Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 08/10/2026 **Time:** 11:45 AM - 01:15 PM

Course Code: ISE2502 **Course Name:** Information Retrieval

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	46	44

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Define n-gram tokenization.	2	CO1	BT1
2	State four applications of Information Retrieval	2	CO1	BT1
3	State why stopwords should be removed.	2	CO1	BT2
4	Define Logarithmic term frequency.	2	CO2	BT1
5	Explain word tokenization and sentence tokenization.	2	CO2	BT2

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	Describe various Lemmatization Techniques.	10	CO1	BT2
7	Explain the types of stemmer provided by Python’s NLTK .	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Explain various structured formats in which unstructured text data can be transformed	10	CO1	BT2
9	Explain inverted index and how dictionary can be constructed in inverted index.	10	CO1	BT2

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
10	In vector space model, explain how cosine similarity is used to retrieve documents relevant to a query.	10	CO2	BT2
11	Consider the Document Term Matrix given below for three documents D1, D2 and D3 with weighting scheme -term frequency (TF- raw count). Covert this Document Term Matrix into Document Term Matrix with TF-IDF weighting scheme.	10	CO2	BT3

	Term1	Term2	Term3
D1	2	1	0
D2	1	0	2
D3	0	3	1

PART E: Answer Any 1 Following Questions. 10 M				
Q.N.	**Questions**	**M**	**CO**	**BT**
12	Explain B25 model with example.	10	CO2	BT2
13	Compute the TF-IDF of terms: data, mining, machine in documents D1, D2 and D3 D1: data mining data data D2: machine mining mining D3: machine data machine For computation of TF consider raw count term frequency.	10	CO2	BT3