

National University of Computer & Emerging Sciences FAST-Karachi Campus CS4051- Information Retrieval Ouiz#1

Dated: February 13, 2023		Marks: 20
Time: 20 n	nin.	
Std-ID:	Sol	

Question No. 1

Outline some of the drawback of Boolean Model of Information Retrieval. [5]

There are several drawbacks of Boolean Model for IR.

From Users prospective: Users need training on query formulations, they need to understand Boolean queries. They need to have some clear idea about what features are there in the relevant documents.

From System's prospective: The IR systems based on Boolean model considers all terms with same importance and independent of each other. It is based on exact matching and result-set is flat (that is all documents are equally ranked).

Question No.2

In an IR System there were 60 relevant documents for a given query "q". The system returned 108 documents in response to the same query. If 50% documents in the result-set are relevant, compute the Precision and Recall of the system? [5]

```
we know,

precision = (relevant-retrieved) / (total-retrieved)

=> precision = (relevant-retrieved) / (result-set) ---------- eq(A)

recall = (relevant-retrieved) / (total-relevant) ---------- eq(B)

we need to find total relevant documents in result-set,

(result-set) = 108 documents

(relevant documents in result-set) = 108 X .50 = 54= (relevant-retrieved)

From eq(A) .... precision = 54/108 = 0.5

From eq(B) .... recall = 54/60 = 0.9
```

Ouestion No.3

Answer these questions in 2-3 lines of text. [2X5]

a. How position index is related to document size?

A position index is directly related to document size. Large documents have more positional index entries.

b. In permuterm index, what key will be searched if we need to execute the in-place wildcard query **tri*ph**?

The permuterm index ph\$tri* will be searched. The idea is to get all the documents that contains string with prefix "tri" and suffix "ph"

c. Consider the wildcard query **tri*ph** once again, what Boolean query on a bi-gram index would be generated for this query? Can you think of a term that matches the query?

The equivalent bi-gram query will be (tr AND ri AND ph) –the trivial term would be triumph.

d. Why post-processing is required sometime in information retrieval? Why it is not recommended?

In information retrieval sometime the retrieval system uses relax criteria for retrieval and hence get some false positive. The post processing step is essential for removing these false positive results from the final result-set. The complexity is directly related to number of candidates in the results-set, larger resut-set need more processing and hence it is not recommended.

e. Define context sensitive spelling correction. Give an example.

Context-sensitive spelling correction is the task of correcting spelling errors which result in valid words. For example, in the sentence "I want a peace of cake", peace is a valid word in isolation but an error in this context (should be piece). This type of spelling error corrected by inspecting the entire context of the text.