**Building and Evaluating an End-to-end Information Retrieval (IR) System**

**Computing Lab Examination, CSE, IIT Kharagpur**

**Full Marks: 75 Time allotted: 2 hours**

1. **Using the Vector Space Model (VSM).** In this model, the term frequency and term weight values for each document (and query) can be represented (and stored) as vectors; hence the name vector space model. **[5 x 6 = 30]**
2. Download and *preprocess* (as discussed in class) the IR resources (queries, document collection and relevance judgments) that have been previously provided to you. **[5]**
3. As earlier, build inverted indices for each term in the collection. Now apply the vector space model for document retrieval as follows. For each term-document pair *(t, d)*, compute the term frequencies tf*t,d* where tf*t,d* is the number of times *t* appears in *d*. **[5]**
4. For each term *t*, compute the Inverse Document Frequency idf*t* as follows.

where *N* is the total number of documents in the collection and df*t* is the document frequency of *t*, i.e. the number of documents in which *t* occurs. **[5]**

1. Compute the tf-idf (also called term weights in this context) for *every* term-document pair as follows (the small hyphen between tf and idf in the tf**-**idf term is **not** a minus sign). **[5]**
2. Compute the overlap score for *every* query-document pair as described below. **[5]**
3. Then, for each query, find and output the ten documents that have the highest overlap score with the query. **[5]**
4. **Evaluation of a ranked list: Mean Average Precision (MAP).** MAP is one of the popular metrics to evaluate ranked lists with binary relevance levels. **[5 + 25 + 5 = 35]**
5. Convert relevance judgments in the *qrels* files 1 and 2 to 1 (relevant), and 0 as non-relevant. **[5]**
6. Compute Average Precision (AP) for each query *q* as follows for the ranked result list containing ten documents for each query derived in Question 1f. **[25]**

where *k* is a position in the ranked list (1 <= *k* <= 10), *rel(k)* is the relevance of the document at rank *k* with respect to *q* (i.e. *rel(k)* is either 0 or 1), and *Precision(k)* is given by

1. Compute MAP for your result set (i.e. each query accompanied by the top-10 documents) as follows:

where *Q* is the set of all queries.

1. **Evaluation of a ranked list: Discounted Cumulative Gain (DCG@10)**. DCG (and its normalized variant nDCG) is the most widely used metric today for evaluating ranked result lists with graded relevance levels (0, 1, 2 in this case). For computing DCG, use the original relevance judgments of 0, 1, and 2 (and **not** the binary levels as in Q. 2). For each query *q*, find the DCG@10 (Discounted Cumulative Gain) for its graded ranked list of 10 documents as follows

where *reli* is the graded relevance judgment (0, 1 or 2) of the document at Rank *i* in the result list. Compute the average DCG@10 for all the queries in *Q*. **[10]**