

BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI  
(MID SEMESTER EXAMINATION SP/2025)

CLASS: BTECH  
BRANCH: AIML

SEMESTER : VI  
SESSION : SP/2025

SUBJECT: AI317 INFORMATION RETRIEVAL

TIME: 02 Hours

FULL MARKS: 25

**INSTRUCTIONS:**

1. The question paper contains 5 questions each of 5 marks and total 25 marks.
  2. Attempt all questions.
  3. The missing data, if any, may be assumed suitably.
  4. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates.
- 

		CO	BL
Q.1(a)	Write the requirements of a good information retrieval system?	[2]	1 K1
Q.1(b)	What are the steps involved in web crawling, and why is priority queue-based crawling sometimes preferred over FIFO-based crawling?	[3]	1 K2
Q.2(a)	How does the document unit size of a collection affect the Precision and Recall measures of an information retrieval system?	[2]	1 K2
Q.2(b)	Write the True or False for the given statements with justifications. Statement 1: In a Boolean retrieval system, stemming never lowers precision. Statement 2: Stemming should be invoked at indexing time but not while processing a query.	[3]	1 K2
Q.3(a)	What is the size in MB of an incidence matrix created for a corpus containing 100000 documents and having a vocabulary of 40000 tokens with an average length of 8 characters per token. You can assume 1 character requires 1 byte while a truth value is stored as an integer requiring 2 bytes per integer.	[2]	1 K3
Q.3(b)	We have a two-word query. For one term the postings list consists of the following 17 entries (which have skip pointer of floor( $\sqrt{\text{length}}$ ), where length is equal to the total number of entries), [3, 5, 9, 15, 24, 39, 60, 68, 75, 81, 84, 89, 92, 96, 97, 100, 115] And for the second term, it is the eight entries (which have no skip pointer) [3, 5, 89, 95, 97, 99, 100, 101] (a) How often is a skip pointer followed (i.e., p1 is advanced to skip(p1))? Note: p1 is the pointer that points to the posting list of the first term. (b) Work out how many comparisons would be done to intersect the two postings list with the following two strategies (1) Using standard postings list (no skip pointer in any of the lists). (2) Using a postings list stored with skip pointers	[3]	2 K3
Q.4(a)	What are the different approaches to answering phrase queries and discuss their disadvantages.	[2]	2 K2
Q.4(b)	Write an algorithm to find the Levenshtein's (Edit) distance between two strings. Use your algorithm to find the proximity between the terms "alice" and "paris".	[3]	2 K3
Q.5(a)	Why Single-pass in-memory Indexing (SPIMI) is preferred over Blocked sort-based Indexing (BSBI)?	[2]	2 K2
Q.5(b)	Explain the working of the Logarithmic Merge dynamic indexing system. Prove that such an indexing system requires $\Theta(T \log(T/n))$ time to construct the index where T is the number of postings and n the number of documents.	[3]	2 K4