

Name: _____

Roll Number: _____

Date: 09-09-2024

Quiz -1

Time allowed: 25 mins

NLP-B

Total Marks: 18

Q1. Identify the different types of morphemes (bound & free, or reduplication etc.) in the following words: (7)

1) Unhappiness

2) chit chat

3) overwhelm

4) بے خوابی

5) reconsider

6) beautiful

7) پرانا

Q2. What do the following regex return. (3)

- /geor?ge*?/
- /\\$[0-9]+\.[0-9][0-9]/
- /a\.{24}z/

Q3. State the difference between stemming and lemmatization. Provide three examples where they provide different results. (3)

Name: _____

Roll Number: _____

Q4. Assume you have the following training corpus, include `</s>` in your corpus like any other token. Perform all steps. (5)

`<s> i want to eat thai food </s>`

`<s> we ate pakistani food </s>`

`<s> i ate apples </s>`

`<s> they ate thai food </s>`

Calculate the probability of the following *test sentence*. Include `</s>` in your counts just like any other token.

`<s> i ate thai food </s>`

Find the ***Unigram and Bigram*** probability of the given test sentence.
Calculate the perplexity of these both as well.