Printed Pages:02 **Sub Code:KCS072**

Paper Id: 2 3 1 3 5 1 **Roll No.**

B.TECH.
(SEM VII) THEORY EXAMINATION 2022-23
NATURAL LANGUAGE PROCESSING

Time: 3 Hours ***Total Marks: 100***

Note: Attempt all Sections. If you require any missing data, then choose suitably.

SECTION A

1. Attempt *all* questions in brief. 2x10 = 20

(a) Explain Language Modelling.
(b) What are the different issues and problems in NLP?
(c) Explain Treebank Corpus.
(d) Differentiate between syntax and semantics.
(e) Explain Context Free Grammars.
(f) How spelling corrections work? Explain.
(g) Explain Frequency and Amplitude.
(h) Explain transcription.
(i) What is Log–Spectral Distance?
(j) Explain TF and IDF.

SECTION B

2. Attempt any *three* of the following: 10x3 = 30

(a) We are given the following corpus:
<s> I am Sam </s>
<s> Sam I am </s>
<s> I am Sam </s>
<s> I do not like green eggs and Sam </s>
Using a bigram language model with Laplace smoothing, what is P(Sam |am)? Include <s> and </s> in your counts just like any other token.

(b) Explain Context Free Grammar. Draw tree structure for the following sentence: *I need to fly between New Delhi and Mumbai.*

(c) Explain nearest neighbour algorithm using contextual word embeddings. Take example to explain.

(d) **Write short notes on:**
(i) Short-Time Fourier Transform
(ii) Linear Predictive Coding (LPC).

(e) Explain Viterbi Search Algorithm with an example.

SECTION C

3. Attempt any *one* part of the following: 10x1 = 10

(a) Write an algorithm for parsing a finite-state transducer using the pseudo code. Explain the algorithm with an example.

(b) Differentiate between Bottom-up and Top-down parsing. What algorithms are used for each of these types of parsing?

4. Attempt any *one* part of the following: 10 x1 = 10

(a) Draw a tree structure for the following statement:
Rachael Ray finds inspiration in cooking her family and her dog
Write the Context Free Grammar (CFG) for the above statement.

(b) Explain Cocke-Kasami-Younger (CKY) algorithm. Explain it with an example.

5. Attempt any *one* part of the following: 10x1 = 10

(a) Explain about Word Sense Disambiguation and Distributional semantics. Also explain about the path length problem in with an example.

(b) Define the issues in Information Content-BasedSimilarity Measures with example. Elaborate on the approachused in similarity measure.

6. Attempt any *one* part of the following: 10x1 = 10

(a) Define Articulatory Phonetics. Also explain production and classification of speech sounds in detail. Give examples.

(b) Write regular expressions for the following languages.
(i) the set of all alphabetic strings;
(ii) the set of all lower-case alphabetic strings ending in a b;
(iii) the set of all strings from the alphabet a,b such that each a is immediately preceded by and immediately followed by a b;
Explain how these regular expressions will be used in speech processing.

7. Attempt any *one* part of the following: 10x1 = 10

(a) Explain different feature extraction and pattern comparison techniques used in speech analysis. Also elaborate on likelihood distortions, and spectral distortion, in detail.

(b) Explain hidden Markov model with Baum-Welch parameter re-estimation. Also elaborate on its implementation issues.