PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

TEST 1

Winter Semester: 2021 - 22

Course Code: CSE 232

Course Name: Information Retrieval and Organization

Program & Sem: B.TECH & VI Sem

Date: 26th April 2022

Time: 1:30 PM to 2:30 PM

Max Marks: 30

Weightage: 15%

Instructions:

(i) Read the all questions carefully and answer accordingly.

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries TWO marks.   **(5Qx 10M= 10M)**

1. Define information retrieval   (C.O.No.1) [Knowledge's level]
2. What are the applications of IR?   (C.O.No.1) [Knowledge's level]
3. Give the functions of information retrieval system   (C.O.No.1) [Knowledge's level]
4. Define relevance.   (C.O.No.2) [Knowledge's level]
5. What is meant by stemming?   (C.O.No.2) [Knowledge's level]

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries FIVE marks.   **(2Qx5M=10M)**

6. Explain Boolean model for IR   (C.O.No.2) [Comprehensive level]
7. Describe the architecture of Information Retrieval system with a neat diagram.
    (C.O.No.1) [Comprehensive level]

Part C [Problem Solving Questions]

Answer the Question. The question carries TEN marks.   **(1Qx10M=10M)**

8. Find Nuclear AND Treaty using Boolean Model.   (C.O.No.2) [Application level]

	Nuclear	NonProfile	Treaty	Iran
D1	0	0	1	0
D2	1	0	1	1
D3	0	0	0	0
D4	0	0	0	1
D5	1	1	1	0
D6	0	0	0	1
D7	1	0	0	0
D8	0	1	0	1

Find Nuclear AND Treaty using Boolean Model ?

PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

TEST 1

Winter Semester: 2021 - 22

Course Code: CSE232

Course Name: Information Retrieval and Organization

Program & Sem: B.TECH VI Sem

Date: 10th May 2022

Time: 2:00 PM-3:00 PM

Max Marks: 30

Weightage: 15%

Instructions: *Read the all questions carefully and answer accordingly.*

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries 2 marks.   (5Qx 10M= 10M)

Q.NO1. Give any two historical developments of Information Retrieval.

[2](C.O.No.1)[Knowledge's level]

Q.NO2. Differentiate between data retrieval and Information Retrieval.

[2](C.O.No.1)[Knowledge's level]

Q.NO3. List any two issues in information retrieval system   [2](C.O.No.1)[Knowledge's level]

Q.NO4. How can we represent the queries in Boolean model?   [2](C.O.No.2)[Knowledge's level]

Q.NO5. What are the disadvantages of Boolean model?   [2](C.O.No.2)[Knowledge's level]

Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries 5 marks.   (2Qx5M=10M)

Q.NO6. Summarize the impact of WEB on Information Retrieval.

[5](C.O.No.1)Comprehensive level]

Q.NO7. Discuss the indexing, retrieval and ranking process in Information Retrieval with a diagram.   [5](C.O.No.1)Comprehensive level]

Part C [Problem Solving Questions]

Answer all the Questions. Each question carries 10 marks.   (1Qx10M=10M)

Q.NO8.   Query : Advance and Structure Not Analysis. Solve the query by applying Boolean model
[10](C.O.No.2)[Application level]

Terms/Doc	D1	D2	D3	D4	D5
English	1	0	0	0	0
Tutorial	1	0	0	0	0
Fast	1	0	0	0	0
Track	1	0	0	0	0
Book	0	0	1	0	0
Semantics	0	1	1	1	0
Advance	0	0	0	1	0
Analysis	0	0	0	0	1
Learning	0	1	0	0	0
Latent	0	1	0	0	1
Index	0	1	1	1	0
Structure	0	0	0	1	1

Query : Advance And Structure Not Analysis ? using Boolean Model

PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

TEST 2

Winter Semester: 2021 - 22

Course Code: CSE232

Course Name: Information Retrieval and Organization

Program & Sem: B.Tech. VI Sem

Date: 01st June 2022

Time: 1:30 PM to 2:30 PM

Max Marks: 30

Weightage: 15%

Instructions:

(i) Read the all questions carefully and answer accordingly.

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries Two marks.   (5Qx 2M= 10M)

1. Compare Boolean and Vector Space Retrieval Model   (C.O.No.2) [Knowledge's level]

2. What is the need of tf-idf weighting model and write the formula for calculating the score?

(C.O.No.2) [Knowledge's level]

3. Write the sparse matrix for the given document   (C.O.No.2) [Knowledge's level]

$$\begin{bmatrix} 0 & 3 & 0 & 4 \\ 0 & 5 & 0 & 7 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 9 & 1 \end{bmatrix}$$

4. What is meant by Collection Frequency?   (C.O.No.2) [Knowledge's level]

5. Why Euclidean distance is BAD idea in Vector Space Model?   (C.O.No.2) [Knowledge's level]

Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries Five marks.   (2Qx5M=10M)

6. Generate your own scenario for disease diagnosis and write the confusion matrix and also calculate the Precision and Recall.   (C.O.No.2) [Comprehensive level]

