Final Project (25 Pts)

Data Collection (5 Pts)

Each group is to compile Frequently Asked Questions (FAQs) and their answers about one or more topics. You are to submit the questions, their answers and the topics in text files.

Training Data

The compiled questions from all the groups will serve as training data for the group project. You will be given three files. The first will contain the questions with each question on a new line. The second will contain the answers to the questions with each answer on a new line. For each question in the first file, the third file will contain the topic from which the question was derived with each topic on a new line.

Topic Modeling (10 Pts)

Given a question, your program should be able to print the topic associated with the question. Note that the question used to test your program may not have appeared in the training data exactly. For instance, the training data may contain "Who is the president of Ghana" but the test question might be "Who is Ghana's head of state?". If the topic for "Who is the president of Ghana?" is "Ghana government", we expect the topic for "Who is Ghana's head of state?" to be "Ghana government".

Rubric

- Write up describing 3 ideas (3 pts)
- Implementation of 3 ideas (3 pts)
- Passing of tests (4 pt)

Question Answering (10 Pts)

Given a question, your program should be able to print an answer to the question. The notice provided about the nature of the test questions under Topic Modeling applies here. Thus, if the answer to the question "Who is the president of Ghana?" is "Nana Akufo Addo", then the answer to "Who is Ghana's head of state?" should be "Nana Akufo Addo".

Rubric

- Write up describing 3 ideas (3 pts)
- Implementation of 3 ideas (3 pts)
- Passing of tests (4 pt)

Submission

- Due date is 15th December, 2018 by 11:55am.
- There should be a folder called group_project in your github repository for the course. The folder should contain your write ups and implementation.
- Each member of the group must have their group_project in their repository by the due date.
- Your main program should be called main.py. It should take two arguments: the first is the task and the second is a file containing questions. For topic modeling, we will run your program thus: *python main.py topic questions_file.txt* and we expect you to output a file called topic_results.txt which has the topic for each question on a new line. For question answering, we will run your program thus: *python main.py qa questions_file.txt* and we expect you to output a file called qa_results.txt which has the answer for each question on a new line.
- You will be sent a link to peer review the contributions of your group members. This will be factored in computing the project grade for the individual.