Quiz 1 : Information Retrieval

Max Marks 30 Time 1 hr

Write only relevant sentences in short. Lengthy and irrelevant answers may attract negative marking.

1. (A) Compare (both similarities and differences) Boolean and Vector Space Representations of Documents and Queries in terms of accuracy, efficiency and ease of implementation of document-search**.** **[5]**

(B) In a collection of 10K documents the following words are present in the number of documents as shown in the table :

oasis 400	place 3000	desert 800
beneath 800	water 800	come 800
people 400	vast 200	around 800
sand 400	world 1000	region 800

Calculate tf-idf term vector for the following document after stemming and removing stop words :

Doc1 : An oasis is a place in a desert where water comes out from beneath the ground.

Doc2 : Most people think of desert as a vast sandy region but only 20% of the world's deserts are sandy. **[5]**

(C) Compare the two documents in terms of similarity and distance **[2]**

2. (A) Two IR Systems returned 15 documents for a user query. Out of six relevant documents in the corpus the following returned ranked documents were relevant

System 1 : 1, 2, 5, 8, 12

System 2 : 1, 3, 5, 11

Calculate interpolated precision at 11 recall points (i.e. 0.0, 0.1, 0.2, 0.3,……. 0.9, 1.0) and draw the recall – precision graph for both the systems. **[5]**

(B) Which system is better? Give quantitative values for your arguments. **[3]**

3. Explain and Give Example (i) Skip Pointers, (ii) Zipf's law (iii) Semantic Indexing (iv) E-measure (v) Dictionary and Posting (vi) Cross lingual IR (vii) Document clustering in IR (viii) Focus drift (ix) Normalization with respect to document length (x) Automatic stop word list creation. **[10]**