Supervisor's review of master thesis

Author of the review:	RNDr. Pavel Pecina Ph.D.
Author of the thesis:	Eric Lief
Title of the thesis:	Deep contextualized word embeddings from character language models for neural sequence labeling

The presented thesis of Eric Lief deals with natural language processing tasks which can be solved by sequence-to-sequence (also called sequence tagging or sequence labeling) models based on deep neural networks. This area has been recently influenced by development of contextualized word representations. The goal of the thesis is to explore this area and investigate the effect of using different combinations of embeddings on selected sequence-to-sequence tasks in Portuguese. The tasks explored in the thesis include: part-of-speech tagging, named entity recognition, and multiword expression identification.

The thesis is structured into seven chapters and several appendices including a rich bibliography counting about 40 items. After the introduction, the author presents on overview of modern sequence tagging models (Chapter 2) and description of models used in the experiments presented in the thesis (Chapter 3). Chapters 4-5 include description of the author's experiments conducted for part-of-speech tagging (Section 4), named entity recognition (Section 5), and multiword expression identification (Section 6). The conclusions are given in Chapter 7. The main text spans the total of 87 pages.

The thesis is well structured, written in English without typos and grammatical errors. The text occasionally contains minor mistakes (e.g. in referring to Table 14 instead of Table 15 on page 44). The theoretical background provided in Chapters 1 and 2 is quite brief and shallow and not all details are explained (e.g. not all the mathematical symbols and formulas in Figures).

The number and scale of the experiments is large. The explored model architectures were properly tuned, applied and analyzed. The main drawback is the fact that the author did not test his results for statistical significance. Some differences of the resulting scores are very small and probably not statistically significant.

I agree with the author that the main contribution of the thesis is the "thorough exploration of different experimental setups with embeddings applied to sequence labeling tasks". Though it does not introduce any surprising findings and ground-breaking results, the experiments were designed and conducted correctly. I recommend the thesis to be defended.

Prague 28.1.2019