SRN | | | | | | | | | | | |

PES UNIVERSITY

PES University, Bangalore

(Established under Karnataka Act No. 16 of 2013)

UE18CS334

Aug 2021: END SEMESTER ASSESSMENT (ESA) B.TECH VI SEMESTER

UE18CS334: NATURAL LANGUAGE PROCESSING

Time: 3 Hrs | Answer All Questions | Max Marks: 100

1.	a)	Show any three possible alignments for converting **STALL** to **TABLE** indicating operations? Can we have more than one optimal alignment? Use minimum edit distance algorithm to find the optimal alignment and the cost involved. (Consider the cost of insertion and deletion as 1 and substitution cost as 2). The entries in the table must be explained properly.	3+1 +6
	b)	Give one example of non-word error and real-word error each. Justify the statement "The noisy channel model is a kind of Bayesian Inference."	4
	c)	Define the following with an example: 1. BOW model 2. Term Frequency-Inverse document Frequency 3. Lexical Ambiguity	2*3
2.	a)	Consider a corpus D = {New, higher, education, policy} of four words. And following are the counts of trigrams, bigrams and unigrams in the corpus: New higher new 0; New higher higher 0; New higher education 2; New higher policy 2 Higher new 0; Higher higher 0; Higher education 2; Higher policy 2 New 2; higher 3; education 3; policy 2 Use the following Backoff (BO) smoothing to find the P_{BO}(w | new higher) where $w \in D$. Note that Trigram and bigram version of backoff is defined as $$P_{BO}(z|x,y) = \begin{cases} P^*(z|x,y) & if\ c(x,y,z) > 0 \\ \alpha(x,y)P_{BO}(z|y) & else\ if\ c(x,y) > 0 \\ P^*(z) & otherwise \end{cases}$$ $$P_{BO}(z|y) = \begin{cases} P^*(z|y), & c(y,z) > 0 \\ \alpha(y)P^*(z) & otherwise \end{cases}$$	6

SRN | | | | | | | | | | | |

Where $P^*(t) = P(t) - \frac{1}{8}$ where t can be a unigram, bigram or trigram etc.

And $P(t) = \frac{c(t)}{\sum c(t)}$

Where $\alpha(x, y)$ is a parameter that passes the left-over mass to the lower order n-grams and can be computed by *normalizing the probabilities* (ie the sum of all probabilities of trigrams would be one and so is for bigram and unigram.)

b) Consider two words, "**Tagore**" and "**wrote**". In the transition table, rows represent the conditioning event, P(Vb|N)=0.40. <s> is "start of the sentence". 2+3

Construct the Markov model graphically. Also, determine the POS tags assigned to the sentence "Tagore wrote" using an HMM. The tag transition distribution and the word likelihood distribution are given below. In the transition table, rows represent the conditioning event, i.e P(Vb|N)=0.40. <s> is "start of the sentence". Show all steps neatly, either by using a matrix or on the graph.

	N	Vb
<s>	0.60	0.40
N	0.10	0.40
Vb	0.50	0.20

	Tagore	wrote
N	0.50	0.20
Vb	0.10	0.30

c) State whether true or false: 5

1. Maximum Entropy Markov Model and Maximum Entropy models are both sequence labeling model.
2. In a MEMM, the posterior probability P(T|W) is computed directly in contrast to HMM where Likelihood is estimated, where W is a sequence of words and T is a sequence of tags.
3. Conditional Random fields uses per-state exponential model for the conditional probabilities of next states given the current state.
4. Label bias problem occurs because the transitions leaving a given state compete globally.
5. For Named Entity Recognition, one has to label the spans of the text so. As to identify the named entities.

