COMP 7402 Computer Systems Technology January 2019

Cryptography and Cryptanalysis

Assignment #1

Due: To be completed by January 24, 1700 hrs. This is an individual assignment.

Task 1:

- To design and implement an application that will generate a frequency count of all letters in a text file. You will then use the output from this application and generate a graph of the relative distributions of the letters.
- Once the frequency count data is available you can import it into spreadsheet to generate the graphs.
- **Constraints:**
    - You may use any language of your choice.
    - Your program will read the words from a user-specified text file and write out the frequency count to another user-specified file in CSV (comma-separated values) format:

      letter,count

    - Your application must either prompt the user for the filenames, or specify them as command line arguments.
- For your test data select two English works of literature (Alice in Wonderland, Moby Dick) and generate graphs of the relative distributions of the letters for both works.
- Compare the two graphs and make observations on the characteristics of each.
- Also verify that the sum of the probabilities of each distribution is 1. In other words confirm that:

$$P(M) = \sum_{i=a}^{z} P(m_i) = 1$$

Task 2:

- Assume that for this task we will be using a simple substitution cipher such as the Caesar cipher (single constant offset or "key").

- The objective of this task is to compute the following probability distributions for one of the text files above:

  P(M); P(K); and P(C)

- Note that in this case the code for the Caesar cipher has been provided to you. Use it to generate the ciphertext and then graph the relative distribution to determine P(C).

- Then calculate the following conditional probabilities (we will restrict the task to the six most frequent letters):

  $\mathbf{P(M=e|c_i)}$ $\quad c_i \in C$
  $\mathbf{P(M=t|c_i)}$ $\quad c_i \in C$
  $\mathbf{P(M=a|c_i)}$ $\quad c_i \in C$
  $\mathbf{P(M=i|c_i)}$ $\quad c_i \in C$
  $\mathbf{P(M=o|c_i)}$ $\quad c_i \in C$
  $\mathbf{P(M=n|c_i)}$ $\quad c_i \in C$

- I would suggest writing a program (or use a spreadsheet) to perform this task since doing it manually will be time consuming.

- Examine the conditional probabilities and comment on whether or not the results are helpful in identifying the plaintext/ciphertext pairings.

To Be Submitted Electronically:

- Submit a zip file containing all the code and documents as described below in the sharein folder for this course under "**Assignment #1**".
- Submit a complete, zipped package that includes your report, tools that you used, and any supporting data (dumps, etc), and references.
- Test results, complete with supporting data such as screen shots in PDF format.
- Hand in complete and well-documented design work and documents in PDF format.
- Also provide all your **source code** and an **executable.**
- You are required to demo this assignment in the lab.

Assignment #1 Evaluation:

Documentation (explanation, user guide, etc):	10 / 10
Testing and Supporting Data (Report):	20 / 20
Functionality:	40 / 40
Total:	70 / 70