
CSCI 493.66 Unix tools

>
>
>

----

with emphasis on string algorithms and Bioinformatics

>

Teaching staff
Saad Mneimneh, **HN 1090F**
Office hours: **Wed 2-4 or by appointment**

Textbook
Understanding Unix/Linux programming, Bruce Molay

Lectures
Lecture 1: Introducing the course
Lecture 2: Standard and terminal Input/Output with *more*
Lecture 3: *more* and tty Examples in C, C++, and Perl
Lecture 4: *who* and utmp Raw data with C/C++ structs, structure alignment/padding, C/C++ structs in Perl
Lecture 5: Online sampling of words, exercise: better who (convert time), writing cp (read and write files)
Lecture 6: Directories and files using *ls*
Lecture 7: More on *ls*, file permissions and special bits, *stat*, *passwd* struct, *getpwuid* and *getgrgid*
Lecture 8: Sorting with priority queues, *grep* and regular expressions
Lecture 9: String matching with DFAs (first 2 pages of this), matching regular expressions with NFAs, online material
Lecture 10: String matching with suffix trees, here
Lecture 11: String alignment with dynamic programming, here
Lecture 12: *pwd*, inodes and the Unix file system, file system functions
Lecture 13: Disk files and Device files, *fnctl*, *open* and *creat* flags, atomic system calls
Lecture 14: Controlling terminal device settings, *termios* struct, *tcgetattr* and *tcsetattr*
Lecture 15: Curses library, signals, alarams and signal handlers
Lecture 16: Better signaling, periodic timers, sigaction, atomic and re-entrent functions
Lecture 17: Putting 15 and 16 together in an example, asynchronous input using O_ASYNC flag and *aio_read* with SIGIO
Lecture 18: Creating processes with *fork*, *execvp*, and *wait*
Lecture 19: Z algorithm here (chapter 1), periodic strings, tandem repeats, compression
Lecture 20: Maximal repeats with suffix trees, here
Lecture 21: a suffix tree library, reference
Lecture 22: String folding, here
Lecture 23: Project discussions
Lecture 24 (and on): open

Grading policy
Homework 50%
Final Project 40%
Report 10%

Homework
Homework 1 Due 02/19/09 more.cc, sampler.h, sampler.cc, sampler2.cc, more.pl (a basic more in perl)
Homework 2 Due 02/26/09 who_std.c, who_sys.c, note, who.cc (sorted by time), who.pl, ls.cc
Homework 3 perl NFA Due 03/12/09 regexp.h, regexprun.h (run NFA), regexp.cc (build NFA, missing additions of dots and !), regexp.pl (missing in2post and build NFA), command line options not implemented for both, align.cc, align.pl, note
Homework 4 Due 04/02/09 pwd.cc, pwd.pl, mesg.c, note, lock.cc, note
Homework 5

Projects
Looking for patterns with regular expressions that make mistakes
Finding exact/approximate tandem repeats using heurisic string compression
Suggesting webpages using uniform online sampling of text
Heuristic primer selection using alignment and folding

Learning goals
This course satisfies the following learning goals set by the department: 1b, 2a, 2b, 2c, 3a, 4. It also fulfills credits for the Bioinformatics concentration.