Name: Trinadha R Muppala

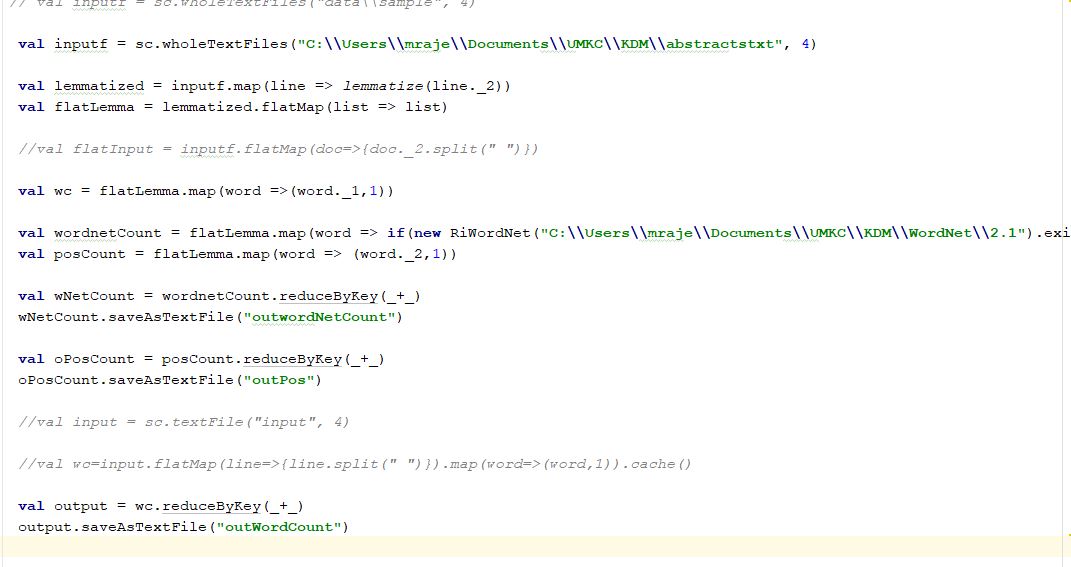
ClassID : 7

1. Find out word count, words exist in word net and find count of those words, words POS count, word count after lemmatization

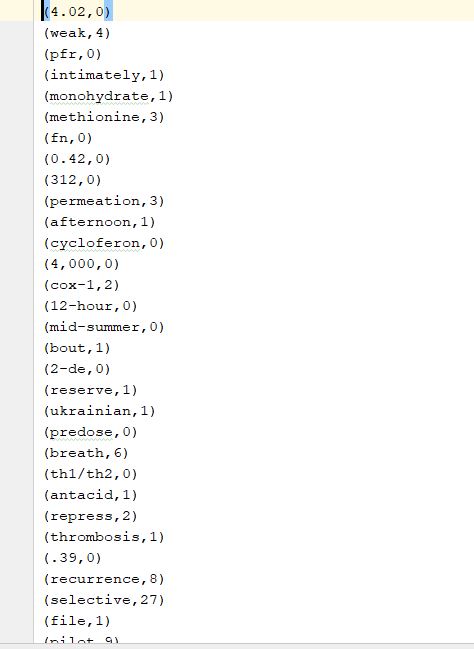
Function to do lemmatization and POS from coreNLP. Create List with both lemmatized word and POS



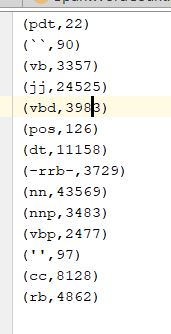
Code to read all the abstracts and find out lemma count, pos count, check word exist in wordnet then set 1 else set word to 0.



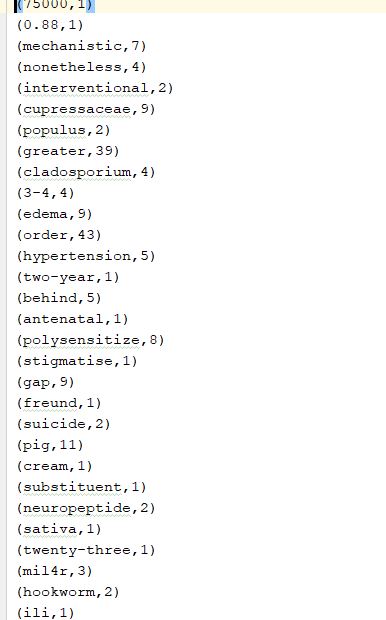
Wordnet word count output, word count is zero for words not exist in wordnet



