Programming for AI -Fall 2023

Assignment#1

Date: 08 Aug, 2023

Due date: 06 Aug, 2023, 05:00 AM **Total Marks:60**

Instructions:

- **All questions must be answered within a single notebook or .py file.**
- **Follow the file naming conventions: Name your submission file as RollNo.ipynb or RollNo.py (e.g., i22_xxxx.ipynb, where xxxx is your Roll Number).**
- **Use headings to distinguish each question in the notebook.**
- **Ensure that your solutions are free of errors.**
- **Late submissions will not be accepted and will be given a zero.**
- **No external libraries are allowed for this assignment.**
- **Any form of plagiarism will result in a zero for both parties involved.**
- **AI-generated content is prohibited. Detection of such content will lead to a zero score.**

Question 1 [20 Marks]

In this question, you have been given two text files in Urdu. The first file contains an Urdu dictionary, which consists of a list of words. The second file contains sentences that do not have spaces between the words and are difficult to read.

آجخودبخشہوں

This sentence, without proper word segmentation, is difficult to understand. However, with proper word segmentation, the sentence can be separated into individual words:

آج خود بخش ہوں

This makes the sentence much easier to read and understand.

This task involves performing maximum matching using the provided Urdu dictionary. You need to insert spaces appropriately in sentences that lack spaces between words. Here's a breakdown of the process:

- Utilize the Urdu dictionary to identify and match words from the sentences.
- Apply the maximum matching approach to find the longest words in the dictionary that match the sentence.
- Insert spaces after the matched words to separate them.

Question 2 [20 Marks]

This question is about automatic spelling correction in **Urdu** . Studies have shown that the majority of typing errors are caused by (let's assume the dictionary contains the word "پاکستان "):

1. Omitting one letter, e.g., the input word is "پاکتان".

2. Adding an extra letter, e.g., the input word is "پاکستانف".

3. Mistyping one letter, e.g., the input word is "پاکسطان".

4. Transposing two adjacent letters, e.g., the input word is "پاسکتان".

With the aid of a dictionary, word processing systems can often detect and automatically correct these kinds of spelling errors.

The Problem:

You are to write a program that recognizes the four kinds of errors described above.

The Input:

The input should be the file wrong.txt which contains wrong words which you have to corrected.

The Output:

Print each input word to be spell checked. Then, if the word is in the dictionary, print CORRECT. If the word is not in the dictionary, then find each word in the dictionary (in the order provided in the dictionary) for which the given input word might be a misspelling, and print the appropriate message from the following list:

ONE LETTER OMITTED FROM *word*

ONE LETTER ADDED TO *word*

ONE LETTER DIFFERENT FROM *word* TWO

LETTERS TRANSPOSED IN *word*

where *word* is a dictionary word. If the input word is not CORRECT and none of the above messages apply, then print UNKNOWN.

Note that two or more of the above messages might be applicable to an input word, and that one message might apply for more than one dictionary word. Note, however, that for a given input word and given dictionary word, at most one of the above messages apply. For each input word, you are to process the dictionary words in the order provided in the input and print all messages that are valid.

Leave a blank line after the output for each input word. Follow the format illustrated in Sample Output.

پاکستان پاکتان پاکستانف پاکسطان پاسکتان orooji	پاکستان CORRECT پاکتان ONE LETTER OMITTED FROM پاکستان پاکستانف ONE LETTER ADDED TO پاکستان پاکسطان ONE LETTER DIFFERENT FROM پاکستان پاسکتان TWO LETTERS TRANSPOSED IN پاکستان orooji UNKNOWN

Question 3 [20 Marks]

SEE THE GIVEN IPYNB NOTEBOOK