

Autora: Luana Balador Belisário

CLASSIFICADOR DE SUBJETIVIDADE: INSTRUÇÕES DE USO

O método utiliza léxicos de palavras previamente rotuladas para classificar as sentenças desejadas. O programa, basicamente, lê palavra por palavra de cada léxico e armazena essas palavras em dois vetores, em ordem alfabética, um de palavras subjetivas positivas e outro de subjetivas negativas. É importante que as palavras/expressões positivas estejam em um documento diferente das negativas, além de que as palavras/expressões devem estar uma embaixo da outra (separadas por uma quebra de linha, ou seja, um '\n') e sem espaços desnecessários (no início ou no fim da palavra/expressão), como também não podem apresentar caracteres desnecessários como ponto, vírgula, aspas, parênteses, entre outros. Adicionar também à pasta do classificador "Lexicon-based subjectivity classifier for portuguese" o documento contendo o texto a ser lido e classificado. O programa classificará o texto por parágrafo, ou seja, cada vez que houver uma quebra de linha no texto ou quando encontrar o final do documento. Cada parágrafo vai ser classificado da seguinte forma:

- 0 (Objetivo neutro);
- 1 (Subjetivo positivo);
- -1 (Subjetivo negativo).

O resultado será gerado em outro arquivo "resultado.txt" contendo as classificações dos parágrafos ou frases na ordem em que elas se apresentam no arquivo de leitura. Os arquivos de texto a serem utilizados devem estar no formato ".txt" e similares.

PARA USAR O CLASSIFICADOR

No terminal, no diretório do classificador, digitar "gcc classifier.c -o class" e dar enter, depois digitar "./class". O programa começará a ser executado e serão solicitados os nomes dos arquivos correspondentes aos léxico de palavras positivas, o de palavra negativas e o arquivo de leitura, respectivamente e um de cada vez. Então o programa classificará o texto e gerará o arquivo "resultado.txt" no diretório do classificador. Se já houver um arquivo com esse nome no diretório o programa irá sobrescrevê-lo.

Esse programa foi feito com base no método baseado em léxico de classificação de subjetividade proposto no artigo "Comparing Approaches to Subjectivity Classification: A Study on Portuguese Tweets" dos autores Silvia M.W. Moraes, André L.L. Santos, Matheus Redecker, Rackel M. Machado, e Felipe R. Meneguzzi.