End Semester Examination / C3 Evaluation

Information Retrieval

Instructions

- The paper is divided into three sections: A, B, and C.
- Section A: All questions are compulsory. (20 Marks)
- Section B: Attempt any 2 out of 3 questions. (10 Marks)
- Section C: Attempt any 1 out of 2 questions. (10 Marks)
- Show all steps for numerical problems to earn partial credit.
- Time Allowed: 3 hours.

Section A: (25 Marks)

All questions are compulsory in this section. Each question carries 5 marks.

Q1. Discuss the trade-offs between precision and recall in information retrieval. Propose a strategy to balance these trade-offs in a large-scale web search engine.

Q2. Describe the design of an IR system where the Corpus has documents in two languages, say, English and Hindi, and the query can be in any language and the system should offer documents of both the languages as the output. How to give more emphasis for the output in one language ?

Q3. Explain the concept of relevance feedback. How does the Rocchio algorithm implement relevance feedback in the vector space model? Include its mathematical formulation.

Q4. Discuss the role of probabilistic language models in information retrieval. Compare unigram and bigram models and explain their use in ranking documents.

Q5. Explain how neural network-based methods, such as BERT, are transforming traditional information retrieval systems. Discuss one challenge in applying neural approaches to IR.

Section B: (10 Marks)

Attempt any 2 out of 3 questions in this section. Each question carries 5 marks.

Q1. A retrieval system retrieves 50 documents for a query. Of these,15 documents are relevant, and there are 25 relevant documents in total for the query.(

(a) Calculate precision, recall, and F1 score for the system.

(b) Calculate E score, assuming suitable values of alpha and beta. Compare it with F1 score in (a).

Q2. Term-Document Matrix: Given a document-term matrix:

$$A = \begin{matrix} 4 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 2 & 3 \end{matrix}$$

Perform Singular Value Decomposition (SVD) to calculate the first singular value and its corresponding singular vectors. Explain how these singular vectors can be used for Latent Semantic Analysis (LSA).

Q3. Query Similarity: Two queries, Q1 and Q2, are represented as vectors in a 3-dimensional term space:
Q1 = [0.5, 0.7, 0.1], Q2 = [0.2, 0.9, 0.4].
Calculate the cosine similarity between the two queries and interpret the result in the context of information retrieval.

Section C: (10 Marks)

Attempt any 1 out of 2 questions in this section. Each question carries 10 marks.

Q1. Relevance Feedback and Query Expansion: A retrieval system processes a query and retrieves documents based on cosine similarity. Assume the term frequencies for the initial query and a relevant document are: Query: Q = [2, 1, 0] Relevant Document: D = [1, 3, 1]
(a) Rewrite the query using Rocchio's formula:
$Q_{new} = \alpha Q + \beta D$
where $\alpha = 1$, $\beta = 0.8$.
(b) Explain how this query expansion technique improves retrieval performance. Also, discuss one limitation of relevance feedback in practical applications.

Q2. Assume that documents are coloured images and the query is also an image. Which model is applicable to such an Image Retrieval System. Explain the design of such a system, comparing it with Document Retrieval System.