APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

08 PALAKKAD CLUSTER

SECOND SEMESTER M.TECH DEGREE EXAMINATION, MAY 2016

Computer Science and Engineering

08 CS 6022
INFORMATION RETRIEVAL

Max. Marks : 60 Duration : 3 Hours

Answer ALL six questions. Part (a) of each question is compulsory.

Answer EITHER part (b) *or* part (c) of each question.

Q.No. Marks

Module I

1 a. What role does logical view of documents play in information retrieval? 3

Answer b or c

b.The vector space model of document retrieval uses the notion of an information space. How is term weighting used to position objects in the space? Why is term weighting important for effective document retrieval? 6

c. A new IR algorithm yields the following answer to query q1 (relevant docs are marked with a bullet):

01. d425	06. d615	11. d193
02. d87	07. d512	12. d715
03. d56 •	08. d129 •	13. d810
04. d32	09. d4	14. d5
05. d124	10. d130	15. d3 •

(i) Explain the evaluation metrics for the system. 2

(ii) Give the formula for mean average precision (MAP), and illustrate the metric by calculating System's MAP. 2

(iii) Draw a precision-recall curve for the system. Explain how you arrived at your result. 2

Module II

2a.Explain the effect of IDF and TF factors on inter cluster dissimilarity and intra cluster similarity. 3

Answer b or c

b. Consider two documents D1 = (0.5, 0.8, 0.3) and D2 = (0.9, 0.4, 0.2) indexed by three terms, where the numbers represent term weights. Given the query Q = (1.5, 1.0, 0) indexed by the same terms, find the cosine measures for the two documents. 6

c.Boolean model is more a data retrieval model than an information retrieval model. Justify your answer with an example. 6

Module III

3a. Discuss the concept of hubs and authorities in multimedia indexing. 3

Answer b or c

b. Explain how correlated terms are built form local document set through association clustering. 6

c. Discuss page rank algorithm in the context of a small web consisting of three pages A, B and C, whereby page A links to the pages B and C, page B links to page C and page C links to page A. Assume the damping factor is 0.5.Find the page rankings also. 6

Module IV

4a. Discuss how XML retrieval is performed. 3

Answer b or c

b.Explain how R tree helps in mapping objects into f-D space to do clustering. 6

c.Explain the application of GEMINI method in colour images. Discuss its application in medical field. 6

Module V

5a. Explain the relevance of nearest neighbour in information retrieval 4

Answer b or c

b. Explain the concept of decision tree with an example. 8

c. Consider following data set.Apply Naive Bayesian equation to find the probability of players playing if weather is sunny. 8

Weather	play
Sunny	no
overcast	Yes
Rainy	yes
sunny	Yes
Sunny	Yes
Overcast	yes
Rainy	no
Rainy	no
Sunny	Yes
Rainy	yes
Sunny	no
Overcast	Yes
overcast	Yes
Rainy	no

Module VI

6a. Explain agglomerative clustering method. Discuss the complexity of algorithm also. 4

Answer b or c

b. Consider the following data set consisting of the scores of two variables on each of seven individuals. Using k means clustering group this data set into two clusters. 8

Subject	A	B
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

c. Discuss in detail expectation maximization algorithm. 8