



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.

Index the Temporalia dataset in order to allow for keyword searches and publication date range searches. Parse the raw content, analyze it, identify the publication date and then write a program to index the documents. The index should be able to answer standard keyword queries as well as temporal queries.

Temporal queries are of the format – *keywords @ year-year*

Note, we do not need arbitrary time range queries to be supported. Simple year based intervals are enough. You have to complete the following subtasks:

1. Constructing a valid index. **(20 points)**
2. Support simple keyword queries and implement a query processor to isolate time intervals and keywords for the given input format. **(15 points)**
3. Return the right number of documents for each query in the evaluation script. If you pass all checks in our validation script you get another **15 points**

Dataset: http://l3s.de/~fernando/datasets/Temporalia_Sample3.tar.gz

Tutorial: <http://pharos.l3s.uni-hannover.de:7080/tir/lectures/lecture-lucene.pdf>