



Review report of a final thesis

Student: Bc. Kryštof Slavík
Reviewer: Ing. Filip Křikava, Ph.D.
Thesis title: Record and Replay debugging in R
Branch of the study: System Programming

Date: 4. 6. 2018

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 5.</i>
1. Difficulty and other comments on the assignment	<i>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</i>
<i>Criteria description:</i> Characterize this final thesis in detail and its relationships to previous or current projects. Comment what is difficult about this thesis (in case of a more difficult thesis, you may overlook some shortcomings that you would not in case of an easy assignment, and on the contrary, with an easy assignment those shortcomings should be evaluated more strictly.)	
<i>Comments:</i> The author had to learn the internals of R not only to understand all the corner cases of R itself, but also to understand how R interpreter and byte-code compiler work and how they can be extended with the recording and replaying functionality.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
2. Fulfilment of the assignment	<i>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</i>
<i>Criteria description:</i> Assess whether the thesis meets the assignment statement. In Comments indicate parts of the assignment that have not been fulfilled, completely or partially, or extensions of the thesis beyond the original assignment. If the assignment was not completely fulfilled, try to assess the importance, impact, and possibly also the reason of the insufficiencies.	
<i>Comments:</i> The result of this work is R extended with record and replay functionality. While the written part is something that could have been improved, the actual work has been successfully executed.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
3. Size of the main written part	<i>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</i>
<i>Criteria description:</i> Evaluate the adequacy of the extent of the final thesis, considering its content and the size of the written part, i.e. that all parts of the thesis are rich on information and the text does not contain unnecessary parts.	
<i>Comments:</i> The size of the written part is sufficient.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
4. Factual and logical level of the thesis	<i>80 (B)</i>
<i>Criteria description:</i> Assess whether the thesis is correct as to the facts or if there are factual errors and inaccuracies. Evaluate further the logical structure of the thesis, links among the chapters, and the comprehensibility of the text for a reader.	

Comments:

I tested RRnR from FIT's Gitlab on a few examples and it worked on all of them without any problems. Well done!

The written part is where I have some concerns.

1. The main thing is that it is a bit hard to follow. It must be said that the chosen language is English and I understand that carving sentences in a non-native language is not easy. Also I must say that the level of English is very good! I only wish it was less technical. Most of the details are not that relevant. They would suit better as comments in the source code. What is missing is a better description of the concepts, explaining what were the problems, possible solutions and realized solution. In the current state it reads like a description of the source code. Many of the code listings do not add much value. It would be better to have more examples and use them to illustrate the difficulties. A lot of interesting work has been done and it is a pity that it is not better presented.

2. The evaluation focuses a lot on performance and less on the actual use of RRnR to find bugs. While it is great that overhead is low, I don't think it is that important (mozilla rr is often used to run overnight fuzzing the program's input). What I would like to see are more examples of how RRnR helps to find the root cause of bugs. And if the focus is on the overhead, it would be good to discuss the overhead when replaying/recording vignettes since they present much more realistic R code than the shootout benchmarks.

3. Finally, it would improve the reading experience if the Section 3 (which takes 37/66 pages) is broken into at least two sections preventing the heavy use of subsubsubsections.

<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
5. Formal level of the thesis	95 (A)

Criteria description:
Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 26/2017, Article 3.

Comments:
Formally, the written part is good. It could have been improved by tuning code listing output and merging some paragraphs together - there are sections where almost each sentence is a paragraph on its own (cf. 3.3.1.1).

<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
6. Bibliography	95 (A)

Criteria description:
Evaluate the student's activity in acquisition and use of studying materials in his thesis. Characterize the choice of the sources. Discuss whether the student used all relevant sources, or whether he tried to solve problems that were already solved. Verify that all elements taken from other sources are properly differentiated from his own results and contributions. Comment if there was a possible violation of the citation ethics and if the bibliographical references are complete and in compliance with citation standards.

Comments:
The references are adequate. I would only change a bit the way Section 2 is written (focus on which features of R are relevant for this thesis) for which the following reference would have been handy:

- Morandat, Hill, Osvald, Vitek. Evaluating the Design of the R Language. ECOOP, 2012

<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
7. Evaluation of results, publication outputs and awards	95 (A)

Criteria description:
Comment on the achieved level of major results of the thesis and indicate whether the main results of the thesis extend published state-of-the-art results and/or bring completely new findings. Assess the quality and functionality of hardware or software solutions. Alternatively, evaluate whether the software or source code that was not created by the student himself was used in accordance with the license terms and copyright. Comment on possible publication output or awards related to the thesis.

Comments:
This is a very interesting topic and to my knowledge the first attempt to bring record/replay functionality into R. Being a heavy R user myself, I would love to have something like RRnR built in R. I just wish the actual use of the debugger was more in line of what mozilla rr does. Yet with some work, there is definitely a publication potential. At least the author should consider the R journal and a demo at next UserR conference. What would be a good start is to polish its usage, create a few videos demonstrating its use (like the mozilla rr demonstrations) and publish it in the R-bloggers and R mailing lists to spread the word.

<i>Evaluation criterion:</i>	<i>No evaluation scale.</i>
8. Applicability of the results	

Criteria description:
Indicate the potential of using the results of the thesis in practice.

Comments:
This work has a potential to have an impact in the R world. See the point 7.

<i>Evaluation criterion:</i>	<i>No evaluation scale.</i>
9. Questions for the defence	

Criteria description:
Formulate any question(s) that the student should answer to the committee during the defence (use a bullet list).

Questions:

1. What is the overhead of recoring / replaying of the vignettes?
2. The reverse execution in mozilla rr is one of its killer feature. While the actual reverse execution might require quite some engineering effort (well beyond this thesis), would be possible to simulate it by replaying the code from the beginning with a break point N lines before the current one?
3. The "running example" (cf. Listing 1) in the thesis could be debugged easily without the need to record / replay. Could you present a better one on which one see the true power of RRnR?
4. Can you give some intuition of what are the unhandled corner-cases in which replaying of the 26% tested vignette fails?

Evaluation criterion:

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

10. The overall evaluation

95 (A)

Criteria description:

Summarize the parts of the thesis that had major impact on your evaluation. The overall evaluation **does not** have to be the arithmetic mean or any other formula with the values from the previous evaluation criteria 1 to 9.

Comments:

Very interesting topic and well executed thesis which could have an impact in the R community. The minor issues with the textual part are over-weighted by the implementation.

Signature of the reviewer: