



Please complete the programming assignment using Java. If you are extremely unfamiliar with Java you may use another language that has a Lucene port. The evaluation script will be in Java and requires you to have the latest JVM installed.

Problem 1.

The objective of this task is to build a query expansion system. For a given query, using the index from Assignment 1 retrieve results and rank according to the statistical language model. Then take the top 1000 documents and build a word embedding model. Using the built word embedding model, expand the query with the top 5 similar terms. Now retrieve and rank (using the language model) results using the new query. Compute precision at 5, 10 using the available ground truth data from Assignment 3.