L3S Research Center
Temporal Information Retrieval, WiSe 2016-17
Dr. Avishek Anand
Programming 2, due: **5 December 2016**

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.
Implement the probabilistic language model explained in the lecture. You will need to implement two versions with different smoothing techniques - Jelinik-Mercer and Dirichlet.
Remember to allow for the passing of parameters for each ranking model. Retrieve all results from the index for a given query and then rank.
Create the following two programs:

1. LMDirichlet – input parameters are the query, top k results to return (integer value) and Dirichlet prior.
2. LMMercer – input parameters are the query, top k results to return and smoothing parameter α.

Queries are of the format – *keywords* @ *year-year* or just keywords.
The output should be the documents ids in rank order. Output each id in a new line.
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. Implement LMDirichlet (**10 points**)
2. Implement LMMercer (**10 points**)
3. Return the correct top k results for a sample set of queries. (**30 points**)

Dataset: `http://l3s.de/~fernando/datasets/Temporalia_Sample3.tar.gz`
Tutorial: `http://pharos.l3s.uni-hannover.de:7080/tir/lectures/lecture-lucene.pdf`