Pos count output

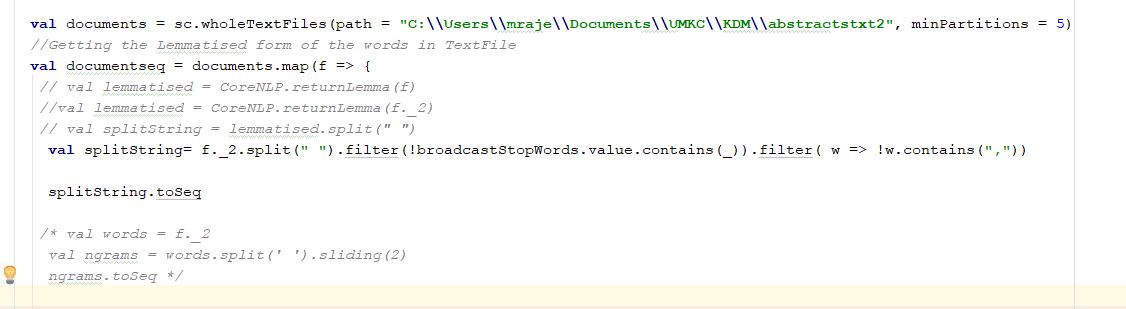


Wordcount output



1. Top TF\_IDF words and word2vec synonyms

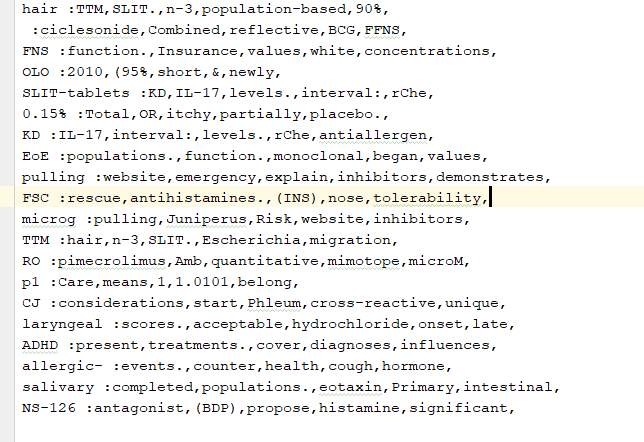
Read all the abstract files



Find out top TF\_IDF words and get synonyms from word2vec model



Top TF IDF words word2vec

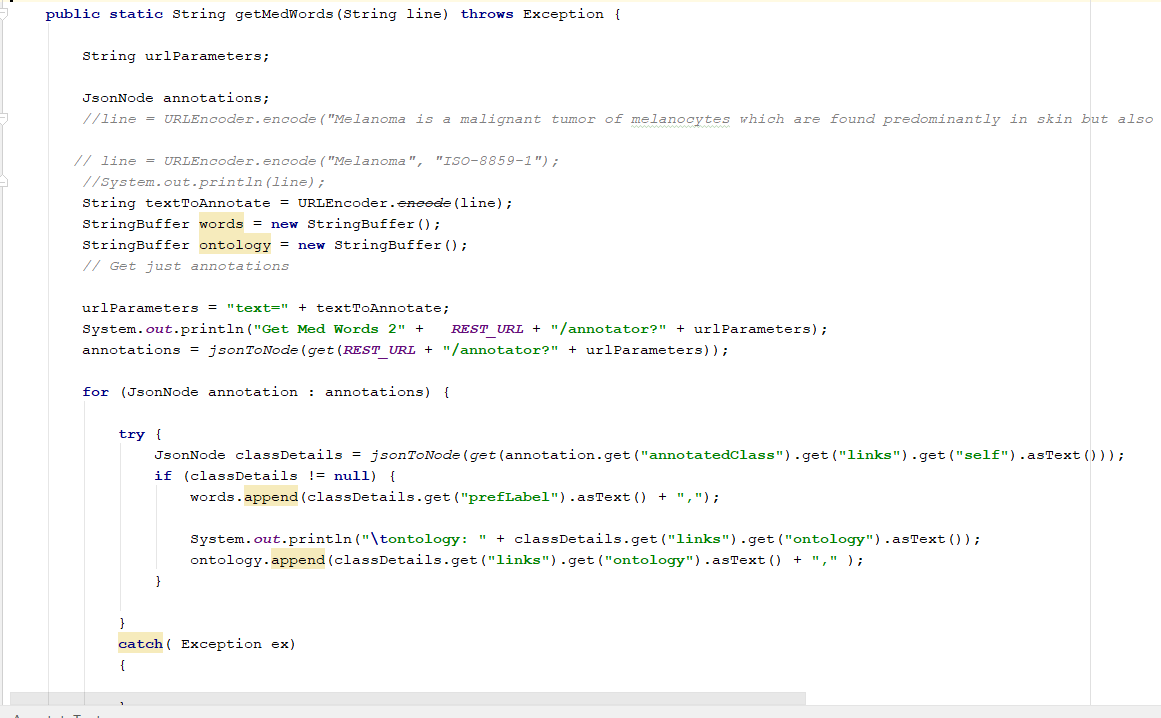


1. Medical word statics

Find all the medical words and write those to file.



