UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2023

Course: Big Data Search   Semester: VI
Program: B.Tech CSE spl in Big Data   Time   : 03 hrs.
Course Code: CSBD 3008   Max. Marks: 100

Instructions: Kindly stick to the question and support every answer with proper explanation.

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	List the advantage of Lucene	4	CO2
Q 2	Define Stemming and Lemmatization process	4	CO1
Q 3	State the configuration of Kafka with Elastic Search	4	CO4
Q 4	List challenges of text search.	4	CO1
Q 5	State Split Brain Concept	4	CO3

SECTION B
(4Qx10M= 40 Marks)

Q 6	Write a SCALA program to generated inverted index of any ebook.	10	CO4
Q 7	Demonstrate the Schema less architecture for Full Text Search	10	CO3
Q 8	Illustrate the architecture of Lucene. Paraphrase the merit and demerits of the same	10	C02
Q 9	Design the Vector Space Model for the following datapoints
D1: Tropical Freshwater Aquarium Fish
D2: Tropical Fish, Aquarium Care, Tank Setup.
D3: Keeping Tropical Fish and Goldfish in Aquariums, and Fish Bowla
D4: The Tropical Tank Homepage – Tropical Fish and Aquariums.

OR | 10 | C02 |

Design the TDIM for the following datapoints

.

	Process Control Block	Process Scheduling	CPU utilization	Deaklock in Operating System	Disk Scheduling Algorithm	Critical Section
Process	1	1	0	0	0	1
Kernel	0	0	0	1	0	0
CPU	1	1	1	0	0	0
Scheduling	0	1	0	0	1	0
Deadlock	0	0	0	1	0	0

SECTION-C
(2Qx20M=40 Marks)

Q 10	Illustrate the various algorithms used to perform text search. Explain each algorithm with proper example.	**20**	**CO1**

Q 11	Write detail note to appraise the role of the following:
	   1. RestFull API
	   2. Compression posting List
	                OR
	Write detail note to appraise the role of the following:
	   1. Boolean Retrieval Model
	   2. Issue of Auxiliary and main Indexer
		10+10	**CO4**