

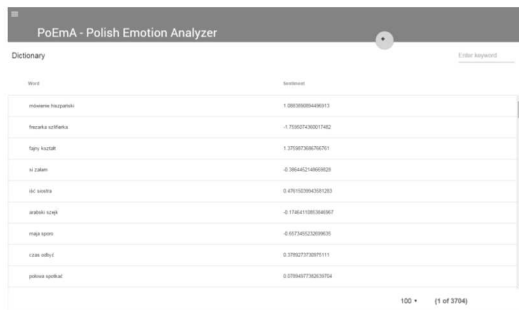
poema - Polish Sentiment Analyzer Demo

Antoni Sobkowicz*

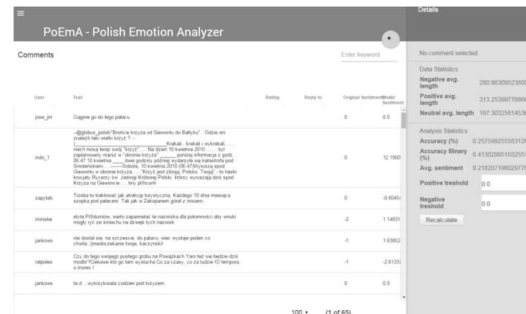
*Ośrodek Przetwarzania Informacji - Państwowy Instytut Badawczy
al. Niepodległości 188b, 00-608 Warszawa, Poland
antoni.sobkowicz@opi.org.pl

Abstract

poema is easy to use software for sentiment analysis, mainly in Polish language. Software is based on automated sentiment analysis methods presented in authors' previous works. poema enables sentiment analysis based on machine learning models and social networks. Program is designed for users who want quick and simple tool to aid their research. poema allows user to evaluate sentiment of large quantities of texts using various sentiment analysis models, both simple dictionary based and machine learning based, enclosed in simple web based application. Users can modify sentiment dictionary and use custom one to fine-tune analysis process, customize models and check detailed information about analyzed text. One of poema key features is an ability to process sentiment in social networks, provided texts contain information required to build such network. poema consists of four basic modules, responsible for program functionality. Additional demonstration module and tools are available as well. *Sentiment Dictionary Editor* provides simple way to display sentiment dictionary. poema provides predefined set of terms described with sentiment weights, gathered from automatic analysis of twitter messages, that can be filtered and each sentiment term can be changed. User can add and remove each entry to make dictionary suitable for their need. Full, custom, user-defined dictionaries can be uploaded and used in model training and sentiment analysis. Custom dictionaries must be defined in CSV files. Dictionaries can be also automatically created from sentiment annotated texts in CSV format. *Model Selection Module* allows user to train dedicated models or select existing, pre-trained ones. It provides selection of ML (machine learning based) and simple dictionary based models that can be trained using built-in or custom training data sets. Some models are usable only for polish language based training data, and some are language independent, meaning they will work for input data in the same language as training data. *Text Analysis Component* allows batch text preview and analysis. It is mostly designed to use with comments and twitter messages (that have defined user-author) but can be also used with any text. In addition to basic text sentiment estimation, module can also display detailed results (for example word-to-word interaction information, bigram detection and emotion of each individual word). User texts can be uploaded as json or CSV files in specified format. Analysis outcome can be exported as CSV and json files for off-line analysis. *Social Network Analysis* module allows evaluating of social network properties, such as in- and out-degree, user (text author) total received sentiment (sentiment of texts directed to users), total generated sentiment (sentiment of texts written by user) and sentiment of each interaction between user. It also provides visualization of users and interactions. Social network data can be exported to one of the popular network analysis software formats for off-line analysis. poema will be freely available to general public from Winter 2015. Availability will be announced on <http://lil-devel.opi.org.pl/ai-web>



(a) Dictionary view component.



(b) Text analysis component view.