

Nirma University
Institute of Technology
Additional test, May-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 Write applications of Multimedia retrieval. How natural language processing is used in Multimedia Retrieval? [07]
- B Write regular expression for following definitions. [07]
1. Regular expression for an email address
2. the set of binary numbers that are multiples of 3(ϵ , "0", "00", "11", "000", "011"...)
3. regular expression to filter spam mails
- C For the following document collection, Use and depict the Probabilistic model and what are the returned results for query "wood steel truck" [07]
D1 : "shipment of wood damaged in a fire"
D2 : "delivery of steel arrived in a steel truck container"
D3 : "Shipment of wood in a truck"
- D Given the following short movie reviews, each labeled with a genre, either comedy or action. [07]
1. fun, couple, love, love **comedy**
2. fast, furious, shoot **action**
3. couple, fly, fast, fun, fun **comedy**
4. furious, shoot, shoot, fun **action**
5. fly, fast, shoot, love **action**
Test data D: fast, couple, shoot, fly
compute the most likely class for D. Assume a naive Bayes classifier and use add-1 smoothing for the likelihoods.
- E Explain direct relevance feedback and indirect relevance feedback with example. Discuss Rocchio method for relevance feedback. [07]