7. Explain the process of initial stages of text preprocessing.   (C.O.No.2)[Comprehensive level]

Part C [Problem Solving Questions]

Answer the Questions. The question carries Ten marks.   (1Qx10M=10M)

8. Compute the cosine similarities between (a) docs 1 and 2 (b) docs 3 and 4 (without using weights)

Doc 1: Information Retrieval Systems   Doc 2: Information Storage

Doc 3: Digital Speech Synthesis Systems   Doc 4: Speech Filtering, Speech Retrieval

PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING

END TERM EXAMINATION

Winter Semester: 2021 - 22

Course Code: CSE 232

Course Name: Information Retrieval and Organization

Program & Sem: B.TECH & VI Sem

Date: 30th June 2022

Time: 09:30 AM to 12:30 PM

Max Marks: 100

Weightage: 50%

Instructions:
   (ii) *Read the all questions carefully and answer accordingly.*

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries FIVE marks. **(5Qx 5M= 25M)**

Q.NO1. What is the boolean Model? Give the explanation to the following example. To Find "Brutus AND Caesar AND Calpurnia"?

	Anthony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello.	Macbeth
Anthony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
Mercy	1	0	1	1	1	1
Worser	1	0	1	1	1	0

(C.O.No.1) [Knowledge's level]

Q.NO2. Consider a document containing 100 words wherein the word Dog appears 3 times, find Term Frequency(TF). Now, assume we have 10 million documents and the word Dog appears in one thousand times of these, find Inverse Document Frequency(IDF) and TF-IDF.   (C.O.No.2) [Knowledge's level]

Q.NO3. What is SVM? Explain in detail with real-time applications.   (C.O.No.3) [Knowledge's level]

Q.NO4. Explain in detail search engine architecture with a neat sketch.   (C.O.No.4) [Knowledge's level]

Q.NO5. What is User-Based Collaborative Filtering? with the following example find the Rupam(User) rating for the Spider–Man Film.

User Name	Titanic	Avatar	Batman	Spider-Man
Nasur	5	2	5	4
Lokesh	2	5	-	3
Rupam	5	1	5	?

(C.O.No.5) [Knowledge's level]

Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries TEN marks. **(3Qx10M=30M)**

Q.NO6. D1: BITS Pilani Goa Campus.
        D2: Presidency University.
        D3: The IIT Delhi.
        Q: BITS Pilani.

In the above two documents and one query, find the relevant document by using the cosine similarity.
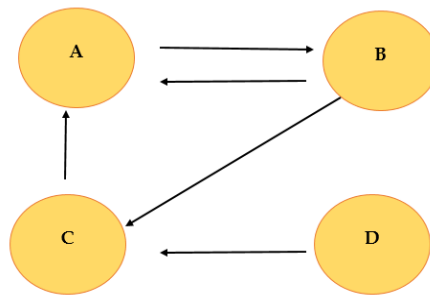
(C.O.No.2) [Comprehensive level]

Q.NO7. Explain in detail about Brute Force Method in the following example. Find the pattern found position.

P	R	E	S	I	D	E	N	C	Y	U	N	I	V	E	R	S	I	T	Y
U	N	I	V	E	R	S	I	T	Y										

(C.O.No.3) [Comprehensive level]

Q. NO8. Find the Page Rank of below graph. Damping Factor(D)=0.76   (C.O.No.4) [Comprehensive level]

Part C [Problem Solving Questions]

Answer all the Questions. Each question carries FIFTEEN marks.   (3Qx15M=45M)

Q.NO9. a) Explain in detail about the Components of IR. (8)
        b) Write a Short note on Characterizing the Web for Search. (7)   (C.O.No.1) [Application level]

Q.NO10. Document collection contains N = 20 documents, including:
- d1: "Presidency College"
- d2: "Presidency University campus"
- d3: "Presidency Campus"
- The query is: "Presidency University Campus"
- User has indicated R = 6 relevant documents for this query
- Query terms: t1 = "Presidency", t2 = "University", t3 = "Campus"
- Document frequencies of query terms in relevant documents and overall
- collection is given as follows:
- rt1 = 3, Nt1 = 13
- rt2 = 4, Nt2 = 15
- rt3 = 2, Nt3 = 12

To find the relevant document among the three documents, Use the Binary Independence Model(BIM) with relevance and without log.   (C.O.No.2) [Application level]

Q. No11. We have data from the questionnaires survey (to ask people's opinions) and objective testing with two attributes (acid durability and strength) to classify whether a special paper tissue is good or not.
- Here are four training samples.

Name	Acid Durability	Strength	Class
Type 1	7	7	Bad
Type 2	7	4	Bad
Type 3	3	4	Good
Type 4	1	4	Good

Now the factory produces a new paper tissue that passes laboratory tests with X1 = 3 and X2 = 7. Without another expensive survey, can we guess what the classification of this new tissue is using KNN Classifier?

(C.O.No.3) [Application level]