

Exam.Code:0920

Sub. Code: 6367 ✓

2054

B.E. (Computer Science and Engineering)

Eighth Semester

Elective V

CS-803C: Information Retrieval and Management ✓

Time allowed: 3 Hours

Max. Marks: 50

**NOTE:** Attempt five questions in all, including Question No. 1 which is compulsory and selecting two questions from each section.

x-x-x

- Q1. a) What are the steps involved in query optimization?  
b) Explain the purpose of stemming and its significance in text processing.  
c) How do wildcard queries differ from exact and fuzzy queries?  
d) Explain the mathematical formula for computing cosine similarity between two vectors.  
e) How does the size of the document corpus affect the scalability of near duplicate detection algorithms? (5x2)

**Section - A**

- Q2. a) What is an inverted index in the context of information retrieval? How does it differ from other indexing techniques? What are its key components and the process of building an inverted index?  
b) What are Skip pointers? Why are skip pointers not useful for queries of the form  $a$  OR  $b$ ? (2x5)
- Q3. How does the size of a term vocabulary affect the efficiency of an inverted index? How are posting lists used in conjunction with the term vocabulary during query processing? Also explain the caching play in optimizing the retrieval of posting lists. (10)
- Q4. a) Explain the difference between rule-based and statistical methods for spelling correction.  
b) How do n-gram indexes handle partial matches and approximate string matching? Also list the evaluation metrics used to assess the performance of n-gram indexes. (2x5)

P.T.O.

(2)

**Section-B**

- Q5. a) What is vector space modelling? How does it represent documents and queries?  
b) Explain any method for learning the linear classifier weights for ad hoc Information Retrieval. (2x5)
- Q6. a) Explain the basic principle behind random projection as a dimensionality reduction technique and how it preserves the pairwise distances between high-dimensional vectors?  
b) How does the Naive Bayes model calculate the probability of a document belonging to a particular class? (2x5)
- Q7. Write short notes on:  
a) Search Engine Optimization  
b) Rank algorithms in Web Crawling (10)

x-x-x