Report from Dagstuhl Seminar 14331

Querying and Reasoning Under Expressive Constraints

Edited by

Michael Benedikt[1], Carsten Lutz[2], and Balder Ten Cate[3]

1 **University of Oxford, GB,** michael.benedikt@gmail.com
2 **Universität Bremen, DE,** clu@informatik.uni-bremen.de
3 **LogicBlox Inc. and University of California – Santa Cruz, US,** balder.tencate@gmail.com

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 14331 "Querying and Reasoning Under Expressive Constraints" which took place from August 10th to August 14th, 2014. The seminar aimed to bring together researchers in databases, knowledge representation, decidable fragments of first-order logic, and constraint satisfaction to identify and discuss common themes and technique as well as complementary ones, identify future research issues, and foster cooperation and cross-fertilization between the communities.

Seminar August 10–14, 2014 – http://www.dagstuhl.de/14331
1998 ACM Subject Classification H.2.1 Database Management: Logical Design, I.2.4 Artificial Intelligence: Knowledge Representation, Formalisms and Methods, F.4.1 Mathematical Logic and Formal Languages: Mathematical Logic
Keywords and phrases Integrity constraints, Open-World Query Answering, Ontology-Based Data Access, Knowledge Representation, Automated Reasoning, Decidable Fragments of First-Order Logic
Digital Object Identifier 10.4230/DagRep.4.8.1
Edited in cooperation with Cristina Civili (University of Rome "La Sapienza", IT)

1 Executive Summary

Michael Benedikt
Carsten Lutz
Balder ten Cate

Motivation

Query answering in the presence of expressive constraints and logical rules is a topic that has drawn attention from several different research communities. In databases, the interaction of constraints and queries arises in the context of query optimization – for example, how to make use of integrity constraints such as inclusion dependencies and functional dependencies in running a query more efficiently. The topic is also central to the more recent database topics of data integration and data exchange, where constraints are used in the specification of schema mappings. In the area of knowledge representation, the interaction of constraints and queries plays a great role as well – particularly in ontology-based query answering.

The work in these areas is closely related also to another fundamental topic in theoretical computer science, namely decidable fragments of first-order logic. In particular, many of the

query answering and query analysis techniques used in recent work within databases and knowledge representation have close links to static analysis of guarded logics, a family of logics that arose out of work by the modal logic and finite model theory communities.

The seminar focused on the convergence of interest of the databases, knowledge representation, and computational logic communities. Its goal was to make visible the connections between these distinct communities, to look at tools and algorithms in one community that can be applied within others, to understand which formalisms and techniques are most promising from the perspective of practical applications, and to propose new ways to combine techniques across communities.

Overview and Outcome

The week started with three overview lectures from well-known authorities in databases, description logics, and decidable fragments of first-order logic. These talks introduced the necessary background for participants and raised research themes that would be explored in later talks. The week then proceeded with a wide-ranging series of talks by participants. In addition to finite model theory, description logics, and databases, there were also talks concerning the interaction of querying problems with constraint satisfaction. The presentations included theoretical work as well as system demonstrations and discussion of practical obstacles to efficient querying with constraints. There were two presentations by participants from industry (IBM and LogicBlox), describing products that implement integrity constraint-based approaches to entity resolution and data analytics, respectively. There was also a presentation on the status of constraint-based reasoning within the W3C endorsed query language SPARQL. In addition to the formal talks, the seminar had an open discussion session, which included a mention of some major open problems and directions to be explored for the communities, as well an attempt at mapping the distinct vocabularies of the different communities.

A main outcome of the discussion was a desire for further interaction between the communities. There were a number of proposals put forward for how to achieve this, including co-location of a KR-related conference with a database conference like VLDB or SIGMOD/PODS. Another outcome was a collection of topics that were particularly worth pursuing by all communities. The handling of inconsistency in databases was one of these – both further investigation of the most widely-used approach for inconsistency-handling, based on repair and consistent query answering, and the examination of alternative approaches. The notion of repair tied into the question of investigating the relationship of data uncertainty and constraints. Markov logic networks (MLNs) are likely to play a role in reconciling "hard" integrity constraints with probabilities, although the interplay of probabilistic data and classical approaches to integrity constraints will involve a more general revision of the major computational problems with uncertainty in mind. Another topic identified for future work was the notion of incremental checking of constraints. Incremental computation was alluded to in several talks, but there appears to be a need to take a more holistic look at models for incremental computation and their application in constraint maintenance. The recent activity within dynamic complexity makes the topic of incremental computation within constraint handling particularly ripe for revisiting.

Conclusion

We believe that the seminar was very successful in bringing together the involved communities and in promoting interaction and exchange between them. Similarities as well as differences

between the communities' research efforts became clearly visible and the participants conceived the seminar as a significant step forwards in bridging the gap and raising mutual awareness. Many participants expressed interest in a followup event.

2 Table of Contents

3 Overview of Talks

3.1 Open-World Finite Query Answering Under Number Restrictions

Antoine Amarilli (Telecom ParisTech, FR)

Open-world finite query answering (QA) is the problem of deciding, given a database instance, a set of constraints and a query, whether the query holds over all possible finite completions of the instance satisfying the constraints. It is used to reason over incomplete information and find out if a query is entailed by constraints given non-exhaustive data. Though finite QA is in general undecidable under expressive constraint languages, decidable cases are known: the guarded fragment, which cannot express number restrictions such as functional dependencies, or the guarded fragment with number restrictions but on a signature of arity only two.

