1. Distributed Indexing

Many parallel tasks, such as index construction, that run on distributed computer clusters are based on the MapReduce model.

1. Sketch a generic MapReduce schema and explain the input and output of the map as well as the reduce step.
2. Given is the following document collection:
   - D1: to be or not to be
   - D2: strive to be the best
   a) How can MapReduce be used to count the occurrences of all terms in this document collection? Describe the map/reduce input and output.
   b) Modify your described MapReduce schema so that it generates an inverted index.