



Reg. No. :

| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

Question Paper Code : 40887

B.E./B.Tech. DEGREE EXAMINATION, APRIL/MAY 2018

Seventh Semester

Computer Science and Engineering

CS 6007 – INFORMATION RETRIEVAL

(Regulations 2013)

Time : Three Hours

Maximum : 100 Marks

(Codes/Tables/Charts to be permitted, if any, may be indicated)

Answer ALL questions

PART – A

(10×2=20 Marks)

1. Give any two advantages of using artificial intelligence in information retrieval tasks.
2. How does the large amount of information available in web affect information retrieval system implementation ?
3. Can the tf-idf weight of a term in a document exceed 1 ? Why ?
4. Consider the two texts, "Tom and Jerry are friends" and "Jack and Tom are friends". What is the cosine similarity for these two texts ?
5. How do spammers use cloaking to server spam to the web users ?
6. Can a digest of the characters in a web page be used detect near duplicate web pages ? Why ?
7. Compute the Jaccard's similarity for the two list of words (time, flies, like, an, arrow) and (how, time, flies).
8. How is cross-lingual retrieval typically implemented ?
9. Give an example for a Dendrogram.
10. What is the use of the Expectation-Maximization algorithm ?



11. a) List and describe the components of an Information Retrieval (IR) system in detail.

(OR)

- b) Identify and describe the components of a search engine with a neat diagram.

12. a) Explain Latent Semantic Indexing in detail. Why is Latent Semantic Indexing used in information retrieval tasks ?

(OR)

- b) Discuss the Query likelihood model in detail and describe the approach for information retrieval using this model.

13. a) Illustrate the various challenges in XML retrieval with appropriate examples.

(OR)

- b) Draw the basic crawler architecture and describe its components.

14. a) How does MapReduce work ? Illustrate the usage of MapReduce programming model in Hadoop.

(OR)

- b) Discuss the design of a Question-Answer engine with the various phases involved. How can the performance of such an engine be measured ?

15. a) Discuss Naïve Bayes Text classification in detail and how it can be implemented ?

(OR)

- b) How is k-means clustering algorithm used in information retrieval ? Discuss the distance measure used in the algorithm and how the value for k can be fixed ?