



### Final Exam

Department: Information Systems  
Course Name: Information Retrieval/Data Storage and Retrieval  
Course Code: (IS322)/ (IS313)  
Instructor(s): Dr. Ehab Ezzat Hassanein

Semester: Spring 2021-2022  
Date: 29-06-2022

Duration: 2 Hours  
Total Marks: 60 Marks

#### تعليمات هامة

- حيازة التليفون المحمول مفتوحاً داخل لجنة الامتحان يعتبر حالة غش تستوجب العقاب وإذا كان ضروري الدخول بالمحمول فيوضع مغلق في الحقيبة.
- لا يسمح بدخول سماعة الأذن أو الطوئوت.
- لا يسمح بدخول أي كتب أو ملازم أو أوراق داخل اللجنة والمخالفة تعتبر حالة غش.

1-

a. Describe and draw the Classic Search Model and show and explain the areas of misconception and misinformation. Explain why do we need query refinement [5 marks]

b. List and Describe the steps needed before the Inverted Index construction [5 marks]

2.

a. Recommend a query processing order for ( River OR Nile ) AND ( France OR Paris ) AND ( Egypt OR Cairo ) given the following postings list sizes: [6 marks]

Term	Postings Size
Egypt	100000
Cairo	40000
River	100000
Nile	4000
France	20000
Paris	40

b. Compare web crawler with web scraper. [4 marks]

3-

a. Compute the cosine similarity between the following documents that makes up the whole collection  $N=3$ , given the term row frequency in each Document [5 marks]

Term	doc1	doc2	doc3
Information	10000	1000	0
Systems	1000	10	10
FCI	0	0	100
Cairo	100	100	1000

b. Explain why it is better to use biwords than positional Intersect when searching for a phrase like "Mohamed Salah". Suggest a scheme that is suitable for handling such phrase among other phrases. [5 marks]

4.

a. Explain the following Information retrieval equation [6 marks]

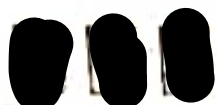
$$\text{Score}(q,d) = \sum_{t \in q \cap d} \text{tf} \cdot \text{idf}_{t,d}$$

- ✓ Explain its components ( $\text{Score}(q,d)$ ,  $t$ ,  $\text{tf}$ ,  $\text{idf}$ )
- Why do we multiply  $\text{tf}$  and  $\text{idf}$
- Why do we sum the  $\text{tf} \cdot \text{idf}_{t,d}$

b. What is a spider trap (or crawler trap)? [4 marks]

5.

Given

 = the relevant documents

$\frac{1}{3}, \frac{2}{5}, \frac{3}{6}$

and the following two rankings



- Compute the recall for each item of each ranking [2 marks]
- Compute the Precision for each item of each ranking [2 marks]
- Compute the Average Precision for each ranking. [6 marks]

6. Using Java, implement a Jaccard similarity function that measures the similarity between sentences/documents and provide a similarity score based on how similar the sentences. [10 marks]

$$\frac{1+1+1}{3} = 1 \quad / \quad \frac{\frac{1}{4} + \frac{2}{5} + \frac{3}{6}}{3} = 0.38$$