Total No. of Questions : 10]

SEAT No :

P3142

[Total No. of Pages :2

[5154]-708
B.E.(I.T.)
INFORMATION STORAGE AND RETRIEVAL
(2012 Course) (Elective-III) (414463 C) (Semester-II)

Time : 2½ Hours *Max.Marks:70*

Instructions to the candidates:

1) ***All questions are compulsory.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Assume suitable data, if necessary.***

Q1) a) Show how single link clusters may be derived from the dissimilarity coefficient by thresholding it. **[5]**

b) You are developing a text processing system for use in an automatic retrieval System. Explain the following parts: **[5]**
Removal of high frequency words.
Suffix stripping.
Detecting equivalent stems.

OR

Q2) a) Find the similarity of following query with D1,D2,D3, using vector model. **[6]**

Query	keywords	
q	ant, dog	
document	Text	Terms
D1	ant ant bee	ant bee
D2	dog bee dog hog dog ant dog	ant bee dog hog
D3	cat gnu dog eel fox	cat dog eel fox gnu

b) Write a short note on user oriented measures to evaluate the performance of the system. **[4]**

Q3) a) Writr a note on "Ontology based information sharing". **[5]**

b) Explain the concept of hash addressing. **[5]**

OR

P.T.O.

Q4) Consider a reference collection and its set of example information request. If q is the information request and a set Rq= (d3,d5,d9,d25,d39,d44,d50,d70, d80,d120). Now consider new retrieval algorithm has been designed and has been evaluated for information request q returns, ranking of the documents in the answer set as. **[10]**

1.d120	6.d9	11.d38
2.d84	7.d58	12.d48
3.d50	8.d129	13.d230
4.d6	9.d143	14.d113
5.d8	10.d25	15.d3

The documents that are relevant to the query q are underlined. Calculate precision and recall for the documents that are relevant to the query q.

Q5) a) Describe the architecture of distributed IR. **[8]**

b) What do you understand by multimedia query language? Explain various query predicates. **[8]**

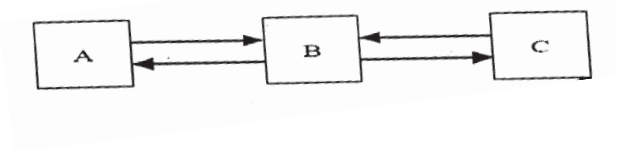
OR

Q6) a) What are the issues in distributed IR computing? Write the techniques used to address these issues. **[8]**

b) Write a note on MULTOS. **[8]**

Q7) a) Write a short note on web data mining. **[8]**

b) What is page ranking? Calculate page rank of following web pages. Assume damping factor 7.0. **[10]**



OR

Q8) a) Explain centralized and distributed architecture of a search engine. **[10]**

b) What is web crawling? Explain the techniques used by web crawlers to crawl the web. **[8]**

Q9) a) What is content based recommendation? **[8]**

b) Explain semantic web in details. **[8]**

OR

Q10)a) Define Recommender system? Explain in brief collaborative filtering **[8]**

b) Discuss trends and research issues involved in web. **[8]**

♦ ♦ ♦