

National University of Computer & Emerging Sciences  
FAST-Karachi Campus  
Information Retrieval (CS317)  
Quiz#2

Marks: 20

Std-ID: \_\_\_\_\_ Sol \_\_\_\_\_

**What are some of the advantages of Vector Space Model over Boolean Model for IR? [5]**

- Vector Space Model is simple and yet based on rigor mathematical formalism of linear algebra.
- Term weights not binary.
- It produces continuous degree of similarity between query and documents.
- It produces rank retrieval based on possible relevance scores.
- It allows partial matching too.

**How Rocchio's algorithm works for relevance feedback? Explain.**

## Rocchio's Algorithm

Step1: It computes centroids for both relevant and non-relevant documents from the examples (training dataset).  $\mu_{(\text{relevance})}$  and  $\mu_{(\text{non-relevance})}$

Step2: It computes the distance of  $x$  from  $\mu_{(\text{relevance})}$  and  $\mu_{(\text{non-relevance})}$ . Then  $x$  can be assigned the label as relevance or non-relevance based on the minimum distance from the either one.

### Question No.3

Comments on the following statement as TRUE or FALSE with justification (1-2 line explanations). 5 X 2 marks each.

1. Inverse Document frequency (Idf) of a term is always finite.

TRUE. Inverse Document Frequency (idf) is computed as  $\log(N/df)$ , where  $N$  is the total number of document in the collection. The count of document that contains term  $t$  is always between 0 and  $N$ , hence idf will always be a finite value.

2. Dynamic snippet summary is computed based on query term.

TRUE. Dynamic snippet summary is computed by using query term through key-word in context index.

3. An ideal Discount Cumulative Gain is the maximum value for a rank list.

TRUE. The IDCG is an absolute ranking list as per user relevance through gold standard hence the maximum value possible for any ideal rank list.

4. Direct relevance feedback is less reliable.

FALSE. Direct relevance feedback is taken by explicitly asking users about the results. The users gave feedback by inspecting document/summary of document. Hence it is quite reliable.

5. Mean Average Precision(MAP) generally computed as an average on a bunch of queries.

Mean Average Precision is computed as (sum of Average Precision for all queries) over total queries in the test set. Hence it is always for a fixed number of queries in the test-set.