

Tribhuvan University
Institute of Science and Technology
2076

Bachelor Level / seventh-semester / Science

Computer Science and Information Technology(CSC413)

(Information Retrieval)

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

The figures in the margin indicate full marks.

Full marks: 60

Pass marks: 24

Time: 3 hours

Attempt any TEN questions.

1. How IR in web search is different from other IR systems? Discuss IR architecture with suitable diagram.(2+4)

2. Assume that document Space is defined by four terms: Network, CSIT, Nepal, TU, and Graduate.

And we have three documents containing the following terms:

Doc1: CSIT Nepal

Doc2: TU CSIT

Doc3: CSIT TU Nepal

If the Query is "CSIT Nepal", find top 2 document retrieved by Boolean space model.(6)

3. What is meant by stop word removal? Explain text normalization with suitable example.(1+5)

4. Suppose that table below list all the documents retrieved by an algorithm. If total number of relevant documents is 6, calculate the value of recall, precision, and F-score.(6)

SN	Doc ID	relevant
1	D1	no
2	D2	no
3	D3	yes
4	D4	no
5	D5	yes
6	D6	yes
7	D7	no
8	D8	no
9	D9	yes

5. Why query expansion is important? Discuss query expansion techniques with examples.(1+5)

6. Why Hits algorithm is used? Discuss its working with example.(2+4)

7. How Bots are different from spiders? Describe simple and multithreaded spidering algorithm.(1+5)

8. How text categorization is different from clustering? Explain nearest neighbor categorization algorithm.(1+5)

9. Differentiate collaborative filtering from content based filtering? Discuss content based recommender system with its strengths and drawbacks.(2+4)

10. Why TF-IDF weighting is important in information retrieval? Explain with suitable example.(6)

11. How information extraction differs from information retrieval? Discuss role of XML in information extraction.(6)

12. Write short notes on:(3+3)

a) Latent Semantic Indexing

b) Spiders