

Exercise Sheet 1

Submit until Wednesday, October 31 at 4:00pm

Exercise 1 (5 points)

Please register at our course system Daphne (using your RZ Account + password for authentication). Make sure that your data is correct, in particular, that you are reachable under the specified e-mail address.

Exercise 2 (5 points)

Implement a simple class *QueryProcessor* that, for a given query with two keywords, computes the ids of all records (from a given text collection) that contain both query words.

Use an inverted index and a linear-time intersect, as explained in the lecture. Consider the implementation advice given in the lecture and the code design suggestions linked on the Wiki. You may use the *InvertedIndex* class written during the lecture.

Write a unit test for every non-trivial method, as explained in the lecture.

Exercise 3 (5 points)

Write a main program *SearchMain* that, using your *QueryProcessor* class, lets the user enter queries with two keywords, and displays k records (where k should be a parameter of your program, default $k = 3$) that contain both query words.

The text collection should be given as a CSV file (one record per line), as in the lecture. For each displayed document, show the URL of the record (first column in the CSV file) and its contents (second column). Optionally, highlight the query terms in the displayed content.

Exercise 4 (5 points)

Check out a working copy of your folder in the SVN repository of the course (see URL on your Daphne page), add your code to a new sub-directory *exercise-sheet-01*, and commit it. Make sure that everything runs through without errors on Jenkins (compiler, unit tests, checkstyle).

Also commit, in that sub-directory, a text file *experiences.txt* where you briefly describe your experiences with the first 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.