SESAMm

Natural Language Processing Exercise

Aim: to develop a simple natural language processing module to do textual sentiment analysis (positive, negative, neutral) either using symbolic techniques, statistical techniques or combination of both. The goal is to focus on natural language processing by proposing interesting ideas to analyze the sentiment of a tweet. A corpus is provided and it is expected to output evaluation metrics, at least the accuracy, at best the F1 score. The developed algorithm will be evaluated on a test corpus which will be kept secret during the exercise. Any programming language can be chosen, any external libraries and any kind of resources can be used (lexical resources, sentiment lexicons, parsers, taggers, etc.).

Expected output:

- Implementation of few natural language processing ideas
- Accuracy and/or F1 score

Provided data:

- CSV file: a corpus of tweets annotated with ternary discrete valence (positive, negative, neutral).

Instructions:

- Any programming language, pick your favorite
- Internet and any source of information
- It is generally advised to reuse an existing library
- The goal is not to create a full-fledged sentiment analyzer but rather put in motion creative ideas to analyze the sentiment
- It is not mandatory to obtain good results but rather show creative abilities in the choice of the algorithm; the cleanliness of the code itself will be evaluated as well