CS 2 Computational Thinking for Scientists
Winter Term 2015

Lab 2: Text Representation: 7-bit ASCII, 8-bit ASCII and Unicode UTF-8

1. Create a text file "abc.txt" containing a few lines of text and end-of-line characters. Use a text editor (visual or terminal based) to view and edit these files; count the number of characters and bytes. Write simple English words, containing characters and symbols as defined in 7-bit printable ASCII.
2. Use a Hex Editor to see their binary and hex representations. Match the number of bytes obtained using Unix commands and what the Hex Editor displays.
3. Use a Hex Editor to create a new text file "def.txt" by entering the hex equivalents (from 32 to 7e) of printable characters and symbols, and display the produced text. Repeat the steps in (1) and (2) about the file size and content "def.txt"

4. Repeat the steps in (1), (2), and (3) for text files containing 8-bit ASCII (ISO) files, containing accented characters.

5. Illustrate properties of Unicode UTF-8 using an Hex Editor and Python IDLE, by writing small text files containing the following languages and scripts:
   a. Arabic
   b. Armenian
   c. Hebrew
   d. CJK (Chinese-Japanese-Korean)

Lab Report: Submit your text files and a brief description of their properties (the number of bytes, encoding schemes, how you obtained them), and submit your report by **January 16, Friday 5pm**. The cs2 web page about lab report contains the submission rules.