Thesis proposal:

**Neural Networks for Part-of-Speech Tagging**

Suitable for a Bachelor’s thesis (15 credits) or a project course

**Background**

Part-of-speech tagging is a central task in language technology. One simple but effective technique for solving this task is based on classification, where the tag of a word is predicted based on the word itself as well as context-based features such as the next word, previous word, or tag of the previous word. A large variety of machine learning models have been employed in this framework. One drawback of many of these is their dependence on a large number of hand-crafted features. Neural networks are an interesting alternative in this context because they are able to learn features by themselves. The goal of this project is to explore the use of neural networks for part-of-speech tagging.

**Project description**

Your task is to replace the machine learning components of an existing part-of-speech tagger with a new model based on neural networks. To do so you will first need to familiarise yourself with the existing literature on part-of-speech tagging and neural networks. A significant part of the project will be devoted to implementation work in Java or Python. After finishing the implementation, you will evaluate the resulting system on standard data sets and compare it to existing systems in the literature. At the end of the project you will prepare a scientific report, documenting and discussing your findings. Depending on the outcome, the project may be expanded into a Master’s thesis.

**Student profile**

You should be familiar with the basic principles of language technology (acquiring and pre-processing textual data) and machine learning (linear classification, logistic regression, standard neural network architectures). You should also be able to program in either Java or Python.

**Contact**

Marco Kuhlmann, marco.kuhlmann@liu.se