Suman Saha

COURSE CODE: 11M1WCI431   MAX. MARKS: 25

COURSE NAME: Advanced Web Mining

COURSE CREDITS: 3   MAX. TIME: 1Hr 30 Min

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.

1. a) What is vector space model for text documents?
   b) Describe TFIDF score for index matrix.
   c) What is cosine similarity between documents?
   d) Explain the ranking of documents with respect to a query, using TFIDF score and cosine similarity.

   [2+2+2+2]

2. a) What is stop-press indexing? Explain the key word search under stop-press indexing.
   b) Explain map-reduce algorithm for constructing large scale inverted index

   [3+4]

3. a) What is the need of ranking of web pages?
   b) Write the Page-rank algorithm and derive its matrix equation.
   c) Write the HITS algorithm for computing rank of the documents. Explain the solution of the recursive matrix equation.   [1+4+5]