

Symbolic Automata Algorithms with Applications

Summary

Symbolic finite automata is a subdomain of computer science where recent advances of application of symbolic derivatives have allowed to redefine the foundations of well established theories to provide better insight into the nature of the problems. One recent application of symbolic automata that have made it into the applications of everyday programmers is the non-backtracking regular expression matching engine available in .Net 7.0. There is a variety of challenges that need to be tackled in order to provide more efficient and secure applications that are based on the symbolic derivatives theory that is currently under active scrutiny among many researchers in the field.

Research field:	Information and communication technology
Supervisor:	Prof. Dr. Juhan-Peep Ernits
Availability:	This position is available.
Offered by:	School of Information Technologies Department of Software Science
Application deadline:	Applications are accepted between October 01, 2022 00:00 and October 23, 2022 23:59 (Europe/Zurich)

Description

The PhD project builds on the results of the application of symbolic derivatives for solving extended regular expression constraints as described in [2]. The focus of the work is on enriching the input language of the symbolic derivatives based regular expression matchers with the goal of facilitating building new applications, e.g. either in the direction of security applications [3], where certain aspects of existing regular expression matching engines can be exploited for discovering weaknesses in existing software (e.g. ReDoS attacks), or for the analysis of large code bases [1], where the previous symbolic automata discovery approach could be extended with novel techniques.

References

- [1] Peleg, H., Shoham, S., Yahav, E., and Yang, H. Symbolic automata for representing big code. *Acta Informatica* 53, 4 (2016), 327–356.
- [2] Stanford, C., Veanes, M., and Bjørner, N. S. Symbolic boolean derivatives for efficiently solving extended regular expression constraints. In *PLDI '21: 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, Virtual Event, Canada, June 20–25, 2021 (2021), S. N. Freund and E. Yahav, Eds., ACM, pp. 620–635.
- [3] Turonova, L., Holik, L., Homoliak, I., Lengal, O., Veanes, M., and Vojnar, T. Counting in regexes considered harmful: Exposing ReDoS vulnerability of nonbacktracking matchers. In *31st USENIX Security Symposium (USENIX Security 22)* (Boston, MA, Aug. 2022), USENIX Association, pp. 4165–4182.

Responsibilities and (foreseen) tasks:

- Use .Net 7 nonbacktracking regular expression library as a research tool to experiment with extensions to the current symbolic derivatives engine.
- Construct case studies involving the applications of symbolic derivatives and evaluate and document them as publications.
- Model systems in terms of constraints involving sequence theory and regular expressions theory (as available in the Z3 SMT solver).
- Contribute to the organization of relevant conferences, workshops organized by the department.
- Support teaching activities of the research group.

Applicants should fulfil the following requirements:

- a master's degree in computer science, informatics and/or artificial intelligence
- a clear interest in contributing to advancing the field of computer science
- excellent command of English
- strong and demonstrable writing and analytical skills
- capacity to work both as an independent researcher and as part of an international team
- capacity and willingness to provide assistance in organizational tasks relevant to the project

The following experience is beneficial:

- Understanding of the core concepts of logic in computer science
- Very good programming skills (e.g. C#, F#)
- Working knowledge automata theory
- Working knowledge of language theory
- Working knowledge of decision procedures and their applications in satisfiability modulo theories solvers

The candidate should submit a research plan for the topic, including the overall research and data collection strategy. The candidate can expand on the listed research questions and tasks, and propose theoretical lenses to be used.

We offer:

- 4-year PhD position in one of the largest, most internationalized and leading computer science research centers in Estonia with a large portfolio of ongoing pan-European and national research and development projects.
- The chance to do high-level research in collaboration with the best specialists in the field
- Opportunities for conference visits, research stays and networking with globally leading universities and research centers in the fields of computer science.

About the department

The mission of the Department of Software Science is to advance internationally and nationally relevant state of the art in research and apply it in bachelor, MSc and doctoral education in the areas of computer science, information systems, data science, artificial intelligence and cyber security with the goal to solve problems the society is facing and support sustainable development.

Additional information

For further information, please contact Juhan Ernits (PhD) juhan.ernits@taltech.ee or visit <https://taltech.ee/tark-varateaduse-instituut>.



To get more information or to apply online, visit <https://taltech.glowbase.com/positions/591> or scan the the code on the left with your smartphone.