Titel:
Context Management in Database Systems with Word Embeddings

Abstract:
Natural language processing focuses recently on the development of learned language models called
word embedding models. Those methods have been shown their usefulness for a wide range of NLP
and information retrieval tasks. In today's complex adaptive systems automatic data integration and
data exploration play an increasingly important role. While word embedding models have already
been used in data integration tasks on tables, they are rather inappropriate for representing the
context-specific meaning of short text values and infrequent named entities in database systems.
Furthermore, trained on documents word embeddings do not fit perfectly for text values in tabular
formats. This thesis thus focuses on two main aspects: On one hand, we want to integrate additional
functionality for text analysis and machine learning into database systems by making use of the
powerful abilities of word embedding models. On the other hand, we want to improve the
representation of text values in database systems generated using word embedding models by
incorporating the context-specific meaning of text values in database systems.