

Reg. No.

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

M.E / M.TECH. DEGREE EXAMINATIONS, MAY 2023
 Second Semester
CP22014 – INFORMATION RETRIEVAL TECHNIQUES
 (Computer Science and Engineering)
 (Regulation 2022)

TIME: 3 HOURS

MAX. MARKS: 100

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Build an Information Retrieval system using the available tool	5
CO 2	Identify and design the various components of an Information Retrieval system	1
CO 3	Measure effectiveness and efficiency of information retrieval techniques	5
CO 4	Use parallel Information Retrieval approaches in real world problems	3
CO 5	Design an efficient search engine and analyze the Web content structure	3

PART- A (20 x 2 = 40 Marks)
 (Answer all Questions)

Q. No.	QUESTION	CO	RBT LEVEL
1.	How does an information seeking session begin in online information systems?	1	2
2.	Outline the role of bucket testing or A/B testing to evaluate Web sites.	1	2
3.	Explain orienteering behavior in web search.	1	2
4.	Demonstrate the need for Query Reformulation.	1	2
5.	Consider a document containing 1000 words wherein the word cat appears 30 times. Assume in a collection of 10 million documents, the word cat appears in one thousands of these. Calculate $tf_{t,d}$, idf_t and $tf-idf_{t,d}$	2	3
6.	Outline the taxonomy of IR models.	2	2
7.	Outline the basic idea in Fuzzy Information Retrieval.	2	2
8.	Explain the ways to address the drawbacks of Boolean model.	2	3
9.	When do you say Mean Reciprocal Rank (MRR) is a good metric? Mention its formula.	3	2
10.	Identify the difficulties in evaluating IR systems.	3	2
11.	Consider a query that retrieves 10 documents namely $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9$ and D_{10} . Let's say 2 out of 10 retrieved documents are relevant. Evaluate precision and recall.	3	5
12.	What do you mean by Precision at k ($P@k$) metric for ranked results?	3	2
13.	Outline the basic idea behind Blocked sort-based Indexing algorithm.	4	3
14.	Can Arithmetic coding attain better compression rates than Huffman coding?	4	4
15.	Differentiate term shading from document sharding.	4	2
16.	Mention the importance of Elias's γ coding.	4	2
17.	Identify the challenges in XML retrieval.	5	3

18.	How do Hubs and authorities play a major role in ranking documents?	5	4
19.	Two web search engines A and B each generate a large number of pages uniformly at random from their indexes. Thirty percent of A's pages are present in B's index, and 50% of B's pages are present in A's index. What is the number of pages in A's index relative to B's?	5	4
20.	How can Near-duplicates and shingling be handled in web content?	5	4

PART- B (5 x 10 = 50 Marks)

Q. No.	QUESTION	Marks	CO	RBT LEVEL
21. (a)	Discuss in detail the high level architecture of an Information Retrieval system.	(10)	1	3

(OR)

(b)	Discuss in detail the importance of Visualization in Search Interfaces for Text Mining.	(10)	1	3
-----	---	------	---	---

22. (a)	Consider the three documents D_1, D_2 and D_3 containing the following terms	(10)	2	4
---------	--	------	---	---

D_1 : ant ant bee

D_2 : dog bee dog hog dog ant dog

D_3 : cat gnu dog eel fox

Evaluate the similarity between these documents using **TF weighting scheme**. For the given query "**ant dog**", evaluate the similarity of the given query to documents.

(OR)

(b)	Outline the steps involved in the construction of an inverted index for the given document collection D_1, D_2, D_3 and D_4	(10)	2	4
-----	---	------	---	---

D_1 : To be is to be. To be is to do

D_2 : To be or not to be. I am what I am

D_3 : I think therefore I am. Do be do be do

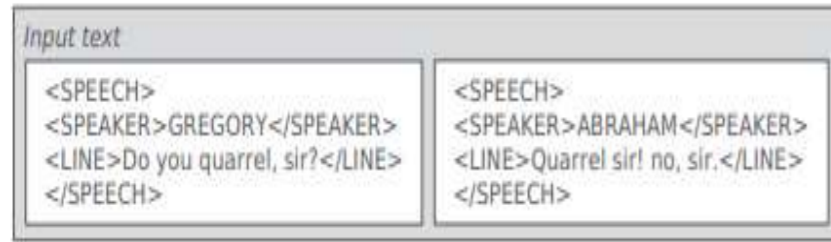
D_4 : Do do do da da da. Let it be, let it be

23. (a)	(i) Identify the need for Reference Collections	(5)	3	3
	(ii) How can BPREF retrieval evaluation work for incomplete information?	(5)	3	3

(OR)

(b)	Illustrate Rocchio method Relevance feedback for the vector model with a suitable example.	(10)	3	3
-----	--	------	---	---

24. (a) Analyze Sort-based index construction algorithm for the following input text (10) 4 4



(OR)

(b) Build a Huffman code tree for the set $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$ with associated probability distribution $\Pr[\alpha_1] = 0.18, \Pr[\alpha_2] = 0.11, \Pr[\alpha_3] = 0.31, \Pr[\alpha_4] = 0.34$ and $\Pr[\alpha_5] = 0.06$ (10) 4 4

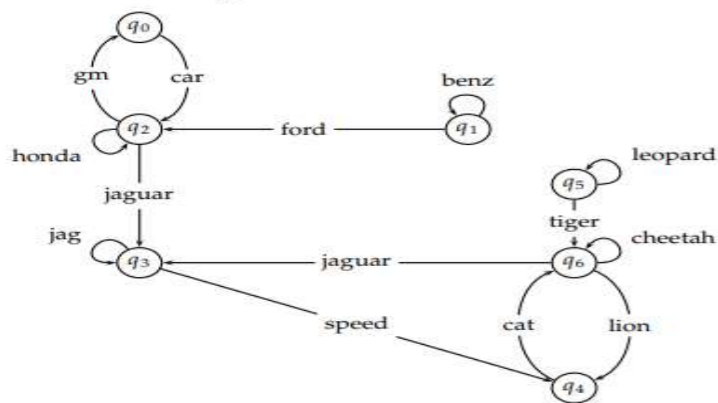
25. (a) Illustrate with a suitable example, how can documents be represented and retrieved in XML retrieval. (10) 5 4

(OR)

(b) Use teleportation rate as 0.14 with its transition probability matrix as (10) 5 5

0.02	0.02	0.88	0.02	0.02	0.02	0.02
0.02	0.45	0.45	0.02	0.02	0.02	0.02
0.31	0.02	0.31	0.31	0.02	0.02	0.02
0.02	0.02	0.02	0.45	0.45	0.02	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.88
0.02	0.02	0.02	0.02	0.02	0.45	0.45
0.02	0.02	0.02	0.31	0.31	0.02	0.31

to compute PageRank vector for the web graph given below.



PART- C (1 x 10 = 10 Marks)
(Q.No.26 is compulsory)

26. Construct a neural network model for information retrieval. (10) 2 4
