

COSC 461 – Fall 2015
Assignment 1
Regular Expression Primer

Your assignment is to create a Python script that uses regular expressions to parse and color hostnames in a log file. Consider that you are the new system administrator at a large data center. System events at the data center are logged to a single file containing the following information (delimited by spaces): the date of the event, the time of the event on that date, the hostname of the customer that generated the event, and a short description of the event. To make the log easier to read, you would like to write a simple Python program that prints the events for each customer as a different color. Your program will read a log file on *standard input* and produce its output on *standard output*. Use the following scheme to color the lines for each customer:

- If the hostname contains the word "mgmt", it is to be considered one of your infrastructure machines and colored **red**.
- If the hostname is made up of the string "notnsa" followed by three numbers, it is one of our top clients, and should be colored **blue**.
- One of our clients decided to name their machines after the phone number of the department responsible for them (formatted as XXX-XXX-XXXX). These should be colored **magenta**.
- Another client named each of their systems after its function. Any hostname made of one these words: "dev", "mail", "db", "web", "app", or "srv", followed by four decimal digits should be colored **green**.
- Another client saw the previous clients idea and decided to grab the last five alpha-numeric characters off of the machine's serial number, added a dash after it, and added one of the words from the previous bullet to describe its function. These lines should be colored **yellow**.
- Your final client has hostnames in the following format: four hexadecimal digits, followed by a dash, four alphabetic characters, another dash, one of the words from the third bullet, and finally a single digit. These should be colored **cyan**.

You will use standard ANSI encoding to color each line. See the colors.py file in the starter package for an example of how to print colored lines in Python. You should use the Python regular expression module ('re') to find and match hostnames from each customer. See the Python website (<https://docs.python.org/2/library/re.html>) for examples on how to use 're'.

After you complete your implementation, you should rigorously test your program to ensure its correctness. We will test your code using the inputs available from the course website as well as a number of other randomly generated log files. We have included a binary executable of our solution in the starter package. This executable was built for the hydra machines and will likely not work on other platforms. You can run it from the starter directory by typing:

```
$ ./bin_soln/regex_primer < sample.log
```

Finally, you will submit a project report. The project report is a short (one page, single-spaced) document that describes:

1. (in your own words) the problem you set out to solve,
2. your approach (i.e. design and relevant implementation details) for solving this problem,
3. how you debugged and tested your solution, and
4. any issues you had in completing the assignment.

Please comment your code so that others (i.e. the grader) can understand it. Comments at the top of the file should indicate your name, this course, and the assignment. You should attempt to match your output as closely as possible to our output.

When you have completed your assignment, you should upload a gzipped tar file (created with `tar cvzf ...`) with your source files and a pdf of your report to the Blackboard course website before the end of the day, anywhere on earth, on Wednesday, September 2nd. (That is, you must submit your project by 6:59am EDT on Thursday, September 3rd to avoid a late penalty). Partial credits will be given for incomplete efforts. However, a program that does not compile or run will get 0 points. Point breakdown is below:

- Reads / writes from standard input and standard output (10)
- Uses Python's `re` module for regular expressions (10)
- Correctly matches hostnames and prints with the appropriate ANSI color codes (40)
- Output is clear and in the appropriate format (10)
- Code is easy to understand (e.g., contains appropriate comments) (10)
- Project report (20)