

Computational Linguistics 1

QUIZ 1

Spring 2022

Answer all the questions in the answer booklet.

Part I: [2 x 5 = 10 marks]

1. Describe briefly any 3 CL/NLP applications that you use on a regular basis.
2. How is a POS tagger useful in a Machine Translation application?

Part II: [3 x 2 = 6 marks]

Lexical density is indicated by the Type-Token Ratio [TTR]. The TTR is the # of Types divided by the # of Tokens. The closer the TTR is to 1 the more lexical variety there is. Calculate the same for each of the following sentences and show which sentences are richer lexically.

- (i) A rose is a rose is a rose.
- (ii) A red rose is not a rose painted in red colour.
- (iii) Not a rose was available on rose day.

Part III: [10 marks]

Read the following text and using the given POS tagset, annotate every token as shown in the example:

Example: I love Hyderabad

I/PRP

love/VBP

Hyderabad/NNP

TEXT: In corpus linguistics, **part-of-speech tagging**, which is also called **grammatical tagging**, is the process of marking up a word in a text or corpus as corresponding to a particular part of speech, based on both its definition and its context.

Part IV: [4 marks]

For each of the following texts, state what kinds of knowledge are needed to understand them.

- a. "I knew he has rectiphobia. But I didn't know he was rectopathic as well".
- b. "This is not what I had ordered," shouted Mary at the waiter.
- c. Murali Vijay played well, but the bowling made it look like he played a brilliant innings, so many length balls, half volleys for him to dispatch at will!
- d. "Con motivo del Día de la Independencia, el primer ministro Shri Rajiv Gandhi, al dirigirse a los activistas de Red Fort, dijo que Jammu y Cachemira es una parte integral de la India..."