**WIA2005: Algorithm Design and Analysis**

**Project Paper: Semester 2, 2017/2018 (30 marks)**

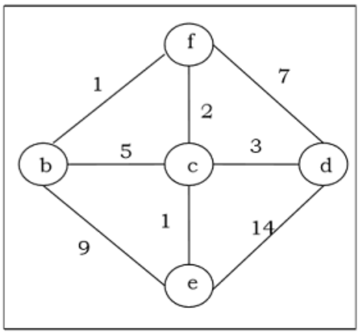
**GENERAL INSTRUCTION:**

- Choose team members that you can work easily with to complete this project
    - Presentation: Week 13/14
    - Report submission: Week 14/15
        - Include source code : raw python files
        - Include the description of the project
            - Team members names
            - Content: Introduction, Method, Description of topic, Implementation (source code and snapshots of input/output), Conclusion, References
- There should be at most nine (6-9) 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

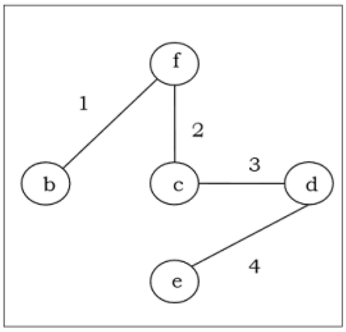
**CHOOSE ONE (1) OF THE FOLLOWING TOPICS:**

**Topic 1: Minimum Spanning Tree (MST)**

1. Get and mark locations of 5- 10 cities in Malaysia
    a. Guide 1: you can use Python Geocoding Toolbox
    Look up: https://pypi.python.org/pypi/geopy#downloads
    b. Guide 2: you can use gmplot
    Lookup: https://github.com/vgm64/gmplot
2. Get the distances between these cities
    a. Guide 1: you can use Python Geocoding Toolbox
    b. Suggestion 2: you should use Google Distance Matrix API
        i. 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
3. Using one of the algorithms for Minimum Spanning Tree (MST), get the minimum distance to travel to all of the cities. Figure 1a shows an example of the distances between 5 cities, while figure 1b shows the minimum distance to travel to all of the cities.



(a)



(b)

Figure 1

4. Plot line between the cities such as in Figure 1 before and after the MST algorithm.
    a. 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/

**Topic 2: String matching (Rabin-Karp) for sentiment analysis**

1. Identify and use 5-10 webpages such as BBC New, Bernama, Al Jazeera and others on topics regarding political/education/economic news in Malaysia.
2. Extract the webpages text and count the number of words in the webpages.
    a. Sometimes a webpage must be converted to the text version before it can be done
        i. Guide 1: You may refer to this website to extract word from a website https://www.textise.net/
    b. Guide 2: You may refer to this website on how to count word frequency in a website https://programminghistorian.org/lessons/counting-frequencies
    c. You can also filter stops words from the text you found
        i. Guide 3: Stops words are such as conjunctions and prepositions. You may refer to this link: https://www.ranks.nl/stopwords
        ii. Program using Rabin-karp algorithm to find and delete the stop words.
3. Plot line/scatter/histogram graphs related to the word count using Plotly (Word count, stop words)
    a. 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/

4. Compare words in the webpages with the positive, negative and neutral English words using Rabin-Karp String Matching algorithm
    a. 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/
    b. Put these words in a text file for you to access them in your algorithm
    c. Words that are not in the list can be considered as neutral
5. Plot histogram graphs of positive and negative words found in the webpages.
    a. Guide 5: Use Plotly
6. Give an algorithmic conclusion regarding the sentiment of those articles
    a. 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.
    b. 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.

## Topic 3: Dictionary and sorting algorithms

1. Create information of 5 000 people with the followings:
    a. Name, address, email, date of birth, salary
    b. Guide 1: Generate and store this information in a text file. You may refer to this link: http://www.pythonforbeginners.com/files/reading-and-writing-files-in-python
2. Extract the information above into a python dictionary so that we can easily access these data:
    a. Guide 2: You may refer to this link on how to create a dictionary: https://docs.python.org/3/tutorial/datastructures.html#dictionaries
3. Create a graphical user interface (GUI) that you can use to modify information as in 1a.
    a. Guide 3: You may use the **Tkinter** module, refer to this link https://pythonspot.com/gui/
        i. Include widgets that you can use to update/edit, delete, add new information of people
4. Sort the data that you have using Quicksort/Mergesort algorithm
    a. Sort based on name, city in address, date of birth and salary
5. Show your answer in a table form in the above GUI:
    a. Guide 4: You may refer to the following link https://stackoverflow.com/questions/9535954/printing-lists-as-tabular-data
6. Send an email message to 10 people in your dictionary:
    a. You may refer to the following link https://www.tutorialspoint.com/python/python_sending_email.htm
7. Can a Dictionary in python be replaced by a database such as MySQL? Discuss.

## Topic 4: Simple Cryptography, encryption using Hash Function

1. Convert your data into an encrypted form:
    a. Take data in string form and convert it to the ascii values
        i. Guide 1: You may refer to the following link https://stackoverflow.com/questions/8452961/convert-string-to-ascii-value-python
    a. Create a script to encrypt your asci data using a simple hash function for example (data_value% a_ big_ prime_ number + a_number_of_ user's_choice.)
        i. Hash your input at least two to three times
    c. Display your new string message
    d. Implement all these using python, coding them in your android mobile phone
        i. Guide 2: It is suggested that you use QPython
            - This is the home page http://www.qpython.org/index.html
            - This is the Guide to start coding http://www.qpython.org/en/guide.html
2. Convert the encrypted data into the original message:
    a. Do the reverse of the above
3. Upload in google/play store for people to use

**Topic 5: Reading Data from an excel file, cleaning it and restore the data in original file**

1. Download a sample of messy data from:
   http://web.fsktm.um.edu.my/~rjry/wia2005/getfile.htm

2. Find out how to read files from excel:
   https://www.datacamp.com/community/tutorials/python-excel-tutorial

3. Clean data (straight forward)
   There are a few category of data: Names, Address, Phone no., dates: Choose one format to standardise the data.
   - For example: Instead of Ali b Ahmad should be written as Ali Ahmad (this can be implemented using string matching functions
   - However, for address, you should use trie data structure

4. Using tries data structure, formulate a method to anticipate what should be the correct way of writing the address.
   - For example, In addresses, sometime "Jalan" is written as "Jln", sometimes "Kuala Lumpur" is written as "KL".
   - Find out how to implement tries:

   https://towardsdatascience.com/implementing-a-trie-data-structure-in-python-in-less-than-100-lines-of-code-a877ea23c1a1

**Topic 6: Drawing 3 dimensional animated object in Python**

1. Your team is required to show:
- a three dimensional animated
  - sorting algorithm:
    - Choose 3 sorting algorithm and at least one must be divide and conquer paradigm
  - FSKTM building, block A the ground floor. From the floor plan given, assume a constant height of the ceiling for the 1st floor for you to be able to draw three dimensional
    - Get the floor plan here:
      - http://web.fsktm.um.edu.my/~rjry/wia2005/getfile.htm

2. Refer to this website to draw three dimensional object:
   https://plot.ly/python/3d-surface-plots/

3. Refer to this website to draw three dimensional animated objects:
   http://geoffboeing.com/2015/04/animated-3d-plots-python/