

Tribhuvan University
Institute of Science and Technology
2077



Bachelor Level / Fourth Year /Seven Semester/Science
Computer Science and Information Technology-(CSc.405)
(Information Retrieval)

Full Marks: 60
Pass Marks: 24
Time: 3 hours.

Candidates are required to give their answers in their own words as far as practicable.

The figures in the margin indicate full marks.

Attempt any TEN questions.

1. How the terminology Web Search and IR related? Justify. Explain the way to represent the document vector with example. (2+4)
2. List any one defect of Boolean retrieval model. Describe the role of cosine similarity and significance of inverse document frequency. (1+3+2)
3. Consider the universe of documents, D1, D2, D3, D4, D5, D6, D7. For a particular query, documents D1, D2, D4 are relevant. However our information retrieval system returns D1, D2, D3, D5, D7. Calculate the Recall, Precision and F - Measure for this query. (6)
4. Explain the ways to expand the query with its significance. How can you correct the misspelling? Describe. (3+3)
5. How does search engine work? Explain with the architecture. (6)
6. Consider the problem of clustering the following documents using K-means with K=2 cosine similarity.
Doc1: "go Longhorns go"
Doc2: " go Texas"
Doc3: "Texas Longhorns"
Doc4: "Longhorns Longhorns"
Assume Doc1 and Doc3 are chosen as initial seeds. Assume simple term frequency weights (no IDF). Show all similarity calculations to cluster the documents, centroid computations for each iteration, and the final clustering. The algorithm should converge after only 2 iterations. (6)
7. What is the role of recommendation engine in IR? Explain about content based recommendation system. (2+4)
8. Describe about brute force algorithm to find the pattern in the given text. (6)
9. Multimedia IR has been emerging in social network and search engine. Justify it with examples and list some issues too. (6)
10. State Zip's Law. How index term are selected to represent the document? Explain. (2+4)
11. How the texts are tokenized? What do morphological analysis and lemmatization refer to? (2+4)
12. Explain the different types of query languages. (6)