FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Review report of a final thesis

Student:	Bc. Bogoljub Jakovcheski
Reviewer:	Ing. Jaroslav Kuchař, Ph.D.
Thesis title:	Domain-Specific NER Adaptation
Branch of the study:	Web and Software Engineering

Date: 28. 1. 2019

Evaluation criterion:	*The evaluation scale: 1 to 4.*
1. Fulfilment of the assignment	*1 = assignment fulfilled,* ***2 = assignment fulfilled with minor objections,*** *3 = assignment fulfilled with major objections,* *4 = assignment not fulfilled*

Criteria description:
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.

Comments:

The topic of this thesis is focused on experiments with domain-specific approaches for the NER. The text of the thesis describes all the specified assignments. However, the quality of the performed steps and results are not at the level of a master thesis. Selected approaches for the data transformation and preprocessing are not sufficiently described (what are the criteria for the "acceptable time", why do you create such tree structure etc.). The selection of categories or entities, parameters of the experiments is not clear. Training and testing of the models on the same datasets is also very trivial approach. The conclusions are thus very questionable.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
2. Main written part	*55 (E)*

Criteria description:
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 text contains all required parts. However, the quality of the presentation is very bad. English should be significantly improved. The text is difficult to understand. Almost every section raises several questions. E.g. how exactly are specific experiments performed, why are such specific parameters selected etc. The structure of the thesis should be revisited too. Only about one quarter is focused on the theory, data preparation and model learning. The rest of the thesis (over 60 pages) describes many experiments. Most of them can be simplified and presented only as a single table (while moving the rest to the appendix section). The results are not convincing, e.g. (page 26, Table 3.5) - why is the total recall equal to the same value when the second value is number one? Similarly for the F1 score. The same situation is with almost every result in the thesis (e.g. Table 3.106 on page 83).

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
3. Non-written part, attachments	*60 (D)*

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 implementation is not easily runnable; it contains many hardcoded settings and file system paths. It is not commented at all. The source codes for all experiments are not available. The quality of the SW is of low quality.

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

4. Evaluation of results, publication outputs and awards — *50 (E)*

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:

As was already mentioned, the thesis contains many experiments with problematic results. The impact of those results is thus questionable. From my point of view, the student did a lot of manual work with the selection of specific entities that can be useful for other experiments. The work itself can be used as initial overview for more advanced experiments.

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:

- Why did you decide for the StanfordNER tool?
- Are all your steps (e.g.dataset preparation) consistent with any other approaches?
- Do you have any comparison with existing general NER approaches? What are the precision, recall and F1 score for those approaches?
- How are the Total values for evaluation metrics computed (e.g. Table 3.5, 3.106)?
- Can you briefly describe how can be the domain-specific NER approach used in practical applications?

Evaluation criterion: — *The evaluation scale: 0 to 100 points (grade A to F).*

6. The overall evaluation — *55 (E)*

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 student studied several theoretical aspects and performed many experiments. Although, the thesis has many issues with the design of the work, with experiments and results; the student did a lot of work.

Signature of the reviewer: