



Supervisor's statement of a final thesis

Student: Bc. Minh Trieu Quang
Supervisor: Ing. Jan Trávníček, Ph.D.
Thesis title: Enhanced suffix arrays implementation and its usage
Branch of the study: Computer Science

Date: 8. 6. 2020

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
1. Fulfilment of the assignment	<u>1 = assignment fulfilled,</u> 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 task was to study an enhanced suffix array as a replacement of a suffix tree to gain performance improvements from more succinct and memory local representation. The suffix array is shown to outperform the suffix tree in the literature and a reference implementation is therefore needed. Representants of algorithms utilising various traversals on a suffix tree were identified and implemented on enhanced suffix arrays. The performance of these algorithms was compared when executed on the suffix tree and the suffix array.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
2. Main written part	95 (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> The definition of a leaf can in some situation match the root. Page 9, line -2, allow -> allows. Page 19, the Boyer-Moore algorithm is stated to have $O(n)$ time complexity of a search phase, however, the worst case is $O(n*m)$. Page 29, the function process is on line 8, not 7. The explanation of algorithms executed on the suffix array would benefit from visualisations to better depict their execution. It feels that some enhancements of the suffix array are closely related to the suffix tree structure. These enhancements would also use some visualisations so that their relation to the suffix trees is more intuitively understandable. The conducted experiments are thorough and compare all algorithms discussed in the thesis. The algorithms are executed on both the enhanced suffix arrays and the suffix trees.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
3. Non-written part, attachments	98 (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.	

Comments:

The implementation is thoroughly tested and well documented.

The implementation itself is a little hard to read since some variable names are just abbreviations (except the common ones i, j, lcp, sa, etc.)

Evaluation criterion:

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

4. Evaluation of results, publication outputs and awards

100 (A)

Criteria description:

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.

Comments:

The resulting implementation can be used as a reference implementation of a suffix array construction and later used in Algorithms library toolkit.

Evaluation criterion:

The evaluation scale: 1 to 5.

5. Activity and self-reliance of the student

5a:
1 = excellent activity,
2 = very good activity,
3 = average activity,
4 = weaker, but still sufficient activity,
5 = insufficient activity
5b:
1 = excellent self-reliance,
2 = very good self-reliance,
3 = average self-reliance,
4 = weaker, but still sufficient self-reliance,
5 = insufficient self-reliance.

Criteria description:

From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations (5a). Assess the student's ability to develop independent creative work (5b).

Comments:

Throughout the student's work on the thesis, the student was diligent and always solved arising issues alone.

Evaluation criterion:

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

6. The overall evaluation

98 (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, the text of the thesis is of high quality, the implementation is thoroughly tested and experimentally evaluated. I recommend evaluating the thesis with grade A (excellent).

Signature of the supervisor: