UNIVERSITÄT
LEIPZIG

Fakultät für Mathematik und Informatik
Institut für Informatik
Wissensrepräsentation
Prof. Dr. Carsten Lutz

Universität Leipzig, Institut für Informatik, 04081 Leipzig

Report on PhD thesis of Albert Gutowski 07.12.2025

Universität Leipzig
Fakultät für Mathematik und Informatik
Institut für Informatik
Augustusplatz 10
04109 Leipzig

Telefon
+49 341 97-32343

Fax
+49 341 97-32299

E-Mail
carsten.lutz@uni-leipzig.de

Web
www.uni-leipzig.de

Postfach intern
212101

With this letter I provide the requested report on the submitted PhD thesis by Albert Gutowski entitled

"Regular Path Queries and Modal Constraints".

One important difference between classical logic and logic in computer science is that the latter often concentrates on finite models. This poses significant technical difficulties because the construction of finite models tends to be rather challenging, often requiring non-trivial combinatorics.

The presented thesis is concerned with such finite model constructions. It offers one main result: *over finite models, UCRPQ containment modulo $ALCO_{reg}$ TBoxes is decidable*. Here, UCRPQ stands for unions of conjunctive regular path queries, an important class of queries situated at the heart of most query languages for graph databases. $ALCO_{reg}$ TBoxes means ontologies formulated in the description logic $ALCO_{reg}$, the extension of the fundamental description logic ALC with nominals and regular expressions over roles.

The most immediate interpretation of the obtained result is thus as a contribution to static query analysis in the area of ontology-mediated querying, where a database is enriched with an ontology that provides domain knowledge. There are, however, also potential applications in graph databases because $ALCO_{reg}$ may be viewed as a natural constraint or schema language for graph databases, and thus the obtained result might also be useful for static analysis in graph databases under constraints / a schema.

The proof of the thesis' main result is rather non-trivial and technically demanding. Because of this, the goal of the thesis is not only to present said proof, but to present it in an as accessible form as possible. To achieve this, the author goes to considerable effort. He decomposes the proof into a series of reductions with the aim of clearly displaying and separating the involved aspects. He starts with the much simpler case of conjunctive queries (CQs) and ALCO TBoxes to first showcase simpler and more digestible versions of the constructions used later for UCRPQs and $ALCO_{reg}$. He transitions to a purely graph theoretic framework based on homomorphisms and bisimulation types early on, to work in a setting that is as simple as possible. And he even seems to choose his main result so that it admits an as clear as possible presentation. Let me expand on the latter.

The technical development in the thesis builds on several articles by the author, along with coauthors: a KR2020 paper that considers finite query entailment for UCQs and the description logics $ALCOI^+$ and $ALCOQ^+$, with + referring to transitive roles; a KR2022 paper that proves decidability of finite query entailment for UCRPQs and ALC TBoxes; and a 2024 paper in PODS/PACMMOD that establishes decidability of finite query containment for UCRPQs and TBoxes formulated in ALCQ and ALCI. While the latter article appears to be most closely related to the content of the thesis, the author leaves out the "Q" (number restrictions) and "I" (inverse roles) parts, supposedly because they make the constructions, which are very complex anyway, even more technical to present. On the other hand, he adds regular expressions on roles to the TBox, so his result is orthogonal to those that can be found in the literature. He also seems to consciously decide to go only for decidability instead of tight upper complexity bounds, also for the reason of achieving a clearer presentation (the algorithm presented in the thesis is non-elemenary while the mentioned articles establish tight 2ExpTime complexity results).

Before describing the content of the individual chapters, let me list the sequence of reductions used in the thesis:

I. Query containment => homomorphism entailment
II. Homomorphism entailment => homomorphism coverage
III. Homomorphism coverage => minimal downset computation
IV. Minimal downset computation => (constructive) satisfiability

Most of these reductions are Turing reductions. All of the above problems come with a crucial parameter: the class of (finite!) graphs that are admitted. This is in fact where the query and TBox language comes in: while for CQs and ALCO it suffices to work with the class of all graphs, UCRPQs and $ALCO_{reg}$ require some carefully annotated (and much more difficult to handle) graphs.

Chapter 1 of the thesis provides an introduction to the studied topic and related results, as well as an overview of the thesis. It is tailored a lot towards finite model constructions. Personally, I would have appreciated a bit more background on querying in description logics, featured prominently in the formulation of the main result, and on graph databases. One almost gets the impression that the author wants to avoid at all cost discussing or even mentioning the extensive literature from those areas.

Chapter 2 introduces the most foundational notions used in the thesis. The author went for an "introduce notions as needed" approach, having Chapter 2 very condensed and introducing other fundamental notions such as UCRPQs and $ALCO_{reg}$ TBoxes only later on when actually used. While this leaves the reader without a central place to look up all of the preliminaries, I nevertheless feel that this works well; it certainly avoids a long and unpleasant to read preliminaries section. The carefully built index also helps.

