CS 114 Unix Tools – Fall 2004

Assignment 3

Handed out: Wed, 13 Oct
Due: Wed, 20 Oct, 12:20pm

Problem 1: Jobs and redirection

Using `|`, but without using AWK or `;`, create one-line commands that:

- (3 points) Appends a list of the currently running processes (in any format) to the end of the file `processes`. This can include background and stopped processes.
- (3 points) Prints to the file `count` the number of *unique* users (and only the number) currently logged in. The commands `who` and `sort -u` may be helpful.
- (3 points) Prints the process ID (and only the process ID) of any `bash` or `-bash` (indicating `bash` run as a login shell) process running on the machine. Be careful to not include processes that just contain the string "`bash`" on their command line, particularly the commands you write to search for processes named `bash`. The `-e` option to `egrep` may be helpful.

Problem 2: Regular expressions

- (3 points) For our purposes, a URL is any string that starts with "`http://`", is followed by a hostname, and then, optionally, ends with a path. A hostname is just a non-empty string containing no whitespace (i.e., spaces and tabs) and no "`/`". A path is any string of non-whitespace characters starting with a "`/`". Examples of URLs include:
  - `http://www.google.com`
  - `http://www.cs.cornell.edu/`
  - `http://www.cs.cornell.edu/courses/cs114/2004fa`

  Write a `sed` command that extracts the hostname from a URL. For example, for the URLs above the command should print
  - `www.google.com`
  - `www.cs.cornell.edu`
  - `www.cs.cornell.edu`

  You may assume the input to `sed` consists of a single line containing only the URL (as if it were output from `echo http://...`).
- (3 points) Look at the man page for `lynx`, a text-based web browser. Write a one-line command using `lynx -source` *`URL`*, `grep` and `sed` that prints only the text between the `<title>` and `</title>` tags of the webpage referred to by *`URL`*. You may assume that both tags occur on the same line in the webpage source and that each tag occurs only once. The command should output at most one line. **Bonus:** use *only* `lynx` and `sed`, not `grep`.

Submitting your assignment

Hand in your answers in class on the due date or email your answers as *ASCII text* to `cs114@cs.cornell.edu` before the deadline. Please don't email me HTML or a Word document; I won't read it.

Late submissions

After the submission deadline, the following penalties apply, unless you get permission from the instructor *first*:

- 10% penalty if you submit within 6 hours of the deadline.
- 25% penalty if you submit within 12 hours of the deadline.
- 50% penalty if you submit within 72 hours of the deadline.
- 100% penalty if you submit more than 72 hours after the deadline.

Getting help

You are free to discuss this assignment with others in the class, but your work should be your own. In particular, copying other peoples answers (or portions thereof) is prohibited. If in doubt, err on the side of caution, or ask the instructor.

If you need help, you should:

1. Take a look at the man pages, lecture notes, and suggested reading.
2. If you have more than just a simple question, please consider coming to my office hours.
3. Go to `cornell.class.cs114` newsgroup and see if your question is already answered. If not, post it there. I will be happy to answer any questions posted to the newsgroup, not only homework-related ones.
4. If you want to keep your question private, e-mail me at `cs114@cs.cornell.edu`.