In dieser Arbeit wird untersucht, inwiefern für Kommentare aus Schachpartien das zugehörige Annotationssymbol gefunden werden kann. Mithilfe einer solchen Zuweisung könnten zu bisher nur mit Kommentaren versehenen Schachspielen zusätzliche Informationen generiert werden, sodass die Spiele besser auswertbar sind und beispielsweise für darauf aufbauende Lernverfahren genutzt werden können. \\\\

Zur Analyse werden überwachte Klassifikationsprobleme sowohl für Kommentare zu Zügen als auch zu Stellungen formuliert. Es werden geeignete Kommentardaten aus Schachdatenbanken extrahiert, mittels Natural Language Processing vorverarbeitet und schließlich in einem für die Klassifikation geeigneten Modell aufbereitet. Hierfür werden vier unterschiedliche Modelle untersucht: eines basierend auf den absoluten Häufigkeiten der Wörter (Count), eines auf den relativen Häufigkeiten der Wörter (TF-IDF), ein auf den Schachkommentaren trainiertes Word Embedding und ein auf Google News vortrainiertes Word Embedding. Für drei verschiedene Datensätze mit gemischt-langen, kurzen und langen Kommentaren werden mehrere Multiklassen-Klassifizierer getestet. \\\\

Während bei den Klassifizierern keine eindeutige Rangfolge erkennbar war, erzielte das Count-basierte Modell die besten Genauigkeitswerte, gefolgt vom TF-IDF-basierten Modell. Bei den beiden Word Embeddings schnitt das auf Google News vortrainierte besser ab als das auf den Schachkommentaren trainierte. Außerdem konnten für die kürzeren Kommentare bessere Genauigkeiten erreicht werden als für gemischt-lange oder lange Kommentare. \\\\

Die erzielten Genauigkeiten sind jedoch insgesamt zu niedrig, um die Modelle für eine Umwandlung von Kommentaren zu Annotationssymbolen nutzen zu können. Eine Verbesserung könnte durch eine Entfernung bestimmter Stoppwörter erreicht werden, die vermutlich besonders bei langen Kommentaren einen negativen Einfluss haben. Ebenso könnte eine Weiterentwicklung des selbst trainierten Word Embeddings zu einer Verbesserung der Genauigkeiten führen, da dieses trotz der niedrigsten Genauigkeitswerte ein vielversprechendes Modell mit sinnvollen semantischen Ähnlichkeiten im Schachvokabular berechnet.

This paper examines how the corresponding annotation symbol can be found for comments in chess games. With the help of such an assignment, additional information could be generated for chess games that have so far only been provided with comments. As a result, these games can be better evaluated and used for learning procedures based on them. \\\\

For analysis, supervised classification problems are set up for both comments on moves and positions. Appropriate comment data are extracted from chess databases, preprocessed by natural language processing and finally transformed into a suitable model for classification. For this purpose, four different models are examined: one based on the absolute frequencies of the words (count), one based on the relative frequencies of the words (TF-IDF), a word embedding trained on the chess comments and a word embedding pre-trained on Google News. For three different data sets with mixed-length, short and long comments, several multiclass classifiers are tested. \\\\

While the classifiers did not show a clear ranking, the count-based model achieved the best accuracy values, followed by the TF-IDF-based model. Of the two word embeddings, the one pre-trained on Google News performed better than the one trained on the chess commentaries. In addition, for the shorter comments better accuracies could be achieved than for mixed-long or long comments. \\\\

However, all the achieved accuracies are too low to use the models for transforming comments into annotation symbols. An improvement could be achieved by removing certain stopwords, which are likely to have a negative effect especially on long comments. Likewise, a further development of the self-trained word embedding could lead to an improvement of the accuracies, since despite the lowest accuracy values it calculates a promising model with meaningful semantic similarities in the chess vocabulary.