International Institute of Information Technology, Hyderabad

Subject: **CL3.101 Computational Linguistics**

Mid Semester Examination

Max. Time: 1 ½ Hours **Max. Marks: 20**

Instructions:

- This exam paper consists of two sections: Section A and Section B. Answer both Sections.
- Section A carries 12 marks and Section B carries 8 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 Analytical Questions

There are six questions. Answer any FOUR questions. **(4 * 3 = 12 marks)**

1. Given the provided data, identify and explain the tokenization issues present. Discuss how different tokenization approaches might affect the analysis of the text.

```
Dr. Jean-Luc Picard M.D., Ph.D., isn't sure if he'll attend the
conference on artificial intelligence@DS9. He's worried about his
flight, which is scheduled for 9:30 a.m., and whether he'll have time
to visit his colleague, Prof.David De'Cruz, before the event.
```

2. (i) Formulate a regular expression to match the provided words.

```
declare, decide, decision, delete, deviate, dawdle
```

(ii) Provide potential words (minimum 10 words) that could match the given regular expression.

```
^st*([a-z]+)([ptkl]?)$
```

3. Compare and contrast the IA (Item-and-Arrangement) model, IP (Item-and-Process) model, and WP (Word and Paradigm) model for the paradigm of the verb "eat". Based on your analysis, identify the most suitable morphological model for explaining the inflections observed in the verb "eat" across its different grammatical categories.

4. Construct a Finite State Transducer (FST) that represents the morphological variations of the word "walk" along with its derived forms: "walks," "walked," "walking," "walker," and "walkathon."

5. Choose any TWO of the following linguistic structures and provide detailed explanations of each, highlighting their significance and impact on language comprehension and analysis.

(i) Garden path sentences (ii) Structural ambiguity (iii) Recursiveness (iv) Ellipsis

6. Identify different clauses present in the paragraph and explain how they contribute to the overall description of the discourse.

(After the rain stopped) (revealing a rainbow (stretching across the sky)), (the children ran outside) (to catch a glimpse of its vibrant colors). (They pointed excitedly at the rainbow) (which arched gracefully over the treetops). (They wondered aloud) (if they could reach the end of the rainbow) (what treasures they would find).

Section-B **Data Annotation** **(1 * 8 = 8 marks)**

7. *Tokenize* the provided text and identify *Lemma (rootword)*, *parts of speech (POS)*, and *chunk* the text.

- Choose one language from the options provided.
- You are also allowed to translate the English version (given as language 1) into any language you are familiar with and do the task.
- Provide the annotation in tab-separated/table format.
- Use BIO format for chunking. B denotes the beginning of a chunk, I denotes inside a chunk, and O denotes outside a chunk.
- Example annotation for the sentence `I saw the children.` is given here:

Token No	Token	Lemma	POS	Chunk
1	I	I	PRP	B-NP
2	saw	see	VF	B-VGF
3	the	the	DET	B-NP
4	children	child	NN	I-NP
5	.	.	PUNC	O

- Use BIS tagset for POS and Chunking.

BIS-POS tags: Common Noun (NN); Proper Noun (NNP); Noun of Space and Time (NST); Pronoun (PR); Personal (PRP); Reflexive (PRF); Relative (PRL); Reciprocal (PRC); Wh-word (PRQ); Demonstrative (DM); Main Verb(VM); Finite Verb (VF); Non-finite Verb (VNF); Infinitive (VINF); Gerund (VNG); Auxiliary (VAUX); Adjective (JJ); Adverb (RB); Postposition (PSP); Conjunction (CC); Coordinator (CCD); Subordinator (CCS); Quotative (UT); Particles (RP); Classifier (CL); Interjection (INJ); Intensifier (INTF); Negation (NEG); Quantifiers (QT); Residuals (RD); Foreign word (RDF); Symbol (SYM); Punctuation (PUNC); Unknown (UNK); Echowords (ECH)

BIS-Chunk tags: NP, VGF, VGINF, VGNN, VGNF, JJP, ADP, NEGP, CCP, FRAGP, BLK

Language 1: English

Deepawali is a festival of light, joy and self-realization. Lets reiterate our resolve to protect the environment on the occasion of Deepotsav celebrated on the day of Amavasya. Plastic waste is against the spirit of this festival of cleanliness and purity. We should illuminate this Deepawali with freedom from plastic waste and pollution. Ever wondered how metal objects, idols and decoration items are prepared?

Language 2: Telugu

దీపావళి వెలుగులు, ఆనందం, ఆత్మ సాక్షాత్కారాల పండుగ. అమావాస్య నాడు జరిగే దీపోత్సవం నాడు పర్యావరణ పరిరక్షణ సంకల్పాన్ని మనం మరోసారి మననం చేసుకుందాం. ప్లాస్టిక్ వ్యర్థాలు శుభ్రత, స్వచ్ఛతల పండుగ స్ఫూర్తికి విరుద్ధం. ఈ దీపావళి నాడు మనం ప్లాస్టిక్ వ్యర్థాలు, కాలుష్య విముక్తి కోసం దీపాలు వెలిగిద్దాం. లోహాలతో తయారుచేసే వస్తువులు, విగ్రహాలు, అలంకరణ వస్తువులు ఎలా తయారవుతాయో మీకు తెలుసా?

Language 3: Hindi

दीपावली का पर्व ज्योति, पुरुषार्थ और आत्म साक्षात्कार का पर्व है। अमावस्या के दिन मनाए जाने वाले दीपोत्सव पर दें पर्यावरण रक्षा के संकल्प को मजबूती । प्लास्टिक कचरा, स्वच्छता और पवित्रता वाले इन पर्वो की भावना के खिलाफ हैं। हमें इस दीपावली को प्लास्टिक और प्रदूषण मुक्त पर्यावरण के नाम उजागर करना चाहिए। कभी सोचा है धातुओं की वस्तुएं, मूर्ति, सजावट के सामान कैसे तैयार होते हैं?

/end