We show that finite QA is decidable under unary inclusion dependencies and functional dependencies. More specifically, we prove that, up to an existing finite closure operation on the dependencies, finite controllability holds: namely, finite QA is equivalent to query answering for arbitrary models (finite and infinite), for which efficient techniques are known. This provides, to our knowledge, the first decidability result for finite QA on arbitrary arity signatures under tuple-generating and equality-generating dependencies with complex interaction.

3.2 The Use of Integrity Constraints at LogicBlox

Molham Aref (LogicBlox – Atlanta, US)

LogicBlox provides an integrated platform for high-performance data management and analytics, based on a declarative language called LogiQL. LogiQL is, at its core, an extension of Datalog that offers native language support for expressing data-intensive tasks such as machine learning and combinatorial optimization. The presentation gives an overview of the platform and the language. In particular, it focuses on the important role of integrity constraints in LogiQL, which are used not only for maintaining data integrity, but also, for example, for the specification of complex optimization problems and probabilistic programming.

3.3 Expressive languages for querying the semantic web

Marcelo Arenas (Pontificia Universidad Catolica de Chile, CL)

Joint work of Arenas, Marcelo; Gottlob, Georg; Pieris, Andreas

The problem of querying RDF data is a central issue for the development of the Semantic Web. The query language SPARQL has become the standard language for querying RDF, since its standardisation in 2008. However, the 2008 version of this language missed some

important functionalities: reasoning capabilities to deal with RDFS and OWL vocabularies, navigational capabilities to exploit the graph structure of RDF data, and a general form of recursion much needed to express some natural queries. To overcome these limitations, a new version of SPARQL, called SPARQL 1.1, was recently released, which includes entailment regimes for RDFS and OWL vocabularies, and a mechanism to express navigation patterns through regular expressions. Unfortunately, there are still some useful navigation patterns that cannot be expressed in SPARQL 1.1, and the language lacks of a general mechanism to express recursive queries.

To the best of our knowledge, there is no RDF query language that combines the above functionalities, and which can also be evaluated efficiently. It is the aim of this work to fill this gap. Towards this direction, we focus on the OWL 2 QL profile of OWL 2, and we show that every SPARQL query enriched with the above features can be naturally translated into a query expressed in a language which is based on an extension of Datalog which allows for value invention and stratified negation. However, the query evaluation problem for this language is highly intractable, which is not surprising since it is expressive enough to encode some inherently hard queries. We identify a natural fragment of it, and we show it to be tractable and powerful enough to define SPARQL queries enhanced with the desired functionalities.

3.4 Inconsistency-tolerant query answering in ontology-based data access

Meghyn Bienvenu (University Paris South, FR)

In this talk, I will give an overview of a recent line of work on inconsistency-tolerant query answering in the setting of ontology-based data access. After reviewing some basic notions related to querying data in the presence of ontologies, I will present various inconsistency-tolerant semantics, discuss their computational properties, and describe a practical method based upon the use of incomplete methods and calls to a SAT solver. At the end of the talk, I will mention some open questions and directions for future research.

References for results mentioned in the talk: [1, 2, 3, 4, 5, 6, 7, 8, 9].

References

1 Lembo et al. Inconsistency-tolerant semantics for description logics. In Proc. of RR (2010).
2 Lembo et al. Query rewriting for inconsistent DL-Lite ontologies. In Proc. of RR (2011).
3 Rosati. On the complexity of dealing with inconsistency in description logic ontologies. In Proc. of IJCAI (2011).
4 Bienvenu. First-order expressibility results for queries over inconsistent DL-Lite knowledge bases. In Proc. of DL (2011).
5 Bienvenu. On the complexity of consistent query answering in the presence of simple ontologies. In Proc. of AAAI (2012).
6 Lukasiewicz, Martinez, and Simari. Inconsistency handling in datalog+/- ontologies. In Proc. of ECAI (2012).
7 Lukasiewicz, Martinez, and Simari. Complexity of inconsistency-tolerant query answering in datalog+/-. In Proc. of OTM (2013).

8 Bienvenu and Rosati. Tractable approximations of consistent query answering for robust ontology-based data access. Proc. of IJCAI (2013).
9 Bienvenu, Bourgaux, and Goasdoué;. Querying inconsistent description logic knowledge bases under preferred repair semantics. In Proc. of AAAI (2014).

3.5 Complexity of Constraints: a short introduction to the universal-algebraic approach

Manuel Bodirsky (Ecole Polytechnique – Palaiseau, FR)

The complexity of Constraint Satisfaction Problems (CSPs) is in multiple ways connected to the theory of databases. It is still an open problem whether the class of all CSPs over a fixed finite domain exhibits a complexity dichotomy: Feder and Vardi conjectured that all such CSPs are in P or NP-complete. In this talk I will give a short introduction to a universal-algebraic approach to this conjecture, including a description of a conjecture about the boarder betwen NP-complete and polynomial-time tractable CSPs. We also present a universal-algebra description of those CSPs that can be solved by a Datalog program. Finally, we give an outlook on how these techniques can be used to study the complexity of (well-behaved) classes of CSPs where the domain is infinite, which creates a link to the subsequent talk of Florent Madelaine on the logic of MMSNP.

