

Nirma University
Institute of Technology
Semester End Examination *(IR/RPR), May 2022
B. Tech. in CSE, Semester-VI
2CSDE70 NATURAL LANGUAGE PROCESSING

Roll /
Exam No.

19BCE745

Supervisor's initial
with date

J. R. S.

Time: 3 Hours

Max. Marks : 100

Instructions:

1. Attempt all questions.
2. Figures to right indicate full marks.
3. Use section-wise separate answer book.
4. Draw neat sketches wherever necessary.

Section - I

Q-1. Do as directed.

[18]

CO-1

- A** Why natural language processing is difficult? Explain with example. [06]
B How ELIZA chatbot works? Explain the concept of substitution in regular expression with example. [06]

OR

- B** How POS tagging is helpful in natural language processing application? Is it necessary to perform POS tagging in all natural language processing application? [06]
C Compare word Normalization and word tokenization with example. [06]

OR

- C** Compare Natural language processing and machine translation. [06]

Q-2. Do as directed.

[16]

CO-2

- A** Write regular expression for following definitions. [08]
1. The language starting with a but not having consecutive b's.
2. All strings which can be used as identifier in C language.
3. the language over $\Sigma = \{a\}$ having even length of the string
4. the language L over $\Sigma = \{a, b\}$ such that all the string do not contain the substring ab.
- B** Explain and Perform stemming on the following paragraph using porter stemmer algorithm. (Mention the steps). [08]
"The European Union (EU) is a political and economic union of 27 member states that are located primarily in Europe. The EU has developed an internal single market through a standardised system of laws that apply in all member states in those matters, and only those matters, where members have agreed to act as one."

Q-3. Do as directed.

CO-3

- A** What do you mean by word sense disambiguation in NLP? Explain traditional Lesk algorithm with example. [08]
- B** For the following document collection, Use, depict and explain the Boolean model. What are the Returned results for these query "schizophrenia AND drug" [08]
- Doc 1 :breakthrough drug for schizophrenia
- Doc 2 :new schizophrenia drug
- Doc 3 :new approach for treatment of schizophrenia
- Doc 4 :new hopes for schizophrenia patients

Section - II

Q-4. Do as directed.

CO-1

- A** What is the importance of word embedding in Natural language processing? Explain one hot representation technique with example and write its drawbacks. [18]
- B** Discuss the concept of selective read, selective write and selective forget with example. [06]
- C** Discuss lexical analysis, syntax analysis and semantic analysis in formal languages [06]
- OR**
- C** Compare Continuous bag of words, Skip gram and glove model for word embedding. [06]

Q-5. Do as directed.

CO-2

- A** For the following document collection, use and depict the Vector space model and write the returned results for the query "orange lemon price" [08]
- D1 : "sales of orange affected in market"
- D2 : "Expense of lemon increased due to lemon price"
- D3 : "increment of orange price in market"
- B** Build Co-occurrence matrix for the below given corpus. Explain the importance of PMI in distributional representation. [08]
- Kerala is famous for its ancient history.
 - Kerala is a place with rich culture and traditional values.
 - Kerala is a beautiful place with beautiful beaches.

OR

Apply multinomial Naïve bayes text classification and predict the class of below given table.

2055E70

[08]

chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	?
chills	runny nose	headache	fever	flu?
Y	N	Mild	Y	N
Y	Y	No	N	Y
Y	N	Strong	Y	Y
N	Y	Mild	Y	Y
N	N	No	N	N
N	Y	Strong	Y	Y
N	Y	Strong	N	N
Y	Y	Mild	Y	Y

Q-6. Do as directed.

CO-4

[16]

A Compare RNN, LSTM and GRU with example.

[08]

B How Sequence learning problems are different then classification problems? Explain with example.

[08]