PUNE VIDYARTHI GRIHA'S
COLLEGE OF SCIENCE & TECHNOLOGY

Final Examination Sept-2020
T.Y.BSc.Computer Science (Semester-VI)

Subject – Information Retrieval
(USCS604)

Date:

1. **Information retrieval is querying of _____ textual data.**
a) structured
b) **unstructured**
c) formatted
d) None

2. **_____are indexed units in the incidence matrix.**
a) **Terms**
b) Collection
c) Information
d) Data

3. **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

4. **The standard way of quantifying the similarity between two documents d1 and d2 is to compute the _____of their vector representations.**
a) sine similarity
b) cot similarity
c) **cosine similarity**
d) None

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

6. **_____fraction of tChe returned results are relevant to the information need.**
a) Proximity
b) Posting Merge
c) **Precision**
d) Posting list

7. **A dictionary of terms is sometime also referred as_____**
a) **Corpus**
b) Collection

c) Lexicon
d) none of the above

8. A model of information retrieval in which we can pose any query in which search terms are combined with the operators AND, OR, and NOT
a) Ad Hoc Retrieval
b) Ranked Retrieval Model
c) **Boolean Information Model**
d) D Proximity Query Model

9. The number of times that a word or term occurs in a document is called the:
a) **Term Frequency**
b) Vocabulary Lexicon
c) Proximity Operator
d) D Indexing Granularity

10. Given two strings s1 and s2, the edit distance between them is sometimes known as the:
a) **Levenshtein distance**
b) isolated-term distance
c) porter stemmer algorithm

11. _____nodes that can be reached from the giant SCC but cannot reach it.
a) In
b) Out
c) Gcc
d) in-out

12. Postings list should be sorted by:
a) Document Frequency
b) **DocID**
c) TermID
d) Term frequency

13. TREC stands for_____
a) **a.Text Retrieval Conference**
b) b. Text Retrieval Context
c) c.Text Retrieval Congestion
d) d.All the above

14. CLEF stands for_____
a) Cross Language Evaluation Forum
b) Cross lingual evaluating field
c) Cross Language Evaluating Field
d) Cross Language Evaluating Forum

15. A good_____page for a topic links to many authority pages for that topic.
a) Crawler
b) SEO

c) Web
d) **Hub**

16. A large repository of documents in IR is called as
a) **Corpus**
b) Database
c) Dictionary
d) Collection

17. Term document incidence matrix is
a) **Sparse**
b) Depends upon the data
c) Dense
d) Cannot predict

18. Document frequency of a term is the
a) **Number of documents that contain the term**
b) None of the above
c) Number of times the term appears in the document
d) Number of times the term appears in the collection

19. The _____ for finding terms based on a query consisting of k-grams.
a) Document index
b) k-gram index
c) Inverted index
d) Term index

20. _____ mainly encodes numerical and non-text attribute-value data.
a) Data centric XML
b) **text centric XML**
c) both a and b
d) User centric XML

21. Permuterm indexes are used for solving
a) Spelling Checking
b) b. Boolean queries
c) c. Phrase queries
d) **d. Wildcard queries**

22. Each node of the tree is an XML element and is written with an _____
a) Opening tag
b) closing tag
c) both a and b
d) only a

23. The _____consists of a dictionary of terms.
a) Bi-gram index
b) K-gram index
c) Inverted index
d) Incidence index

24. _____includes link building, increasing link popularity by submitting open

directories, search engines, link exchange, etc.

a) **Off Page SEO**
b) In Page SEO
c) Middle Page SEO
d) Both a nd b

25. Best implementation approach for dynamic indexing is

a) Periodic re indexing
b) Using Invalidation bit vector for deleted docs
c) None
d) **Using logarithmic merge**