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

Exercise Sheet 7

Submit until Tuesday, December 10 at 4:00pm

Exercise 1 (5 points)

Copy your web application code from the last exercise sheet to a new folder *exercise-sheet-07*. Then add one or more features of your choice that make meaningful use of cookies. As a minimum, add a drop-down menu that let's the user choose the number of hits to be displayed in response to each query. Remember that choice in a cookie, and make sure that the next time the user visits the website (using the same browser) the number of hits from the last session is used.

Exercise 2 (5 points)

Add a method *decodeUrl* that decodes a URL from your web app to a proper string. Assume UTF-8 encoding for both. For example, *z%C3%BCrich* should be decoded to *zürich*. Apply that method in your server code, on the query string recevied from the client.

Exercise 3 (5 points)

Write a method *iso2utf* that converts a sequence of bytes encoding a string in ISO-8859-1 to a sequence of bytes with a proper UTF-8 encoding representing the same text. In Java, indeed use *byte[]*, since a *String* in Java is already encoded in UTF-16 (2 bytes per character).

Exercise 4 (5 points)

Add an option *–iso* to your server. When the server is run with that option (and only then), use your method *iso2utf* from Exercise 3 to convert the strings from the input file. In Java, make sure that you first read the content of the file into a *byte[]* array, in order to avoid implicit encoding / conversion. After the *iso2utf* conversion, you can store the result in a Java String.

Then run your server on the new file *entities-with-scores.iso-8859-1.txt*, linked on the Wiki. It contains a subset of the entity names from the last exercise sheet, encoded in ISO-8859-1. In your *search-demo.html*, use UTF-8 encoding independent of the *–iso* option. Check that when your server is run with the option *–iso*, the displayed results indeed shows words with Umlauts (ä, ö, ü, etc.) properly. Check what when your server is run without that option, these special characters are shown as invalid UTF-8 (white question mark on a black diamond).

Add your code to a new sub-directory *exercise-sheet-07* 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.