CYBS 3743

Homework #0

Due Feb 6, 2025

Find a string pattern in the bunch of documents

a.k.a. “Regular Expression”

A regular expression is a pattern that the regular expression engine attempts to match in input text. A pattern consists of one or more character literals, operators, or constructs. Such as a MAC address, IP address, time stamp, and so on.

In Linux-like environments, we may use grep, awk, or sed commands to use a pattern matching with some files. In Windows environments, Select-String in PowerShell can be used for those purpose.

* Regex Tutorial – How to write Regular Expressions?  
  ( <https://www.geeksforgeeks.org/write-regular-expressions/> )
* Regular Expression Methods in Python  
  ( <https://www.youtube.com/watch?v=EzeeypMKx7o> )

<Examples with Linux-like commands>

grep -oE "\b([0-9]{1,3}\.){3}[0-9]{1,3}\b" $filename

sed -nE 's/[^0-9]\*(([0-9]{1,3}\.){3}[0-9]{1,3})[^0-9]\*/\1\n/gp' $filename

awk '{while ( match($0,/([0-9]+\.){3}[0-9]+/) ) { print substr($0,RSTART,RLENGTH); $0=substr($0,RSTART+RLENGTH) } }' $filename

<Examples with PowerShell commands>

$filename = "path\to\your\file.txt"  
$ipV4 = "\b([0-9]{1,3}\.){3}[0-9]{1,3}\b"  
Select-String -Path $filename -Pattern $ipV4 | ForEach-Object { $\_.Matches.Value }

For this text processing work, create python scripts, without using external commands above, to find some patterns, such as IPv4 address, in the given documents. Then, create a log file for each transactions

1. Read all configuration files
2. Find all IPv4 addresses in each configuration file
3. Replace it with Class C private IP address
4. Create a log file for each transaction with a time stamp
5. <Extra> find a configuration file, which contains MAC address 96:49:18:8A:D6:B9

All assigned works should be implemented by Python scripts

Please submit your script file, log file, and running result (screen shot on your environment) with a brief explanation.