Thesis Title: Named Entity Recognition in the Biomedical Domain

Named entity recognition (NER) is the task of information extraction that attempts to recognize and extract particular entities in a text. One of the issues that stems from NER is that its models are domain specific. The goal of the thesis is to focus on entities strictly from the biomedical domain. The other issue with NER comes the synonymous terms that may be linked to one entity, moreover they lead to issue of disambiguation of the entities. Due to the popularity of neural networks and their success in NLP tasks, the work should use a neural network architecture for the task of named entity disambiguation, which is described in the paper by Eshel et al [1]. One of the subtasks of the thesis is to map the words and entities to a vector space using word embeddings, which attempts to provide textual context similarity, and coherence [2]. The main output of the thesis will be a model that attempts to disambiguate entities of the biomedical domain, using scientific journals (PubMed and Embase) as the documents of our interest.