3.6 Towards Efficient Reasoning Under Guarded-based Disjunctive Existential Rules

Pierre Bourhis (ENS – Cachan, FR)

Joint work of Bourhis, Pierre; Morak, Michael, Pieris Andreas
Main reference P. Bourhis, M. Morak, A. Pieris, "The Impact of Disjunction on Query Answering Under Guarded-Based Existential Rules," in Proc. of the 23rd Int'l Joint Conf. on Artificial Intelligence (IJCAI'13), IJCAI/AAAI, 2013.
URL http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6504

The complete picture of the complexity of answering (unions of) conjunctive queries under the main guarded-based classes of disjunctive existential rules has been recently settled. It has been shown that the problem is very hard, namely 2 Exptime-complete, even for fixed sets of rules expressed in lightweight formalisms. This gives rise to the question whether its complexity can be reduced by restricting the query language.

Several subclasses of conjunctive queries have been proposed with the aim of reducing the complexity of classical database problems such as query evaluation and query containment. Three of the most prominent subclasses of this kind are queries of bounded hypertree-width, queries of bounded treewidth and acyclic queries.

The central objective of the talk is to understand what whether the above query languages have a positive impact on the complexity of query answering under the main guarded-based classes of disjunctive existential rules.

We show that (unions of) conjunctive queries of bounded hypertree-width and of bounded treewidth do not reduce the complexity of our problem,

Regarding acyclic queries, although our problem remains 2Exptime-complete in general, in some relevant settings the complexity reduces to Exptime-complete; in fact, this requires to bound the arity of the predicates, and for some expressive guarded-based formalisms, to fix the set of rules.

3.7 On the Data Complexity of Consistent Query Answering over Graph Databases

Gaelle Fontaine (University of Chile, CL)

Joint work of Barcelo, Pablo; Fontaine, Gaelle

Graph database applications such as RDF, social networks and scientific databases are prone to inconsistency, mainly due to interoperability issues. This raises the need for understanding query answering over inconsistent graph databases in a framework that is simple yet general enough to accommodate many of its applications. We follow the well-known approach of consistent query answering (CQA), and study the data complexity of CQA over graph databases for the commonly used regular path queries (RPQs) and regular path constraints (RPCs). In this talk we will present the main complexity results and compare those results to the ones obtained in the setting of relational databases.

3.8 SPARQL 1.1 Entailment Regimes

Birte Glimm (Universität Ulm, DE)

Main reference B. Glimm, C. Ogbuji, (eds.), "SPARQL 1.1 Entailment Regimes – W3C Recommendation 21 March 2013," W3C, 2013.
URL http://www.w3.org/TR/sparql11-entailment/

SPARQL is a semantic web query language and protocol for RDF data standardised by the World Wide Web Consortium in 2008. Since 2013, SPARQL 1.1 extends the original standard by several features such as update capabilities, new query language features, new result formats or the ability to also query for implicit knowledge that can be inferred under RDF(S) or OWL semantics. The talk introduces the SPARQL entailment regimes, which allow for retrieving inferred knowledge. The entailment regimes are specified in a very general way by extending SPARQL's standard query evaluation mechanism (simple entailment/subgraph matching). As a consequence, the standard allows for using different semantics and implementation techniques. While the resulting query language is very expressive, it also lacks some commonly used features such as existentially quantified variables as known from conjunctive queries. The main focus of the talk is on the OWL Direct Semantics entailment regime (based on Description Logics) and some optimisations for query evaluation.

3.9 Non-monotonicity in Data Exchange and Ontological Reasoning

André Hernich, (University of Liverpool, GB)

The semantics of a set E of existential rules (a.k.a. tuple-generating dependencies) is typically defined as the set of all its models. The answer to a query over E is then defined as the set of all tuples that are answers to the query in all models of E. While this notion of query answer is quite robust for positive queries (e. g., it has several natural alternative characterizations, including a procedural one in terms of the chase), it is not so clear whether this semantics is the "right" one for queries that involve negation, or if we allow existential rules extended with negative body literals. Indeed, in the context of data exchange where existential rules serve as a specification of how to translate a given source database into a target database, it has been argued that the model-theoretic, or open world based semantics leads to counter-intuitive answers for non-monotone queries. This triggered several proposals of alternative semantics based on variants of the closed world assumption. Furthermore, existential rules extended with negative body-literals, and corresponding semantics have been studied recently in the context of ontology-based data access.

In this talk, I'll discuss ways of how to deal with negation in queries and/or existential rules. The first part of the talk focuses on closed-world approaches from data exchange, whereas the second part will deal with existential rules extended with negative body literals.

3.10 Beyond DL-Lite: Pay-as-you-go query answering

Ian Horrocks (University of Oxford, GB)

The attractive features of DL-Lite are well known: data can be left in legacy/scalable data stores, and data complexity is low (AC0). However, these benefits come at the cost of a severely constrained ontology language. Many applications seem to require more expressive ontologies – certainly many existing ontologies do not satisfy the relevant constraints. In this talk I will survey techniques for (empirically) scalable query answering in cases where more expressive ontologies are used, focusing on recent work on enhanced materialisation-based techniques.

