

ECE560

Computer and Information Security

Fall 2023

Shell Proficiency and Data Manipulation

Tyler Bletsch
Duke University

Motivation

- Everyone needs to manipulate data!
- Attackers need to:
 - Scan target environment for assets
 - Catalog and search target assets for possible vulnerabilities
 - Inspect binaries for specific instruction patterns
 - Extract specific data for processing by other tools (e.g. extracting password hashes from a user database)
- Defenders need to:
 - Scan own environment for assets *and* malicious entities
 - Catalog own inventory and compare against known vulnerabilities
 - Inspect traffic and data for known attack signatures
 - Extract specific data for processing by other tools (e.g. summarizing login failures to update a firewall blacklist)

Fundamental approach: UNIX Philosophy

- Combine simple tools to get complex effects
- Each tool does one thing and does it well
- Basic format of information is *always* a byte stream and *usually* text
- Core ingredients:
 - Shell (e.g. bash)
 - Pipes and IO redirection
 - A selection of standard tools
- Bonus ingredients:
 - SSH trickery
 - Regular expressions (HUGE!)
 - Terminal magic (color and cursor control)
 - Spreadsheet integration
 - More...



```
/[\\w._%+~]+@[\\w.-]+\\. [a-zA-Z]{2,4}/
```

40	41	42	43	44	45	46	47
90 :1	30 90 :1	30 90 :1	30 90 :1	30 90 :1	30 90 :1	30 90 :1	30 90 :1
91 :1	91 :1	31 91 :1	31 91 :1	31 91 :1	31 91 :1	31 91 :1	31 91 :1
92 :1	32 92 :1	32 92 :1	32 92 :1	32 92 :1	32 92 :1	32 92 :1	32 92 :1
93 :1	33 93 :1	33 93 :1	33 93 :1	33 93 :1	33 93 :1	33 93 :1	33 93 :1
94 :1	34 94 :1	34 94 :1	34 94 :1	34 94 :1	34 94 :1	34 94 :1	34 94 :1
95 :1	35 95 :1	35 95 :1	35 95 :1	35 95 :1	35 95 :1	35 95 :1	35 95 :1
96 :1	36 96 :1	36 96 :1	36 96 :1	36 96 :1	36 96 :1	36 96 :1	36 96 :1
97 :1	37 97 :1	37 97 :1	37 97 :1	37 97 :1	37 97 :1	37 97 :1	37 97 :1



The bash shell and common Unix tools

The bash shell

- Shell: Most modern Linux systems use **bash** by default, others exist
 - We'll use bash in our examples
- Side-note: You can get a proper UNIX shell on Windows using Windows Subsystem for Linux.
 - PowerShell is Microsoft's answer to bash...it's fine.



Shell basics review



- **Standard IO:** stdin, stdout, stderr
- **Pipes:** direct stdout of one to stdin of another

```
ls | sort -r # sort files reverse order
```

- **File redirection:** direct any stream to/from a file

```
ls > file_list.txt # save ls to a file (note: no columns!)
```

```
gzip -dc < archive.gz | wc -c # how big is this file uncompressed?
```

```
find -iname dog.* 2> /dev/null # suppress stderr
```

- **Tab completion:** **ALWAYS BE MASHING TAB!!!!!!!!!!!!!!**

- Once = complete, twice = list.
- Not just to save time – it INFORMS you

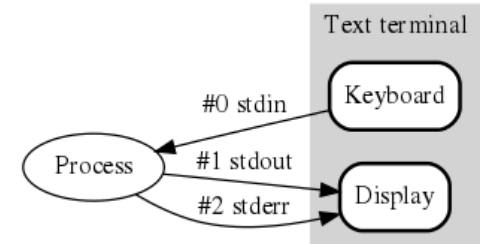
- **Semicolon** for multiple commands on one line

```
make ; ./myapp
```

- Can use **&&** and **||** for short-circuit logic

```
make && ./myapp
```

(Based on **return value** of program, where 0 is success and nonzero is error)



Tab tips

Doesn't complete = doesn't exist
Completes with a space = file exists, unique
Completes with a slash = directory exists, unique
Completes with neither = multiple choices
(hit tab more to see)

Stuff from Homework 0 that I assume you know

- echo
- cat
- head
- tail
- less
- grep
- diff
- wc
- sort
- find



Note: The guy who did the Lynda video, [Scott Simpson, has more videos](#). See [Learning Bash Scripting](#) for examples of some of the stuff in this lecture.

Bash syntax

- Expansions:
 - **Tilde (~)** is replaced by your home directory (try `echo ~`).
`~frank` expands to *frank's* home directory.
 - **Braces** expand a set of options: `{a,b}{01..03}` expands into 6 arguments:
`a01 a02 a03 b01 b02 b03`
 - **Wildcards**: `?` matches one char in a filename, `*` matches many chars,
`[qwe0-3]` matches just the chars q, w, e, 0, 1, 2, or 3.
 - Non-trivial uses! Find all Makefiles two dirs lower: `*/*/Makefile`
 - **Variables** are set with `NAME=VALUE`. Values are retrieved with `NAME`.
Names usually uppercase. Fancy expansions exist, e.g. `FILENAME%.*` will get filename extension; [see here for info](#). Variables can be made into *environment variables* with `export`, e.g. `export NAME=VALUE`.
- Quotes:
 - By default, each space-delimited token is a separate argument (different `argv[]` elements)
 - To include whitespace in a single argument, quote it.
 - Use **single quotes** to disable ALL expansion listed above: `'|{001}'`
 - Use **double quotes** to allow variable expansion only: `"NAME is |{001}"`
 - Or **backslash** to escape a single character: `\$1.21`

Bash syntax (2)

- Control and subshells

for NAME **in** WORDS... ; **do** COMMANDS; **done**

- Execute commands for each member in a list.

while COMMANDS; **do** COMMANDS; **done**

- Execute commands as long as a test succeeds.

