



**Sample Question Format**

**(For all courses having end semester Full Mark=50)**

**KIIT Deemed to be University**  
**Online End Semester Examination(Autumn Semester-2021)**

**Subject Name & Code:**

Natural Language Processing (NLP)-IT-3035

**Applicable to Courses: B.Tech**

**Full Marks=50**

**Time:2 Hours**

**SECTION-A(Answer All Questions. Each question carries 2 Marks)**

**Time:30 Minutes**

**(7×2=14 Marks)**

<b><u>Question No</u></b>	<b><u>Question Type (MCQ/SAT)</u></b>	<b><u>Question</u></b>	<b><u>CO Mapping</u></b>	<b><u>Answer Key (For MCQ Questions only)</u></b>
<b><u>Q.No:1</u></b>	In linguistic morphology  _____ is the process for reducing inflected words to their root form. a) Rooting b) Stemming c) Text-Proofing d) Both Rooting & Stemming	Question -1 on concept 1	CO1	b
	Which of the following is(are) application(s) of NLP ?  a. Email Filter b. Alexa c. Google Translate d. All of the Above	Question -2 on concept 1	CO1	d
	Which of the following is an	Question -3 on concept 1	CO1	a

	<p>application of NLP ?</p> <p>a. Text analytics b. Google Assistant c. Cortana d. Speech Recognition</p>			
	<p>_____ can resolve the meaning of words having more than one meaning.</p> <p>a. POS Tagging b. HMM Model c. Word Sense Disambiguation d. None of these</p>	Question -4 on concept 1	CO1	c
<b><u>Q.No:2</u></b>	<p>In the spam detection case, false positives are documents</p> <p>a. that are indeed spam but our system incorrectly labeled as non-spam b. that are indeed non-spam that our system correctly labeled as non-spam c. that are indeed spam that our system correctly said were spam d. that are indeed non-spam that our system correctly said were spam</p>	Question -1 on concept 2	CO2	d
	<p>In the spam detection case, false negatives are documents</p>	Question -2 on concept 2	CO2	a

	<p>a. that are indeed spam but our system incorrectly labeled as non-spam</p> <p>b. that are indeed non-spam that our system correctly labeled as non-spam</p> <p>c. that are indeed spam that our system correctly said were spam</p> <p>d. that are indeed non-spam that our system correctly said were spam</p>			
	<p>The assumption that the probability of a tag is dependent only on the previous tag, rather than the entire tag sequence is known as</p> <hr/> <p>—.</p> <p>a. Unigram assumption</p> <p>b. Bigram assumption</p> <p>c. Trigram assumption</p> <p>d. N-gram assumption</p>	Question -3 on concept 2	CO2	b
	<p>The assumption that the probability of a tag is dependent only on the previous 2 tags, rather than the entire tag sequence is</p>	Question -4 on concept 2	CO2	c

	<p>known as</p> <hr/> <p>—.</p> <p>a. Unigram assumption b. Bigram assumption c. Trigram assumption d. N-gram assumption</p>			
<b>Q.No:3</b>	<p>To keep a language model from assigning zero probability to the unseen events is a concept known as</p> <hr/> <hr/> <p>.</p> <p>a. Stemming b. Normalizing c. Smoothing d. Perplexity</p>	Question -1 on concept 3	CO3	c
	<p>A language model can achieve low perplexity by choosing a</p> <hr/> <p>vocabulary and assigning the unknown word a</p> <hr/> <p>probability.</p> <p>a. Large, Low b. Small, High c. Large, High d. Small, Low</p>	Question -2 on concept 3	CO3	b
	<p>The parameters of an N-gram model are estimated by getting counts and then the counts are normalized</p>	Question -3 on concept 3	CO3	c

	<p>so that they lie between 0 and 1 in the following methodology :</p> <p>a. Hidden Markov Model b. Multinomial Bayes estimate c. Maximum Likelihood estimate d. Chain Rule of Probability</p>			
	<p>_____ models are the class of probabilistic models that assume we can predict the probability of some future unit without looking too far into the past.</p> <p>a. Bayes Learning b. Maximum Likelihood c. Markov d. Named Entity Recognition</p>	Question -4 on concept 3	CO3	c
<b><u>Q.No:4</u></b>	<p>_____ is a probabilistic sequence model that given a sequence of units computes a probability distribution over possible sequences of labels and choose the best label sequence.</p> <p>a. POS-tagging b. Classification</p>	Question -1 on concept 4	CO4	c

