End Semester Examination	
Max. Marks: 60	Duration: 2Hr 10 min.
Class: SYMCA	Semester: IV
Course Code: MCAE45B	Date: 23/05/2021
Subject: **Natural Language Processing**	Time: 8.00 PM to 10.10 PM

Instructions: (1) All questions are compulsory.

(2) Use of scientific calculator is allowed.

(3) Assume any necessary data but justify the same.

Q. No.	Questions	Max Marks	CO-BL-PI
	Question Number- 2		
1	**You are building a model distribution for an infinite stream of word tokens. You know that source of this stream has a vocabulary of size 1000. Out of these 1000 words you know of 100 words to be stop words each of which has a probability of 0.0019. With only this knowledge, discover the maximum possible entropy of the modeled distribution. (Use log base 10 for entropy calculation)** **OR** **What is parsing? Explain Top Down and Bottom Up Approach of parsing with suitable examples.**	6 M	3-4-2.3.2
2	**Explain with suitable examples following relationship between word meanings:** **1. Homonymy** **2. Polysemy** **3. Hypernymy** **4. Hyponymy** **5. Antonymy**	5 M	1-2-2.2.3
3	**How Baum Welch algorithm is useful in Hidden Markov Models?**	4 M	3-3-2.2.2
	Question Number- 3		
1	**Consider for the sentence "SEE BOB IN AUSTIN", PCFG given below:**	6 M	2-3-2.3.2

S -> NP VP	0.6	PropNoun -> DALLAS	0.2
S -> VP	0.4	PropNoun -> ALICE	0.2
NP -> NP PP	0.4	PropNoun -> BOB	0.3
NP -> PropNoun	0.6	PropNoun -> AUSTIN	0.3
VP -> Verb	0.3	Verb -> ADORE	0.5

VP -> Verb NP	0.3	Verb -> SEE	0.5
VP -> VP PP	0.4	Prep -> IN	0.4
PP -> Prep NP	1.0	Prep -> WITH	0.6

Using the given PCFG, Identify its equivalent CNF.

2	**What are the major tasks executed using Distributional semantics? Explain any one of them.** **OR** **What is Semantic Analysis? Why Semantic Analysis is difficult? In list various approach to perform Semantic Analysis.**	4 M	4-3-3.4.3
3	**Define the topic modeling? Explain its use in terms of Historic archaeology studies.**	5 M	4-2-2.4.1

Question Number- 4

1	**Suppose you are using Gibbs sampling to estimate the distributions, θ and β for topic models. The underlying corpus has 3 documents and 5 words, {rain, satellite, forecast, weather, image} and the number of topics is 2.** **At certain point, the structure of the documents looks like the following.** **Doc1: weather[1] forecast[1] image[1] forecast[1] weather[1] weather[1] forecast[1] image[1]** **Doc2: weather[1] forecast[1] forecast[2] rain[2] image[1] satellite[2] forecast[1] weather[1]** **Doc3: rain[2] forecast[2] satellite[2] forecast[2] rain[2] rain[2] satellite[2] forecast[2]** **Superscripts 1 and 2 denote whether the word is currently assigned to topics T1 and T2 respectively. ŋ= 0.2 and α = 0.1.** **Use the above structure to estimate $\beta^{(2)}_{weather}$ and $\beta^{(1)}_{weather}$ and $\theta^{doc2}t1$ and $\theta^{doc2}t2$ at this point.**	6 M	2-5-2.4.1
2	**Write a short note on: (Any Three)** 1. **Text summarization** 2. **N-gram language Model** 3. **Wordnet** 4. **Aspect based sentiment analysis**	9 M	1-2-2.2.2

Q. No.	Questions	Max Marks	CO	BL	PI
Q. 1	Explain how AI follow "act rationally" approach	2	CO 1	L3	1.3.1
Q. 1	How AI Adaption will be useful	2	CO 1	L3	1.3.1

Q. 3	List the how 5 components of AI with reference to Tic–Tac– Toe Game	5	CO 1	L3	1.3.1
Q. 4	Formulate the Problem for "Shortest Path Algorithm, using defined concepts for problem analysis	3	CO 1	L3	1.3.1 2.1.2
Q. 5	How do ANN works in similar way to BNN	5	CO 2	L3	1.3.1 2.1.2
Q. 6	Why activation function is important in ANN	3	CO 2	L2	1.3.1 3.1.1