Reg No.:_____ Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

B.Tech Degree S7 (R, S) Examination November 2024 (2019 Scheme)

Course Code: ADT 453

Course Name: INFORMATION EXTRACTION AND RETRIEVAL

Max.Marks:100 **Duration:3 Hours**

PARTA

Answer all questions, each carries 3 marks. Marks

1		Define information retrieval.	(3)
2		Write the formal characterization of IR models.	(3)
3		Compare boolean and vector models.	(3)
4		Construct a Boolean query that retrieves documents containing the words "machine learning" and "classification " but excludes any documents with the word "neural networks" present.	(3)
5		Differentiate phrase and proximity.	(3)
6		Summarise the term single word queries.	(3)
7		Define meta data.	(3)
8		Enumerate different text formats.	(3)
9		Define crawler.	(3)
10		What do you mean by web directories?	(3)

PARTB

Answer any one full question from each module, each carries 14 marks.

Module I

11	a)	Discuss the trends in IR.	(6)
	b)	Differentiate data retrieval and information retrieval.	(8)

OR

12	a)	Briefly explain taxonomy of IR models.	(6)
	b)	How can you explain the logical view of a document?	(8)

Module II

13	a)	Explain alternative algebraic IR model in detail.	(10)
	b)	Define precision and recall in retrieval evaluation.	(4)

OR

14	a)	Explain the classic IR models.	(14)

Module III

15 a) Briefly describe query protocol. (9)

b) What are the evaluation measures of TREC. (5)

OR

16 What are ISI and CACM collections? Explain in detail. (14)

Module IV

17 What does sequential searching mean? Give two instances to illustrate your point. (14)

OR

18 In detail explain different Mark-up languages with example. (14)

Module V

19 a) Explain in detail, the centralized and Distributed architecture. (10)

b) Write a note on the user interfaces in web search. (4)

OR

20 a) Identify various characteristics of web. (8)

b) Explain page ranking algorithm. (6)

****