an optimal algorithm for any *f.u.s.*. The question of finding polynomial cases has also been addressed [Kikot *et al.*, 2011].

References

[Abiteboul *et al.*, 1994] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases*. Addison Wesley, 1994.

[Baader *et al.*, 2007] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, second edition, 2007.

[Baader, 2003] F. Baader. Terminological cycles in a description logic with existential restrictions. In *IJCAI*, pages 325–330, 2003.

[Baget *et al.*, 2011a] J.-F. Baget, M.-L. Mugnier, S. Rudolph, and M. Thomazo. Walking the Complexity Lines for Generalized Guarded Existential Rules. In *IJCAI*, pages 712–717, 2011.

[Baget *et al.*, 2011b] Jean-François Baget, Michel Leclère, Marie-Laure Mugnier, and Eric Salvat. On Rules with Existential Variables: Walking the Decidability Line. *Artif. Intell.*, 175(9-10):1620–1654, 2011.

[Calì *et al.*, 2009] Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A General Datalog-Based Framework for Tractable Query Answering over Ontologies. In *PODS*, pages 77–86. ACM, 2009.

[Calvanese *et al.*, 2007] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Tractable reasoning and efficient query answering in description logics: The *dl-lite* family. *J. Autom. Reasoning*, 39(3):385–429, 2007.

[Chortaras *et al.*, 2011] Alexandros Chortaras, Despoina Trivela, and Giorgos B. Stamou. Optimized query rewriting for OWL 2 QL. In *CADE*, pages 192–206, 2011.

[Fagin *et al.*, 2005] R. Fagin, P. G. Kolaitis, R. J. Miller, and L. Popa. Data Exchange: Semantics and Query Answering. *Theor. Comput. Sci.*, 336(1):89–124, 2005.

[Gottlob *et al.*, 2011] Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological queries: Rewriting and optimization. In *ICDE*, pages 2–13, 2011.

[Grau *et al.*, 2012] Bernardo Cuenca Grau, Ian Horrocks, Markus Krötzsch, Clemens Kupke, Despoina Magka, Boris Motik, and Zhe Wang. Acyclicity conditions and their application to query answering in description logics. In *KR*, 2012.

[Kikot *et al.*, 2011] Stanislav Kikot, Roman Kontchakov, and Michael Zakharyaschev. Polynomial conjunctive query rewriting under unary inclusion dependencies. In *RR*, pages 124–138, 2011.

[König *et al.*, 2012] Mélanie König, Michel Leclère, Marie-Laure Mugnier, and Michaël Thomazo. A sound and complete backward chaining algorithm for existential rules. In *RR*, pages 122–138, 2012.

[Krötzsch and Rudolph, 2011] M. Krötzsch and S. Rudolph. Extending Decidable Existential Rules by Joining Acyclicity and Guardedness. In *IJCAI*, pages 963–968, 2011.

[Marnette, 2009] Bruno Marnette. Generalized schema-mappings: from termination to tractability. In *PODS*, pages 13–22, 2009.

[Mugnier, 2011] M.-L. Mugnier. Ontological Query Answering with Existential Rules. In *RR*, pages 2–23, 2011.

[Pérez-Urbina *et al.*, 2009] Héctor Pérez-Urbina, Ian Horrocks, and Boris Motik. Efficient query answering for OWL 2. In *International Semantic Web Conference*, pages 489–504, 2009.

[Rosati and Almatelli, 2010] Riccardo Rosati and Alessandro Almatelli. Improving query answering over dl-lite ontologies. In *KR*, 2010.

[Thomazo *et al.*, 2012] Michaël Thomazo, Jean-François Baget, Marie-Laure Mugnier, and Sebastian Rudolph. A generic querying algorithm for greedy sets of existential rules. In *KR*, 2012.

[Thomazo, 2012] Michaël Thomazo. From el to tractable existential rules with complex role inclusions. In *Description Logics*, 2012.

[Venetis *et al.*, 2012] Tassos Venetis, Giorgos Stoilos, and Giorgos B. Stamou. Incremental query rewriting for OWL 2 QL. In *Description Logics*, 2012.

Application for IJCAI'13 Doctoral Consortium

Michaël Thomazo

- Michaël Thomazo
- Born 24/02/1987, French citizen
- Ph.D Student (3rd year) at the University of Montpellier 2
- Supervisors : Jean-François Baget and Marie-Laure Mugnier
- Contact : thomazo@lirmm.fr - Homepage : www.lirmm.fr/∼thomazo
- Good English proficiency (TOEIC : 980/990)

Motivation for participation

I began my Ph.D at the University of Montpellier in October 2010, and I am thus approaching towards the end of my Ph.D, whose defense should take place either this summer or next fall. I submitted a paper at IJCAI'13 of which I am the sole author, and I co-authored a paper at RR'12 which received the best paper award, and as such will be presented at the IJCAI'13 track on Best Papers in Sister Conferences.

