“Zipf's law was originally formulated in terms of quantitative linguistics, stating that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc.: the rank-frequency distribution is an inverse relation.” (see [Wikipedia](https://en.wikipedia.org/wiki/Zipf%27s_law))

If this rule holds for a text, then the graph of the frequency function is a line in a coordinate system where both coordinate axes are logarithmic.

The task will be to graphically check this rule for a natural text, that is, plot the frequencies in a loglog coordinate system. To split the text and remove the punctuation marks you should use the built in functions split and strip.

The sample text should be read from a text file, its name will be given as a command line argument. The program shall be run from command line with the sample text in test.txt (contained in the same folder as the program) as follows:

python3 5BAdrianSmith.py test.txt

For testing some text samples are attached. Zipf's law is perfectly implemented by the perfect.txt file in the text samples. There are smaller files with Hungarian and larger with English text.