

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(VSM) for IR? [5]

- ### Question No.2

Differentiate between following pairs of terms.

Direct Relevance Feedback <ul style="list-style-type: none"> - Explicit (Direct) feedback requires user to marks document relevant. It is slow process and does not scale to large systems. - It is more reliable and generally save from query drift. 	Indirect Relevance Feedback <ul style="list-style-type: none"> - Implicit (indirect) feedback does not bother user for explicit actions. It is fast and can be possible for large IR system. - It is less reliable and possibly introduce a problem of query drift.
Evaluation of Ranked Retrieval <ul style="list-style-type: none"> - Result-set is ranked list, contains documents in decreasing order of relevance. - Evaluation of such a list for IR is different as it need to cater with the decreasing notion of relevance in the process. - Usually used mean average precision, or normalized cumulative gain. 	Evaluation of Unranked Retrieval <ul style="list-style-type: none"> - Result-set is flat all the documents are equally ranked for a given query. - Set oriented evaluation is used for such a result set. - Usually precision and recall is used in this scenarios.

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 for relevance.

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. Indirect relevance feedback is less reliable.

TRUE. Indirect relevance feedback is taken by implicitly by tracking users interactions from the results. Hence it is less 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.