**Information Retrieval
WS 2012/2013**

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

Exercise Sheet 7

Submit until Wednesday, December 12 at 4:00pm

For the following exercises, copy your code from the last *exercise-sheet-06* to this *exercise-sheet-07* and then extend / modify it.

Exercise 1 (5 points)

Extend your web application from Exercise Sheet 6 such that the user can specify (via UI elements of your choice) the number of suggestions (in the autocompletion box below the search field) and the number of hits (for the highlighted word / query from the autocompletion box).

Exercise 2 (5 points)

Extend the web application further such that these numbers are stored in a cookie, and will be (displayed and used) automatically the next time the user comes to the website.

Optional task: set the expiration date of the cookie to 1 minute after the last user action (that is, if the user goes to the website again 1 minute after that or later, the default number of suggestions and hits will be used again).

Exercise 3 (5 points)

Add a method *utf8Repair* to your class *InvertedIndex* that takes an arbitrary given string (with possibly incorrect UTF-8 sequences here and there) and outputs a string that contains only valid UTF-8 sequences. Try to make as few changes to the given string as possible.

It is not necessary that you achieve the minimal number of changes possible (depending on the input that might be a very hard problem), but you should avoid extreme / naive solutions like overwriting all bytes with a value < 128 (which gives a perfectly valid UTF-8 string, but one which probably has nothing to do with the given string anymore).

Exercise 4 (5 points)

Modify your web application from Exercise Sheet 6 such that it reads the file *wikipedia-sentences. with-utf8-errors.csv* linked on the Wiki. It contains UTF-8 errors of all kinds here and there, and you should use your method from the previous exercise to fix them.

Commit your code to our SVN, in a new sub-directory *exercise-sheet-07*, and make sure that everything (including checkstyle) runs through without errors on Jenkins. Also commit a text file *experiences.txt* with your feedback. As a minimum, say how much time you invested and if you had major problems, and if yes, where.