

B.E./B.TECH. Degree Examination, December 2020
Semester - VI
CS16002 – INFORMATION RETRIEVAL TECHNIQUES
(Regulation 2016)

Time: Three hours

Maximum : 80 Marks

Answer **ALL** questions

PART A - (8 X 2 = 16 marks)

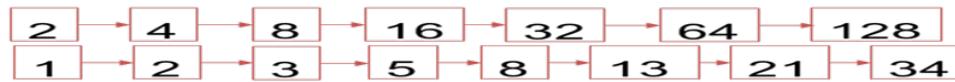
1. Opinion mining is _____
 - a) the science of using text analysis to understand the drivers behind public sentiment.
 - b) a set of rules for optimizing a website for search engines
 - c) to use information about whether or not those results are relevant to perform a new query.
 - d) is a ranking of the documents in the collection in order of decreasing probability of relevance
2. How do you measure the success of a web site?
 - a) Number of Visitors
 - b) Average Time On Page
 - c) Bounce Rate
 - d) All of the above
3. Information retrieval is _____
 - a) a process through which a computer system can respond to a user's query for text-based information on a specific topic
 - b) a binary decision criterion
 - c) All of the above
 - d) None of the above
4. Assume there are 8 relevant documents in total in the collection RRNNNNNNRRNR NNNRNNNNR where Rs and Ns represents relevant (R) and nonrelevant (N) returned documents in a ranked list of 20 documents retrieved in response to a query from a collection of 10,000 documents. Calculate the precision of the system on the top 20?
 - a) 0.3
 - b) 0.03
 - c) 3.0
 - d) 0.003
5. Discuss Probability Ranking Principle.
6. Give reasons why does behavioral targeting seem to be successful in search engines.
7. Discuss the need for Search Engine Optimization.
8. Illustrate with an example cluster hypothesis principle.

PART B - (4 X16 = 64 marks)

09. (a) (i) Illustrate the various components of an Information Retrieval system. (12)
 (ii) Suppose $N = 1$ million, compute idf_t for the term “animal” whose df_t is 10000 (4)

(OR)

- (b) (i) How web search engine works? Detail on the architecture of web search engine. (12)
 (ii) Examine the query “Brutus AND Caesar OR Calpurnia” for any given posting lists



10. (a) Consider the three documents (16)

D1 : ant

D2 : dog dog

D3 : cat dog fox

A) Find out the similarity between these documents using

- a. No weighting scheme and
 b. Term frequency weighting scheme

B) For the given query “ant dog”, find out the similarity of the given query to documents

(OR)

- (b) A collection consists of the following documents: (16)

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.

Use Latent Semantic Indexing to rank these documents for the query “gold silver truck”

11. (a) (i) Explain how social networks can be extracted from the web. (8)
 (ii) How are searching techniques applied in social networks? Illustrate the working of various search techniques in social networks. (8)

(OR)

- (b) (i) Illustrate Collaborative filtering with a suitable example. (8)
 (ii) When is opinion mining preferred? With various real world scenarios, illustrate the usage of various strategies in opinion mining. (8)

12. (a) Justify the need for Hierarchical Agglomerative clustering (HAC) algorithm along with its variants for various applications. (16)

(OR)

- (b) (i) Illustrate Naive Bayes text classification algorithm for the problem given below (8)

Phase	Doc	Words	Class
Training	1	Chinese Beijing Chinese	C
	2	Chinese Chinese Shanghai	C
	3	Chinese Macao	C
	4	Tokyo Japan Chinese	J
Test	5	Chinese Chinese Chinese Tokyo Japan	?

- (ii) With suitable example and algorithm, illustrate the working of K-means (8) clustering and discuss on the various similarity measures that can be used for modeling.