



Review report of a final thesis

Student: Martin Hron
Reviewer: Ing. Jan Trávníček, Ph.D.
Thesis title: Backreferences in practical regular expressions
Branch of the study: Computer Science

Date: 7. 6. 2020

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
1. Fulfilment of the assignment	1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled
<i>Criteria description:</i> Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.	
<i>Comments:</i> The thesis topic is unarguably important and interesting. The research section of the thesis is broad and implemented approaches were wisely selected. The evaluation section is also very thorough.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
2. Main written part	99 (A)
<i>Criteria description:</i> Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies? Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 26/2017, Art. 3. Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.	
<i>Comments:</i> Throughout the text, there are only few notable issues: NFAs are defined as epsilon-NFAs, which might be a bit confusing. Definition 4.1 is using undefined terms $o(x)$ and $c(x)$ that can be later understood from the context, however, they should be explained within the definition or sooner. Page 66, line 30 should be "greater than or equal to a".	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
3. Non-written part, attachments	100 (A)
<i>Criteria description:</i> Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.	
<i>Comments:</i> The implementation is solid, easy to read, and understandable. I have not identified any issues within the code.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
4. Evaluation of results, publication outputs and awards	100 (A)
<i>Criteria description:</i> Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.	
<i>Comments:</i> The implementation of regex matcher through simulation of memory automata does not reach the efficiency of some other regex matcher implementations. However, the implementation seems to be immune to catastrophic backtracking in contrast to all other regex matcher implementations.	

Evaluation criterion:

No evaluation scale.

5. Questions for the defence

Criteria description:

Formulate questions that the student should answer during the Presentation and defence of the FT in front of the SFE Committee (use a bullet list).

Questions:

Naive backtracking algorithm is prone to infinite loops owing to epsilon-transition loops. The case exists in general, however, MFA constructed for a regex pattern is a special case. Can the patterns for MFA construction, used in Thompson-like algorithm, be modified to produce MFA without epsilon-transition loops? If so, maybe the backtracking algorithm would not have to memorize already visited configurations.

Have you considered other graph searching algorithms that might beat the backtracking, for instance, some traversals guided with heuristics?

With the MFA simulated in its nondeterministic form and strong relation between the regex pattern AST and nondeterministic MFA, couldn't the implementation avoid construction of MFA completely and only simulate its behaviour on the regex pattern AST?

Elaborate whether epsilon-transition removal would anyhow affect execution performance.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. The overall evaluation

99 (A)

Criteria description:

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.

Comments:

All in all, I evaluate the quality of both the thesis text and implementation as excellent. I recommend the thesis for defence and I propose to evaluate it with grade A (excellent). I would also like to propose the thesis for the Dean's award.

Signature of the reviewer: