

## Question Bank for IAE-II Information Retrieval October-2024

1. Attempt ALL

2X5

- a) What is a term frequency and normalized term frequency? Write down their equations
- b) What is meant by query processing?
- c) What is multimodality in the multimedia retrieval system
- d) Explain Precision and Recall.
- e) What do you mean by Latency and Network Efficiency with respect to distributed IR
- f) Describe the difference between relevance feedback and query expansion in terms of user interaction
- g) Write the challenges of relevance feedback.
- h)

Document1	1	2	3	4	5	6	7	8
Relevance?	1	1	0	1	0	0	1	1

Calculate precision after each document retrieved

- i) Explain F-measure and E-measure of performance metrics
- j) Explain contingency table (confusion matrix) with example.
- k) Contrast recall and precision.
- l) Contrast automatic local analysis, automatic global analysis
- m) Explain inverted indexing with illustration.
- n) Explain  $tf - idf$  weighting.
- o) Write a few sentences for the vector space model for scoring

2. Attempt the following

(5 marks each)

- a) Give the formula for mean average precision (MAP). Calculate the mean average precision (MAP) following scenario.

Document1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Relevance?	1	1	0	1	0	0	1	1	0	1	1	1	0	0	1

Document2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Relevance?	1	0	0	1	1	0	1	0	1	0	0	0	1	1	0

- b) Write the challenges of relevance feedback.
- c) What is pattern matching, and why is it important in information retrieval?
- d) Discuss the concept of scoring in information retrieval. What is term weighting, and why is it important?
- e) What do you mean by Cumulative gain and discounted cumulative gain? Here are 10 ranked documents judged on 0-3 relevance scale: 3, 2, 3, 0, 0, 1, 2, 2, 3, 0. Calculate the cumulative and discounted cumulative gain.
- f) Here are 4 documents:  $d_1, d_2, d_3, d_4$  with the following parameters

i	Ground Truth		Ranking Function <sub>1</sub>		Ranking Function <sub>2</sub>	
	Document Order	$r_i$	Document Order	$r_i$	Document Order	$r_i$
1	d4	2	d3	2	d3	2
2	d3	2	d4	2	d2	1
3	d2	1	d2	1	d4	2
4	d1	0	d1	0	d1	0

Calculate the normalized discounted cumulative gain (NDCG)

- g) How is relevance assessed and justified in the context of information retrieval? Explain the factors that influence relevance assessment.
- h) Write short notes on the following
- 1) Multimedia Information Retrieval
  - 2) Distributed Information Retrieval