# 1 Assignment 2: Information Retrieval

1. In the classical Rocchio method for relevance feedback, searchers are given an option to give feedback of a binary nature (either positive or negative relevance feedback). Consider designing a system where users can give feedback in a range 1 to 5, where 1 means very non-relevant and 5 means very relevant. Modify the Rocchio approach to allow for such a feedback mechanism. **(5 marks)**

2. 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)**

3. 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 and id and a set of terms. Any duplicate terms in a query is 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 an entered their queries 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 that captures the main ideas in your approach.

   Identify advantages and disadvantages of your approach (briefly).

   **(10 marks)**