Automatic Extraction of Grammatical Multiword Expressions for Lithuanian

Lexical association measures

MLE, PMI, T, DICE, LL

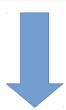


Supervised Machine Learning

Naïve Bayes, Bayesian Network and Random Forest algorithms



- Candidate list
- Lexical association measures
 - MLE (Maximum Likelihood Estimation
 - PMI (Pointwise Mutual Information)
 - T (Student's t score)
 - DICE (Dice's coefficient)
 - LL (Log-likelihood score)
- Reference list
- Evaluation against the reference list



- Supervised machine learning algorithms
- Filters
 - SMOTE (Synthetic Minority Oversampling Technique)
 - Resample
- · 10-fold cross validation

Justina Mandravickaitė, Michael Oakes, Tomas Krilavičius