

Nirma University
Institute of Technology
Class test, Feb-2021
B. Tech(CSE), Semester-VI
2CSDE70 Natural language processing

Roll No.

Supervisor's initial with date:

Time: 1:15 hours

Max. Marks: 35

Instructions: 1. Attempt all questions.

2. Figures to right indicate full marks.

3. Draw neat sketches wherever necessary.

4. Assume suitable data wherever necessary and clearly indicate it.

Q-1. Do as directed.

- A Discuss most difficult problems of natural language processing. [05]
- B Write regular expression for following definitions. [06]
1. The set of all lower-case alphabetic strings ending in a b.
2. All strings which can be used as identifier in C language.
- C Apply porter stemmer algorithm on following paragraph. [06]
“A paragraph is a series of related sentences developing a central idea, called the topic. Try to think about paragraphs in terms of thematic unity: a paragraph is a sentence or a group of sentences that supports one central, unified idea. Paragraphs add one idea at a time to your broader argument.”
- D Apply bigram probability and find out highest probability pair from [06]
following data.
<s> I am Sam </s>
<s> Sam I am </s>
<s> I do not like green eggs and ham </s>
- E Discuss and Apply Maximum Matching Word Segmentation Algorithm [06]
on “Thetabledownthere”.
- F Apply Boolean model on following statement. [06]
Consider these documents:
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
For the document collection, Use and depict the Boolean model and what are the Returned results for query “schizophrenia AND drug”