Detecting Multi-Word Expressions using Supervised ML Methods

Name: Yaakov HaCohen-Kerner

Affiliation: Jerusalem College of Technology, Jerusalem, Israel

WG3: Statistical, Hybrid and Multilingual Processing of MWEs
The General Algorithm

Automatic detection of Multi-Word Expressions (MWEs) in different languages using supervised ML methods

(1) Indication of MWEs that appear in the CORPORA

(2) Definition and programming of features related to the classification task of whether a sequence of words is a MWE

(3) Computing the features for all possible sequences of words

(4) Applying a variety of supervised ML methods in order to obtain optimal results for detection of MWEs

(5) Analyzing the results and concluding
(1) "PoS Tags of individual tested words"
(2) "Sequences of PoS Tags"
(3) "PoS Tags of individual words before/after the tested sequence of words"
(3) "Sequences of PoS Tags before/after the tested sequence of words"
(4) Function (stop) words: e.g., normalized frequencies of function words in the tested sequence of words
(5) Quantitative features: statistical measures concerning the tested sequences of words
(6) prefixes or suffixes of words tokens in the tested sequence of words