3.11 Ontology-based data access with Ontop and databases: rewriting and optimisations

Roman Kontchakov (University of London, GB)

Joint work of Kontchakov, Roman; Zakharyaschev, Michael

In the ontology-based data access (OBDA) paradigm, an ontology defines a high- level global schema and provides a vocabulary for user queries, thus isolating the user from the details of the structure of data sources (which can be relational databases, triple stores, datalog

engines, etc.). The OBDA system transforms user queries into the vocabulary of the data and then delegates the actual query evaluation to the data sources.

In this talk, we focus on the rewriting algorithm and optimisation techniques implemented in the OBDA system Ontop in the context of relational databases [8, 6]. In particular, we discuss the tree- witness rewriting, which considers all possible ways of splitting the query into fragments that are mapped onto the elements present in the database and the labelled nulls derived by the axioms of the ontology; the latter fragments are called *tree- witnesses* [5, 7].

From the theoretical point of view, tree witnesses over *OWL 2 QL* ontologies give rise to exponential UCQ rewritings and can be used to obtain exponential lower bounds even for non-recursive datalog rewritings [1, 2, 3, 4]. We also remark that over *OWL 2 EL* ontologies, the same approach results in (recursive) datalog rewritings of polynomial size [7].

In practical scenarios, it appears that the tree witnesses do not produce too many choices and the size of the rewriting can be significantly reduced by taking account of the integrity constraints (in particular, inclusion dependencies) from datasources. These observations confirm that the more general Semantic Query Optimisation methods are important for making OBDA systems efficient [8].

References

1 Bienvenu, M., Kikot, S., and Podolskii, V. V. Succinctness of query rewriting in OWL 2 QL: the case of tree-like queries. In *Informal Proc. of the 27th Int. Workshop on Description Logics (DL 2014)*, volume 1193 of CEUR-WS, pages 45–57, 2014.

2 Gottlob, G., Kikot, S., Kontchakov, R., Podolskii, V. V., Schwentick, T., and Zakharyaschev, M. The price of query rewriting in ontology-based data access. *Artif. Intell.*, 213:42–59, 2014.

3 Kikot, S., Kontchakov, R., Podolskii, V., and Zakharyaschev, M. Exponential lower bounds and separation for query rewriting. In *Proc. of the 39th Int. Colloquium on Automata, Languages, and Programming (ICALP 2012), Part II*, volume 7392 of *LNCS*, pages 263–274. Springer, 2012.

4 Kikot, S., Kontchakov, R., Podolskii, V., and Zakharyaschev, M. On the succinctness of query rewriting over shallow ontologies. In *Proc. of the 23rd EACSL Conf. on Computer Science Logic and the 29th ACM/IEEE Symposium on Logic in Computer Science (CSL-LICS 2014)*, pages 57:1–57:10. ACM, 2014.

5 Kikot, S., Kontchakov, R., and Zakharyaschev, M. Conjunctive query answering with OWL 2 QL. In *Proc. of the 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2012)*, pages 275–285. AAAI Press, 2012.

6 Kontchakov, R., Rezk, M., Rodríguez-Muro, M., Xiao G., and Zakharyaschev, M. Answering SPARQL queries under the OWL 2 QL entailment regime with databases. In *Proc. of the 13th Int. Semantic Web Conf. (ISWC 2014), Part I*, volume 8796 of *LNCS*, pages 552–567. Springer, 2014.

7 Kontchakov, R., and Zakharyaschev, M. An Introduction to description logics and query rewriting. In *The 10th Reasoning Web Summer School*, volume 8714 of *LNCS*, pages 195–244. Springer, 2014.

8 Rodríguez-Muro, M., Kontchakov, R., and Zakharyaschev, M. (2013). Ontology-based data access: Ontop of databases. In *Proc. of the 12th Int. Semantic Web Conf. (ISWC 2013)*, volume 8218 of *LNCS*, pages 558–573. Springer, 2013.

3.12 Schema-Agnostic Query Rewriting in SPARQL 1.1

Markus Krötzsch (TU Dresden, DE)

Joint work of Bischof, Stefan; Krötzsch, Markus; Polleres, Axel; Rudolph, Sebastian
Main reference S. Bischof, M. Krötzsch, A. Polleres, S. Rudolph, "Schema-Agnostic Query Rewriting in SPARQL 1.1," in Proc. of the 13th Int'l Semantic Web Conf. (ISWC'14), LNCS, Vol. 8796, pp. 585–600, Springer, 2014; pre-print available from author's webpage.
URL http://dx.doi.org/10.1007/978-3-319-11964-9_37
URL http://korrekt.org/page/Schema-Agnostic_Query_Rewriting_in_SPARQL_1.1

SPARQL 1.1 supports the use of ontologies to enrich query results with logical entailments, and OWL 2 provides a dedicated fragment OWL QL for this purpose. Typical implementations use the OWL QL schema to rewrite a conjunctive query into an equivalent set of queries, to be answered against the non-schema part of the data. With the adoption of the recent SPARQL 1.1 standard, however, RDF databases are capable of answering much more expressive queries directly, and we ask how this can be exploited in query rewriting. We find that SPARQL 1.1 is powerful enough to "implement" a full-fledged OWL QL reasoner in a single query. Using additional SPARQL 1.1 features, we develop a new method of schema-agnostic query rewriting, where arbitrary conjunctive queries over OWL QL are rewritten into equivalent SPARQL 1.1 queries in a way that is fully independent of the actual schema. This allows us to query RDF data under OWL QL entailment without extracting or preprocessing OWL axioms.

