Ministry of Higher Education and Scientific Research
University of Basrah
College of Education for Pure Sciences
Department of Computer Science

Subject: Information Retrieval
Stage: M. Sc.
Exam. Time: 3 Hours
Exam Date: 1 1 / 5 /2023

Final Examination of First Attempt (الدور الاول) In Studying Year 2022-2023

for

Note: Answer All Questions (calculator is permission to use)

Q1: (18 marks: a= 10, b= 2+6) :

a) Determine whether the following statement is **True or False with Correct the false statement** (if any)

1- In a Boolean retrieval system, stemming never lowers recall.

2- Stemming increases, the size of vocabulary.

3- Stemming should be invoked at indexing time but not while processing a query.

4- Boolean queries are useful for information retrieval tasks that require semantic analysis and understanding the meaning of documents.

5- Vector space model can be extended to handle document relevance feedback by modifying the query vector based on the user's feedback on the initial set of retrieved documents.

b- Answer the following questions:

1- In Porter's algorithm, for example : "replacement" → "replac" but "cement"→ "cement" . Why? Explain your answer

2- If the following documents:

D1: " Basrah university includes twenty two colleges containing eighty three scientific departments"

D2: "scientific departments in Basrah university contain a lot of students"

D3: " Basrah university and colleges is located in center of Basrah"

Query: "Basrah university colleges"

Compute **Vector Space Model** to rank the retrieval of the query.

Q2 (8 marks)

a) if the following three documents:

D1: "Ahmed played the football with the sword"

D2: “Ali and Ahmed ripped football”

D3:” Ahmed took the sword”

Query:” Ahmed and football and sword”

Rank the documents according to **Unigram LM for IR** (ignore stop words)

Q3: (13 marks: a= 8, b=5)

a) Given the query “Iraqi team students” and the following term-frequencies for the two documents *doc1* and *doc2*

	Iraqi	team	attend	English	students	course
doc1	5	4	3	3	0	5
doc2	2	2	0	2	1	3

Calculate **the unsmoothed query-likelihood** for both documents.

(i) Describe two ways in which **smoothing affects** the retrieval of these documents

(ii) Is smoothing more **important** for long or short queries? Justify your answer.

b) Given the query “happy person smiles”, show how **a unigram language modelling** approach would rank the documents above. Choose a suitable form of smoothing and include all your works. State any other assumption made.

Q4: (14 marks: a=4, b=10)

a) Edit distance can be used for spelling correction in search queries.

(i) Define **Edit Distance**

(ii) As an example of how to calculate edit distance efficiently, show how **dynamic programming** can be used to calculate the edit distance between ***able*** and ***belt.***

Q4: b) Choose the correct answer (10 **marks**):

1- Steps of indexing are performed in following order:

a- Stop-ward elimination, tokenization, stemming b- tokenization, stemming, stop-ward elimination c- tokenization, stop-ward elimination, stemming d- stemming, tokenization, stop-ward elimination

2- In information retrieval most common words such as articles, prepositions etc. are removed from tokens by using

a- Stemming b- stop-ward elimination c- indexing d- ranking

3- Data stored in a table is a form of ---------

a- Unstructured data b-structured data c- semi-structured data d- none of the above

4- Following are the example of classical model of IR

a- The Boolean model b- the vector model c- set-based model d- all options are correct

5- Given the document containing the sentence “I **left my left bag at my home**” the number of tokens in the sentence is

a- 8 b- 4 c-5 d-1

Q 5: (17 marks: a=8, b= 5, c=4)

a) Query : “ president lincolin” . Compute **Dirichlet Smoothing** and why it is a good choice for many IR tasks?

tf	**15**
cf	**160,000**
tf	**25**
cf	**2400**
|d|	**1800**
∑	**10**
µ	**2000**

b) Compute **Page Rank** in matrix form. write equations if necessary

c) Write simple arithmetic for **HITS algorithm**

With Good Luck

Instructor *Head of Dept.*

Asst/ prof Dr. Khawla Hussein Ali ***Prof. Dr. Hamid Ali Al-Asady***