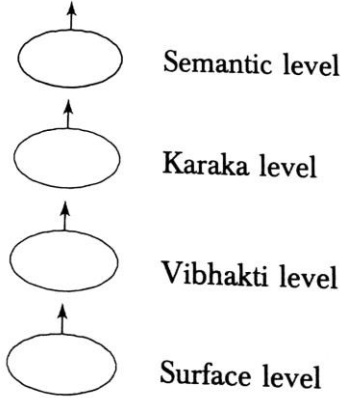


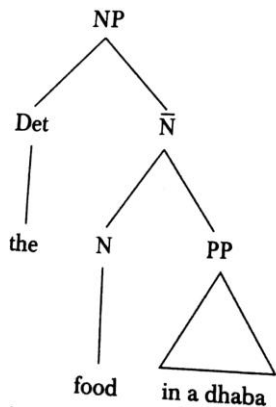
Internal Assessment Test 1 – November 2021

Sub:	Natural Language Processing				Sub Code:	18CS743	Branch:	ISE		
Date:	13/11/2021	Duration:	90 min's	Max Marks:	50	Sem/Sec:	VII A, B & C	OBE		
Answer any FIVE FULL Questions								MARKS	CO	RBT
1.	Define Natural Language processing. What are its purposes? List and explain different levels of processing involved in it.						[10]	CO2	L1	
	<p>It is the understanding and generation of Natural languages by generating computational models of natural languages.</p> <p>Purposes:</p> <ol style="list-style-type: none"> To develop automated tools for language processing To gain a better understanding of human communication <p>Different Levels of processing:</p> <ol style="list-style-type: none"> Lexical Analysis: It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of text into its small units. Syntactic Analysis: It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. It creates an appropriate order words and symbols. Semantic Analysis: Establishes the exact meaning of the sentence by linking or mapping one interface to another. Established the relationship between words, phrases, signs and symbols. Pragmatic Analysis: It involves the understating the sentence or other text in the context of overall world knowledge. It established the actual meaning of the sentence. 									
2.	“Processing Indian languages using NLP is far more challenging”, justify the statement with suitable examples.						[10]	CO2	L2	
	<p>Processing Indian languages are far more challeing as Indian languages differes from English in more than many ways:</p> <p>Differences between Indian languages and English</p> <ol style="list-style-type: none"> Indic scripts have a non-linear structure. Unlike English, Indian languages have SOV(subject-object-Verb) as default sentence structure. Indian languages have a free word order i.e words within sentence can be freely moved without changing the meaning of the sentence. Rich set of morphological variants as languages have evolved over centuries. Indian Languages uses post-positions case markers instead of pre-positions. Indian languages makes extensive and productive use of complex predicates. Indian languages use verb complexes consisting of sequence of verbs. Auxiliary verbs provide information about tense, aspect , modality. 									

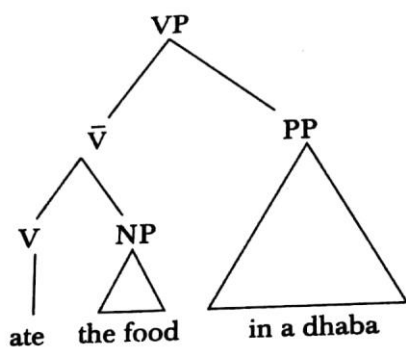
3.	<p>Explain Karaka Theory of Paninian Grammar (PG). Identify different karaka's in the following Hindi sentence:</p> <p>“Maa ne aangan me thaali se khaana uthakar ko bachche ko diyaa”</p>	[10]	CO1	L3
	<p>Levels of Paninian Grammar:</p>  <p>Karaka literally means CASE, these case relations are based on the way the word group participates in the activity denoted by the verb group. Karaka relations are assigned based on the roles players by various participants in main activity.</p> <p>Various karaka's are (case marker in hindi)</p> <ol style="list-style-type: none"> 1. Karta (Subject) case marker: 'ne' or Φ 2. Karma (Object) case marker: 'ko' or Φ 3. Karana (instrument) case marker: 'dwara' or 'se' 4. Sampradana (Beneficiary) case marker: 'ko' or 'ko liye' 5. Apadana (Seperation) case marker: part that serves as separation 6. Adhikaran (Locus) case marker: (support in space or time) 7. Sambandh (Relation) 8. Tadarthya (Purpose) 			
4.	<p>What is Statistical Language Modeling and explain the features of n-gram model. Find the probability of the test sentence using the corpus given below using bi-Gram modeling:</p> <p>Test Sentence:</p> <p>The Arabian Knights are collection of fairy tales</p> <p>Corpus:</p> <p>The Arabian Knights</p> <p>These are collection of the fairy tales of the east</p> <p>The stories of the Arabian Knights are translated in many languages</p>	[10]	CO1	L3
	<p>n-Gram modelling:</p> <p>n-Gram predicts the probability of a word by considering all the previous words by the conditional probability given previous n-1 words.</p> $P(w_i/h_i) \approx P(w_i/w_{i-n+1} \dots w_{i-1})$ <p>It makes use of the markov model, if the model limits the previous words to one only then it is known as bi-gram model. Proability of a sentence is the product of bi-gram probability of all words in it, which is given as below:</p> $P(s) \approx \prod_{i=1}^n P(w_i/w_{i-1})$			