During my Ph.D training, I already had once the opportunity to attend a doctoral consortium during the KR conference, last year in Roma. It has been for me an excellent opportunity to meet an experienced researched in my field, and to meet other Ph.D. students from across the world. Both parts were very stimulating : while discussions with my mentor provided me with great feedback on my research topic, I also had the chance to discover other research topics, as well as to introduce others to my topic, in a very friendly fashion. This has been a very good experience. I wish to renew it at IJCAI'13, especially since I expect the backgrounds of students to be even more varied than they was at KR.

Michaël Thomazo
4, rue du Faubourg Boutonnet
34 090 Montpellier
Phone: (+33) 6 23 61 96 73
Email: thomazo@lirmm.fr

Date of birth: 24/02/1987
Nationality: French

Objective: *Application for the IJCAI'13 Doctoral Consortium*

Education

2010 – Now UNIVERSITY OF MONTPELLIER, LIRMM, Montpellier, France.
Ph.D Student, now in 3rd year, member of GraphIK INRIA team-project.

2006 – 2010 ÉCOLE NORMALE SUPÉRIEURE DE CACHAN, France.
Master at Master Parisien de Recherche en Informatique, Summa Cum Laude.

2003 – 2006 LYCÉE LOUIS-LE-GRAND, Paris, France.
"Classes préparatoires" (intensive training in mathematics/physics in order to take competitive exams to French "Grandes Ecoles").

Research Interships/Visits

2012 (6 weeks) KIT, AIFB, Karlsruhe, Germany.
Research visit by Sebastian Rudolph, funded by DAAD.

2011 (2 months) IWATE UNIVERSITY, KANEIWA LABORATORY, Morioka, Japan.
JSPS Summer Program (Japan Society for the Promotion of Science).

2010 (5 months) LIRMM, GRAPHIK TEAM, Montpellier, France.
Master Thesis, Knowledge Representation and Reasoning.

2009 (8 months) RWTH AACHEN, INSTITUT I1, Aachen, Germany.
Algorithmics.

2008 (5 months) TOKYO UNIVERSITY, TSUJII LABORATORY, Tokyo, Japan.
Natural Language Processing.

Teaching

2010 – 2013 UNIVERSITY MONTPELLIER 2.
Teaching Assistant.
▷ *Introduction to Algorithmics - L1.* ▷ *Co-advisor of Master level research project.* ▷ *Bureautics - L1.* ▷ *Algorithmics, Complexity, Calculability - M1.*

2009 – 2010 UNIVERISTÉ PARIS OUEST NANTERRE LA DÉFENSE.
Lecturer.
▷ *Mathematics, L1(Economy).*

2008 – 2009 LYCÉE LOUIS-LE-GRAND.
"Colleur".
▷ *Oral Interrogation, Mathematics, Bachelor level.*

Languages

French : native.
English : autonomous, 980/990 at TOEIC in 2008.
German : oral communication.
Japanese : intermediary skills. JLPT 3 in 2009.

Publications

2012 Michaël Thomazo, Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, *A Generic Querying Algorithm for Greedy Sets of Existential Rules*, KR'12.

Marie-Laure Mugnier, Geneviève Simonet, Michaël Thomazo, *On the Complexity of Entailment in Existential Conjunctive First Order Logic with Atomic Negation*, Information and Computation.

Mélanie König, Michel Leclère, Marie-Laure Mugnier, Michaël Thomazo, *A Sound and Complete Backward Chaining Algorithm for Existential Rules*, RR'12.

Michaël Thomazo, *From $\mathcal{EL}$ to Tractable Existential Rules with Complex Role Inclusions*, DL'12.

2011 Jean-François Baget, Marie-Laure Mugnier, Sebastian Rudolph, Michaël Thomazo, *Walking the Complexity Lines for Generalized Guarded Existential Rules*, IJCAI'11.

Jean-François Baget, Marie-Laure Mugnier, Michaël Thomazo, *Towards Farsighted Dependencies for Existential Rules*, RR'11.

Jean-François Baget, Marie-Laure Mugnier, Michaël Thomazo, *Notions de dépendance pour les règles existentielles*, IAF'11.

Conference Attendance

Attendance to RR'11, DL'12, KR'12, RR'12.
Attendance to 8th Reasoning Web Summer School, Vienna, Austria.

Miscellaneaous

French national titles in fencing.
Handball player at ENS Cachan (2006 - 2008).
Member of the literary prize of ENS Cachan in 2008.
Co-organisator of LIRMM Ph.D Student seminary (2011-2012).