


Course No:  
 Course Title: Topics in CIS  
 Date: 15 / 01 / 2012  
 No. of Questions: 3 Questions  
 Time: 2 hours  
 Using Calculator (YES)

University of Palestine  
  
 Final Exam  
 1<sup>st</sup> semester 2011/2012  
 Total Grade: 50

Instructor Name: Mrs. Eman Alajrami  
 Student No.: \_\_\_\_\_  
 Student Name: \_\_\_\_\_  
 College Name: \_\_\_\_\_  
 Dep. / Specialist: \_\_\_\_\_  
 Using Dictionary (No)

| First Question | No. of Branches (1) | 10 Marks |
|----------------|---------------------|----------|
|----------------|---------------------|----------|

(A)

(5)

Put ( √ ) or (X) for each of the following statements:

- 1) The edit distance between two strings is the minimum number of characters insertions, deletions, and replacements needed to make them equal. ( )
- 2) Exact matching is accomplished by assigning binary weights to index term in queries and documents ( )
- 3) The most common measures of IR system performance are precision and recall ( )
- 4) Proximity queries can be ranked in the same way if the ranking technique does not depend on physical proximity. ( )
- 5) A pattern is a set of syntactic features that must occur in a text segment. ( )
- 6) Term re-weighting is to add new terms to query from relevant documents. ( )
- 7) Document preprocessing is a process of controlling the size of the vocabulary ( )
- 8) Specificity of index is the no. of different topics indexed. ( )
- 9) Lexical Analysis Convert an input stream of characters into stream words or tokens. ( )
- 10) Building the index in main memory is feasible because swapping would be unbearable. ( )

| Second Question | No. of Branches (1) | 10 Marks |
|-----------------|---------------------|----------|
|-----------------|---------------------|----------|

(A) Define the following briefly:

(5)

1. Proximity

---



---

2. Relevance Feedback

---



---

3. Block Addressing

---



---

4. posting

---



---


5. Inverted file

---



---

Course No:  
Course Title: Topics in CIS  
Date: 15 / 01 / 2012  
No. of Questions: 3 Questions  
Time: 2 hours  
Using Calculator (YES)

University of Palestine  
  
Final Exam  
1<sup>st</sup> semester 2011/2012  
Total Grade: 50

Instructor Name: Mrs. Eman Alajrami  
Student No.: \_\_\_\_\_  
Student Name: \_\_\_\_\_  
College Name: \_\_\_\_\_  
Dep. / Specialist: \_\_\_\_\_  
Using Dictionary (No)

Third Question No. of Branches (6) 30 Marks

Answer the following questions:

1. a) What is the *Levenshtien distance* for the query words ‘flo aer’ and ‘flower’?  
\_\_\_\_\_  
\_\_\_\_\_

b) If stemming is applied in an information retrieval system what effect is it likely to have on the precision and the recall of the system? Explain your answer and include appropriate examples of stemming.  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_


2. a) Consider a query vector  $v_q$ , two documents returned by an information retrieval system that a user considers relevant with vectors  $v_1$  and  $v_2$ , and three documents returned considered irrelevant with vectors  $v_3$ ,  $v_4$ , and  $v_5$ . Compute a modified query using the Standard Rochio equation with  $\alpha = \beta = \gamma = 1$ .

$v_q = (2, 1, 0, 0)$   
 $v_1 = (0, 4, 0, 2)$        $v_2 = (0, 3, 0, 1)$   
 $v_3 = (1, 0, 2, 0)$        $v_4 = (0, 1, 4, 0)$        $v_5 = (1, 1, 0, 0)$

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

b) Why should the evaluation metrics of precision and recall be applied to the residual collection after query modification with the Standard\_Rochio equation?  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**Course No:**  
**Course Title: Topics in CIS**  
**Date: 15 / 01 / 2012**  
**No. of Questions: 3 Questions**  
**Time: 2 hours**  
**Using Calculator (YES)**

**University of Palestine**  
  
**Final Exam**  
**1<sup>st</sup> semester 2011/2012**  
**Total Grade: 50**

**Instructor Name: Mrs. Eman Alajrami**  
**Student No.:** \_\_\_\_\_  
**Student Name:** \_\_\_\_\_  
**College Name:** \_\_\_\_\_  
**Dep. / Specialist:** \_\_\_\_\_  
**Using Dictionary (No)**

3. Table 1 shows a frequency table for three documents (D1, D2 and D3) in an information retrieval system. A user, who is interested in sport, in particular football and rugby, enters the query 'sport, football, rugby' into this system.

|           | <b>D1</b> | <b>D2</b> | <b>D3</b> |
|-----------|-----------|-----------|-----------|
| Sport     | 4         | 1         | 3         |
| Politics  | 1         | 5         | 0         |
| Leisure   | 0         | 3         | 0         |
| Football  | 3         | 1         | 0         |
| Rugby     | 0         | 1         | 5         |
| Economics | 1         | 4         | 1         |

*Table 1: a frequency table for three documents D1, D2 and D3*

What results would the system return if it used the Boolean model of information retrieval with query words joined with AND unless otherwise specified? You should explain your answer.

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

4. For the corpus { ABLE, BEATABLE, FIXABLE, READ,READS,READABLE, READING, RED, ROPE, RIPE } how many successor varieties does the string "READABLE" have? What are they?

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---



