Mock Examination SEP -2020
Information Retrieval

1) A data structure that maps terms back to the parts of a document in which they occur is called an

   a) Postings list
   b) Incidence Matrix
   c) Dictionary
   d) Inverted Index

2) Which is not a valid statement for Vector Space model?

   a) Documents and queries are represented as vectors in a common vector space.
   b) Retrieval is based on the sine similarity between the query vector and the document vector.
   c) The size of the vector is equal to the size of the dictionary.
   d) A query can also be viewed as a vector of very short document.

3) Information retrieval systems have much in common with
   a)Filing systems
   b)Transaction systems
   c)Database systems
   d)Management systems

4) _____is a way of measuring the importance of website pages.
   a) Querying
   b) **Page Rank**
   c) Link Analysis
   d) HITS

5) The standard approach to information retrieval system evaluation involves around the notion of:

   a) Quantity of documents in the collection
   b) **Relevant and non relevant documents.**

c) Accuracy
d) user happiness

6) _____ is a page that contains actual information on a topic.
a) **authority**
b) Hub
c) Hyperlinks
d) Image

7) The number of documents in the collection that contain a term t is called as ____
a) Document Index di_t
b) **Document frequency df_t**
c) Document Inverse din_t
d) Document Incidence Matrix dim_t

8) A good_____page for a topic links to many authority pages for that topic.
a) Crawler
b) SEO
c) Web
d) **Hub**

9) PM stands for _____
a) Cost per migrating
b) Cost per making
c) Cost per manage
d) **Cost per mil**

10) For taking decisions data must be
a) Very Accurate
b) Massive
c) **Processed Correctly**
d) Collected from diverse sources

11) SEO stands for _____

a) Search engine order
b) Search engine organizer

c) Search engine option
d) **Search engine optimization**

12) Which of the following is a technique for context sensitive spelling correction:
   a) the Jaccard Coefficient
   b) Soundex algorithms
   c) <u>**k-gram indexes**</u>
   d) Levenshtein distance

13) Operational information is
   a)Haphazard
   b)Well organized
   c)Unstructured
   d)Partly structured

14) A deadlock can be broken down by
   a) Committing one or more transaction
   b) Aborting one or more transactions
   c) **Rolling back one or more transactions**
   d) Terminating one or more transactions

15) The list of web pages that a web crawler has queued up to index is called the:

   a) Web Page Queue
   b) Seed set
   c) URL Filter
   d)**URL Frontier**

16) Variable size postings lists is used when
   a) More seek time is desired and the corpus is dynamic
   b) Less seek time is desired and the corpus is dynamic
   c) Less seek time is desired and the corpus is static
   d) **More seek time is desired and the corpus is dynamic**

17) The _____ is a wild card that represents one or more characters
   a) question mark
   b) **asterisk**
   c) exclamation mark
   d) dollar sign

18) _____fraction of the returned results are relevant to the information need.

    a) Proximity
    b) Posting Merge
    c) Posting list
    d) Precision

19)_____ is the task of chopping documents into the pieces.
a) Ranked
b) Wild card
c) Tokenization
d) Boolean retrieval

20) Permuterm indices are used for solving:

a. Algebraic queries
b. Boolean queries
c. Phrase queries
d. Wildcard queries