Implementation:

Task1:

In this task I have implemented Lucene code. Most of the code remains the same as provided in 'https://blackboard.neu.edu/bbcswebdav/pid-15032873-dt-content-rid-24244906_1/courses/CS6200.15344.201810/HW4.java'.

Since the format in which the ranked list of documents should be written into a file is different from the standard format, I have modified it as required.
Also my implementation expects a query file to be present in the same directory as the code, from which each query will be fetched and ranked accordingly.

Task2:

In this task I have implemented BM25 algorithm. It is used to generate scores and will be ranked accordingly.

To generate a list of ranked documents based on the scores calculated the following steps needs to be followed,

Step1:
Unigram inverted index is generated using the corpus generated in the third assignment.

Step2:
Query is fetched from the query text file which will be present in the same directory as the code.

Step3:
For each query fetched the following steps needs to be repeated,

- A list of unique query terms are maintained
- Assuming the relevance information is zero since nothing is specified.
- Calculate K using K = k1((1-b)+ b*dl/avdl)
    - dl and avdl values are computed using the generated inverted indices and
    - where k1 = 1.2
      b = 0.75
- Calculate the score of this document for the given query term using BM25 algorithm and for every term given in the query the summation value is considered.
- The list of documents is ranked based on the scores calculated.
- Top 100 documents are printed in the following format,
  'query_id Q0 doc_id rank BM25_score system_name'