Reg No.:_____   Name:_____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
THIRD SEMESTER MCA (Second Year Direct) DEGREE EXAMINATION, DECEMBER 2018
Course Code: RLMCA301
Course Name: WEB DATA MINING

Max. Marks: 60   Duration: 3 Hours

PART A
Answer all questions, each carries 3 marks.

1	What are the three types of web mining tasks?	(3)
2	Give the general IR system architecture with a neat diagram.	(3)
3	Differentiate supervised learning from unsupervised learning?	(3)
4	What is Stemming?	(3)
5	How will you classify the data using decision tree learning technique?	(3)
6	Explain with a neat diagram the meta-search engine architecture.	(3)
7	What is the importance of precision and recall in the mining process?	(3)
8	What are the steps involved in data preparation for web usage mining?	(3)

PART B
Answer six questions, one full question from each module and carries 6 marks

Module I

9   What are the steps involved in KDD process?   (6)

OR

10   Illustrate the basic GSP algorithm for mining sequential patterns with the   (6)
           transactional database (sorted by customer id and transaction time) given
           below giving emphasis on the join and prune steps. Minimum support is 2.

Customer id	Transaction Time	Transaction(items bought)
10	Nov 20, 2018	C
	Nov 28, 2018	I
20	Nov 20, 2018	A,B
	Nov 21, 2018	C
	Nov 27, 2018	A,D,F,G
30	Nov 29, 2018	C, E,G,H
40	Nov 29, 2018	C
	Dec 1, 2018	C,D,G,H
	Dec 2, 2018	I
50	Dec 2, 2018	I

Module II

11   Explain the support vector machine learning system which is widely used in   (6)
           web page classification.

OR

| 12 | a) | Explain hierarchical clustering method and its types. | (3) |
| | b) | What are strengths and weaknesses of K- means algorithm? | (3) |

Module III

| 13 | Explain the concepts of Boolean model and Vector Space model for Information Retrieval (IR) System. | (6) |

OR

| 14 | What are different forms of user queries for query operations module in an IR system? | (6) |

Module IV

| 15 | What are the various web page pre-processing tasks? | (6) |

OR

| 16 | Give any two bitwise scheme (coding and decoding techniques) for inverted index compression | (6) |

Module V

| 17 | What is web spamming? Briefly describe content spamming method. | (6) |

OR

| 18 | What is web crawling? Give a basic crawler algorithm. | (6) |

Module VI

| 19 | Describe any two types of pattern discovery and analysis techniques employed in the web usage mining domain. | (6) |

OR

| 20 | What are the key elements of web usage data pre-processing? | (6) |

****

.