

Supervisor's statement of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Bc. David Přihoda
Supervisor: Ing. Ivo Lašek, Ph.D.
Thesis title: Distributed Conversion of RDF Data to the Relational Model
Branch of the study: Knowledge Engineering

Date: 4. 6. 2017

<p><i>Evaluation criterion:</i></p> <p>1. Difficulty and other comments on the assignment</p> <p><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.)</p> <p><i>Comments:</i> The student had to study a new domain of Linked Data and Semantic Web. Based on the gained knowledge he had to implement a working application suitable for scalable parallel processing on a Hadoop cluster, which is not a straightforward task.</p>	<p><i>The evaluation scale: 1 to 5.</i></p> <p>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</p>
<p><i>Evaluation criterion:</i></p> <p>2. Fulfilment of the assignment</p> <p><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.</p> <p><i>Comments:</i> The assignment was completely fulfilled.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</p>
<p><i>Evaluation criterion:</i></p> <p>3. Size of the main written part</p> <p><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.</p> <p><i>Comments:</i> The extent of the thesis is adequate. The content is relevant and well organized.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</p>
<p><i>Evaluation criterion:</i></p> <p>4. Factual and logical level of the thesis</p> <p><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.</p> <p><i>Comments:</i> I haven't spotted any factual errors. The thesis is logically structured into chapters. It provides a good introduction in basic semantic web concepts and in parallel data processing using Apache Spark. The work is adequately set in the context of related work. The design and implementation is clearly described and evaluated. I especially liked the part of the thesis describing use cases that provides good examples of a real world usage of the implemented application.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>100 (A)</p>
<p><i>Evaluation criterion:</i></p> <p>5. Formal level of the thesis</p> <p><i>Criteria description:</i> Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 14/2015, Article 3.</p> <p><i>Comments:</i> I have no remarks.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>100 (A)</p>
<p><i>Evaluation criterion:</i></p> <p>6. Bibliography</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>90 (A)</p>

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 thesis contains a well elaborated research part summarized mainly in the Related Work chapter. Citations are correctly used. Some citations contain minor formal errors.

Evaluation criterion:

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

7. Evaluation of results, publication outputs and awards

100 (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:

During the work on the thesis the student managed to develop a working distributed application that has a wide range of use cases. It can be reused to transform many information rich RDF datasets to a more standard relational form supported by a wide variety of data analytics and visualization tools. This is well demonstrated especially on the use case I. described in the thesis. The student has proven the scalability of the developed solution when he used his framework to process some of the largest available RDF datasets.

Evaluation criterion:

No evaluation scale.

8. Applicability of the results

Criteria description:

Indicate the potential of using the results of the thesis in practice.

Comments:

Many comprehensive publicly available datasets are exposed primarily in an RDF form. However, the toolset available to process, analyze and visualize RDF data is very limited. RDF2X enables standard enterprise tools to analyze these datasets and thus make these datasets available to much wider audience.

Evaluation criterion:

The evaluation scale: 1 to 5.

9. Activity and self-reliance of the student

9a:
1 = excellent activity,
2 = very good activity,
3 = average activity,
4 = weaker, but still sufficient activity,
5 = insufficient activity
9b:
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:

Review student's activity while working on this final thesis, student's punctuality when meeting the deadlines and consulting continuously and also, student's preparedness for these consultations. Furthermore, review student's independency.

Comments:

The overall student's activity was excellent. He was able to implement first versions of the tool quite quickly which left enough space for thorough testing and performance tuning on real world datasets. The student was able to work independently and to proactively come with his own ideas and improvements.

Evaluation criterion:

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

10. The overall evaluation

100 (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:

The overall quality of the thesis and the developed application is high. The student was able to implement a scalable solution that can be used as a standalone application as well as in a highly scalable setup on a Spark cluster. On real use cases the student managed to prove the scalability and flexibility of the implemented solution. The quality of the thesis text is also very good. The whole problem and its solution is clearly described in an understandable form. I recommend the thesis for the defense and propose the overall grade A.

Signature of the supervisor: