

1 Assignment 2: Information Retrieval

1. Given a query that a user submits to an IR system and the top N documents that are returned as relevant by the system, devise an approach (high level algorithmic steps will suffice)

- to suggest query terms to add to the query. Typically, we wish to give a large range of suggestions to the users capturing potential intended query needs, i.e., *high diversity* of terms that may capture the intended query context/content. **(10 marks)**

2. Consider the following scenario: a company search engine is employed to allow people to search a large repository. All queries submitted to the system are recorded. A record that contains the id of the user and the terms in the query is stored. The order of the terms is not stored and neither is any timestamp. Each entry in this record is effectively an id and a set of terms. Any duplicate terms in a query are ignored.

The designers of the search engine, decide to use this information to develop an approach to make query term suggestions for users, i.e., at run time, once a user has entered their query terms, the system will suggest potential extra terms to add to the query.

Given the data available, outline an approach that could be adopted to generate these suggested terms. A brief outline is sufficient to capture the main ideas in your approach.

The designers of the system wish to take into evidence in previous queries and also any similarities between users.

Identify the advantages and disadvantages of your approach (briefly).

(10 marks)