FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Supervisor's statement of a final thesis

Student:	Bc. Bogoljub Jakovcheski
Supervisor:	Ing. Milan Dojčinovski
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 main goal of the thesis is to investigate a domain-specific adaptation of Named Entity Recognition (NER) systems. The student had to: 1) get familiar with the NER technology, 2) investigate available datasets for domain-specific NER learning, 3) develop NER datasets for different domains, 4) train an adapted, domain-specific NER model, and 5) evaluate the developed models. The student fulfilled the requirements of the thesis with two minor objections:
- the text contains some grammatical issues
- the experiments and the results could be presented/summarized in a more clear way. Also, some exceptional results from the experiments (see comment below) require more discussion.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
2. Main written part	*70 (C)*

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 thesis is well structured into the following chapters: introduction, background an related work, the method, experiments, and conclusion. The sections logically follow and text behind the individual chapters is easy to understand and follow (apart from the experiments section).

The student clearly identified the related work (section 1.2) and provides appropriate citations. Considered software and data are used according to the associated license.

There are two main issues with the writing part of the thesis:

- grammar issues - there are grammatical/language mistakes which can be found in the thesis. However, these issues are not crucial for the overall quality of the thesis, neither they introduce big problems in reading/understanding the contents.

- experiments - the experiments chapter forms the main part of the thesis (roughly 60 pages). The experiments chapter is quite lengthy, and the student had to consider reducing/merging the experiments. The student describes each experiment and the results from it, which is often usually short and repetitive. While there is an overall summary for the individual group of experiments, the student could focus more on better presentation of the findings from the experiments.
Also, the student does not discuss the 0.0 values from some experiments - e.g. what could be the reason for such results?

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

3. Non-written part, attachments — *85 (B)*

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:

Overall, the developed solution is of good quality.
The student developed scripts for working with the dataset (see my comment below on the main challenge), developed and populated domains, trained models and evaluated the models. In addition, the student has created two additional datasets for evaluation and used them in the experiments (section 3.3.5).

The technologies used in the process of the development of the solution are appropriate and well selected.

The size of the data introduces a challenge - the student trained models on a large-scale dataset (DBpedia NIF) where the size (over 4 million of documents) and the format (NIF) is an obstacle, which the student successfully managed to overcome.

The student could use a more systematic approach for the development of the domains.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
4. Evaluation of results, publication outputs and awards	*85 (B)*

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:

The main outputs from the thesis are:
- domain definitions - defined with ontology classes (entity types)
- domain specific NER models - trained on several datasets

The thesis generates novel insights/findings, however, their publication in a scientific article would require further work - especially, in the experiments section and the domains specification.

Evaluation criterion:	*The evaluation scale: 1 to 5.*
5. Activity and self-reliance of the student	*5a:* ***1 = excellent activity,*** *2 = very good activity,* *3 = average activity,* *4 = weaker, but still sufficient activity,* *5 = insufficient activity* *5b:* *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:
From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations (5a). Assess the student's ability to develop independent creative work (5b).

Comments:

The student regularly attended the meetings and consulted the current state of the work. The student was always well prepared with a prepared set of questions on the identified problems.
There were several challenges around the dataset and its format which the student managed to overcome alone.
More active consultation and support were required in the last phase of the thesis - writing the thesis.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
6. The overall evaluation	*79 (C)*

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 main goal of the thesis was to develop models for domain-specific classification of named entities.
In general, the student fulfilled the requirements of the thesis topic with two minor issues: the student had to put more attention on the written part (the grammar) and on the experiments - better presentation of the experiments and summarization of the results. However, these issues are not crucial and do not significantly influence the final quality of the thesis.
Considering all the comments above, I recommend assessment of the thesis with grade C.

Signature of the supervisor: