<1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111>

CS317- Information Retrieval

Quiz # 1

Dated: March 04, 2021 **Student ID: ____Sol____**

Marks: 20 **Time: 15 min.**

Q1: What do we mean by Extended Boolean Retrieval Model? [5]

The idea of Extended Boolean Retrieval model is to use term weighting and partial matching for Boolean retrieval. Hence it uses term frequency in each document to represents the weight of document with respect to query. Hence document with most frequent query terms inthem rank higher of the result-set.

Q2: Indicate True/False with a brief justification. [5]

a. Stemming increases recall in a Boolean Retrieval Systems.
False. Stemming only add documents in the retrieval set against a query.

b. The modern trends in information retrieval now do not consider any word as stop word?
True. Yes, stop-word only help a smaller percentage of storage but cost a lot of opportunity in getting good retrieval efficacy. Hence the trends in favor of not using any stop word list.

c. Positional queries can be easily supported by a bi-word index.
False. No bi-word index only covers a distance of 0 between words, thus cannot be used to support positional queries.

Q3: There are 16 relevant documents in a collection for a given query "q". The precision of the query is 0.40 and recall of the query is 0.25. How many documents are in the results-set (number of documents returned by the system against the query)?

we know,

precision = (relevant-retrieved) / (total-retrieved)
=> 0.4 = (relevant-retrieved) / (result-set)
=> (result-set) = (relevant-retrieved) /0.4 ---------- eq(A) similarly,
recall = (relevant-retrieved)/ (total-relevant)
=> 0.25 = (relevant-retrieved)/ 16
=> relevant-retrieved= 0.25 * 16 = 4 hence
eq(A) => result-set = 4 /0. 4 = 10

Result-set has 10 documents.

<22222222222222222222222222222222222222222222222222222222222222222222222222>

CS317- Information Retrieval

Quiz # 1

Dated: March 04, 2021 **Student ID: ____Sol____**

Marks: 20 **Time: 15 min.**

Q1: What are the limitations of Boolean Retrieval Model? [5]

- Exact matching is used hence no support for partial matching.
- All the results from the retrieval are at the same level, no ranking of the results.
- All features are equally weighted – that is all terms have same importance.
- Users has to formulate a Boolean query to get its required documents.

Q2: Indicate True/False with a brief justification. [5]

a. Stemming is fast and increase precision.

False. The statement I partially true, stemming is fast but it does not increase precision. It only increases retrieved documents.

b. Skip pointers are not useful for queries of the form x OR y.

In queries of the form (x OR y), it is essential to visit every entry of posting lists of two terms x and y. Hence skip pointers are not any helpful in speeding up the process.

c. Proximity queries can be easily supported by a positional index?

Proximity queries are of the form x y /k, we need to find documents contains both x and y, k words apart. Positional index is very helpful in determining the k for these two terms.

Q3: Consider the posting lists for two terms T1 with skip pointer and T2 is simple.

T1	
T2	3 5 89 95 97 99 100 101

Briefly justify your answers:

a. Using standard postings lists. Finds the number of comparisons performed.
There are a total of 19 comparisons with standard posting list.

b. How many postings comparisons will be made by this Intersect(skip) algorithm while intersecting the two lists?
The skip pointer only follows one time. There is no change in total number of comparisons. 19 comparisons are done.