International Institute of Information Technology, Hyderabad

CL3.101 Computational Linguistics

End Semester Examination- 2nd May 2024

Max. Time: 3 Hr **Max. Marks: 60**

Instructions:

- This exam paper consists of two sections: Section A and Section B. Answer both Sections.
- Section A carries 30 marks and Section B carries 30 marks
- Any attempt to answer extra questions in Section A will not be considered for grading.
- Wherever required, use linguistic gloss to explain the concept

Section A: Choose and respond to six questions

Each question carries a value of 5 marks and comprises two sub-questions. (6*5=30)

1. a) Explain Item and Arrangement, Item and Process, and Word and Paradigm approaches to morphological modeling. (2.5 marks)

   b) Discuss the advantages and disadvantages of each approach. (2.5 marks)

2. a) Discuss the challenges involved in POS tagging with examples (2 marks)

   b) How does defining tagsets help in addressing these challenges? (3 marks)

3. a) Define dependency analysis in syntactic parsing. (2.5 marks)

   b) How does it differ from constituency analysis? (2.5 marks)

4. a) What are Multi-Word Expressions (MWEs)? (1.5 marks)

   b) Describe the types and give examples of MWEs. (3.5 marks)

5. a) Describe the process of chunking and its role in sentence analysis. (2.5 marks)

   b) How are minimal phrases formed during chunking? (2.5 marks)

6. a) Given a text document containing dates in various formats (e.g., MM/DD/YYYY, DD-MM-YYYY, YYYY/MM/DD), write ONE regular expression to extract all dates. (2.5 marks)

   b) Write ONE regular expression to split a text document into sentences, assuming that sentences end with a period, exclamation mark, or question mark followed by a space or newline character. (2.5 marks)

7. a) Discuss the applications of speech processing technology in real-world scenarios. (2.5 marks)

b) What are the challenges involved in speech recognition? and explain the linguistic contributions to those challenges. (2.5 marks)

8. a) Define Hidden Markov Models (HMM) and explain their components. (2.5 marks)

b) Discuss the advantages of Conditional Random Fields over other sequence labeling models. (2.5 marks)

Section B: Data Annotation

A set of sentences are given below. Annotate them as asked in Q. 9-13. Each question carries a value of 6 marks. Answer all. (5*6=30)

Sentences:

(a) The beautiful girl whom I saw, built a sandcastle on the beach yesterday.

(b) The quick brown fox jumps over the lazy dog.

(c) Despite the initial hiccups caused by the unexpected power outage at the Bengaluru Tech Summit, the Infosys team's never say die attitude ensured they completed their presentation to potential investors on time.

(d) While trekking in the Munnar Hills of Kerala, Rajiv Sharma accidentally revealed the secret about Priya's surprise Diwali party planned for next week.

9. Analyze sentences (a) and (b) using morphological analysis, POS tagging, chunking.

10. Identify Multi-Word Expressions from sentences (c) and (d) and annotate the type.

11. Identify Named Entities from sentences (c) and (d) and annotate the type.

12. Draw X-bar schema for sentences (a) and (b).

13. Draw Dependency Tree with syntactic relations for sentences (a) and (b).

/End