3.13 Infinite CSP and dichotomy : a quest for nice logics with dichotomy

Florent R. Madelaine (Clermont University, FR)

In this talk, we recall 2 logics MMSNP and MMSNP2 that have been studied in the context of finite domain Constraint Satisfaction Problems (CSP), that correspond naturally to 2 Ontology Based Data Access query languages (cf. talk of Frank Wolter).

The first one, MMSNP, was introduced by Feder and Vardi in their influential 93 paper in which they motivated the CSP dichotomy conjecture. Though too expressive to capture only finite domain CSP, it is known to be Ptime equivalent to the class of finite domain CSP. Consequently, any MMSNP problem would also be either in Ptime or NP-complete, if the conjecture holds. The logic MMSNP enjoys a simple combinatorial definition in terms of forbidden patterns problems via vertex coloured obtructions, and via some suitable normal form it is possible to decide effectively interesting properties such as containment.

The second one, MMSNP2, extends MMSNP. The original definition was combinatorial and involved an extension of forbidden patterns problems via edge coloured obstructions. An equivalent and more logically flavoured definition is to define MMSNP2 as GMSNP where guarded second order quantification is used in lieu of monadic quantification. It is open whether this logic enjoys a dichotomy.

Both logics fall within the framework of CSP with infinite domains, given by a template that is omega-categorical, a framework extensively studied by Manuel Bodirsky et al. (cf. his talk). This could be used to attack the above open question.

3.14 All–instances termination of chase is undecidable

Jerzy Marcinkowski (University of Wroclaw, PL)

Joint work of Gogacz, Tomasz; Jerzy Marcinkowski, Jerzy
Main reference T. Gogacz, J. Marcinkowski, "All-instances Termination of Chase is Undecidable," in Proc. of the 41st Int'l Colloquium on Automata, Languages, and Programming (ICALP'14), LNCS, Vol. 8573, pp. 293–304, Springer, 2014.
URL http://dx.doi.org/10.1007/978-3-662-43951-7_25

We show that all–instances termination of chase is undecidable. More precisely, there is no algorithm deciding, for a given set $\mathcal{T}$ consisting of Tuple Generating Dependencies (a.k.a. Datalog$^{\exists}$ program), whether the $\mathcal{T}$-chase on D will terminate for every finite database instance D. Our method applies to Oblivious Chase, Semi-Oblivious Chase and – after a slight modification – also for Standard Chase. This means that we give a (negative) solution to the all–instances termination problem for all version of chase that are usually considered.

3.15 Efficient Separability of Regular Languages by Subsequences and Suffixes

Wim Martens (Universität Bayreuth, DE)

Main reference W. Czerwinski, W. Martens, T. Masopust, "Efficient Separability of Regular Languages by Subsequences and Suffixes," in Proc. of the 40th Int'l Colloquium on Automata, Languages, and Programming (ICALP'13), LNCS, Vol. 7966, pp. 150–161, Springer, 2013.
URL http://dx.doi.org/10.1007/978-3-642-39212-2_16

When can two regular word languages K and L be separated by a simple language? We investigate this question and consider separation by piecewise- and suffix-testable languages and variants thereof. We give characterizations of when two languages can be separated and present an overview of when these problems can be decided in polynomial time if K and L are given by nondeterministic automata.

3.16 Query Rewriting and Optimization for Ontological Databases

Giorgio Orsi (University of Oxford, GB)

Joint work of Gottlob, Georg; Orsi, Giorgio; Pieris, Andreas
Main reference G. Gottlob, G. Orsi, A. Pieris, "Query Rewriting and Optimization for Ontological Databases," ACM Transactions on Database Systems (TODS), 39(3):25:1–25:46, 2014.
URL http://dx.doi.org/10.1145/2638546

Ontological queries are evaluated against a knowledge base consisting of an extensional database and an ontology (i. e., a set of logical assertions and constraints which derive new intensional knowledge from the extensional database), rather than directly on the extensional database. The evaluation and optimization of such queries is an intriguing new problem for database research.

In this talk, we discuss two important aspects of this problem: query rewriting and optimisation [1]. Query rewriting consists of the compilation of an ontological query into

an equivalent first-order query against the underlying extensional database. We present a set of optimization techniques for query rewriting under rather general classes of ontological constraints which is well-suited for practical implementations. In particular, we show how a conjunctive query against a knowledge base, expressed using linear and sticky existential rules, that is, members of the recently introduced Datalog+/- family of ontology languages, can be compiled into a union of conjunctive queries (UCQ) against the underlying database. Ontological query optimisation, in this context, attempts to improve this rewriting process so to produce possibly small and cost-effective UCQ rewritings for an input query.

