Name : --- Department: ----- Academic no.:----------
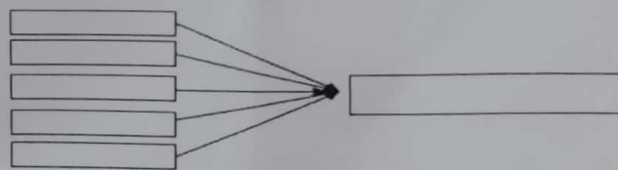
Answer the following questions:

Questions 1:

1) An inverted index arranges data in a sorted order as per

   a) the documents

   c) the frequency of each term

   b) the frequency of each document

   d) the terms

2) What is the disadvantage of Boolean retrieval model?

   a) Easy to implement

   b) Difficult to rank output

   c) Difficult to process a query

   d) It is one of the complex retrieval model

3) A data structure that maps terms back to the parts of a document in which they appear is called

   a) Lexicon

   c) Inverted index

   b) Dictionary

   d) All of the above

4) In thesaurus-based query expansion technique, for every term t in a query, we expand the query using the synonyms and related words of t taken from the thesaurus. This can increase _____

   a) precision

   c) performance

   b) recall

   d) mean average precision

5) Which of the following is a non-decreasing function of the number of documents retrieved?

   a) precision

   c) accuracy

   b) recall

   d) mean average precision

6) In tokenization language issues we may use

   a) package splitter

   c) compound splitter

   b) symmetric splitter

   d) all of the above

7) The study of identifying the meaningful parts of words, i.e. prefixes, suffixes and roots.

   a) lemmatization

   c) morphological analysis

   b) stemming

   d) both a,b

8) Each item in the list is called as_____.

a) Items

b) Posting

c) Query

d) Information

9) etr term is called _____k-grams wildcard query.

a)3

b)4

c)1

d)2

10)Term document incidence matrix is

a) Sparse

b) Depends upon the data

c) Dense

d) Cannot predict

11) ....is usually provided by most modern information retrieval systems.

a) Information and knowledge retrieval

b) Information or knowledge retrieval

c) Iinformation and data retrieval

d) Information or data retrieval

12) Which of the folllowing is correct search statement for "Select all items that discuss software and not storage or hardware in the item"?

a) SOFTWARE OR (STORAGE NOT HARDWARE)

b) SOFTWARE OR STORAGE NOT HARDWARE

c) SOFTWARE AND NOT STORAGE OR HARDWARE

d) SOFTWARE OR NOT STORAGE AND HARDWARE

Questions 2:

a)

1) Write down the entries in the permuterm index dictionary that are generated by the term bord.

2) If the given Query is:"bo*d" which entry from the previous permuterm index dictionary will be used to look up with?

b) Complete the following term-document incidence matrix for these documents collection.

Doc 1   breakthrough drug schizophrenia

Doc 2   new schizophrenia drug

Doc 3   new approach treatment schizophrenia

Doc 4   new hopes schizophrenia patients

	Doc 1	Doc 2	Doc 3	Doc 4
Approach				
breakthrough				
drug				
hopes				
new				
patients				
schizophrenia				
treatment				

what are the returned results for this query: drug AND Not Schizophrenia