FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Supervisor's statement of a final thesis

Supervisor:	prof. Ing. Jan Holub, Ph.D.
Student:	Bc. Dominika Draesslerová
Thesis title:	Bioinformatics index tool for elastic degenerate string matching
Branch / specialization:	Computer Science
Created on:	9 February 2023

Evaluation criteria

1. Fulfillment of the assignment

▸ [1] **assignment fulfilled**
[2] assignment fulfilled with minor objections
[3] assignment fulfilled with major objections
[4] assignment not fulfilled

All the tasks of the assignment were fulfilled.

2. Main written part

80 / 100 (B)

The written part could be better. There are problems with English that make reading harder. Some parts should be more self-contained. Some notions should be stated directly in the text instead of referencing the source literature. Some statements should be more precise. The algorithms should also be described verbally in addition to the pseudocode.

3. Non-written part, attachments

90 / 100 (A)

The student solved the problem of EDS indexing using the full-text index BIO-FMI. BIO-FMI was initially implemented as a proof of concept and to evaluate experiments in the paper by Procházka & Holub published at conference DCC2014. The BIO-FMI had to be reimplemented to be stable and to allow all three edit operations: replace, delete, and insert. The initial implementation allowed only the operation replace.
She studied the source formats of biological data and made a tool to generate pseudo-real data for testing.
She also found a way to efficiently implement EDS strings in the BIO-FMI index so that for a given pattern, one can find all occurrences in the pangenome representing a large number of genomes of the same species.
She managed to run alternative indices like LZ-RLBWT and r-index that were not working

out of the box.
Finally, she ran experiments on both real data and pseudo-real comparing BIO-FMI, LZ-RLBWT, and r-index, comparing for EDS and ALN (aligned data) format.

4. Evaluation of results, publication outputs and awards 90 /100 (A)

The student created an efficient tool for indexing pangenome represented as EDS. The results show that EDS is an efficient representation of pangenome (more capable than degenerate strings and faster than variation graphs). The EDS indexing is then a promising tool. More testing on real data is planned. We will prepare a publication afterwards.

5. Activity of the student

▸ [1] **excellent activity**
[2] very good activity
[3] average activity
[4] weaker, but still sufficient activity
[5] insufficient activity

We did meet regularly. The student was always ready and have partial goals fulfilled.

6. Self-reliance of the student

▸ [1] **excellent self-reliance**
[2] very good self-reliance
[3] average self-reliance
[4] weaker, but still sufficient self-reliance
[5] insufficient self-reliance

The student came with her own ideas and some issues she discussed with bioinformaticians.

The overall evaluation 85 /100 (B)

I am happy with the non-written results of the thesis, which is the ground for a nice scientific publication. The written part could be better. Therefore I have to downgrade the evaluation to 85 B.

Instructions

Fulfillment of the assignment

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.

Main written part

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. 52/2021, 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.

Non-written part, attachments

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.

Evaluation of results, publication outputs and awards

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.

Activity of the student

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.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

The overall evaluation

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.