CS458 Natural Language Processing Course Project.

Project Title : Documents Specific Question Answering with Sentiment Analysis.
Implemented By:
Name : Abdul Aziz Muhammad Ibrahim Isa.
Roll No: P17-6143.

Project Description:
The directory " P176143_NLP_Project " contains two things.
1 : The python file " questions.py " which contains the implementation of the project.
2 : The directory " corpus " which contains datasets (txt files) that I am using for this project.
All the datasets are contents of Wikipedia pages regarding different topics. This program is platform independent therefore it can run on any operating system but it takes a command line argument which is basically the directory containing all the datasets. This command is used to run the program.

" python questions.py corpus "

The code works as first it takes all the data from the corpus directory and then we formulate the data.It then prompts the user for a query Then we calculate TF-IDF values of each file (I am using 7 files inside corpus) by multiplying the idf values of each word with the term frequency in that specific file. We then then determine the top file, which basically means the topic of the query as each file contains data from a specific topic, by TF-IDF values. When we get to the top file we then extract all the sentences from the file and calculate its IDF value based on the same word that also occurs in the query. The sentence with the highest IDF value is the answer to the query. If there are two or more sentences having highest IDF value, then we apply query term density to get the best possible result. After getting the output we then apply sentiment analysis to the query with help of the Wordnet dictionary.

This image contains several inputs with the corresponding outputs:

```
C:\Users\abdul\Desktop\questions>python questions.py corpus
Query: What are the types of supervised learning?
Topic of Query: machine_learning
Answer: Types of supervised learning algorithms include Active learning , classification and regression.

Performing Sentiment Analysis on query.......
Positive: 42.99 %
Negative: 57.01 %

C:\Users\abdul\Desktop\questions>python questions.py corpus
Query: When was Python 3.0 released?
Topic of Query: python
Answer: Python 3.0 was released on 3 December 2008.

Performing Sentiment Analysis on query.......
Positive: 52.98 %
Negative: 47.02 %

C:\Users\abdul\Desktop\questions>python questions.py corpus
Query: How do neurons connect in a neural network?
Topic of Query: neural_network
Answer: Neurons of one layer connect only to neurons of the immediately preceding and immediately following layers.

Performing Sentiment Analysis on query.......
Positive: 53.64 %
Negative: 46.36 %
```