This is a joint work with Georg Gottlob and Andreas Pieris. The prototype IRIS+/- engine is available at: https://bitbucket.org/giorsi/nyaya.

References

1 Georg Gottlob, Giorgio Orsi, and Andreas Pieris. *Query Rewriting and Optimization for Ontological Databases*. ACM Transactions on Database Systems (TODS), 39(3):25:1–25:46, 2014.

3.17 Decidable Logics for Managing Change in Graph Databases

Magdalena Ortiz (TU Wien, AT)

Joint work of Ahmetaj, Shqiponja; Calvanese, Diego; Ortiz, Magdalena; Simkus, Mantas
Main reference S. Ahmetaj, D. Calvanese, M. Ortiz, M. Simkus, "Managing Change in Graph-Structured Data Using Description Logics," in Proc. of the 28th AAAI Conf. on Artificial Intelligence (AAAI'14), pp. 966–973, AAAI, 2014.
URL http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8238

We study different reasoning problems relevant for managing data that evolves as a result of operations (carried out by users or applications). The problems range from ensuring the satisfaction of a given set of integrity constraints after a given sequence of updates, to deciding the (non-)existence of a sequence of actions that would take the data to an (un)desirable state, starting either from a specific data instance or from an incomplete description of it. Our goal is to identify settings that are expressive enough for realistic application domains, yet are computationally manageable and allow for decidable reasoning. To this aim, we focus on graph-databases, understood broadly as relational databases restricted to unary and binary predicates, and rely on decidable fragments of first order predicate logic whose computational properties are well understood, such as description logics and the two variable fragment of first order predicate logic. We consider an elegant action language in which actions are finite sequences of insertions and deletions of nodes and labels, and use the mentioned decidable logics for describing integrity constraints and (partial) states of the data. We then formalize the data management problems mentioned above as a static verification problem and several planning problems, and we provide algorithms and tight complexity bounds for the formalized problems.

3.18 Generic finite realisations of binary overlap specifications

Martin Otto (TU Darmstadt, DE)

Main reference M. Otto, "Finite Groupoids, Finite Coverings and Symmetries in Finite Structures," arXiv:1404.4599v2 [math.CO], 2014.
URL http://arxiv.org/abs/1404.4599v2

We present a generic construction of finite hypergraphs and relational structures based on reduced products with Cayley graphs of suitable groupoids. To this end we need to obtain groupoids whose Cayley graphs have large girth not just in the usual graph-theoretic sense, but with respect to the length of cycles formed by cosets w.r.t. subsets of generators. Reduced products with such Cayley graphs are sufficiently generic to provide highly symmetric and locally acyclic finite realisations of specifications given in terms of binary local overlap patterns (like tgd-type extension requirements or guarded bisimulation types). These in turn yield corresponding hypergraph coverings and finite model constructions as well as a new proof of Herwig-Lascar style extension properties for partial isomorphisms that lift local to global symmetries. Suitable finite coverings of controlled acyclicity and Herwig-Lascar extension properties have previously found several applications in the finite model theory of guarded logics. Of special interest here are: finite model properties for guarded logics, finite controllability of conjunctive queries w.r.t. guarded constraints, and characterisation theorems for guarded logics in the style of classical preservation properties that also work in restriction to just finite models.

3.19 Certain Answers to Well-Designed SPARQL Queries

Reinhard Pichler (TU Wien, AT)

The need to answer queries over many possible worlds arises in many different settings such as data integration/exchange or in ontology based query answering. The generally agreed approach to query answering in such settings is to compute the so-called certain answers, i. e., answers that one gets by answering the query over any possible world.

In this talk, we study the problem of computing certain answers for the semantic web query language SPARQL. More precisely, we focus on an important fragment of SPARQL – the so-called *well-designed* SPARQL – under OWL2-QL entailment. The main challenge comes from the non-monotonicity of the OPTIONAL operator. We thus first define an intuitive semantics of certain answers based on the subsumption relation – building upon the work of Arenas and Pérez [1]. To actually compute the certain answers of well-designed SPARQL queries under QOL2-QL entailment, we follow the DL-Lite approach of Calvanese et al. [2], i. e.: we start by showing that the certain answers can be obtained from the (in general infinite) canonical model of the RDF graph and the ontology. This fact is then used to show the correctness of a rewriting-based query evaluation procedure.

Acknowledgement. The results presented here stem from unpublished joint work with Shqiponja Ahmetaj, Wolfgang Fischl, Mantas Simkus, and Sebastian Skritek. This research has been funded by the Vienna Science and Technology Fund (WWTF) through project ICT12-015.

References

1 Marcelo Arenas and Jorge Pérez. Querying semantic web data with SPARQL. In *Proc. PODS 2011*, pages 305–316. ACM, 2011.
2 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.

3.20 Polynomial Combined Rewritings for Existential Rules

Andreas Pieris (University of Oxford, GB)

We consider the scenario of ontology-based query answering where a conjunctive query is evaluated against a database enriched with intensional knowledge via an ontology. It is generally accepted that true scalability of query answering in this setting can only be achieved by using standard relational database management systems (RDBMSs). An approach to query answering that enables the use of RDBMSs is the so-called polynomial combined approach. We investigate this approach for the main guarded- and sticky-based classes of existential rules, and we highlight the assumptions on the underlying schema that are sufficient for the polynomial combined first-order rewritability of those classes. To the best of our knowledge, this is the first work which explicitly studies the polynomial combined approach for existential rules.

