IT3041 - IRWA

Assignment 01

Part A

You are provided with a folder that contains five text files. Each file corresponds to a group and contains the names of students belonging to that group.

- Group1.txt
- Group2.txt
- Group3.txt
- Group4.txt
- Group5.txt

01) Identify the students who belong to both Group 2 and Group 3, but not in Group 5

(5 Marks)

02) Identify the students who belong to both Group 1 and Group 4, but not in Group 5

(5 Marks)

03) Combine the lists of students from the two tasks above to form a new group (5 Marks)

You are provided with four text files, each containing the descriptions of contestants for different competitions held at the school:

- Sing.txt (Singing competition contestants)
- Dance.txt (Dancing competition contestants)
- Draw.txt (Drawing competition contestants)
- Drama.txt (Drama competition contestants)

Each file includes descriptions and names of students participating in the respective competition.

04) Using the inverted index method, find in which competitions are the students from Group H performing. (25 Marks)

Part of Speech (POS) tagging is a fundamental task in Natural Language Processing (NLP) that involves assigning each word in a text to its appropriate grammatical category. These categories, known as parts of speech, describe the syntactic role that a word plays in a sentence.

05) Identify the competitions in which, the Group H students are not performing and find the proper nouns in that competition description using POS in NLTK in Python. In POS the proper nouns are represented by NNP. (10 Marks)

Part B

Create a game called the **Quote Master**, where the user needs to write a similar quote for a randomly given quote by the computer.

01) The game should be developed as below. The computer will randomly select one quote from the following quotes.

- Dream big and work hard
- Love deeply and worry less
- Stay positive and stay focused

Then the user has to create and input a quote by himself. When the quote which was created by the user is sufficiently similar to the quote given by the computer, the user will win the game.

- Similarity between two quotes should be measured using the combination of TF-IDF and Cosine Similarity.
- If the Cosine Similarity is greater than 0.5, the user will win the game.

Random module in Python will be helpful for you.

```
Import random

L=[2,4,5,6,7]

random.choce(L) -------> 4 (Randomly selected)
```

(30 Marks)

02) Extend the above game such that the user must input a similar quote within 15 seconds.

- If the user can give the input quote within 15 seconds, the user will win, and the game will end.
- If the time exceeds, the user will lose, and the game will end.
- If the user inputs a quote within the given time but loses, a new quote will be given by the computer, and the user must follow the same steps. But the time will not be renewed. Which means the time will with the remaining time and ended after the 15 seconds.

Time module in Python will be helpful for you.

```
import time

start_time = time.time()
```

x=10

y=20

print(x+y)

end_time =time.time()

print(end_time - start_time) <--Time taken for the given operation with seconds.

(20 Marks)

- Submit two different jupyter notebooks with their HTML version files (4 files).
- Rename all files as follows:
    - For the Part A - **StNo_PA.ipynb and StNo_PA.html**
    - For the Part B - **StNo_PB.ipynb and StNo_PB.html**