d) Discuss in brief about the classic tasks of information extraction. 4

3. a) Given a mini grammar: 8

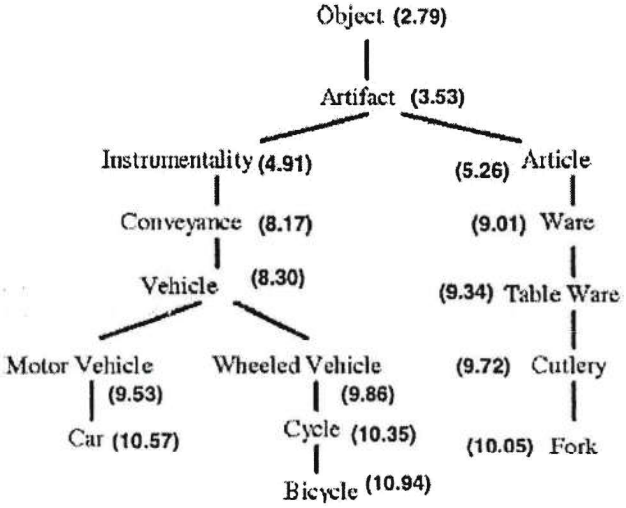
$S \rightarrow NP\ VP$	[0.80]	$Det \rightarrow the$	[0.40]
$NP \rightarrow Det\ N$	[0.30]	$Det \rightarrow a$	[0.40]
$VP \rightarrow V\ NP$	[0.20]	$N \rightarrow sportsman$	[0.01]
$V \rightarrow ran$	[0.05]	$N \rightarrow mile$	[0.02]

The values in square brackets represent the probability associated with the respective rules. Apply Probabilistic CKY algorithm to parse the following sentence:

"The sportsman ran a mile"

Construct the table and show the steps neatly.

SRN | | | | | | | | | | | |

	b)	Why dependency parsing is required when phrase structures are generated using CFGs? With an example discuss how dependency relations are defined? When a dependency tree is said to be projective?	2+3 +1
	c)	Consider the following text: "*Swastika M says she will always be grateful to Amartya S. The actress revealed that the Nobel Laureate helped her calm down when she was scared by a thunderstorm while travelling on a plane.*" 1. Detect all the mentions in the above text. 2. Identify the coreferential chains. 3. What is the result of setting Anaphoricity classifier's threshold too low? 4. Is there any nested mention in the above text?	1+3 +1+ 1
4.	a)	Suppose the figure below is Wordnet Noun Taxonomy with their Information Content (IC) in brackets:  What is the 1. IC of LCS(Car, Cycle)? 2. Resnik Similarity between "Bicycle" and "Fork"? 3. Jiang-Conrath Similarity between "Car" and "Article"? 4. Lin Similarity between "Vehicle" and "Tableware"? What are the two classes of algorithms to measure the word similarity?	1+2 +2+ 2+1
	b)	Do as directed. 1. Give a clear example of semantic relatedness but no semantic similarity between the terms. 2. Distributional Algorithms use the structure of the dictionary to define the word similarity. (True/False) 3. A _____feature vector extracted from a window of three words to right and left of the target word (in bold) is [He, PRN, was, VBD, playing, VB, well, ADV, and, CC, expected, VBD] for the sentence "He was playing **superbly** well and expected to enter National Level team." 4. _____is one way to build lexicons. 5. What is the difference between a term-document matrix and term-term matrix?	5

SRN | | | | | | | | | | | |

	c)	What is Skip Gram with negative sampling approach? Show that $$L_{CE} = -[\log \sigma(c_{pos} \cdot w) + \sum_{i=1}^{k} \log \sigma(-c_{neg_i} \cdot w)$$ Also find the derivative of L_{CE} wrt w and interpret it. (L_{CE} is the loss function for one word w with one positive concept c_{pos} and i negative concepts c_{negi} and $\sigma(c.w) = \frac{1}{1+e^{-c.w}} = P(+\|w,c)$. P(+|w, c) is the prob that c is the real context word of w.)	2+2 +2+ 1
5.	a)	How Recurrent Neural networks (RNNs) are used as language models (LMs)? Draw the figure showing the training regime of RNNs as LMs.	4+2
	b)	With a neat diagram, discuss the architecture of a Transformer and how it is used in BERT?	2+2
	c)	Consider the following 4X4 matrix and a 2X2 kernel What is the output feature map after convolution operation if stride=2? And what would be the order of output matrix if stride=1? 10 15 9 14 16 2 17 11 12 10 3 18 5 9 16 13 1 1 0 1 Briefly indicate how convolutional networks are used at character level for text input.	2+1 +3
	d)	Answer the following and rectify the statement in case it is false. 1. Language models before BERT were unidirectional. (True/False) 2. ULMfit is also a transformer-based LM. (True/False) 3. Neural Factoid question answering is a single stage process of ---------- and ---------- .	4