Master's Internship

XPath-like query logics: a proof theoretical complexity analysis

Advisors and location

Sylvain Schmitz
Department of Computer Science
University of Warwick
Coventry, UK

David Baelde
Laboratoire Spécification et Vérification
École Normale Supérieure de Cachan
Cachan, France

The internship may take place in Warwick, or be co-located between Warwick and Cachan.

Context

XPath is a widely employed language for selecting XML elements. The two main computational problems associated with XPath are

querying: given an XML document and an XPath query, return the set of nodes selected by the query;

satisfiability: given an XPath query (and optionally a schema for the kind of XML documents we want to consider), does there exist an XML document on which this query would select at least one node?

Both problems have been studied in depth for many fragments of XPath [1]. When considering queries without *data tests*, these problems can be tackled using tree automata techniques, and we are on well-trodden grounds. Adding data tests, i.e. the ability to check whether two attributes of two XML elements are the same or not, complicates matters considerably. Nevertheless, several results already give a good picture:

- XPath queries can be resolved in linear time (in the document size) for a large fragment [2], and
- depending on the *navigation* primitives allowed in the fragment, the complexity of the satisfiability problem ranges from ExpSpace to Ackermannian to undecidable [4].

A Different Approach

An issue with the results on the complexity of satisfiability is that each fragment has its own decidability proof, using either some form of a *data tree automaton* or "model massaging" techniques. We hope to unify these results, by relying on a more syntactic approach based on proof systems for data logics.

As a first step in that direction, we propose to study a modal logic that can be seen as a small but interesting fragment of XPath. The first goal of the internship will be to **define a sequent calculus** that is sound and complete for the query logic. The purpose here is different from that of previous work on axiomatisations of XPath [3], in that data should be handled, and in that the proof system should be sufficiently well structured to be helpful for the complexity analysis of provability by means of proof search. This is the second goal: **prove decidability and provide complexity bounds** for provability in the query logic, by applying classical proof search arguments such as (a variant of) the subformula property, and a monotonicity property, which would allow to re-use techniques from infinite-state systems verification [6, 5].

This work is a first step in a more ambitious programme that would be a natural topic for a PhD. There is funding for such a PhD at ENS Cachan, under the supervision of Sylvain Schmitz and as part of the ANR project Prodaq (starting in 2015), whose purported aim is the investigation of the proof theory and complexity of data logics.

Prerequisites

This is a broad topic that encompasses techniques from different communities. The student needs to know some basics in complexity and logic. Some exposure to sequent calculus and modal logic, and having followed MPRI course 2.9.1 are a plus, but the internship is also a great opportunity to learn the necessary background.

Contact

Sylvain Schmitz	David Baelde
Phone: (+33/0) 147 407 542	Phone: (+33/0) 147 407 780
Email: schmitz@lsv.ens-cachan.fr	Email: baelde@lsv.ens-cachan.fr

References

[1] Benedikt, M. and Koch, C., 2009, XPath leashed. *ACM Computing Surveys*, 41 (1):3. doi:10.1145/1456650.1456653.

[2] Bojańczyk, M. and Parys, P., 2011, XPath evaluation in linear time. *Journal of the ACM*, 58(4):17. doi:10.1145/1989727.1989731.

[3] ten Cate, B., Litak, T., and Marx, M., 2010, Complete axiomatizations for XPath fragments *Journal of Applied Logic*, 8(2):153–172. doi:10.1016/j.jal.2009.09.002.

[4] Figueira, D., 2013, On XPath with transitive axes and data tests. In *PODS 2013*, pages 249–260 doi:10.1145/2463664.2463675.

[5] Lazić, R. and Schmitz, S., 2014, Non-Elementary Complexities for Branching VASS, MELL, and Extensions. In *CSL-LICS 2014*, article 61 http://arxiv.org/abs/1401.6785.

[6] Schmitz, S. and Schnoebelen, Ph., 2012, *Algorithmic Aspects of WQO Theory*. Lecture Notes. http://cel.archives-ouvertes.fr/cel-00727025.