ODD SEMESTER EXAMINATION, 2024 – 25

IV Year (VII Sem) B.Tech.: CS&E/E&CE

INFORMATION RETRIEVAL

Duration: 3:00 hrs   **Max Marks: 100**

Note: - Attempt all questions. All Questions carry equal marks. In case of any ambiguity or missing data, the same may be assumed and state the assumption made in the answer.

Q 1.	Answer any two parts of the following.	(10x2= 20)
	a) (i) Define the term vocabulary and postings lists.	(5 marks)
	(ii) Define dictionaries with a suitable example.	(5 marks)
	b) Define index-construction. Describe index construction for the following documents:	(10 marks)
	Document 1: The quick brown fox jumped over the lazy dog.	
	Document 2: The lazy dog slept in the sun.	
	c) Explain the working of Boolean retrieval method in information retrieval.	(10 marks)
Q 2.	Answer any two parts of the following.	(10x2= 20)
	a) (i) Explain term frequency and document frequency in ranked retrieval.	(5 marks)
	(ii) Explain relevance feedback with a suitable example. .	(5 marks)
	b) Describe the different ways of computing scores in a complete search system.	(10 marks)
	c) Explain in brief vector space model. Describe the ranking for the following documents :-	(10 marks)
	D1: Dog runs behind cat	
	D2: Cat runs behind rat	
	Query: Cat	
Q 3.	Answer any two parts of the following.	(10x2= 20)
	a) (i) Define Naïve Bayes theorem in information retrieval.	(5 marks)
	(ii) Explain the concept of text classification.	(5 marks)
	b) Consider making a language model from the following training text: "*The martian has landed on the latin pop sensation ricky martin*"	
	i. Under a MLE-estimated unigram probability model, what are P(the) and P(martian)?	
	ii. Under a MLE-estimated bigram model, what are P(sensation|pop) and P(pop|the)?	(10 marks)
	c) Derive a ranking function for query terms in binary independence model.	(10 marks)
Q 4.	Answer any two parts of the following.	(10x2= 20)
	a) (i) Define the term web crawling and indexes.	(5 marks)
	(ii) Explain in brief language modeling.	(5 marks)
	b) Differentiate between flat clustering and hierarchical clustering.	(10 marks)
	c) Explain why kNN handles multimodal classes better than Rocchio	(10 marks)
Q 5.	Answer any two parts of the following.	(10x2= 20)
	a) (i) Define index-compression.	(5 marks)
	(ii) Explain why relevance feedback has been little used in web search.	(5 marks)
	b) Explain Bayesian Network approaches to IR.	(10 marks)
	c) Explain in detail the concept of latent semantic indexing.	(10 marks)

****