	c. Hidden Markov Model d. Sentiment Analyser			
	Spam Detection is an important commercial application of _____.  a. Text Standardization b. Text Categorization c. Text Summarization d. Text Generation	Question -2 on concept 4	CO4	b
	Examples of sentiment analysis are :  a. Review of a movie b. Extracting consumer preference c. Editorial Text toward a political candidate d. All of these	Question -3 on concept 4	CO4	d
	Naive Bayes classifier is an example of _____.  a. Generative classifier b. Discriminative classifier c. Multinomial classifier d. None of these	Question -4 on concept 4	CO4	a
<b>Q.No:5</b>	The probability of seeing the observations from time $t + 1$ to the end, given that we are in	Question -1 on concept 5	CO5	b

	<p>state <math>i</math> at time <math>t</math> (and given the automaton <math>\lambda</math>):</p> <p>a. Forward probability b. Backward probability c. Forward-backward probability d. Viterbi-path probability</p>			
	<p>The task of determining which sequence of variables is the underlying source of some sequence of observations is called the</p> <hr/> <p>—.</p> <p>a. Encoding b. Likelihood c. Decoding d. Learning</p>	Question -2 on concept 5	CO5	c
	<p>In</p> <hr/> <p>_____ each probability expresses likelihood of an observation being generated from a state.</p> <p>a. Transition probability b. Forward probability c. Emission probability d. Backward probability</p>	Question -3 on concept 5	CO5	c
	<p>Observation likelihoods in an Hidden Markov</p>	Question -4 on concept 5	CO5	d

	<p>Model are also known as</p> <hr/> <p>— .</p> <p>a. Transition probabilities b. Forward probabilities c. Backward probabilities d. Emission probabilities</p>			
<b><u>Q.No:6</u></b>	<p>A</p> <hr/> <p>grammar expresses the ways that symbols of the language can be grouped and ordered together.</p> <p>a. Generative grammar b. Non-generative grammar c. Context-free grammar d. None of these</p>	Question -1 on concept 6	CO6	c
	<p>A context-free grammar is in</p> <hr/> <p>form if each production is either of the form <math>A \rightarrow BC</math> or <math>A \rightarrow a</math>.</p> <p>a. Baker Naus Form b. Chomsky Normal Form c. 2-Naus Form d. 3-Naus Form</p>	Question -2 on concept 6	CO6	b
	<p>A sentence has</p> <hr/> <p>ambiguity if a</p>	Question -3 on concept 6	CO6	b



	<p>particular constituent can be attached to the parse tree at more than one place.</p> <p>a. Coordination b. Attachment c. Lexical d. Scope</p>			
	<p>In _____, instead of choosing the best token to generate at each timestep, we keep k possible tokens at each step.</p> <p>a. Forward-backward b. CKY parsing c. Beam search d. Semantic Analysis</p>	Question -4 on concept 6	CO6	c
<b><u>Q.No:7</u></b>	<p>Rule-based Machine Translation comprises of following two types :</p> <p>a. Statistical-based and Transfer-based b. Interlingua-based and Phrase-based c. Statistical-based and Phrase-based d. Interlingua-based and Transfer-based</p>	Question -1 on concept 7	CO6	d
	Statistical	Question -2 on concept 7	CO6	d

	<p>Machine Translation (SMT) learns _____.</p> <p>a. from data and patterns b. patterns c. translational patterns d. mappings of strings of words and their probabilities</p>			
	<p>Which type of Machine Translation produces a representation that any language can work with ?</p> <p>a. SMT - Statistical Machine Translation b. EBMT - Example Based Machine Translation c. RBMT - Rule-based Machine Translation d. KBMT - Knowledge-based Machine Translation</p>	Question -3 on concept 7	CO6	c
	<p>Example-based Machine Translation (EBMT) learns _____.</p> <p>a. from probabilities b. rules c. translational patterns</p>	Question -4 on concept 7	CO6	c

