Chair for Algorithms
and Data Structures
Prof. Dr. Hannah Bast
Björn Buchhold

Information Retrieval
WS 2013/2014

http://ad-wiki.informatik.uni-freiburg.de/teaching

UNI FREIBURG

Exercise Sheet 5

Submit until Tuesday, November 26 at 4:00pm

Exercises 1 - 3 should be implemented as methods of a class *ErrorTolerantSearch*. See the TIP file on the Wiki for further specifications.

Exercise 1 (5 points)

Implement a method that constructs a q-gram index for a given input set, as explained in the lecture. Convert all q-grams to lower case, like we converted all words to lower case in Lecture 1.

Exercise 2 (5 points)

Implement a method that computes the union of an arbitrary given number of inverted list of a q-gram index, as explained in the lecture. Pay attention to keep duplicates in the result or count them, as also explained in the lecture.

Exercise 3 (5 points)

Implement a method that finds all input strings that are within a given PED from a given prefix. As explained in the lecture, first use the q-gram index to determine a subset of candidate phrases. For each candidate phrase then compute the exact PED using the code provided on the Wiki. Rank the matching strings by their score.

Exercise 4 (5 points)

Implement a program *ErrorTolerantSearchMain* that accepts arbitrary queries by the user, and outputs the top-10 matches together with their scores. Use the new data set linked on the Wiki. Also output the numbers and timings asked for on the table on the Wiki, and put your numbers there for the queries *perf* and *uniwersity*. As threshold for the PED, take $\lfloor |x|/4 \rfloor$, where x is the query word. That is, the longer the query, the more errors are allowed.

Add your code to a new sub-directory *exercise-sheet-05* of your folder in the course SVN, and commit it. Make sure that *compile*, *test*, and *checkstyle* run through without errors on Jenkins. As usual, also commit a text file *experiences.txt* where you briefly describe your experiences with this exercise sheet and the corresponding lecture. As a minimum, say how much time you invested and if you had major problems, and if yes, where.