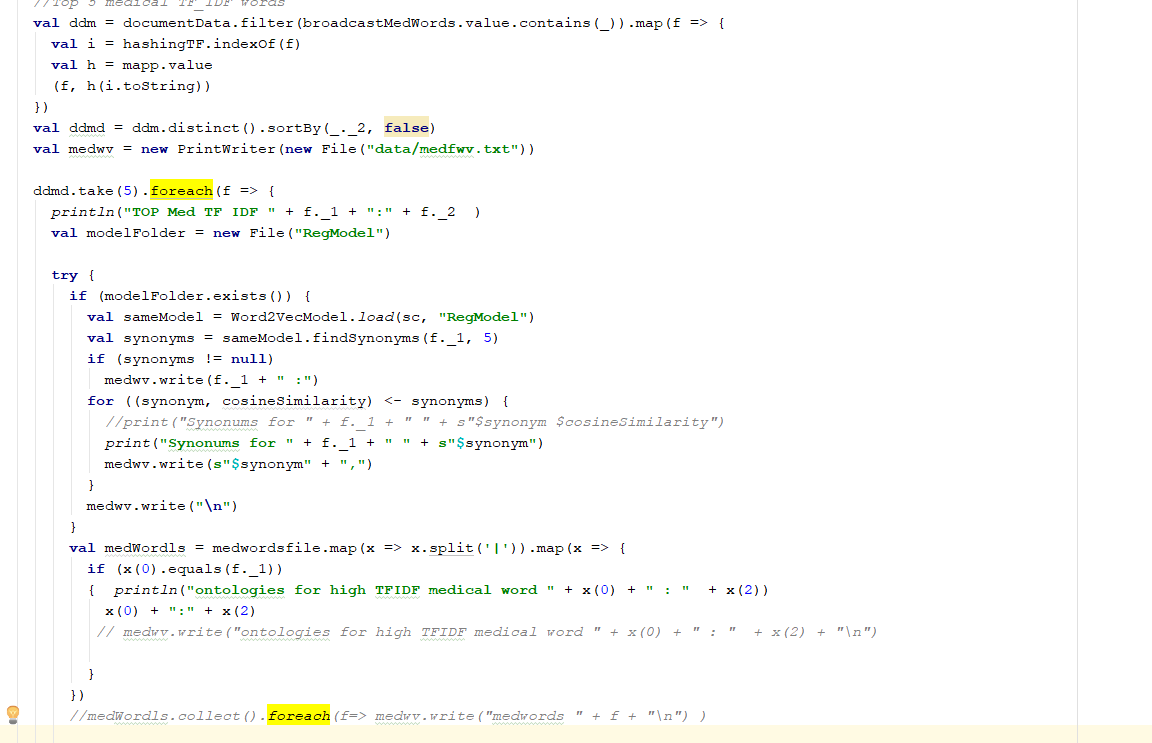
GetMedical words using BioPortal:



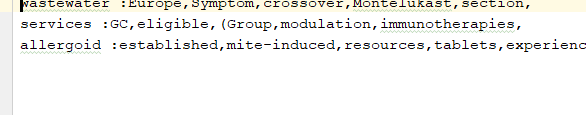
Medical words file output:



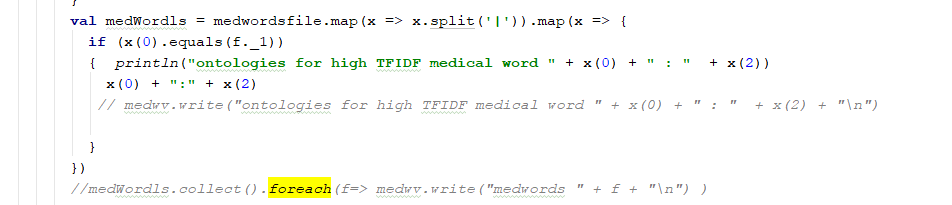
Read medical words file, Broad cast all the medical words and get the list of TF\_IDF in the medical list get Top TFIDF word 2 vec



Output of TOP\_TF\_IDF words word2vec



1. Find ontology of the Top TFIDF words. Ontology saved to the medical words list file get the ontology for matching top TF\_IDF word



Ontology :

