Question Paper Code: 59218

B.E./B.Tech. DEGREE EXAMINATION, MAY 2021

Elective

Computer Science and Engineering

15UCS918- INFORMATION RETERIVEL

(Regulation 2015)

Duration: 1:45 hrs Maximum: 50 Marks

PART A

(Answer any Ten Questions 10 x 2 Mark = 20 Marks)

1. A _____ is the topic about which the user desires to know more. **(CO1-U)**

a. Effectiveness b.ad hoc retrieval

c. Information need d. precision

2. _____ is the process of selecting how to organize the work of answering a query so that the least total amount of work needs to be done by the system. **(CO1-U)**

a. Intersect b. Query optimization

c. Posting merge d. Conjunctive query

3. A group of related documents against which information retrieval is employed is called---

a. Corpus b. Text Database **(CO1-U)**

c. Index Collection d. Repository

4.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: **(CO2-U)**

a. Ad Hoc Retrieval b. Ranked Retrieval Model

c. Boolean Information Model d. Proximity Query Model

5. A measure of similarity between two vectors which is determined by measuring the angle between them is called: **(CO2-U)**

a. Cosine similarity b. Sin similarity

c. Vector similarity d. Vector scoring

6. Major drawback suffered by Boolean model is due to the fact **(CO2-U)**

a. Model based on set theory and Boolean algebra

b. Queries are specified as Boolean expression

c. Retrieval strategy is based on binary decision criteria

d. Model predict that each document is either relevant or non-relevant

7. A web server communicates with a client (browser) using which protocol: **(CO3-U)**

a. HTML b. HTTP c. FTP d. Telnet

8. The basic operation of a web browser is to pass a request to the web server. This request is an address for a web page and is known as the **(CO3-U)**

a. UAL: Universal Address Locator
b. HTML: Hypertext Markup Language
c. URL: Universal Resource Locator
d. HTTP: Hypertext transfer protocol

9. The list of web pages that a web crawler has queued up to index is called the: **(CO3-U)**

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

10. Link analysis is one of many factors considered by _____ in computing a composite score for a web page on any given query. **(CO4-U)**

a. Web application b.web search engines

c. Web graph d. PageRank

11. The _____ of a node will depend on the link structure of the web graph. **(CO4-U)**

a. Matrix b. PageRank c. teleport d. probability

12. A good _____ is one that points to many good authorities. **(CO4-U)**

a. Specific page b.hub page c. authority page d.web page

13. _____ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. **(CO5-U)**

a. Data mining b. Text mining
c. File mining d. Deep mining

14. The process of breaking out long-form text into sentences and words called? **(CO5-U)**

a. Stem b. Cluster c. Bag d. Tokens

15. Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences, **(CO5-U)**

a. True b. False c. Can be true or false d. Can not say

PART – B (Answer any Three Questions 3 X 10 = 30 Marks)

1. Demonstrate the working of IR architecture with a neat diagram **(CO1-U)**

2. Sort and rank the documents in descending order according to the similarity values: Suppose we query an IR system for the Query Q="Gold silver truck" and Demonstrate the working of IR architecture with a neat diagram **(CO2-App)**

   The database collection consists of three documents (D = 3) with the following content

   D1="Shipment of gold damaged in a fire"

   D2="Delivery of silver arrived in a silver truck"

   D3="Shipment of gold arrived in a truck"

   i) Calculate Term document frequency and Inverse document frequency for given document and query

   ii) Calculate Similarity Coefficient between query (Q) with all three documents (D1,D2 and D3)

3. Design and develop a Web search Architecture and the components of search engine and its issues. **(CO3-APP)**

4. i)Define Link Analysis and explain in detail. (5) **(CO4-U)**
   ii) Describe in detail about HUBS and Authorities.(5) **(CO4-U)**

5. Develop an algorithm for classification using K Nearest Neighbor. Illustrate the algorithm with a relevant example. **(CO5-U)**