17749

BRAINWARE UNIVERSITY

Term End Examination 2024-2025
Programme – M.Tech.(CSE)-AIML-2024
Course Name – Intelligent Information Retrieval
Course Code - MTA20111
(Semester II)

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) In Boolean model, correctly identify the main operations for information retrieval.

a) AND, OR, NOT
b) Summation
c) Multiplication
d) Comparison

(ii) Examine the options to correctly identify an example of stop word removal during text normalization

a) "running" -> "run"
b) "Running quickly" -> "run quickly"
c) "The quick brown fox" -> "quick brown fox"
d) "I am learning" -> "learning"

(iii) Analyze and select the correct statement about Naive Bayes classification?

a) Naive Bayes is always the best choice for classification tasks due to its simplicity.
b) Naive Bayes works well when the features are independent but struggles with highly correlated features.
c) Naive Bayes performs poorly on small datasets because of its simplicity.
d) Naive Bayes is most useful when the class distributions are very complex.

(iv) Analyze and select the best approach to mitigate the problem of data sparsity in an N-gram model.

a) Increasing the model's order (e.g., from bigram to trigram)
b) Adding more training data
c) Using a deep neural network to replace the N-gram model
d) Ignoring rare N-grams in the training data

(v) Explain the main goals of XML retrieval.

a) To rank documents based on metadata alone
b) To retrieve documents based on their XML structure and tag-based content
c) To analyze the HTML content of a webpage
d) To normalize the queries and documents for improved search results

(vi) Identify the correct option that explains one key characteristic of user profiles in information retrieval.

a) They are used to evaluate the effectiveness of search algorithms
b) They store information about the user's search history, preferences, and behavior
c) They provide an overall ranking of all available documents
d) They are used only for improving query processing speed

(vii) Choose the correct approach to integrate similarity measures and probabilistic models for document ranking.

a) Apply similarity measures to classify documents and use probabilistic models to adjust the ranking based on likelihood of relevance
b) Use similarity measures to expand the query and probabilistic models to predict user preferences
c) Rely only on similarity measures to rank documents based on content and ignore probabilistic models
d) Combine similarity measures with probabilistic models to rank documents without adjusting for relevance scores

(viii) Choose the correct technique to improve both recall and precision in an information retrieval system that combines query expansion and relevance feedback.

a) Expanding the query with highly relevant terms based on relevance feedback, while removing irrelevant terms
b) Expanding the query by adding random terms to explore all possible relevant documents
c) Using only query expansion to improve recall and relying solely on precision metrics for evaluation
d) Expanding the query by increasing the weight of each term uniformly based on relevance feedback

(ix) Choose the correct option that identifies the entity represented by vector space model in information retrieval.

a) Documents and queries as vectors in a multi-dimensional space
b) Documents and queries as trees in a hierarchical structure
c) Documents and queries as a list of keywords
d) Documents and queries as a set of logical rules

(x) Identify the primary goal of text classification in information retrieval.

a) To group similar documents together
b) To assign predefined labels to documents based on their content
c) To rank documents by their relevance to a query
d) To remove irrelevant documents from the collection

(xi) Choose the correct option to estimate the index size of a corpus with a large number of documents in a web search engine.

a) By counting the number of pages in the corpus
b) By calculating the average number of words per document and the number of unique terms
c) By calculating the average document length and multiplying it by the number of queries
d) By considering the frequency of search engine requests

(xii) Choose the option that describes the most efficient way to handle the indexing of frequently updated documents in a web search engine.

a) Only index documents that are updated once a week
b) Re-crawl and re-index documents regularly to ensure they are up-to-date

c) Ignore documents that are frequently updated
d) Only index static content

(xiii) Identify key advantage of using an inverted index in information retrieval systems

a) It allows for efficient storage of documents
b) It speeds up the search process by allowing fast access to term locations in documents
c) It eliminates the need for relevance ranking
d) It stores documents in a compressed format to save space

(xiv) Determine the indexing strategy to most likely reduce the size of an inverted index.

a) Using a larger vocabulary
b) Using a higher term frequency threshold
c) Storing each term separately in the index
d) Storing terms with low frequency in the index

(xv) Select the option correctly identifying the typical role of the singular values during analysis of text using matrix decomposition techniques like SVD in the context of LSI.

a) They represent the importance of each term in the document collection
b) They determine the weights of terms in the query
c) They represent the strength of relationships between terms and documents
d) They are used to rank the documents based on term frequency

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Explain the advantage of applying dimensionality reduction (e.g., using SVD) in the context of the Vector Space Model to improve performance. (3)
3. Define text classification in Information Retrieval. (3)
4. Explain briefly the methodology behind implementation of a query expansion technique to enhance web search results (3)
5. Briefly explain the approach to be adopted for critically evaluating the effectiveness of a web search engine's ranking algorithm (3)
6. Analyze the benefits and limitations of using N-gram models in information retrieval. (3)

OR

Analyze the strategy undertaken by Laplace Smoothing to prevent overfitting in a Naive Bayes classifier. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Explain briefly the conecpt of relevance feedback and the technqiues used in it to improve search results. (5)
8. Compare Relevance Feedback and Query Expansion in terms of their strengths and weaknesses. (5)
9. Evaluate the impact of some ranking algorithms in modern information retrieval. (5)
10. Evaluate the challenges in handling near-duplicate content on the web and suggest potential solutions. (5)
11. Describe the procedure in which the Boolean Model represents queries and documents using logical operators. (5)
12. Explain briefly the advantage of using hierarchical clustering over flat clustering in document clustering? (5)

OR

Explain in brief the primary approaches of Latent Semantic Indexing (LSI) in improving information retrieval. (5)

****