

(Information Systems)
MSc(SS) DEGREE EXAMINATION MAY 2026
(Eighth Semester)

Branch – **SOFTWARE SYSTEMS (five year integrated)**

MAJOR ELECTIVE COURSE – III INFORMATION RETRIEVAL

Time: Three Hours

Maximum: 50 Marks

SECTION-A (5 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** marks

(5 x 1 = 5)

- 1 Which factor directly influences document length normalization in TF-IDF weighting?
(i) Collection frequency (ii) Term rank
(iii) Vocabulary size (iv) Vector magnitude of the document
- 2 Identify the key feature that distinguishes the Extended Boolean Model from the classical Boolean model.
(i) Strict binary matching (ii) Exact AND/OR evaluation
(iii) No ranking mechanism (iv) Partial matching with term weights
- 3 Find the classifier that assumes conditional independence among features given the class label.
(i) kNN (ii) Naive Bayes
(iii) Support Vector Machine (iv) Rocchio
- 4 Which algorithm assigns separate hub and authority scores to web pages?
(i) PageRank (ii) SALSA
(iii) HITS (iv) TrustRank
- 5 What technique extracts semantic relationships such as “works-for” or “located-in” from text?
(i) Relation Extraction (ii) Named Entity Recognition
(iii) Query Expansion (iv) Co-reference Resolution

SECTION - B (15 Marks)

Answer **ALL** Questions

ALL Questions Carry **EQUAL** Marks

(5 x 3 = 15)

- 6 a. Explain the process of index term selection using a thesaurus.
OR
b. Discuss tokenization with a simple example.
- 7 a. Explain the major retrieval models used in IR systems.
OR
b. Illustrate the process of ranked retrieval.
- 8 a. Apply query expansion to improve a given search query.
OR
b. Determine how Rocchio algorithm updates the query vector.

Cont...

- 9 a Explain how near-duplicate web pages are classified and detected.
OR
b Illustrate content-based recommendation with a simple example.
- 10 a Determine how co-reference resolution helps in integrating web documents.
OR
b Assume a web page containing product details. Explain how relation extraction can be applied.

SECTION -C (30 Marks)

Answer ALL questions

ALL questions carry EQUAL Marks (5 x 6 = 30)

- 11 a Interpret the effectiveness of Porter's stemming algorithm in improving retrieval performance.
OR
b Elucidate the role of stop-word removal in improving indexing efficiency.
- 12 a Analyze the effectiveness of TF-IDF weighting in ranked retrieval systems.
OR
b Survey the limitations of Boolean matching in large-scale information retrieval.
- 13 a Assess the accuracy of Naive Bayes and kNN classifiers for text categorization.
OR
b Justify the need for dimensionality reduction using LSI and PCA.
- 14 a Justify the need for duplicate detection in modern web search engines.
OR
b Evaluate the role of relevance feedback in information filtering systems.
- 15 a Formulate a strategy to improve accuracy in Named Entity Recognition.
OR
b Develop a web mining framework for information integration.

Z-Z-Z

END