Nirma University

Institute of Technology
Class test, February 2023
*B. Tech. in Computer Science and Engineering, Semester-VI
2CSDE53 INFORMATION RETRIEVAL SYSTEMS

Roll / Exam No.		Supervisor's initial with date	

Time: 1 Hour 15 Minutes   Max. Marks : 35

Instructions: 1. Attempt all questions.
                   2. Figures to right indicate full marks.
                   3. Assume suitable data wherever necessary.

Q-1 Are the following statements true or false? Justify your answer.   **[9]**
CLO2

   a. Stemming increases the size of the vocabulary.
   b. In a Boolean Retrieval system stemming never lowers precision.
   c. Skip pointers are useful in the context of indexing.
   d.

Q-2 Shown below is a portion of a positional index in the format: term: <doc1:   **[6]**
CLO3 <position1, position2, . . . >; doc2: <position1, position2, . . .>; etc.

angels: 2: <36,174,252,651>; 4: <12,22,102,432>; 7: <17>;
fools: 2: <1,17,74,222>; 4: <8,78,108,458>; 7: <3,13,23,193>;
fear: 2: <87,704,722,901>; 4: <13,43,113,433>; 7: <18,328,528>;
in: 2: <3,37,76,444,851>; 4: <10,20,110,470,500>; 7: <5,15,25,195>;
rush: 2: <2,66,194,321,702>; 4: <9,69,149,429,569>; 7: <4,14,404>;
to: 2: <47,86,234,999>; 4: <14,24,774,944>; 7: <199,319,599,709>;
tread: 2: <57,94,333>; 4: <15,35,155>; 7: <20,320>;
where: 2: <67,124,393,1001>; 4: <11,41,101,421,431>; 7: <16,36,736>;

In following queries, each expression within quotes is a phrase query.
Which of the above document(s) match the query? Write proper explanation
of your answer.
a. "fools rush in"
b. "fools rush in" AND "angels fear to tread"

Q-3 Which of the following representation will result into sparse matrix? Justify   **[6]**
CLO2 your answer. Discuss mechanism to deal with sparsity issue.
(i) Boolean representation (ii) Weighted(normalized) TF representation (iii)
TF-IDF representation

Q-4 "The term appearing in a large number of documents in the collection, is   **[6]**
CLO3 probably not significant or discriminative." Give your remarks on this
statement with appropriate justification.

Q-5
CLO1
What are the issues with bioword indexing mechanism when used to [4] handle phrase query ? Justify your answer.

Q-6
CLO1
Recommend a query processing order for [4]

(tangerine OR trees) AND (marmalade OR skies) AND (kaleidoscope OR eyes) given the following postings list sizes:

Term	Postings size
eyes	213312
kaleidoscope	87009
marmalade	107913
skies	271658
tangerine	46653
trees	316812