BEFORE ANSWERING THE QUESTION PAPER THE CANDIDATE SHOULD ENSURE THAT THEY HAVE BEEN SUPPLIED THE CORRECT QUESTION PAPER. COMPLAINTS IN THIS REGARD, IF ANY, SHALL NOT BE ENTERTAINED AFTER THE EXAMINATION.

NOTE: QUESTION NO. 1 IS COMPELLARY AND ATTEMPT ANY TWO QUESTIONS FROM EACH SECTION. ALL QUESTIONS CARRY EQUAL MARKS.

1. (5x4=20)
   (a) What are the various problems associated in information retrieval? Do you have any solutions to handle them, is any then explain them.
   (b) Why the index is constructed? Discuss different types of index known to you.
   (c) What is inverse document frequency? Explain its use.
   (d) Define the query likelihood model. Give some applications of this model.

SECTION - A

2(a) Give the steps for processing Boolean queries. Describe the benefits of Boolean query processing.
   (10)
   (b) Describe inverted index. How it is different than Bi-word index?
   (10)
3(a) Explain blocked sort-based indexing method. Why it has excellent scaling properties?
   (10)
   (b) Describe the dictionary compression mechanism. Explain its relation with response time of IR system and memory related issues.
   (10)
4(a) Explain the benefit of scoring in IR. Describe the weighted zone scoring method.
   (10)
   (b) What were the various issues with Jaccard for calculating score and how it can be overcome with tf-idf and vector space model.
   (10)

SECTION - B

5(a) What is inexact top K document retrieval? Explain it and suggest some algorithm or heuristics for this.
   (10)
   (b) Write short notes on Champion lists.
   (10)
6(a) Define cloaking and spamming, explain the concept of cloaking used by spammers.
   (10)
   (b) Describe the page rank and explain the page rank computation method.
   (10)
7(a) Describe the need of language model for information retrieval. Explain various types of language models used in IR.
   (10)
   (b) What is URL frontier? Explain with the help of diagram.
   (10)