AWH ENGINEERING COLLEGE, CALICUT

DEPARTMENT OF MCA

RLMCA301 WEB DATA MINING

S5 MCA SECOND SERIES DEC 2021

TOTAL MARKS: 30   **TOTAL HOURS:1.5**

PART A

ANSWER ALL QUESTIONS

EACH QUESTION CARRIES 3 MARKS

1. Give the general IR system with architecture?
2. Explain any three types of user queries with example.
3. What is Stemming?
4. What are the various pre-processing task for text?

PART B

ANSWER ANY ONE OUT OF 2 IN EACH SET OF QUESTIONS

EACH QUESTION CARRIES 6 MARKS

5. Explain the concepts of Vector Space model for IR System.
                                    OR
6 .What are the different form of user query operations module in IR System?

7 .Explain Statistical Language Model.
                                    OR
8.   a) TF-IDF Scheme of Vector Space model
     b) Explain Boolean Model

9.What are the various web page pre-processing tasks?
                                    OR
10. What is Duplicate detection? Describe n - grams duplicate detection technique.
.