

USN

--	--	--	--	--	--	--	--	--	--

**RV COLLEGE OF ENGINEERING**  
**Autonomous Institution affiliated to VTU**  
**V Semester- Model Question Paper**  
**Natural Language Processing [IS355TBC]**  
**(2022 SCHEME)**

**Time: 03 Hours****Maximum Marks: 100****Instructions to candidates:**

1. Answer all questions from Part A. Part A questions should be answered in first three pages of the answer book only.
2. Answer FIVE full questions from Part B. In Part B question number 2 is compulsory. Answer any one full question from 3 and 4, 5 and 6, 7 and 8, and 9 and 10.

**PART-A**

1	1.1	Write a tag pattern to cover noun phrases that contain gerunds, e.g., the/DT receiving/VBG end/NN, assistant/NN managing/VBG editor/NN.	2
	1.2	Write a function unknown() that takes a xCorpus as its argument, and returns a list of unknown words that occur on a corpus.	2
	1.3	Consider a list sent =['Take', 'care', 'of', 'sense', 'look', 'palace', 'Bengaluru'] Print the length of the words in the list	2
	1.4	Write a program to following paragraph and perform to remove all the stop words	2
	1.5	Define the metric Recall and F-score (F-Measure)?	2
	1.6	Give examples for syntactic agreement?	2
	1.7	What does likelihood ratio indicates?	2
	1.8	Write procedure to extract structured data from unstructured text?	2
	1.9	Print sentences using NP-Chunker with exactly 2 Verbs and 4 Nouns	2
	1.10	Briefly explain the term IOB (Inside-Outside and Begin) tags with example?	2

**PART-B (Maximum subdivisions is limited to 4 in each question)**

UNIT-I			
2	a	What is the difference between Programming Language and Natural Language, Explain the same with example?	8
	b	Explain the role of regular expressions in Natural Language Processing (NLP). How are regular expressions used to address common challenges encountered in NLP tasks such as text preprocessing, tokenization, and pattern matching?	8

UNIT-II			
3	A	What is supervised classification? And explain the same by taking Gender Identification problem?	10
	B	Discuss the application of Naive Bayes classifiers in text classification tasks, including sentiment analysis or spam detection.	6
		<b>OR</b>	
4	a	Explains how tagged corpora are utilized in part-of-speech tagging, named entity recognition, and syntactic parsing	8

	b	Describes different methods and techniques used in NLP to determine the category or part of speech of a word.	8
--	---	---	---

UNIT-III			
5	a	Apply Recursive descent parsing to convert below sentence to a tree. "Tom ate an ice-cream in the hotel". Assume and write the Grammar adopted.	10
	b	Describes various approaches and algorithms used for Named Entity Recognition, including rule-based methods and statistical models	6
OR			
6	a	Mention the applications of below i. Named Entity Recognition ii. Inverse document frequency iii. Information extraction iv. Confusion matrix v. Information gain	10
	b	Write the CFG for the following sentences and execute using NLTK: "The sun sets in the west." "Birds sing melodiously in the morning." "Children play happily in the park." "She reads books in the park"	06

UNIT-IV			
7	a	Illustrate the challenges in Machine Translation in general. Write atleast two challenges which are specific to Indian Languages	8
	b	Take example documents with few lines and show how TFIDF can be used in text classification	8
OR			
8	a	Identify the NLP applications used in market intelligence with justification.	8
	b	Explain the various types of question answering systems in detail	8

UNIT-V			
9	a	Define what statistical translation is and explain its key principles	6
	b	Discuss how homonymy, polysemy, and synonymy pose challenges for NLP tasks such as word sense disambiguation, semantic analysis, and information retrieval.	10
10	a	Describes the key components of the Vector Space Model, including term frequency-inverse document frequency (TF-IDF) weighting and cosine similarity	8
	b	Discusses common challenges encountered by IR systems with user queries, such as ambiguity, vagueness, and lack of specificity.	8

**Q.P.Code**

**Course code:**

**Course Title: Natural Language Processing**

**PART-A**

<b>Q.No</b>	<b>1.1</b>	<b>1.2</b>	<b>1.3</b>	<b>1.4</b>	<b>1.5</b>	<b>1.6</b>	<b>1.7</b>	<b>1.8</b>	<b>1.9</b>	<b>1.10</b>
<b>BT</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>2</b>
<b>COs</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>2</b>	<b>1</b>

**PART-B**

<b>Question No</b>		<b>BT level</b>	<b>COs Addressed</b>		<b>Question No</b>		<b>BT level</b>	<b>COs Addressed</b>
<b>2</b>	<b>a</b>	<b>2</b>	<b>1</b>		<b>3</b>	<b>a</b>	<b>3</b>	<b>2</b>
	<b>b</b>	<b>3</b>	<b>2</b>			<b>b</b>	<b>2</b>	<b>2</b>
<b>4</b>	<b>a</b>	<b>3</b>	<b>3</b>		<b>5</b>	<b>a</b>	<b>3</b>	<b>3</b>
	<b>b</b>	<b>4</b>	<b>3</b>			<b>b</b>	<b>2</b>	<b>2</b>
<b>6</b>	<b>a</b>	<b>2</b>	<b>2</b>		<b>7</b>	<b>a</b>	<b>2</b>	<b>3</b>
	<b>b</b>	<b>3</b>	<b>3</b>			<b>b</b>	<b>3</b>	<b>3</b>
<b>8</b>	<b>a</b>	<b>4</b>	<b>4</b>		<b>9</b>	<b>a</b>	<b>2</b>	<b>2</b>
	<b>b</b>	<b>3</b>	<b>2</b>			<b>b</b>	<b>3</b>	<b>2</b>
<b>10</b>	<b>a</b>	<b>3</b>	<b>3</b>			<b>a</b>		
	<b>b</b>	<b>2</b>	<b>1</b>			<b>b</b>		

Signature of Scrutinizer:

Signature of Chairman

Name:

Name:

