BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – B.Tech.(CSE)-DS-2022
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) What is the primary purpose of an Information Retrieval System (IRS)?
a) To retrieve all documents from a database
b) To retrieve relevant documents based on a query
c) To delete irrelevant documents
d) To store structured data

(ii) Which of the following components is essential in an inverted index?
a) Query processor
b) Term dictionary and postings list
c) URL resolver
d) PageRank algorithm

(iii) In Boolean retrieval, what is the result of a query using the AND operator?
a) All documents containing any of the query terms
b) Only documents containing both query terms
c) Only documents containing one of the query terms
d) Documents excluding the query terms

(iv) What is the primary advantage of index compression in IR systems?
a) Reducing storage space and improving retrieval speed
b) Improving query performance but increasing storage
c) Deleting unnecessary documents
d) Removing stop words from queries

(v) Which component helps in handling spelling errors in queries?
a) Query expansion
b) Tolerant retrieval
c) Document indexing
d) Term frequency normalization

(vi) What is an example of a stop word?
a) Information
b) The
c) Retrieval
d) Index

(vii) What is the main purpose of a text retrieval system?
a) Image classification
b) Storing large images
c) Retrieving relevant text documents
d) None of the these

(viii) Which component is essential for constructing an inverted index?

a) Term vocabulary
b) Postings list
c) Both a and b
d) None of the these

(ix) Boolean retrieval models rely on which type of logic?
a) Propositional logic
b) Predicate logic
c) Fuzzy logic
d) None of the these

(x) In text retrieval, what does "tolerant retrieval" refer to?
a) Handling spelling variations and errors
b) Ignoring stop words
c) Storing data efficiently
d) Using boolean operators

(xi) Which model should you develop to classify news articles into predefined categories?
a) k-Means
b) Naïve Bayes
c) DBSCAN
d) PAM

(xii) Plan a feature extraction strategy for a text classification model. Which method would be most effective?
a) Using only stopwords
b) Applying TF-IDF to extract important words
c) Using raw text without preprocessing
d) Removing all words except nouns

(xiii) How would you utilize k-NN for text classification?
a) Use it for classification by finding the k nearest neighbors
b) Apply it for text clustering
c) Use it to generate word embeddings
d) None of the above

(xiv) How does Naïve Bayes classification compare with k-NN in terms of computational efficiency?
a) Naïve Bayes is more computationally expensive than k-NN
b) k-NN is faster as it requires model training
c) Naïve Bayes is faster as it computes probabilities instead of distances
d) Both have the same computational complexity

(xv) Contrast supervised and unsupervised learning in the context of text classification.
a) Supervised learning requires labeled data, while unsupervised does not
b) Unsupervised learning always gives better accuracy
c) Both require labeled data for training
d) There is no significant difference between them

Group-B
(Short Answer Type Questions) 3 x 5=15

2. Describe the role of TF-IDF in text classification. (3)
3. Explain how DBSCAN differs from K-Means in text clustering. (3)
4. Explain the PAM algorithm and its role in clustering. (3)
5. Compare Boolean retrieval and query expansion-based retrieval. (3)
6. Explain the structure of the web as a graph. (3)

OR

Explain the role of anchor text in web search. (3)

Group-C
(Long Answer Type Questions) 5 x 6=30

7. How does BM25 differ from the TF-IDF ranking model? (5)
8. How does the KL-divergence retrieval model rank documents? (5)
9. What is tolerant retrieval in IR? Discuss different tolerant retrieval techniques. (5)
10. Explain the importance of feature selection in text classification and discuss common methods used. (5)
11. Explain the term "relevance feedback" in IR and its impact on retrieval performance. (5)
12. Explain the need for index construction in IR systems. How does index compression improve performance? (5)

OR

Explain how inverted indexing improves the efficiency of retrieval models. (5)

****