**FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE**

Review report of a final thesis

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**

The work contains a very high number of long passages literally taken (copy & paste) from other works: scientific articles, blogs and websites. The copied passages can be found in Chapters 1, 2 and 3, e.g. not only in the Chapter 1 devoted to mapping the current state of the Art, but also in Chapters 2 and 3, which would contain the original work of the student. Most of the copied sources are not cited in the Bibliography.

2. Main written part $0/100$ (F)

Cannot be evaluated, the number of copied passages is very high.
I tested the passages at random. After finding 16 publications from which the student copied passages, without in most cases mentioning the source or mentioning the source in the Bibliography. It is therefore possible that other unmarked paragraphs may be copied from other works.

The numbering of sources corresponds to my numbering, not to the numbering of the publications in Bibliography.

Selected pages and numbers of copied lines:

Chapter 1: Background and related works (Pages 3-17)
page 4... 12 lines copied, source 008
page 5... 14 lines copied, sources 008 and 009
page 6.... 13 lines copied, source 009
page 11 ... 5 lines copied, source 10.

Str. 13... 30 lines copied, source 007
Str. 14... 17 lines copied, source 007
Str. 15... 7 lines copied, source 007
page 16... 16 lines copied, sources 014 and 013

Chapter 2: Analysis and Implementation (Pages 19 to 33)
page 19 ... 16 lines copied, source 005
page 22... 17 lines copied, sources 002, 015 and 016
page 25... 25 lines copied, source 006
page 27... 3 lines copied source 001
page 29... 32 lines copied, source 001
page 30... 3 lines copied, source 001

Chapter 3: Testing and Results (Pages 35 to 43)
Str. 38... 13 lines copied, source 003
Str. 39... 14 lines copied, source 004
Str. 40... 36 lines copied, source 011
Page 41... 14 lines copied, source 011

Resources used in the thesis (copy& paste), but not included in the Bibliography;
[001] https://www.tutorialspoint.com/spring/spring_overview.htm
[002] https://www.visual-paradigm.com/guide/development/what-is-rest-api/
[003] https://softwaretestingfundamentals.com/smoke-testing/
[004] https://en.wikipedia.org/wiki/Unit_testing
[005] https://www.guru99.com/functional-requirement-specification-example.html
[006] https://zapier.com/engineering/how-to-cli/
[007]   https://citeseerx.ist.psu.edu/viewdoc/download?
doi=10.1.1.537.8904&rep=rep1&type=pdf
[008] https://en.wikipedia.org/wiki/Web_Ontology_Language
[009] https://en.wikipedia.org/wiki/Resource_Description_Framework
[010]   https://www.researchgate.net/publication/240615157_Linked-
Data_Aware_URI_Schemes_for_Referencing_   Text_Fragments/link/
00b7d5243cc214e06c000000/download
[011] https://www.tutorialspoint.com/junit/junit_test_framework.htm
[013]   https://www.researchgate.net/publication/
235299230_DBpedia_and_the_live_extraction_of_structured_data_from_Wikipedia
[014]   Dbpedia   Live   Extraction,   https://www.researchgate.net/publication/
220830431_DBpedia_Live_Extraction
[015] https://messaggio.com/glossary/rest/
[016] https://nataliekate.github.io/restful_api/

Only one resource copied in the thesis is cited in Bibliography
[012] http://downloads.dbpedia.org/wiki-archive/data-set-2014.html

There is also an appendix to this review report, where the identical passages in the articles and in the student's work are marked in color.

3. Non-written part, attachments   30 /100 (F)

The student tried to developed tool for parsing and extracting information from Wikipedia XML dumps. There is a lack of experiments. Due to the high amount of copied text, it is difficult to approve the originality of the solution.

4. Evaluation of results, publication outputs and awards 0/100 (F)

In my opinion, due to the high amount of copied text, the results are not usable.

The overall evaluation 10/100 (F)

The work contains a large number of passages copied literally from other works (copy& paste): scientific articles, websites and blogs. Copied sources are not listed in Bibliography, except one. The scope of copying is large and the work cannot be perceived as a the original student's work. Even if the passages were marked and cited, the scope of copying is so great that it would not be possible to recommend the work for the defense. There is also an appendix to this review report, where the identical passages in the articles and in the student's work are marked in color. I do not recommend the thesis for defense.

Questions for the defense

Be so kind, please, and explain the high number of passages in your work copied from other texts.

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.

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.