



Supervisor's statement of a final thesis

Supervisor: doc. Ing. Filip Křikava, Ph.D.
Student: Bc. Michael Vrána
Thesis title: Automated data analysis pipelines
Branch / specialization: System Programming
Created on: 5 June 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

2. Main written part

75 /100 (C)

The thesis is written in English, and the level is quite good (I am not a native speaker, so I refrain from typos).

One more proofread would improve the clarity of the text.

I have four comments:

1. Properly use of programming language terminology could clarify the meaning in several places. For example, when talking about the call by need, first-class environments, and non-standard evaluation (which is used in two different contexts).
2. The assessment could have been improved with more than one pipeline implemented and more experiments. That will also help to streamline the API. However, I understand that it would require significantly more time which could only be gained by a better-prepared thesis assignment.
3. The chapters could be more systematic, focusing on the problem, why it is needed, and the possibilities, and only then showing what has been done. That would make it more to the point.
4. There could be fewer references to the existing targets solution. Clear design goals address the targets's shortcomings, and there is no need to refer to targets throughout the text.

3. Non-written part, attachments

100/100 (A)

The R programming language is great for data analysis and statistics but not so much for general-purpose programming.

The lack of type annotations and type-checking facilities makes it hard to grow and refactor any non-trivial implementation.

Michael delivered a working system that feels like idiomatic R with good SE practices.

4. Evaluation of results, publication outputs and awards

95/100 (A)

The result of the thesis is a tool for running data analysis pipelines which is at the core of the evaluation of much of the research today. Concretely, at the PRL laboratory, this is what we have done for nearly all our papers.

The assignment resulted from our frustration with the existing tooling, which was often more of a hindrance than a help.

Michael's work addresses those shortcomings and shall help us in future evaluation endeavors.

He also managed well the balance between the features of the language, the runtime, and the complexity of the implementation.

Eventually, its maintenance will fall on the members of the PRL lab, and we wanted it to be manageable.

The next step should be to clean up the documentation, extract pieces of the thesis in standalone vignettes (literate-programming style documents that demonstrate the typical use cases), and publish the package to CRAN.

5. Activity of the student

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

Since the first discussion when we put together the outline of the thesis, he was active and focused.

He does get things done.

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

Michael is one of the students every advisor would like to have—intelligent, critical thinking, and reliable.

He quickly learned the R programming language and familiarized himself with the R ecosystem and data analysis pipelines.

He kept refactoring the project as the requirements were evolving without losing hope of finishing on time.

The overall evaluation

95 /100 (A)

Working on this thesis required to acquire rather good knowledge of the R programming language and its idioms and patterns that are used to craft DSL.

The result is a system for defining an executing reproducible data analysis pipelines in distributed environment, suitable to be used for the work that we do at the PRL lab.

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.