The short Chapter 3 is concerned with Reduction I from the list above, for the case of CQs, ALCO TBoxes, and the corresponding class of all graphs.

Chapter 4 presents Reduction II for the class of all graphs, thus corresponding to the CQ/ALCO case. The construction prepares for the general case where graphs are annotated, but the proofs are simpler here. The main idea is to partition the graph into an inner part, which intuitively is the homomorphic image of the left-hand query, and an outer part. The inner part is then treated by explicit enumeration in the reduction while the outer part is treated in the homomorphism coverage instances produced by the reduction.

Chapter 5 presents Reduction III, for all classes of graphs. It also introduces downsets and their minimal version, as well as certain lemmas and constructions that are used in the remainder of the thesis: the one-step bisimulation lemma, ravelings, and trace-based graph transformations.

Chapter 6 then provides Reduction IV for the case of all graphs. This finishes the "showcase" proof that finite CQ containment relative to an ALCO TBox is decidable because the (constructive) satisfiability problem for the class of all graphs, because it is also shown that every `yes' instance of the satisfiability is witnessed by a modal of bounded size. The chapter introduces a notion of global representatives of modal profiles, provides a combinatorial construction of raveled graphs, and shows how to obtain bounded-size witnesses for the satisfiability problem. While not the only non-trivial construction, the raveling construction arguably is the core of the decidability proof for query containment.

Chapter 7 constitutes another intermediate step before addressing the main result of the thesis. It introduces reachability-annotated graphs, which (in a refined form) play a role in the rest of the thesis, and establishes the crucial Reduction IV for them, again as a showcase. An important difference to Chapter 6 is that global representatives need not exist, and the authors propose SCC-reachability annotations to deal with that. This makes both the reduction and the solution of the satisfiability problem considerably more technical and tedious.

The substantial Chapter 8 then introduces ranked graphs in which each edge is annotated with a positive integer, its rank. The reachability annotations and SCC-reachability annotations from Chapter 7 are also still in the picture. Then Reduction IV and a corresponding version of the satisfiability problem are solved, again being more subtle and challenging than the previous version. Crucial ingredients include a last edge appearance record for the reduction and universal topological orders for the satisfiability.

Chapter 9 connects the ranked reachability machinery to finite automata, introducing graph classes that incorporate automaton-state information as annotations. One centerpiece is the construction of permutation semiautomata built using first appearance records, which induce a notion of "rank" on transitions. The chapter then goes on to showing that the minimal downset problem and the satisfiability problem are decidable for automaton-state annotated graphs, by reduction to the corresponding version of these problems studied in Chapter 8. In essence, this chapter thus presents a "lifting" step, porting the results from Chapter 8 to the annotated graphs introduced in this chapter.

Chapter 10 revisits Reduction II, but now for the "final" class of graphs introduced in Chapter 9. Intuitively, it is thus the counterpart of 4 for UCRPQs and ALO_{reg} TBoxes in place of CQs and ALCO TBoxes. The proof strategy is similar, yet the details are more involved.

Chapter 11 states and proves the main theorem again and completes the full reduction chain. It also provides a summary of the full algorithm as a pipeline of reductions and explains where the major complexity blowups come from. In addition, it discusses modifications of the algorithms that might serve to reduce its time complexity to elementary, possibly obtaining even the optimal 2ExpTime upper bound. The modifications are non-trivial and while it sounds as if working out the details is only a matter of effort and routine, it does not become entirely clear whether this is really the case. To be fair, though, the author does not formally claim the improved result and it is appreciated that he lays out the envisioned modifications in quite some detail.

Chapter 12 concludes and briefly explains how it might be possible to extend the presented proofs to qualified number restrictions and inverse roles (separately).

The thesis is written in a very thoughtful way. As already explained above, even the main result seems to have been chosen (no Q, no I, no optimal complexity) to enable a didactic and "digestible" presentation. The notation is chosen extremely carefully as well and used very consistently. I was not able to spot any typos. There are overviews at the beginning of most chapters, and the notions relevant for the chapter are recalled. At the end of chapters, there are very helpful summaries. Examples are crafted in an extremely careful way and help a lot to understand the intricate constructions. They are even designed to be readable with a black and white printout while colors are user to further support the reader who can see them. All of this must have been a significant amount of work, and the author deserves appreciation for taking it on. The result is a highly readable exposition of an intricate and highly combinatorial proof. Only in Chapters 9 and 10 I had the feeling that the guidance through the chapters could have been even more explicit; but this is really nitpicking.

I deem the submitted thesis as sufficient to grant a PhD.

With best regards,

Carsten Lutz

Carsten Lutz