Ex/PG/CSE/T/128F/2024

MASTER OF COMPUTER SCIENCE AND ENGINEERING

First Year, Second Semester Examination, 2024

Text Analytics

Time- Three Hours **Full Marks-100**

Answer Question No. 1 and Any Three Questions from the rest

All the answers under same question number should be answered together

1. **Answer all the questions** 10 X 4=40

a. How can we employ add-1 smoothing technique while estimating the maximum likelihood? Justify with an example.

b. Why do we prefer cosine measure instead of Euclidian distance in case of calculating vector space proximity? Justify your answer with respect to query and document.

c. What do we mean by intents in a multi-party conversation? Give an example.

d. Why accuracy is not a proper measure for evaluating the performance of any information retrieval (IR) system? What are the metrics used in IR evaluation?

e. Does idf have any effect on ranking for single-term queries? Justify your answer.

f. Why do we perform passage re-ranking in a QA system? What is MRR?

g. Is supervised content selection is preferred in document summarization? Justify with proper reasons.

h. Why do we prefer to use Point-wise Mutual Information (PMI)? Justify with an example.

i. What is the role of proximity matches in case of query feature identification?

j. What are the roles of Barge-in and Grounding in a conversation? Justify with example.

[Turn over

2. a. What are the differences between Boolean and Vector Space model? What are the roles of a tf based model and an idf based model for ranked information retrieval? Why do we use Jaccard coefficient?

   b. What is relevance feedback query? State and explain Rocchio SMART algorithm for calculating a relevance feedback query using Vector Space model. What is the relation between original and modified query in SMART algorithm?

(3+5+2) + (2+6+2) =20

3. a. State different unsupervised content selection techniques for text summarization. Why information ordering and sentence fusion are required in summarization? The following are three reference summaries along with a system generated summary. What are the scores of ROUGE-2 and ROUGE-3 evaluation schemes?

- Human 1: We are the great Indian citizen who can devote for the country.
- Human 2: You are the great Indian citizen who can identify the proper value for the country.
- Human 3: We are really proud to be the great Indian citizen who can devote for the nation.
- *System answer: We are the great Indian citizen who can sacrifice their lives for the country.*

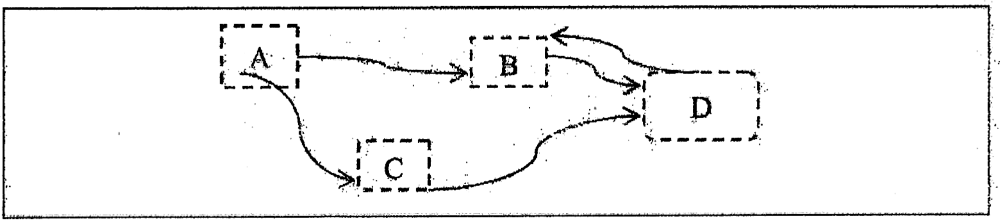
   b. Write down and explain the steps of formulating query words from a question.

(6+3+6)+5=20

4. a. What do you mean by extent lists? Write down a Page Rank algorithm and explain it with a suitable example. What do you mean by inverted index?

   b. Using Page Rank algorithm, in different iterations, identify the ranks of the webpages A, B, C and D as shown in the following figure.

(2+6+2)+10 = 20



5. a. What are the differences between Natural Language Generation and Natural Language Understanding? Write an algorithm to measure the polarity of a phrase? Justify with an example.

b. Define ROUGE-2 and state its use. Consider the following table which shows the results of four classes, A, B, C and D. What are the Macro-average and Micro-average precision values? Is Micro-averaged score dominated by score on common classes?

Classes	True Positive	False Positive
A	20	20
B	100	900
C	10	10
D	10	10

(3+7) + (3+5+2) = 20

6 a. What do you mean by sentiment co-reference? What do you mean by scaled-likelihood? How does PMI help in identifying sentiments of a phrase? How do we use semi-supervised technique to prepare sentiment lexicon? Give an example to justify.

b. Suppose, three documents are taken from class "X" and two documents from class "Y" and employed them into a Naïve Bayes classifier as training set along with their sentence level constituents. See the following Table. Calculate the probability of the test documents (id6 and id7) to be assigned into a particular class. Show each of the steps.

(3+3+3+3)+8=20

Table 1	Doc	Sentences	Class
Training	id1	Ant Bag Apple	X
	id2	Ant Apple Cat	X
	id3	Ant Dog	X
	id4	Ant Pig Queen	Y
	id5	Pig Cat Queen	Y
Test	id6	Ant Apple Ant Pig Queen	?
	id7	Pig Queen Cat Queen	?