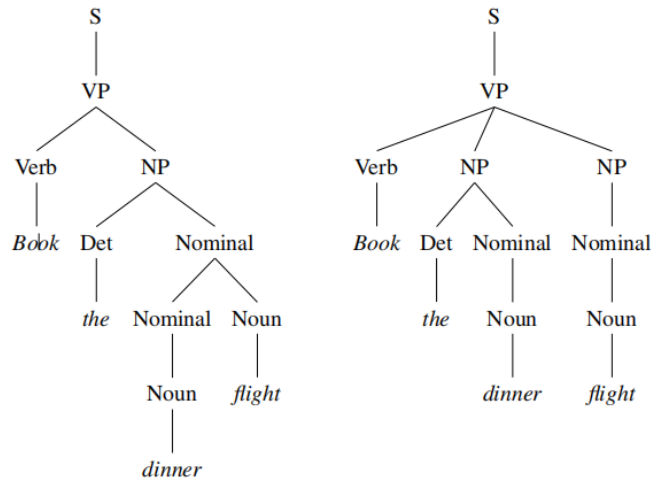
	d. mappings of strings of words and their probabilities			
--	---------------------------------------------------------	--	--	--

**SECTION-B(Answer Any Three Questions. Each Question carries 12 Marks)**

**Time: 1 Hour and 30 Minutes**  
**(3×12=36 Marks)**

<b><u>Question No</u></b>	<b><u>Question</u></b>	<b><u>CO Mapping (Each question should be from the same CO(s))</u></b>
<b><u>Q.No:8</u></b>	<b><u>Q.No:8-1st question</u></b> A. Discuss forward and backward probabilities used in Hidden Markov Model with the help of equations and examples. [8]  B) What is meant by Lexicon? How is it useful in NLP? [4]	CO1 and CO2
	<b><u>Q. No:8-2nd question</u></b> A. Derive Forward-backward algorithm in Hidden Markov Model. [8]  B. Differentiate between open class and closed class of words. [4]	
	<b><u>R. No:8-3rd question</u></b> A. Explain Viterbi algorithm with an example. [8]  <b>B.</b> Define NLP. List any two applications of NLP. [4]	
<b><u>Q.No:9</u></b>	<b><u>Q.No:9-1st question</u></b>  A. Explain POS tagging with an example. [8] B. Identify and describe the ambiguities in the following sentences: i. The man kept the dog in the house.	CO2 and CO3

	ii. Book that flight iii. Time flies like an arrow iv. He crushed the key to my heart [4]	
	<b>Q.No:9-2nd question</b>  A. Explain why CFG is used to represent natural language in parsing. [8] B. Identify the morphological type (Noun phrase, Verb Phrase, Adjective Phrase) of following sentence segments 1. important to Bill 2. looked up the tree [4]	
	<b>Q. No:9-3rd question</b>  A. Explain the different levels of language analysis. [6] B. Identify the head and morphological type (Noun Phrase, Verb Phrase, Adjective Phrase, Adverbial Phrase) of the following sentence segments. i. The president of the company ii. Looked up the chimney iii. Angry as a hippo iv. Rapidly like a bat [6]	
<b><u>Q.No:10</u></b>	<b>Q.No:10-1st question</b> Can statistical techniques be used to perform the task of machine translation? If so, explain in brief. [12]	CO3 and CO4
	<b>Q. No:10-2nd question</b> Explain the Mathematics of Phrase-Based SMT with the help of an example. [12]	
	<b>Q.No:10-3rd question</b>  Explain different approaches to Machine translation. [12]	
<b><u>Q.No:11</u></b>	<b>Q.No:11-1st question</b> Given the following figure on parsing of the sentence “Book the dinner flight” , state which parse would be selected with proper analysis.	CO5 and CO6



Rules	P	Rules	P
S → VP	.05	S → VP	.05
VP → Verb NP	.20	VP → Verb NP NP	.10
NP → Det Nominal	.20	NP → Det Nominal	.20
Nominal → Nominal Noun	.20	NP → Nominal	.15
Nominal → Noun	.75	Nominal → Noun	.75
Verb → book	.30	Nominal → Noun	.75
Det → the	.60	Verb → book	.30
Noun → dinner	.10	Det → the	.60
Noun → flight	.40	Noun → dinner	.10
		Noun → flight	.40

[12]

### Q. No:11-2nd question

Explain Simple Top-down parsing. Show the search tree for “The Old Man cried“

Some of the syntactic categories of a natural language you can use are as follows:

Sentence(S)  
Noun Phrase(NP)  
Determiner(Det)  
Verb Phrase(VP)  
Prepositional Phrase(PP)  
Verb(V)  
Noun(N)  
Adjective (JJ)

[12]

### Q.No:11-3rd question

Perform parsing using simple top down parsing for the sentence “The dogs cried” using the grammar given below:

S->NP VP  
NP->ART N  
NP->ART ADJ N  
VP->V  
VP->V NP

[12]

