Tribhuvan University
Institute of Science and Technology
2079
✡

Bachelor Level / Second Year/ Third Semester/ Science   Full Marks: 60
Computer Science and Information Technology (CSC413)   Pass Marks: 24
(Information Retrieval)   Time: 3 hours.
(NEW COURSE)

Candidates are required to give their answers in their own words as for as practicable.
The figures in the margin indicate full marks.

Section A

Attempt any TWO questions.   **[2×10=20]**

1. How does information retrieval differentiate with data retrieval? Why do we need to stem the word? Describe the pattern of the rules used in Porter Stemmer with example.   [2+2+6]

2. List the limitations of Boolean retrieval model. Rank the query "eat apple" with respective to following documents.   [2+8]
   Doc1 = "eat apple banana grapes"
   Doc2 = "apple banana"
   Doc3 = "eat mango grapes"
   Doc4 = "eat apple appleapple"
   Doc5 = "mango grapes"

3. What is KL divergence? Differentiate between Relevance feedback and Pseudo relevance feedback.   [2+8]

Section B

Attempt any EIGHT questions.   **[8×5=40]**

4. How do you calculate mean average precision? Explain.   [5]

5. Differentiate between dictionary and thesaurus and describe their roles in IR.   [5]

6. Explain the architecture of search engine.   [5]

7. Define Snippet. What are the advantages of CLIR?   [2 + 3]

8. How KNN is used as a classifier? Explain.   [5]

9. What is text shingling? Differentiate between k-means and k-medoids.   [2 + 3]

10. List the limitations of collaborative and content based recommendation.   [5]

11. Describe the process of question and answer processing by a QA system.   [5]

12. Explain the role of Latent Semantic Indexing in IR.   [5]