Indian Institute of Information Technology Sri City, Chittoor

(An Institute of National Importance under Act of Parliament)

Mid Semester Examination - Monsoon 2021

[21 September 2021 - 08:30 AM - 10:30 AM]

Course: **Information Retrieval** Total Marks: **20 + 20 (10 x 1 + 5 x 2 = 20 - MCQs) Marks**

Category: **Set-IR01** Time: **60 + 30 (MCQs) Minutes**

Instructions:

a) The actual examination is scheduled for 60 minutes and another 30 minutes would be given for organizing the answer scripts, scanning and submitting over online.

b) Precisely answer the questions with relevant details. Avoid writing unnecessary explanations.

c) The file to be uploaded should be named as follows:
ABCD-**YYYY**-midsem-IR2021-**AYCOY**.pdf Where **ABCD** is the last 4 digits of your roll number; **YYYY** - year of admission (probably either 2018 or 2019); **AYCOY** - any 5 characters in CAPITAL letters (this may act as a secret key - Do not share with others). Please use hyphen (-) and NOT the "underscore" (_).

d) Submission portal: **http://smartmiss.iiits.ac.in/upload** and choose "**Information Retrieval**" as the course name.

Descriptive Questions: Answer ALL Questions [Please state your assumptions, if any, explicitly in your answer sheet]

1. [5 Marks] Create an inverted index with doc. frequency for the set of documents given below:
   d1: The quick brown fox jumps over a lazy dog
   d2: New film app is released with brown background
   d3: Lazy guys find easy solutions
   d4: Finding new solutions is a hard problem
   d5: Bell bottom is a newly released film

2. [5 Marks] Write an algorithm for merging two postings lists with the worst case asymptotic time and space complexities.

3. [5 Marks] Consider the following Boolean Index

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	1	1	0	0	0	1
Brutus	1	1	0	1	0	0
Caesar	1	1	0	1	1	1
Calpurnia	0	1	0	0	0	0
Cleopatra	1	0	0	0	0	0
Mercy	1	0	1	1	1	1
Worser	1	0	1	1	1	0

Now for each of the following Boolean queries, show their outcome:
   a) Antony AND Caesar AND (NOT Cleopatra)
   b) (Brutus OR Caesar) AND NOT (Antony OR Cleopatra)
   c) (Caesar NOT worser) AND mercy

4. [5 Marks] Consider the wild card query term: co*ion
   a. How do you expand this query in terms of relevant Boolean Operators?
   b. What is the optimal retrieval approach for this type of queries?
   c. Justify your approach in terms of the computational cost.