



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**

Category: **Set-IR02**

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

Time: **60 + 30 (MCQs) Minutes**

Instructions:

- 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.
- Precisely answer the questions with relevant details. Avoid writing unnecessary explanations.
- 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" (_).
- 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]

- [5 Marks] Create a Positional 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
- [5 Marks] Explain biword indexing with an example. Describe pros & cons of this approach and the worst case asymptotic time and space complexities.
- [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:

- Antony AND (NOT Caesar) AND Cleopatra
 - (Brutus OR NOT Caesar) AND (Antony OR Mercy)
 - (Caesar NOT worser) AND (NOT mercy)
- [5 Marks] Consider the wild card query term: op*on
 - How do you expand this query in terms of relevant Boolean Operators?
 - What is the optimal retrieval approach for this type of queries?
 - Justify your approach in terms of the computational cost.