22222222222222222222222222222222222222222222222222

National University of Computer & Emerging Sciences
FAST-Karachi Campus
Information Retrieval & Text Mining (CS567)
Quiz#2 (Sol)

Dated: October 11, 2017 Marks: 30
Time: 30 min.
Std-ID: _____

Question NO. 1

Consider an information need for which there are 4 relevant documents in the collection. Contrast two systems run on this collection. Their top 8 results are judged for relevance as follows (the leftmost item is the top ranked search result): [5]

System 1: R N R N N N R R
System 2: N R N N R R R N

a. What is the MAP of each system?

MAP for System 1: 1/4 * (1 + 2/3 +3/7+4/8) = 0.6488
MAP for System 2: 1/4 * (1/2 +2/5+3/6+4/7) = 0.4928

b. Which system is better? explain your answer.

System 1 is better as it has higher MAP values. Which is a clear indication of presenting relevant documents at higher in the rank list results. Thus satisfying the users information need better and quickly.

Question NO. 2

Differentiate between following pairs of terms**.** [5]

Query expansion	**Query relaxation**
Query expansion (QE) is a process in Information Retrieval which consists of selecting and adding terms to the user's query with the goal of minimizing query-document mismatch and thereby improving retrieval performance.	Query relaxation is the opposite of query expansion. Instead of adding tokens to the query, we remove them. Ignoring tokens makes the query less restrictive and thus increases recall. An effective query relaxation strategy removes only tokens that aren't necessary to communicate the searcher's intent.

Question NO. 3

a. An information retrieval system returned 150 documents from a collection of 400 under a given query. There were 70% non-relevant documents in the retrieved result. The collection contains 35% relevant documents. Find the Precision and Recall for this system. [5]

	Rel	Non-Rel
Ret	45	105
Non-Ret	95	155

Total Returned Docs. =150;
70% non-relevant returned are 150X0.7= 105
The collection contains 35% relevant that is 400X0.35 =140 relevant documents, the query returned 45 relevant (150-105); 155 Non-relevant were not retrieved.
Precision = 45/150 = 0.3
Recall= 45/140 = 0.321

b. Considering the results obtained in part(a), what you can say about the system? [2.5]

The system is supporting both precision and recall equally.

c. What is a break-even point between precision and recall? [2.5]

A break-even point is a point in precision recall curve where precision and recall are equal.

Question NO. 4

What do we mean by relevance feedback? What is the basic procedure for it? What are the different kinds of relevance feedback, known to you? [10]

The idea of relevance feedback (RF) is to involve the user in the retrieval process so as to improve the final result set. The basic procedure is:

1. The user issues a (short, simple) query.
2. The system returns an initial set of retrieval results.
3. The user marks some returned documents as relevant or non-relevant.
4. The system computes a better representation of the information need based on the user feedback.
5. The system displays a revised set of retrieval results.

These are some of the kinds of relevance feedback:

1. Pseudo relevance feedback - (Blind relevance feedback).
2. Indirect relevance feedback -