

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

### Question 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.