5.	<p>Define Binding and Explain the process of Pronoun resolution using Binding principles with suitable example.</p> <p>Binding: A Binds B iff</p> <ul style="list-style-type: none"> b. A C-commands B and c. A and B are co-indexed <p>Binding theory defines relationship between NPs</p> <ul style="list-style-type: none"> a. An anaphor is bound in its governing category b. A pronominal is free in its governing category. c. An R-expression is free. <p>Write suitable example for each</p>	[10]	CO1	L2
6.	<p>What is X-Bar theory? Write the general phrase and sentential structure of X-bar theory, apply the same for the following sentences:</p> <p>“ate the food in a dhaba”</p>	[10]	CO1	L3
	<p>X-bar theory makes the claim that every single phrase in every single sentence in the mental grammar of every single human language, has the same core organization. X-Bar theory is one of the central concept in GB theory. Instead of defining several phrase structures and the sentence structure with separate set of rules, X-bar theory defines them both as maximal projections of some head X. In this manner entities become independent of language. Here’s a tree diagram that shows the general phrase structure of X-bar theory. Noun Phrase(NP), Verb Phrase (VP), Adjective Phrase(AP), and Prepositional Phrase(PP) are maximal projection of Noun(N), Verb (V), Adjective (A) and Preposition(P) respectively.</p> <p>The projection is at two levels – first projection of head at the semi-phrasal level denoted by X’ and then the second maximal projection at the phrasal level denoted by X’’.</p> <div style="text-align: center;"> <pre> graph TD X_double["X̄"] --- Specifier X_double --- X_single["X̄"] X_double --- Modifier X_single --- Head["X"] X_single --- Argument </pre> </div> <p>Example:</p> <p>Sentence:</p>			

1. NP: the food in a dhaba
 $[_{NP} the[_N food]_{PP} [in a dhabha]]$



2. VP: ate the food in a dhaba
 $[_{VP} \bar{V}[_V ate] [_{NP} the food]] [_{PP} in a dhaba]]$



7. Define morphology. Explain the 3 ways of word formation and what are the information sources used in morphological parsing?

[10]

CO2

L1

Morphology is a sub-discipline of linguistics. It studies word structure and the formation of words from smaller units known as morphemes. The goal of the Morphology parsing is to discover the morphemes that build a given word. There are two broad classes of morphemes called stems and affixes.

3 ways of word formation:

- a. Inflection
- b. Derivation
- c. Compounding

Inflection: In inflection root word is combines with a grammatical morpheme to yield a word of the same class as the original stem.

Ex: Egg and Eggs, sing and singing

Derivation: In derivation root word is combined with a grammatical morpheme to yield a word belonging to a different class.

Ex: Compute and Computation

Compounding: Compounding is the process of merging two or more words to form a new word.

Ex: Desktop, Overlook

Information source used in morphological parser:

- a. **A Lexicon:** A lexicon is a list of stems and affixes together with basic information about them.
- b. **Morphotactics:** It deals with the ordering of the morphemes. It describes the way morphemes are arranged or touch each other.

	<p>Ex: Rest-less-ness not rest-ness-less</p> <p>c. Orthographic rules: Spelling rules that specify the changes that occur when two given morphemes combine. Ex: y->ier i.e easy+ ier= easier</p>			
8.	<p>Define the following terms:</p> <p>a. C-Commands</p> <p>b. Θ-Criterion</p> <p>c. AVM (Attribute Value Matrix)</p> <p>d. Add-one smoothing</p>	[10]	CO1	L2

Faculty Signature
Signature

CCI Signature

HOD