

Tribhuvan University
Institute of Science and Technology
 2082
 ☆

Bachelor Level / Fourth Year/ Seventh Semester/ Science
Computer Science and Information Technology (CSC413)
 (Information Retrieval)
(NEW COURSE)

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.

Section A

Attempt any TWO questions.

[2×10=20]

1. In which cases we use CLIR? Explain about HITS and Page ranking as link analysis algorithm. [2 + 8]
2. What do you mean by information bottleneck? Explain the architecture of Question Answering system. [3 + 7]
3. Why SVD task is complex? Describe how Knuth–Morris–Pratt can be used in string matching with example. [3 + 7]

Section B

Attempt any EIGHT questions.

[8×5=40]

4. Define IR system. Describe some related areas. [2 + 3]
5. Why do we need to remove stop word? Does it always be beneficial? Justify with an example. [2 + 3]
6. Discuss about vector space model. [5]
7. Describe about precision, recall and f – measure as evaluation metrics. [5]
8. What is WordNet? Describe about relevance feedback in query processing? [1 + 4]
9. Assume the following dataset. [5]
 Food: “Turky stuffing”
 Food: “Buffalo wings”
 Beverage: “Cream soda”
 Beverage: “Orange soda”
 Apply the Rocchio algorithm to classify a new name “Turky Soda”.
10. Explain the logic behind Expectation Maximization approach in clustering. [5]
11. Describe the working mechanism of collaborative recommendation approach. [5]
12. State Zipf’s law. How to you represent document using TF vector? [2 + 3]