3.21 A Declarative Constraint-Based Framework for Linking Entities

Lucian Popa (IBM Almaden Center, US)

Joint work of Douglas Burdick; Ronald Fagin; Phokion Kolaitis; Popa, Lucian; Wang-Chiew Tan

We investigate the theoretical foundations towards a declarative framework for entity linking and, in particular, for entity resolution, which is the problem of identifying whether two records represent the same real-world entity. The framework that we adopt is based on link-to-source constraints, unlike in some earlier approaches where source-to-link constraints were used to dictate how to generate links. Our approach makes it possible to focus entirely on the intended properties of the outcome of entity linking, thus separating the specification from any procedure that implements it. The link-to-source constraints specify the desired properties of a link relation in terms of source relations and built-in predicates such as similarity measures. A key feature of the link-to-source constraints is that they employ disjunction, which enables the declarative listing of all the reasons as to why two entities should be linked. We also consider extensions that capture collective entity resolution, by allowing inter-dependence between links. We study the semantics and the computational complexity of this declarative framework, and also make connections to some well-known probabilistic approaches for entity resolution, including ones based on Markov Logic Networks.

This is joint work [1] with Douglas Burdick, Ronald Fagin, Phokion Kolaitis, and Wang-Chiew Tan. Much of this work is motivated from Midas [2, 3], a broader research project at

IBM Research – Almaden, focused on developing the high-level languages and tools needed for the large-scale integration of data from public unstructured sources. Midas has been applied to various scenarios, including a data integration scenario in the financial domain, based on regulatory filings archived by the US Securities and Exchange Commission (SEC). A live demonstration of this application shows the importance of entity linking in such integration.

References

1 D. Burdick, R. Fagin, P. G. Kolaitis, L. Popa and W.-C. Tan. *A Declarative Framework for Linking Entities*. Submitted for publication. 2014.

2 D. Burdick, M. A. Hernández, H. Ho, G. Koutrika, R. Krishnamurthy, L. Popa, I. Stanoi, S. Vaithyanathan and S. R. Das. *Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study*. IEEE Data Eng. Bull., 34(3), pages 60–67, 2011.

3 M. A. Hernández, G. Koutrika, R. Krishnamurthy, L. Popa, and R. Wisnesky. *HIL: A High-Level Scripting Language for Entity Integration*. In EDBT, pages 549–560, 2013.

3.22 Linear programming and integer linear programming for logical satisfiability problems

Ian Pratt-Hartmann (University of Manchester, GB)

In this talk, I will give a brief overview of some of the ways in which satisfiability and finite satisfiability problems for logical fragments have been reduced to linear programming and integer linear programming problems. I will argue that such reductions are not merely a useful source of upper complexity bounds, but also a way of thinking about these logics that provides real insight into their structure.

3.23 Query answering with key constraints: certainty, counting, and probabilities

Dan Suciu (University of Washington – Seattle, US)

Main reference P. Koutris, D. Suciu, "A Dichotomy on the Complexity of Consistent Query Answering for Atoms with Simple Keys," in Proc. of the 17th Int'l Conf. on Database Theory (ICDT'14), pp. 165–176, 2014.
URL http://dx.doi.org/10.5441/002/icdt.2014.19

In order to answer a query on a database that violates key constraints one has to considers all possible ways to repair the database. I will discuss three semantics to query answering: certainty (checking if the query is true in all repairs), counting the number of repairs that satisfy the query, and computing the marginal probability that the query is true. In each case there exists a dichotomy theorem for answering conjunctive queries without self-joins: for every query the data complexity is either in PTIME, or it is NP-complete / #P- complete. (For "certainty" the dichotomy holds only for single-attribute keys.)

References

1 Nilesh Dalvi, Dan Suciu. *Management of probabilistic data: foundations and challenges.* PODS, 2007
2 Maslowski, Wijsen. *A dichotomy in the complexity of counting database repairs.* J. Comput. Syst. Sci. 2013
3 Paris Koutris, Dan Suciu. *A Dichotomy on the Complexity of Consistent Query Answering for Atoms with Simple Keys.* ICDT 2014

3.24 Decidable Fragments of First-Order Logic

Lidia Tendera (University of Opole, PL)

In this talk we give an outline of recent work in the quest for expressive fragments of first-order logics with good algorithmic properties. We concentrate on two related problems: the satisfiability problem and the query answering problem (under the open world assumption). We are equally interested in reasoning in unrestricted and in finite models, as in some applications we want to model systems and computations that are essentially finite.

We focus mainly on fragments defined by restricting the number of variables, usage of quantifiers or usage of negation, and on most popular classes of queries.

Presenting the material we highlight most effective techniques used in this context, their advantages and limitations. We also point out some future directions of study.

