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]**

   a. Download and *preprocess* (as discussed in class) the IR resources (queries, document collection and relevance judgments) that have been previously provided to you.   **[5]**

   b. 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]**

   c. For each term *t*, compute the Inverse Document Frequency idf_t as follows.

   $$idf_t = \log_{10} \frac{N}{df_t}$$

   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]**

   d. Compute the tf-idf (also called term weights in this context) for *every* term-document pair as follows (the small hypen between tf and idf in the tf-idf term is **not** a minus sign).
   [5]

   $$tf\text{-}idf_{t,d} = tf_{t,d} \times idf_t$$

   e. Compute the overlap score for *every* query-document pair as described below.   **[5]**

   $$Score(q,d) = \sum_{t \in q} tf\text{-}idf_{t,d}$$

   f. Then, for each query, find and output the ten documents that have the highest overlap score with the query.   **[5]**

2. **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]**

   a. Convert relevance judgments in the *qrels* files 1 and 2 to 1 (relevant), and 0 as non-relevant.   **[5]**

   b. 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]**

   $$AP(q) = \frac{\sum_{k=1}^{10}(Precision(k) \times rel(k))}{\#(relevant\ documents\ for\ q)}$$

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

$$Precision(k) = \frac{\#(relevant\ documents\ retrieved\ till\ rank\ k)}{k}$$

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

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} AP(q_j)$$

where Q is the set of all queries.

3. **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

$$DCG@10 = rel_1 + \sum_{i=2}^{10} \frac{rel_i}{\log_2(i+1)}$$

where *rel$_i$* 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]**