P P SAVANI UNIVERSITY

Third Semester of M.Sc.DS

Examination January 2023

SSDS8511 Natural Language Processing

09.01.2023, Tuesday **Time: 10:00 a.m. To 12:30 p.m.** **Maximum Marks: 60**

Instructions:

1. The question paper comprises two sections.
2. Section I and II must be attempted in separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculators is allowed.

SECTION – I

			CO	BTL
Q - 1	Short Question (Any Five)	**[05]**		
(i)	What is text Summarization?		1	1
(ii)	What is a Bag of Words?		2	1
(iii)	Write a Full form of NLTK.		1	2
(iv)	What are the best NLP Tools?		3	1
(v)	What is a Masked Language Model?		1	4
(vi)	What is Lemmatization ?		1	4
(vii)	What is Stemming ?		3	4
Q - 2 (a)	What is Semantic Analysis? Explain.	**[05]**	1	2
Q - 2 (b)	List any two real-life applications of Natural Language Processing in detail.	**[05]**	2	3
	OR			
Q - 2 (a)	Write a Short Note on Stop Words	**[05]**	3	5
Q - 2 (b)	What is Semantic Analysis? Explain.	**[05]**	2	3
Q - 3 (a)	Write a Note on Regular Expressions.	**[05]**	2	2
Q - 3 (b)	Explain about Regular Grammar.	**[05]**	3	4
	OR			
Q - 3 (a)	What is Parsing in the context of NLP?	**[05]**	3	4
Q - 3 (b)	What is TF-IDF? Explain in Detail.	**[05]**	3	2
Q - 4	Attempt any one	**[05]**	2	2
(i)	Define the terminology in NLP in a detailed manner. Explain			
(ii)	Dependency Parsing in NLP.			

SECTION – II

Q - 1	MCQ (Any Five)	**[05]**		
(i)	Which of the following techniques can be used for keyword normalization in NLP, the process of converting a keyword into its base form?		2	3

A. Lemmatization
B. Soundex
C. Cosine Similarity
D. N-grams

(ii) Which of the following techniques can be used to compute the distance between two-word vectors in NLP? 2 6

A. Lemmatization
B. Euclidean distance
C. Cosine Similarity
D. N-grams

(iii) What are the possible features of a text corpus in NLP? 2 4

A. Count of the word in a document
B. Vector notation of the word
C. Part of Speech Tag
D. Basic Dependency Grammar
E. All of the above

(iv) You created a document term matrix on the input data of 20K documents for a Machine learning model. Which of the following can be used to reduce the dimensions of data? 3 4

1. Keyword Normalization
2. Latent Semantic Indexing
3. Latent Dirichlet Allocation

A. only 1
B. 2, 3
C. 1, 3
D. 1, 2, 3

(v) Which of the text parsing techniques can be used for noun phrase detection, verb phrase detection, subject detection, and object detection in NLP. 3 3

A. Part of speech tagging
B. Skip Gram and N-Gram extraction
C. Continuous Bag of Words
D. Dependency Parsing and Constituency Parsing

(vi) Dissimilarity between words expressed using cosine similarity will have values significantly higher than 0.5 2 3

A. True
B. False

(vii) Which one of the following is keyword Normalization techniques in NLP 2 2

A. Stemming
B. Part of Speech
C. Named entity recognition
D. Lemmatization

Q - 2 (a)	Explain the difference between NLP and NLU?	**[05]**	1	3
Q - 2 (b)	What are unigrams, bigrams, trigrams, and n-grams in NLP?	**[05]**	2	2
	OR			
Q - 2 (a)	Explain Feature Extraction in NLP?	**[05]**	2	4
Q - 2 (b)	How to tokenize a sentence using the **nltk** package?	**[05]**	3	5
Q - 3 (a)	Explain Stemming with the help of an example.	**[05]**	3	3
Q - 3 (b)	Explain Lemmatization with the help of an example.	**[05]**	2	4
	OR			
Q - 3 (a)	What is Parts-of-speech Tagging? Explain in Detail	**[05]**	2	2
Q - 3 (b)	Explain Named Entity Recognition by implementing it.	**[05]**	2	2
Q - 4	Attempt any one/two.			
(i)	How to check word similarity using the spacy package? What is	**[05]**	3	5
(ii)	F1 score in NLP? What is F1 score in NLP?			

****

CO : Course Outcome Number BTL : Bloom's Taxonomy Level

Level of Bloom's Revised Taxonomy in Assessment

1: Remember	2: Understand	3: Apply
4: Analyze	5: Evaluate	6: Create