NLP Assignment 1

Please don't copy from net or other students. Plag will be checked.
Allowed Programming language : Python

Your task in this assignment is to write a python program that accepts as input any text file from the "alt.atheism" and "comp.graphics" folders in 20newsgroups dataset and perform the following tasks.

1) Return the number of sentences and words contained in the file given as input.
2) Return the number of words starting with consonant and the number of words starting with vowel in the file given as input.
3) List all the email ids in the file given as input.
4) Given a word and a file as input, return the number of sentences starting with that word in the input file.
5) Given a word and a file as input, return the number of sentences ending with that word in the input file.
6) Given a word and a file as input, return the count of that word in the input file.

Note: You are allowed to skip the header part (metadata) of all the files for the above questions except for the third question.

Your code will be checked for any file int the above mentioned folders as input. So, don't try to fit your code too closely to a single file.

Submission format: Upload a zip file named **Assignment1_RollNo.py** of your project folder containing the code file and a pdf report file with your implementation details and assumptions made, if any.