Mid-term Examinations **Date: 26 Feb 2023**
CSE508: Information Retrieval **Time: 10-11 am**

Name: **Roll Number:**

Instructions

- It is a close book examination
- All questions are mandatory
- Calculator is not allowed
- There is no negative marking
- MCQ questions (2 marks each) may have more than one correct answers
- Short and long descriptive answers are of 3 and 5 marks, respectively
- There are total 8 questions (3 MCQs, 3 short answers, and 2 long answers)
- Maximum score is 25 marks [3 x 2 + 3 x 3 + 2 x 5]

All the best!

Section A: MCQs [2 marks each]

Q1) Mark all True sentences.

A. Traditional Inverted Indexes are more powerful than Positional Inverted Indexes.
B. Positional Inverted Indexes are more powerful than Traditional Inverted Indexes.
C. Traditional Inverted Indexes require more computation than Positional Inverted I.
D. Positional Inverted Indexes require more computation than Traditional Inverted I.

Q2) Which of the following is true for the query, mercy AND Caesar AND

	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

A. 101100
B. 100111
C. 110011
D. 100011

Q3) Which of the following cannot be input for an information retrieval system.

A. Text
B. Audio
C. Image
D. Gesture
E. None of the above

Section B: Short Answers [3 marks each]

Q4) Describe Information Search, Information Retrieval, and Data Retrieval with an example.

Q5) List the issues/challenges in the information retrieval system.

Q6) What are the benefits of creating an inverted index for indexing documents in an information retrieval system? Explain with an example.

Section C: Detailed Answers [5 marks each]

Q7) What do you mean by web co-pilot? Why do we need the same? List down the benefits and limitations of web co-pilot with a real-world example.

Q8) Google claims to have a better information retrieval system than Microsoft. For a given query, returns the following 5 documents: d1, d2, d3, d4, and d5, with their relevance scores 0, 2, 1, 4, 2, respectively (a higher score indicates more relevant to the given query). New Bing + OpenAI based system returns the following results for the same query: d4, d1, d2, d3, d5. Old Bing returns the following documents: d1, d2, d4, d3, and d5 for the same query. Compare all IR systems based on DCG and NDCG metrics. For your reference, DCG is defined as follows:

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2 i}$$

Additionally, log 3 and log 5 (all on base 2) are 1.59 and 2.32, respectively.