PART-A (20 Marks)

There are 5 parts. Answer any 4.


a. Level Fuzzy Sets
b. Concept-based Retrieval

c. Orthogonality property vs. basis vectors

d. Projection Matrix and its relationship to $RSV_q$ vector
PART-B (55 Marks)

Answer all questions

[12] Q2. Assume that in a document collection there are a total of 5 documents. Let the indexing vocabulary consist of 3 index terms.

a) Assuming that index terms \( t_1 \), \( t_2 \) and \( t_3 \) appear, respectively, in 3, 4 and 2 documents, determine their \( idf \) weights. [3]

b) Let one of the documents be represented by the vector \( d = (0, 3, 5) \), where the elements represent the term frequencies \( (f_j) \). Let \( tf_j \) weights of \( d \), for \( 1 \leq j \leq 3 \), be defined as

\[
\begin{align*}
tf_j &= \begin{cases} 
0, & if \ f_j = 0 \\
0.3 + \frac{(1-0.3)f_j}{k_j} & \text{otherwise}
\end{cases}
\end{align*}
\]

Using results of a) and above information, determine representation of \( d \) given by \( tf \ idf \) weights. Need not normalize the \( tf \ idf \) weights. [3]
c) Assuming the $d$ vector with $tf$ $idf$ weights stands for components of $d$ along term vectors, $G_t = I$ and that a query is given by $q = (1, 0, 0)$, what is the $RSV$ of $d$ with respect to $q$? [1]

d) Assume that

$$G_t = \begin{bmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & -0.6 \\ 0.1 & -0.6 & 1 \end{bmatrix}$$

determine $RSV$ of document $d$ to the same query. [3]
e) Explain the reason for difference in the values of $R_{SV_q}(d)$ in parts (c) and (d). [2]

[20] Q3. Answer all parts of this question, using the following retrieval outputs:

**Method A**

$$(+ - + | + + - | + - - - -)$$

**Method B**

$$(+ - - - | + + + - | + -)$$

a) Find recall and fallout values for Method B, after retrieving 4, 8 and 10 documents. Draw the R/F graph. [6]
b) What is the expected recall and the expected fallout for Method B after retrieving 5 documents (read off from the graph)? [2]

c) What is the expected precision for Method B after retrieving 5 documents. (You need to use the mapping that gives P as a function of G, F, and R). [3]

d) Using the property that $R_{norm}$ is given by the area of R/F above the R/F curve, but below the line represented by the equation $F = 1$, compute $R_{norm}$ (use geometrical method to find the area of the polygon of the graph from part (a)). [3]
e) Compare the performance of Methods A & B using PRR and intuitive (preferred) interpolation at recall levels of 0.25, 0.5 and 0.75. [6]
Q4. [12] (a) For the following subset of rules from a rule-base, show the subtree and indicate the degree to which \textit{bombing} applies to the following 2 documents.

\begin{align*}
\text{device} \& \text{explosion} & \Rightarrow \text{bombing} (0.5, 0.7) \\
\text{grenade} \mid \text{bomb} & \Rightarrow \text{device} \\
\text{shell} & \Rightarrow \text{device} (0.4)
\end{align*}

(i) Show the rule base tree

(ii) $d_1 = (\text{shell}, \text{explosion})$

(iii) $d_2 = (\text{bomb}, \text{explosion})$

[3] (b) Determine all combinations (minimal) of text expressions that have non-zero RSVs.
Q5. a) Contrast Fuzzy Set Retrieval Model to the approach of RUBRIC, by giving 3 differences (there may be more, you need to give only 3). [3]
b) Assuming Fuzzy Set retrieval model where the definitions of operations $\cap$ and $\cup$ are as below, show that the commutative laws hold (i.e., $A \cup B = B \cup A$; $A \cap B = B \cap A$). [5]

$$\mu_{A \cap B} = \max \left( (\mu_A + \mu_B - 1), 0 \right)$$
$$\mu_{A \cup B} = \min \left( (\mu_A + \mu_B), 1 \right)$$