

The 19th International Conference on Web Engineering (ICWE 2019),  
June 11 - 14, 2019 – Daejeon, Korea

Tutorial: Powerful Data Analysis and Composition with the UNIX Shell  
Andreas Schmidt, Steffen Scholz

## Exercise I

### Tasks:

1. Download the book „The Adventures of Tom Sawyer“ from <http://www.gutenberg.org/ebooks/74> (utf-8 format) and store it locally.
2. For cygwin users only: Convert file to Unix Format with command:  
`dos2unix.exe <file>`
3. Browse through the first pages of the book using `less`. Start `less` with the option that line numbers are displayed (for help type `man less`)..
  - a. Go to line 1234 of the file. What is the fifth word?
  - b. Search for the first appearance of Alfred. In which line of the document is he mentioned the first time? What is his last name?
  - c. Go to the end of the document
  - d. Scroll backward 5 pages
  - e. Show only lines containing the word „Tom“
  - f. Quit `less`
4. How many chapters has the book (hint use `grep` with `-a` option)?
5. Count the number of empty lines in the file
6. Execute `grep Tom <file>` and `grep -o Tom <file>`. What is the difference? Compare the number of returned lines using `wc`. Explain the difference.
7. How often does the names „Tom“ and „Huck“ appear in the book?