

Machine Learning Approaches for Temporal Information Extraction: A Comparative Study

Oleksandr Kolomiyets, Marie-Francine Moens

Department of Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
3001 Heverlee, Belgium
oleksandr.kolomiyets@cs.kuleuven.be
sien.moens@cs.kuleuven.be

Temporal expressions are important structures in natural language. In order to understand text, temporal expressions have to be extracted and normalized to ISO-based values. For these purposes rule-based and machine learning techniques were proposed. In this paper we present and compare two approaches for automatic recognition of temporal expressions in free text, based on a supervised machine learning approach and trained on an annotated corpus for temporal information, namely TimeBank. The first approach performs a token-by-token classification following B-I-O encoding. The second one does a binary constituent-based classification of chunk phrases. Our experiments demonstrate that on the TimeBank corpus the constituent-based classification performs better than the token-based one. It achieves F1-measure values of 0.852 for the detection task and 0.828 when an exact match is required, which is better than the state-of-the-art results for temporal expression detection on TimeBank.