3.25 The Automata/Logic Connection for Expressive Guarded Logics

Michael Vanden Boom (University of Oxford, GB)

Joint work of Vanden Boom, Michael; Benedikt, Michael; ten Cate, Balder; Colcombet, Thomas

The guarded fragment (GF) and guarded negation fragment (GNF) are expressive logics, capturing many query classes and integrity constraints of interest in databases and knowledge representation. In this talk, we consider the fixpoint extensions GFP and GNFP of these logics, which are known to have decidable satisfiability and other nice computational and model-theoretic properties.

In particular, these logics have tree-like models, which make them amenable to tree automata techniques. We describe automata that can be used to show optimal complexity bounds for satisfiability. We also describe recent unpublished work demonstrating how these automata can be adapted to decide the following boundedness problem: given phi(X,x) in GFP or GNFP and positive in X, is there a natural number n such that the least fixpoint of the operator defined by phi is reached within n iterations over all structures?

3.26 On Reasoning about Duplicates over Range Restricted Queries with Bag Semantics

Grant Weddell, (University of Waterloo, CA)

We review earlier work on reducing reasoning about duplicate elimination in queries over an object relational data source to reasoning about knowledge base consistency in a dialect of the CFD family of description logics, logics in which unary function symbols replace binary predicate symbols in underlying signatures. We also give an overview of how such capabilities are used in a more general framework for query compilation based on Craig interpolation.

3.27 Presentations, Invariance, and Definability

Scott Weinstein (University of Pennsylvania, US)

Joint work of Weinstein, Scott; Lindell, Steven

We develop the notion of a presentation for abstract relational structures and investigate presentation invariant elementary definability on classes of finite structures. In particular, we focus on linear orderings induced by traversals of simple graphs, and show how they allow us to strictly extend the power of first- order logic in a natural fashion (defining reachability for example). We show how this leads to a new descriptive characterization of logspace in terms of traversal- invariant elementary definability. Continuing, we demonstrate how an elementary partial order presentation can be used to give a very simple definition of tree-width, providing a new normal form for tree decompositions which is especially compelling. We explore algorithmic aspects of this presentation.

3.28 Two-Variable Logic on 2-Dimensional Structures

Thomas Zeume (TU Dortmund, DE)

Joint work of Manuel, Amaldev; Schwentick, Thomas; Zeume, Thomas

In this talk results for the finite satisfiability problem of the two-variable fragment of first-order logic (FO2) over two-dimensional structures have been discussed. Here, two-dimensional structures are structures with two orders, their induced successor relations and arbitrarily many unary relations. Two types of orders have been considered: linear orders and preorders (i. e. equivalence relations whose equivalence relations are linearly ordered).

Participants

- Isolde Adler
Goethe-Universität Frankfurt am Main, DE
- Antoine Amarilli
Telecom ParisTech, FR
- Molham Aref
LogicBlox – Atlanta, US
- Marcelo Arenas
Pontificia Universidad Catolica de Chile, CL
- Michael Benedikt
University of Oxford, GB
- Meghyn Bienvenu
University Paris South, FR
- Manuel Bodirsky
Ecole Polytechnique – Palaiseau, FR
- Pierre Bourhis
ENS – Cachan, FR
- Cristina Civili
University of Rome "La Sapienza", IT
- Gaelle Fontaine
University of Chile, CL
- Birte Glimm
Universität Ulm, DE
- Tomasz Gogacz
University of Wroclaw, PL
- André Hernich
University of Liverpool, GB
- Ian Horrocks
University of Oxford, GB
- Yazmin Angélica Ibáñez-García
Free University of Bozen-Bolzano, IT
- Emanuel Kieronski
University of Wroclaw, PL
- Phokion G. Kolaitis
University of California – Santa Cruz, US
- Roman Kontchakov
Birkbeck, Univ. of London, GB
- Tomer Kotek
TU Wien, AT
- Markus Krötzsch
TU Dresden, DE
- Katja Losemann
Universität Bayreuth, DE
- Carsten Lutz
Universität Bremen, DE
- Florent R. Madelaine
Clermont University, FR
- Jerzy Marcinkowski
University of Wroclaw, PL
- Wim Martens
Universität Bayreuth, DE
- Jakub Michaliszyn
Imperial College London, GB
- Giorgio Orsi
University of Oxford, GB
- Magdalena Ortiz
TU Wien, AT
- Martin Otto
TU Darmstadt, DE
- Reinhard Pichler
TU Wien, AT
- Andreas Pieris
University of Oxford, GB
- Lucian Popa
IBM Almaden Center, US
- Ian Pratt-Hartmann
University of Manchester, GB
- Sebastian Rudolph
TU Dresden, DE
- Vadim Savenkov
TU Wien, AT
- Inanc Seylan
Universität Bremen, DE
- Dan Suciu
University of Washington – Seattle, US
- Balder Ten Cate
University of California – Santa Cruz, US
- Lidia Tendera
University of Opole, PL
- Michaël Thomazo
TU Dresden, DE
- David Toman
University of Waterloo, CA
- Michael Vanden Boom
University of Oxford, GB
- Grant Weddell
University of Waterloo, CA
- Scott Weinstein
University of Pennsylvania, US
- Frank Wolter
University of Liverpool, GB
- Michael Zakharyaschev
Birkbeck, Univ. of London, GB
- Thomas Zeume
TU Dortmund, DE