

Supervisor's statement of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Bc. Jana Čabaiová
Supervisor: Ing. Milan Dojčinovski
Thesis title: Summarizing Linked Open Data Datasets
Branch of the study: Web and Software Engineering

Date: 30. 1. 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 main goal of this thesis is to develop a tool for summarization and comparison of RDF datasets. The tool should enable users to integrate datasets, define domains, and execute domain-specific analysis. In addition, the tool should offer functionality for macro (domain based coverage) and micro (entity type based coverage) visualization and analysis of the data. The thesis requires good knowledge in Web and Semantic Web technologies. The student should get familiar with the current Linked Data access mechanisms such as SPARQL, Linked Data Fragments, HDF, etc. Further, the developed tool had to be validated on a real data from the Linked Open Data cloud.</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 has been fulfilled with one minor objection: the tool does not implement micro visualizations for completeness of information of entity types in a given dataset.</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> In general, the thesis meets the criteria with the following minor objections: - Chapter 1 unnecessary provides information about generally known concepts (RDF, Linked Data principles) or concepts unrelated to the topic (RDF vocabularies, RDFa, RDF serialization formats, etc). - there is no need to provide description on what an analysis and design phases are in software design (intro to Chapter 2, 3) - the results from the comparison analysis between DBpedia and GeoNames need further discussion</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>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>58 (E)</p>

Comments:

The thesis is well structured, however particular parts of the thesis are difficult to understand and they leave many open questions. For example, Chapter 2 does not summarize in details the findings from the analysis of related technologies. Also, the student could have better explained the role of ontology mappings and their usage in the dataset analysis. Further, the student writes that the domains are defined using entities, however, the domains are defined using entity types/classes and not entities. It is also not clearly described how the domains were defined with information retrieved from the Wikipedia page entitled "Featured Articles". The student incorrectly uses the owl:sameAs predicate for indication of equivalent classes between ontologies. Moreover, the work is not well motivated in the introduction section, and the conclusion section does not provide necessary information.

Evaluation criterion:

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

5. Formal level of the thesis

77 (C)

Criteria description:

Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 14/2015, Article 3.

Comments:

The thesis is written in English and it contains quite some grammatical and language mistakes. Also, the student had to pay attention to some typographical aspects - first capital letter when referencing table or figure (Table, not table).

Evaluation criterion:

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

6. Bibliography

80 (B)

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 sources of information are well chosen and several relevant tools have been analyzed. However, the student should have avoided citing sources available on EDUX or Wikipedia.

Evaluation criterion:

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

7. Evaluation of results, publication outputs and awards

75 (C)

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:

The thesis brings new findings. The tool has been validated with two large Linked Data datasets and the current solution faces with some scalability issues - the responsiveness of the tool is a minor issue. Also, some parts of the implementation could be also improved, e.g. for some diagrams, there are no human readable labels, thus the visualized information is useless for the user. Particular parts of the thesis should be improved prior to publishing the outputs on a scientific event.

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:

The tool can be used to generate visualized summary of linked open datasets or compare two available datasets. There are chances that the tool will be used by the Semantic Web community.

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:

During the first period of the supervision the student regularly attended the consultation meetings. However, during the second half of the supervision, the consultation meetings were very rare. At the consultation meetings the student was always well prepared. With regards to the final thesis, only minimal support from the supervisor was requested.

Evaluation criterion:

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

10. The overall evaluation

68 (D)

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 student has shown good capabilities to conduct the development of the software through all its phases. The developed tool will provide good basis for its further development and improvement. However, the student should have paid more attention to the content of the written part. Particular sections require detailed explanation, while some parts are unnecessary. Individual parts of the tool require improvement. However, listed deficiencies are not crucial for the overall contribution of the thesis.

Considering all comments above, I recommend the thesis for defense with grade D.

Signature of the supervisor: