



BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – B.Tech.(CSE)-DS-2022/B.Tech.(CSE)-DS-2023

Course Name – Information Extraction and Retrieval

Course Code - PEC-CSD602C

(Semester VI)

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

- (i) Which of the following is an example of preprocessing in IR?
 - a) Stop-word removal
 - b) TF-IDF weighting
 - c) Query execution
 - d) Index compression
- (ii) Which measure combines precision and recall into a single score?
 - a) TF-IDF
 - b) F1-Score
 - c) Jaccard Index
 - d) Mean Reciprocal Rank
- (iii) What is a document ranking function used for?
 - a) Sorting query terms alphabetically
 - b) Determining the order of relevance of retrieved documents
 - c) Counting the number of documents in a database
 - d) Reducing the number of indexed terms
- (iv) What is the main advantage of an inverted index?
 - a) Faster search and retrieval
 - b) Reduced document storage
 - c) Better document visualization
 - d) Improved user interface
- (v) What is the main goal of query expansion in Information Retrieval?
 - a) Reduce the number of retrieved documents
 - b) Improve search accuracy by adding related terms
 - c) Remove stopwords from queries
 - d) Compress the index size
- (vi) Which of the following is NOT a method of query expansion?
 - a) Synonym-based expansion
 - b) Relevance feedback
 - c) Stemming
 - d) Thesaurus-based expansion
- (vii) What is Relevance Feedback (RF) in IR?
 - a) A technique that automatically reformulates a query based on user feedback
 - b) A method for detecting spam in search results
 - c) A way to store previous queries for faster retrieval
 - d) A process of ranking search results based on popularity

- (viii) In Relevance Feedback (RF), how does the system refine the search query?
- By manually adding new documents
 - By using user-selected relevant documents to modify the query
 - By deleting irrelevant documents
 - By reducing the search scope
- (ix) Which type of relevance feedback does NOT require user interaction?
- Explicit feedback
 - Implicit feedback
 - Pseudo relevance feedback
 - Manual feedback
- (x) Which query expansion technique relies on pre-defined word relationships?
- Relevance feedback
 - Ontology-based expansion
 - Term frequency weighting
 - Document clustering
- (xi) What is Pseudo Relevance Feedback (PRF)?
- A feedback system requiring explicit user input
 - A method that assumes the top retrieved results are relevant and modifies the query accordingly
 - A supervised learning technique for classification
 - A way to detect duplicate documents in a dataset
- (xii) Which of the following best describes query reformulation?
- The process of removing stopwords from queries
 - The restructuring of a query to improve retrieval results
 - The technique of reducing index size
 - A form of query encryption
- (xiii) Which method of query expansion is best suited for domain-specific search engines?
- Thesaurus-based expansion
 - Random keyword insertion
 - Stopword removal
 - Manual query modification
- (xiv) What is the primary disadvantage of using relevance feedback in search queries?
- Increased computational cost
 - Decreased search accuracy
 - Reduction in retrieved documents
 - Loss of index data
- (xv) What is the difference between explicit and implicit feedback?
- Explicit feedback requires user interaction, while implicit feedback is collected automatically
 - Implicit feedback requires user input, while explicit feedback does not
 - Explicit feedback is less accurate than implicit feedback
 - Implicit feedback is only used in Boolean retrieval models

Group-B

(Short Answer Type Questions)

3 x 5=15

- How does query reformulation improve search results? (3)
- What is link spam, and how do search engines combat it? (3)
- Differentiate between explicit and implicit relevance feedback. (3)
- What is PageRank, and how does it work? (3)
- Compare and contrast k-NN and Support Vector Machines (SVM) in text classification. (3)

OR

- Criticize the effectiveness of k-Means clustering for document classification. (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

- What are the different types of compression techniques used in IR? (5)
- How does cosine similarity help in text classification and clustering? (5)
- Discuss the impact of high-dimensional data in text classification and clustering. How can it be reduced? (5)
- What is tolerant retrieval in IR? Discuss different tolerant retrieval techniques. (5)
- Explain how inverted indexing improves the efficiency of retrieval models. (5)

12. Explain the K-Means clustering algorithm with a step-by-step example. (5)

OR

Explain the working principle of the DBSCAN clustering algorithm and its advantages over K-Means. (5)

Library
Brainware University
308, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125
