Please note that there are two more documents describing the submission requirements and the scoring of your solutions (criterias for getting all available points).
Students have to get at least 1/2 of all points in order to pass this lecture.

Assignment 2.1

Program your own web search engine using *Apache Lucene*.
Your program should be able to ...

- crawl arbitrary webpages recursively, starting at a seed URL entered into the Java console. Additionally, a user has to be able to specify the crawling depth (recursion depth).
- parse and index each crawled webpage. You may use *jsoup*[1] in order to get the title and visible text of a webpage.
- print a ranked list of relevant webpages given some query entered into the Java console. The output should contain the first 10 most relevant webpages, their rank, title, URL and relevance score.

During development, please make sure not to spam webpages with a large number of requests. ***7 points***

Assignment 2.2

Extend your web search engine from *Assignment 2.1* such that an additional summary is printed for each search result.
Each summary should ...

- contain a number of excerpts that are relevant to the search query. Each excerpt should contain at least one of the query terms (if possible) and reflect its relationship to the corresponding part of the web page.
- have a configurable number of excerpts. Also, the length of each excerpt should be configurable as well (e.g. the number of words, or sentences).
- be printed in a readable, understandable way.

You may use the *lucene-highlighter* package, which can be found in the `contrib` directory of the Lucene zip-archive. ***3 points***

[1] http://jsoup.org/