

M.E. (Computer Science and Engineering)
Third Semester
CS-8304: Information Retrieval

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

- Q 1. (a) How do positional indices differ from standard inverted indices? [5X2 = 10]
(b) What is the main limitation of using hash tables for large dictionary lookup in IR?
(c) What challenge does the scoring component face when processing very long documents?
(d) What assumption does the Naïve Bayes model make about features?
(e) How does URL frontier management improve crawler efficiency?

SECTION – A

- Q 2. (a) How does query optimization impact retrieval performance? Illustrate with an example comparing Boolean query processing with ranked retrieval. [5]
(b) Analyze how inverted index structure differs when handling XML or JSON-based documents compared to plain text. [5]
- Q 3. (a) Explain, with an example, how skip pointers can be used to accelerate intersection operations in inverted lists. [5]
(b) Design a positional index that can efficiently answer the query: "machine learning near model". Show how it works step-by-step. [5]
- Q 4. (a) Given a dataset of 10,000 words, analyze the storage and lookup efficiency trade-offs between hash-based and tree-based dictionary data structures. [5]
(b) Compare sort-based indexing and in-memory indexing in terms of scalability and performance. [5]

SECTION – B

- Q 5. (a) Two documents have the following term vectors: $D1 = (2, 3, 0)$, $D2 = (1, 1, 2)$. Compute their cosine similarity and interpret the result. [5]
(b) Compare Boolean retrieval and vector space retrieval using a scenario where ranked results are essential. [5]
- Q 6. (a) Describe the key components of an end-to-end information retrieval system and their interaction during query processing. [5]
(b) How can nearest neighbor and random projection techniques be combined to improve high-dimensional document retrieval? [5]
- Q 7. (a) Compare K-Nearest Neighbor and Decision Tree approaches in document categorization tasks. [5]
(b) Design a workflow for a focused web crawler that gathers data only from academic domains (.edu). [5]

x-x-x