The overall objective here is to produce a set of scripts that allows a certain set of functions to be carried out. These functions are for analysing the text in a document (format: .txt)

The code needs to be thoroughly commented.

Existing open source code and libraries may be used as long as it appropriately attributed.

Features:

1. **Wordcount (COMPLETED)**

Generates a list of the most frequent words in a document.

1. **Plot frequency of words (COMPLETED)**

Generates a graph that displays the relative/raw frequency of a word or set of words over the duration of a document.

1. **Part-of-Speech / Syntax analysis / Sentence length (COMPLETED)**

Generates a graph showing a syntactical breakdown of sentences. Also shows the average length of sentences in a document.

1. **Sentiment analysis**

This should generate a graph that traces a particular emotion across a document. This should work on the principle of synonym hops. This method is outlined here: <http://www.seasr.org/documentation/uima-and-seasr/sentiment-tracking-from-uima-data/>

It should perform as follows:

<http://ianmilligan.ca/2012/07/09/was-the-past-a-happy-place/>

1. **Topic modelling**

This should extract a set of topics from a document, like as follows:

<https://de.dariah.eu/tatom/topic_model_python.html>

1. **Delta cluster analysis**

This script should cluster texts based on most frequent wordlists, like the method used in the visual basic used here: <https://files.nyu.edu/dh3/public/TheDeltaSpreadsheets.html>

1. **Zeta analysis**

See https://files.nyu.edu/dh3/public/UsingtheCraigZetaSpreadsheet.html