Introduction

This document contains all the questions for your first assessment of the Artificial Intelligence 2 module. All questions are compulsory, and you must complete all questions in a Jupyter Notebook file. This assessment is worth 15% of the total marks for the Artificial Intelligence 2 module. You have 3 hours to complete this assessment. You must submit your work through Blackboard in the **NLP CA 1 submission 1** link that can be found under the **Assignments & Tests** link on the left of the Blackboard screen.

I emailed you the name of your assigned text file that you will use to complete each task of this CA assessment. You can find the relevant text file inside the **Sample files** folder on Blackboard. **Important: You must use this text file to answer all questions for your CA. You will lose marks if you have not used the correct text file.**

Q1

Download and open the text file assigned to you from Blackboard. You may need to open the text file with utf-8 encoding. Count the number of sentences in the text file, print a sample of the text file contents to the screen.

Q2

Create an NLP document object. Then using a loop, store all the sentences of the document object into an array. Choose one of the sentences from your text document and show the following information using an f-string.

(a) Token
(b) Token POS tag
(c) Token dependency
(d) Explanation of each token POS tag
(e) Token recognised as a stop word or not

Q3

Explain what is meant by a regular expression. Explain this concept using program code to count the number of occurrences of 4 popular words in your assigned text document. Show the number of occurrences of these popular words.

Q4

Explain the concept of POS tags. Then using a loop, show how many of each POS tag are within your assigned text document. For example, your output could include the following:

```
SPACE                                10091
ADJ                                   8165
ADP                                  10808
ADV                                   8740
CCONJ                                 7791
```

Q5

Choose 4 common words in your text document. Then using the rules-based matching technique, demonstrate this concept to find **all combinations** of these words in the entire text. Call the matcher a suitable name. Display the start and end positions of each matching word, as well as each match. Use an f-string to suitably format your display.

And show a total count of all occurrences of these words.

Q6

Now you would like to view some of the words on either side of the output from your rules-based matcher. Amend the output in Q5 to show 5 words before, and 3 words after each matched word. Display the results using suitable spacing with an f-string.

Q7

Using the phrase matching technique, demonstrate the concept of phrase matching using your assigned text document. Use the same 4 common words identified in Q5 to demonstrate how the phrase matching technique can find **all occurrences** of your common words.

Q8

Show the contents of the phrase matcher search result in Q7 including

(a) Start position of the matched phrase
(b) End position of the matched phrase
(c) 5 words before and after the matched phrase

Format your output using an f-string.

Q9

Choose one sentence from your document object. Then explain the concept of lemmatization and demonstrate its use by implementing this technique on your sentence of choice.

Q10

Choose one sentence from your document object. Then demonstrate how the dependency visualiser called displacy can be used to show part-of-speech and dependency tags for this chosen sentence. Configure the displacy outputs including font type, colour, background colour, and distance of 50.

Important Information

Late submissions will not be accepted without a valid medical certificate.

Plagiarism will not be accepted and will result in an automatic mark of zero.

Due Date: Wednesday 18th March at 13:00. You must submit your work through Blackboard using the relevant link. Submit your work as a Jupyter notebook file. A cover sheet must also be submitted with your jupyter notebook file.