

Exercise Sheet 2

Submit until Tuesday, November 5 at 4:00pm

Exercise 1 (10 points)

Copy your code for the class *InvertedIndex* from *exercise-sheet-01* to a new subfolder *exercise-sheet-02*. Then extend the class to incorporate BM25 scores, as explained in the lecture. You find a specification in the file *InvertedIndex.TIP* named on the Wiki.

Try to keep your code as simple and short as possible. As usual, write a unit test for at least one non-trivial input for each non-trivial method = for this exercise, modify your unit tests from the last exercise sheet appropriately.

Exercise 2 (10 points)

Also copy your code for the program *SearchMain* from *exercise-sheet-01* to *exercise-sheet-02*. Make sure that it still works with your new code for the *InvertedIndex* class.

Then use it to inspect the top-5 results for a query of your choice, and manually determine the nDCG@5, as explained in the lecture. Try three different parameter settings for BM25: $k = 0$ and $b = 0$ (unit scores, like implicitly in the last exercise sheet), $k = \infty$ and $b = 0$ (standard tf.idf), and $k = 1.75$ and $b = 0.75$ (the BM25 default). Try to find a query, where the nDCG@5 for unit scores is worse than for the other two. And make sure to pick a query, where two levels of relevance, and hence nDCG@5, make sense. *And pick a query which nobody else has picked!*

Report your results in the table linked on the Wiki, and briefly discuss them in your *experiences.txt* for this exercise sheet (see below).

Add your code to a new sub-directory *exercise-sheet-02* of your folder in the course SVN, and commit it. Make sure that *compile*, *test*, and *checkstyle* run through without errors on Jenkins. Again also commit, now in *exercise-sheet-02*, 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. Don't forget to include the short discussion asked for in Exercise 2.