**WIA 2005: Algorithm Design and Analysis**

**Project Paper: Semester 2, 2018/2019 (25%)**

**GENERAL INSTRUCTION:**

* Choose team members that you can work easily with to complete this project (must be in the same tutorial class)
  + Presentation (5%): Week 14
  + Report submission (5%): Week 14 (31st May, 2019 @ 5.00pm)
    - Include source code : raw python files
    - Include the description of the project with Control flow graph (CFG)
    - Details of members in the group and work assigned (compulsory to all members in developing the program)
    - Content: Introduction, Method, Description of topic, Implementation (source code and snapshots of input/output), Conclusion, References
    - Time complexity of the program code
* There should be 6 of you in a team. Each member must be given important and technical role. Please specify the roles accordingly.
* All programming codes must include the usage of python 2 or 3 language

**FIND THE BEST FLIGHT** (15%)

You have been assigned to develop a system to help the end user to find the best flight to use when there is no direct flight available. The system will recommend the end user the best flight only based on route and political sentiment of the country. The best flight does not necessarily base on the shortest route but also considering the political situation each country the flight transited. Please note that there is no price of ticket involved in this recommendation system.

1. Get and mark locations of 5- 10 major cities in the world (where you can easily gets their main English newspaper online) from Kuala Lumpur
   1. Guide 1: you can use Python Geocoding Toolbox

Look up: <https://pypi.python.org/pypi/geopy#downloads>

* 1. Guide 2: you can use gmplot

Lookup: <https://github.com/vgm64/gmplot>

1. Get the distances between these cities
   1. Guide 1: you can use Python Geocoding Toolbox
   2. Suggestion 2: you should use Google Distance Matrix API
      1. Login to the google developer’s website and follow through the examples. It is important that you know how to use the API key given to you within the code that you are going to use. Refer to this link: https://developers.google.com/maps/documentation/distance-matrix/start
2. Using one of the algorithms for shortest path, get the minimum distance to travel to the destination from Kuala Lumpur by transiting 2-3 cities.
3. Plot line between the cities before and after the algorithm chosen.
   1. Guide1: you can use **google.maps.Polyline**. You can refer to this link:

<https://www.sitepoint.com/create-a-polyline-using-the-geolocation-and-the-google-maps-api/>

1. Extract from the English newspaper of the cities of webpages text and count the number of words in the webpages.
   1. Sometimes a webpage must be converted to the text version before it can be done
      1. Guide 1: You may refer to this website to extract word from a website

<https://www.textise.net/>

* 1. Guide 2: You may refer to this website on how to count word frequency in a website

<https://programminghistorian.org/lessons/counting-frequencies>

* 1. You can also filter stops words from the text you found
     1. Guide 3: Stops words are such as conjunctions and prepositions. You may refer to this link: <https://www.ranks.nl/stopwords>
     2. Program using Rabin-karp algorithm to find and delete the stop words.

1. Plot line/scatter/histogram graphs related to the word count using Plotly (Word count, stop words)
   1. Guide 3: You may refer this link on how to install Plotly and how to use the API keys

<http://www.instructables.com/id/Plotly-with-Python/>

<https://plot.ly/python/getting-started/>

1. Compare words in the webpages with the positive, negative and neutral English words using a String Matching algorithm
   1. Guide 4: Use the following word as positive and negative English words

<http://positivewordsresearch.com/list-of-positive-words/>

<http://positivewordsresearch.com/list-of-negative-words/>

* 1. Put these words in a text file for you to access them in your algorithm
  2. Words that are not in the list can be considered as neutral

1. Plot histogram graphs of positive and negative words found in the webpages.
   1. Guide 5: Use Plotly
2. Give an algorithmic conclusion regarding the sentiment of those articles
   1. Guide 6: If there are more positive words, conclude that the article is giving positive sentiment, if there are more negative words, conclude that the article is giving negative sentiment.
   2. You may try to conclude in different perspectives such as whether the list of positive and negative words above is accurate to be used in the context of the article you extracted the text.
   3. Based on the conclusion, you may say the country has positive or negative political situation.
3. Lastly, calculate the total probability distribution of random routes taken for the end user to travel from Kuala Lumpur to other country.

END OF QUESTION