



Review report of a final thesis

Student: BA. Pragalbha Lakshmanan M.A.
Reviewer: RNDr. Jakub Klímek, Ph.D.
Thesis title: Enrichment of the DBpedia NIF dataset
Branch of the study: Web and Software Engineering

Date: 13. 5. 2019

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
1. 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 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.	
<i>Comments:</i> The assignment was not fulfilled. From the assignment, I would expect that the output of the thesis would be a well documented, configurable, easy to run method (probably in the form of a software tool) for repeatable enrichment of the DBpedia NIF dataset in multiple languages, using various NLP methods suitable for the task (i.e. for each language there is probably a different set of libraries performing various NLP related transformations of the data). Instead, the thesis reads like initial notes from performing one such specific task manually, i.e. to get to know the environment before actually developing a reusable method (e.g. for future DBpedia NIF releases). The enrichment algorithm described was performed only on the English language version of DBpedia NIF dataset, even though the assignment specifically mentions it is to be run on several DBpedia NIF languages. The overall quality of the thesis is too low for the assignment to be considered fulfilled.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
2. Main written part	<i>20 (F)</i>
<i>Criteria description:</i> 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. 26/2017, 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.	

Comments:

The textual part of the thesis is of very low quality and shows that the author lacks basic programming capabilities. The text is also not structured properly. Below is an incomplete list of some of problems in the textual part of the thesis.

Some examples of the formal problems are noticeable right away:

- 9 pages: 8,13, 20-22, 24, 33, 55, 57 are completely missing - this alone is a cause for rejection of the thesis
- Typography: On Page 5, there are 4 headings without any text in between 2, 2.1, 2.1.1, 2.1.1.1 - the reader has no idea why there are so many section levels without some introductory texts
- The thesis contains many errors in English, overflowing texts (end of page 12, page 32, 34, 36, 60)
- There are zero-based heading numbers, e.g. 3.4.0.1 Implementation (page 32)
- Some headings begin with lowercase letters - 3.6.0.2 challenges or 3.7. creation of ...

However, the thesis also has a number of problems with the actual contribution:

- The whole contribution (i.e. Chapter 3) is basically written as a sequential walkthrough of processing of text files in Python, written by a novice programmer, mixed with a description of used technologies.
- Section 3.2 contains a description of used Python libraries. The text is very chaotic and unstructured. Some of the text is repetitive, e.g. in 3.2.1, paragraph 1 and paragraph 2, the text jumps between higher level overview ("RDFlib is a Python library for working with RDF") and low-level function descriptions ("parse() converts into a graphical form...") and Python code snippets.
- Section 3.3 describes how a 60GB input file is split into 5 million smaller files.
 - Pseudocode for counting lines in a file is unnecessary
 - Reading all lines in a file by first counting them and then using the count in a for-cycle (hardcoded in the algorithm on page 29) is just very bad practice
 - Parsing RDF as text, including parsing URIs is also a bad practice - the algorithm is fixed on a particular data file and may not work on a different input (page 29)
- Listing all the filename exceptions in Windows is unnecessary on page 30
- Section 3.4: In 3.4.0.1. - a description splitting a text into sentences is mixed with an explanation of RDF prefix usage. RDF prefixes should have been explained in the RDF section in Background, not repeated here.
- Section 3.8. - The whole section is a comment on one function, detecting links from text, and is very confusing
 - there are hardcoded constants making the code harder to read
 - there are unjustified requirements such as that the link can go from a text of a maximum of 3 words - why? In the experiments section it is stated that this is clearly not enough. But there is no discussion to support this. It seems that the author just written the algorithm for 3 words and then was not able to generalize the approach.
- Section 4 Experiments - the section starts with a generic description of Precision, Recall and F1 score. The author then computes the F1 score of their approach, talking about precision and recall, but there is no description of the actual baseline. It seems like a manual evaluation - a human looks at a wiki article and does tokenization, sentence split, etc. and then compares to the output of the algorithm, but it is not written anywhere. This may be the cause for the evaluation to be performed on only 100 of the 5 million files.
- Missing references - there are many terms used in the thesis without actual references, e.g. in the Introduction, I am missing references for at least these terms:
 - established Semantic Web standards
 - Linked Data
 - DBpedia NIF
 - Linked Open Data cloud

On the other hand, throughout the thesis, there are unnecessarily long quotations (full paragraphs) describing terms like "parsing" or "sentence" (page 31).

Evaluation criterion:

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

3. Non-written part, attachments

20 (F)

Criteria description:

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.

Comments:

The output of the thesis is a set of insufficient Python scripts. They are hardcoded for a single dataset, the whole process is not repeatable.

The proper way of implementing the dataset enrichment methods would be a formal analysis of requirements, design of the solution, implementation description, testing, and proper documentation and evaluation, with the solution published on a publically accessible repository, and ideally with some feedback from the DBpedia NIF community.

Evaluation criterion:

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

4. Evaluation of results, publication outputs and awards

0 (F)

Criteria description:

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.

Comments:

I can see no real reuse of the results of this thesis in the current form.

Evaluation criterion:

No evaluation scale.

5. Questions for the defence

Criteria description:

Formulate questions that the student should answer during the Presentation and defence of the FT in front of the SFE Committee (use a bullet list).

Questions:

- What would it take to run the task on a different release version of the DBpedia NIF dataset?
- What would it take to run the task on a different language version of the DBpedia NIF dataset?

Evaluation criterion:

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

6. The overall evaluation

20 (F)

Criteria description:

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.

Comments:

The thesis is of low quality, badly structured, with unrepeatable results and on top of all that, with pages missing and showing bad coding practice.

Signature of the reviewer: