**Assignment 3: Regular Expressions**

|  |  |
| --- | --- |
| **Date:** | Wednesday, June 3rd |
| **Due:** | Monday, June 8th at 1pm |
| **Assignment Type:** | Group (Up to 6 – may be randomly assigned) |
| **Assignment Title:** | Regular Expressions (Basics) |
| **Style:** | One problem at a time on Discord – WORK TOGETHER |
| **External Sources:** | You are allowed notes, books, and searches |
| **Description:** | This lab covers Regex topics including literals, meta-characters, character classes, and built-ins. Please review the page below for all user input asked for in the problem set given. Also, check out your project! |
| **Points** | 20 |
| **Starting Files** | **mkdir lab3**  **cd lab3**  **cp –r /afs/umbc.edu/users/j/d/jdixon/pub/cs433/lab3/\* .** |

Enter your group member names here:

Rajat Shetty

William Jenkins

Elliot Ressler

Michael Wellen

Ilya Golubs

**Formatting**

1. Create the Perl script “lab3\_1.pl” that will read in everything from the file “lab3\_data.txt” and display it. It should look like this:

1. 'Old Town Road' by Lil Nas X Featuring Billy Ray Cyrus

2. 'Bad Guy' by Billie Eilish

3. 'Talk' by Khalid

4. 'I Don't Care' by Ed Sheeran & Justin Bieber

5. 'Sucker' by Jonas Brothers

6. 'Sunflower (Spider-Man: Into The Spider-Verse)' by Post Malone & Swae Lee

7. 'Wow.' by Post Malone

8. 'Dancing With A Stranger' by Sam Smith & Normani

9. 'Suge' by DaBaby

10. 'Sweet But Psycho' by Ava Max

+ many more

Show both code and result view!!

Michael’s version

#!/usr/bin/perl

use strict;

use warnings;

my $filename = 'lab3\_data.txt';

open(my $fh, '<:encoding(UTF-8)',$filename)

    or die "Could not open file '$filename' $!";

while (my $row = <$fh>) {

    chomp $row;

    print "$row\n";

}

Rajat’s version:

*Same as Michael*

**Adding First Regex Condition**

1. Create the Perl script “lab3\_2.pl” that will read in everything from the file “lab3\_data.txt” and writes it out to a new file, “output.txt” it if it has more than three upper case characters together. It should look like this (in output.txt):

56. 'Here With Me' by Marshmello Featuring CHVRCHES

60. 'Just Us' by DJ Khaled Featuring SZA

61. 'Murder On My Mind' by YNW Melly

63. 'GIRL' by Maren Morris

81. 'Boy With Luv' by BTS Featuring Halsey

84. 'Shotta Flow' by NLE Choppa

87. 'Robbery' by Juice WRLD

98. 'Mixed Personalities' by YNW Melly Featuring Kanye West

Show both code and result view!!

William’s version:

#!/bin/perl

use strict;

use warnings;

my $filename = 'lab3\_data.txt';

open(my $input, '<:encoding(UTF-8)', $filename)

        or die "Could not open file '$filename' $!";

open(my $output, '>', 'output.txt');

while (my $row = <$input>)

{

        chomp $row;

        # Print to stdout.

        print "$row\n" if $row =~ /[A-Z][A-Z][A-Z]/;

        # Print to file.

        print $output "$row\n" if $row =~ /[A-Z][A-Z][A-Z]/;

}

close $output;

close $input;

****

Rajat’s version:

#!/usr/bin/perl

use warnings;

use strict;

my $filename = 'lab3\_data.txt';

my $copyname = 'output.txt';

open(FILENAME,'<',$filename)

or die "Could not open file '$filename' $!";

open(COPYNAME,'>',$copyname)

or die "Could not open file '$copyname' $!";

print("Copying Contents from lab3\_data to output \n");

while(<FILENAME>){

   print COPYNAME $\_;

}

close(FILENAME);

close(COPYNAME);

print "File contents copied successfully!\n";**Arithmetic and Formatting**

1. Create the Perl script “lab3\_3.pl” that will read in everything from the file “lab3\_data.txt” and finds all songs in the top 100 with more than two artists on them. (don't worry about "featuring" at first, but feel free to add it as a challenge)

It should look like this:

24. 'Eastside' by benny blanco, Halsey & Khalid

53. 'Go Loko' by YG, Tyga & Jon Z

67. 'You Stay' by DJ Khaled Featuring Meek Mill, J Balvin, Lil Baby & Jeremih

73. 'Soltera' by Lunay, Daddy Yankee & Bad Bunny

94. 'Bacc At It Again' by Yella Beezy, Gucci Mane & Quavo

Show both code and result view!!

William’s version:

#!/bin/perl

use strict;

use warnings;

my $filename = 'lab3\_data.txt';

open(my $input, '<:encoding(UTF-8)', $filename)

        or die "Could not open file '$filename' $!";

while (my $row = <$input>)

{

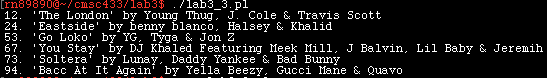
        chomp $row;

        # Print to stdout.

        print "$row\n" if $row =~ /(\'\sby)(.\*\,.\*)(\&)/;

}

close $input;



Rajat’s version:

*I have the same as William*

**Anchors**

1. Create the Perl script “lab3\_4.pl” that will read in everything from the file “lab3\_data.txt” and finds all songs in the top 100 that are between 20-50 in their ranking where their name starts with Look.

The output should look like this:

28. 'Look Back At It' by A Boogie Wit da Hoodie

44. 'Look What God Gave Her' by Thomas Rhett

Show both code and result view!!

**\*I think that the there is a small error - Ilya\***

**\*I think that it should be “print "$row\n" if $row =~ /(^(2[0-9])|(3[0-9])|(4[0-9])|(50))\.\s\'Look/;” “/;” Is missing.\***

William’s version:

#!/bin/perl

use strict;

use warnings;

my $filename = 'lab3\_data.txt';

open(my $input, '<:encoding(UTF-8)', $filename)

        or die "Could not open file '$filename' $!";

while (my $row = <$input>)

{

        chomp $row;

        # Print to stdout.

        print "$row\n" if $row =~ /(^(2[0-9])|(3[0-9])|(4[0-9])|(50))\.\s\'Look/;

}

close $input;



**Adding First Regex Condition**

1. Create the Perl script “lab3\_5.pl” that displays all songs in the top 100 with exactly three words in the title.

You tell me!

Show both code and result view!!

Michael’s version:

It’s a bit hacky to literally check if there’s a “ by” after the title string, so if you have any ideas feel free. It works though.

#!/bin/perl

use strict;

use warnings;

my $filename = "lab3\_data.txt";

open(my $input, '<:encoding(UTF-8)', $filename)

        or die "Could not open file '$filename' $!";

open(my $output, '>', 'output.txt');

while (my $row = <$input>)

{

        chomp $row;

        # Print to stdout.

        print "$row\n" if $row =~ /(')([^\s]\*\s[^\s]\*\s[^\s]\*)('\s[b][y])/

}

close $output;

close $input;



**Finding**

1. Create the Perl script “lab3\_6.pl” that will read in everything from the file “lab3\_data.txt” and finds all songs in the top 100 where any 3 letters is repeated at least two times in any one line ignoring any spaces (case insensitive).

If the line was “This is Your Life” by Hissy Elliott

Then, his from “This” would match “His” from “Hissy”.

Show both code and result view!!

**Ilya’s Solution**

#!/bin/perl

use strict;

use warnings;

my $filename = 'lab3\_data.txt';

open(my $input, '<:encoding(UTF-8)', $filename)

        or die "Could not open file '$filename' $!";

while (my $row = <$input>)

{

        chomp $row;

        print "$row\n" if $row =~ /(\w\w\w)(.\*\1)+/gmi;

}

close $input;