R13

INFORMATION RETRIEVAL SYSTEMS
(Common to CSE and IT)

Time: 3 hours   Max. Marks: 70

PART – A
(Compulsory Question)
*****

1   Answer the following: (10 X 02 = 20 Marks)
   (a)   What are the objectives of information retrieval system?
   (b)   How Jaccard coefficient is defined for query Q and document D_i?
   (c)   What is meant by one-pass clustering?
   (d)   Give Tenfel coefficient for similarity.
   (e)   How to compute K-distance?
   (f)   What is stemming?
   (g)   What is the use of signature files?
   (h)   Quote example for byte-aligned compression.
   (i)   What is the use of non-first normal form approach?
   (j)   How to improve effectiveness of web search engine?

PART – B
(Answer all five units, 5 X 10 = 50 Marks)

UNIT – I

2   Describe the simple term weight model as a retrieval strategy.
OR
3   (a)   Illustrate the basic idea of language model.
   (b)   What is the need of smoothing? Explain any one approach.

UNIT – II

4   Explain the relevance feedback process and discuss the initial use of the vector space model to implement this.
OR
5   (a)   With an example, explain Ward's method.
   (b)   Describe the two stages of logistic regression.

UNIT – III

6   What is word net? Discuss its features and use as a retrieval utility. Give examples for entailment and troponyms.
OR
7   Briefly discuss the four approaches for choosing translations and the need to form these translations into a new query for the target language.

UNIT – IV

8   (a)   Discuss vector space simplifications to improve efficiency.
   (b)   How to remove false positives for signature files?
OR
9   Explain methods to remove near duplicates in documents.

UNIT – V

10   Discuss the features of XML-query language. Explain tracking XML documents.
OR
11   (a)   Explain relevance ranking computation using unchanged SQL with illustration.
   (b)   Demonstrate simple page rank calculation.

*****