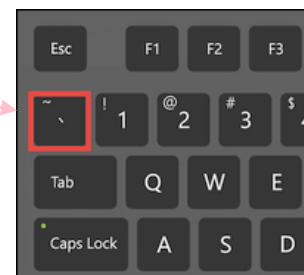
if COMMANDS; **then** COMMANDS;
[**elif** COMMANDS; **then** COMMANDS;] ...
[**else** COMMANDS;] **fi**

- Execute commands based on conditional.

``COMMAND`` or `$(COMMAND)`

- Evaluate to the stdout of COMMAND, e.g.:
`USERNAME=`whoami``

Called a *backtick*



Control flow examples

- Keep pinging a server called 'peridot' and echo a message if it fails to ping.

```
while ping -c 1 peridot > /dev/null ; do sleep 1 ;  
done ; echo "Server is down!"
```

(Can invert by prepending '!' to ping – waits for server to come *up* instead)

- Check to see if our servers have been assigned IPs in DNS:

```
for A in esa{00..06}.egr.duke.edu ; do host $A ; done  
esa00.egr.duke.edu has address 10.148.54.3  
esa01.egr.duke.edu has address 10.148.54.20  
esa02.egr.duke.edu has address 10.148.54.27  
esa03.egr.duke.edu has address 10.148.54.28  
esa04.egr.duke.edu has address 10.148.54.29  
esa05.egr.duke.edu has address 10.236.67.31  
esa05.egr.duke.edu has address 10.148.54.30  
esa06.egr.duke.edu has address 10.148.54.31
```

This stuff isn't just for scripts – you can do it straight on the command line!

Conditionals: [], [[]], (()), ()

- Conditionals

- **Commands:** Every command is a conditional based on its **exit status**

- **Test conditionals:** Boolean syntax enclosed in *spaced-out* braces

- [**STR1 == STR2**] String compare (may need to quote)

- [**-e FILE**] File exists

- [**-d FILE**] File exists and is a directory

- [**-x FILE**] File exists and is executable

- [**! EXPR**] Negate condition described in EXPR

- [**EX1 -a EX2**] AND the two expressions

- [**EX1 -o EX2**] OR the two expressions

- [See here for full list](#)

- Double brackets get you newer bash-only tests like regular expressions:
[[**\$VAR =~ ^https?://**]] VAR starts off like an HTTP/HTTPS URL

- Double parens get you arithmetic:
((**\$VAR < 50**)) VAR is less than 50

- Single parens get you a subshell (various sometimes-useful side effects)₁

What is a script?

- **Normal executable:** binary file in an OS-defined format (e.g. Linux “ELF” format) appropriate for loading machine code, marked with **+x** permission.
- **Script:** Specially formatted text file marked with **+x** permission. Starts with a “**hashbang**” or “**shebang**”, then the name of binary that can interpret it, e.g.:

```
#!/bin/bash
```

or

```
#!/usr/bin/python
```

Note: Modern practice is to use `env` to find the interpreter wherever it lives:
`#!/usr/bin/python` → `#!/usr/bin/env python`

- On execution, OS runs given binary with script as an argument, then any given command-line arguments. No shebang? Defaults to running with bash.
- Example: “`./myscript -a 5`” is run as “`bash ./myscript -a 5`”.
- Can also just run a script with bash manually (e.g. “`bash myscript`”)
- **When should you write a bash script?**
 - When the thing you’re doing is >80% shell commands with a bit of logic
 - Need lots of logic, math, arrays, etc.? Python or similar is usually better.

Examples (1)

- Making an assignment kit for another of my classes:

```
$ echo `ls` > buildkit
```

Dump all the filenames into the would-be script. The echo/backtick makes them space-delimited instead of newline-delimited.

```
$ cat buildkit
```

```
Autograder_rubric.docx Autograder_rubric.pdf byseven.s  
grading_tests homework2-grading.tgz  
HoopStats.s HoopStats.s-cai hw2grade.py  
HW2_GRADING_VERSION Makefile recurse.s
```

```
$ nano buildkit
```

Edit it to add tar command and strip out stuff I don't want to include.

```
$ cat buildkit
```

```
tar czf kit.tgz Autograder_rubric.pdf byseven.s  
grading_tests hw2grade.py HW2_GRADING_VERSION Makefile  
recurse.s
```

```
$ chmod +x buildkit
```

```
$ ./buildkit
```

Mark executable, run, verify tarball was created

```
$ ls -l kit.tgz
```

```
-rw-r--r-- 1 tkb13 tkb13 771264 Sep 14 18:14 kit.tgz
```

Examples (2)

- Script to run the ECE650 “hot potato” project for grading:

```
#!/bin/bash
./ringmaster 51015 40 100 |& tee out-14-rm.log &
./player `hostname` 51015 |& tee out-14-p00.log &
./player `hostname` 51015 |& tee out-14-p01.log &
./player `hostname` 51015 |& tee out-14-p02.log &
...
./player `hostname` 51015 |& tee out-14-p07.log &
./player `hostname` 51015 |& tee out-14-p08.log &
./player `hostname` 51015 |& tee out-14-p09.log &
```

`wait`

Backticks to get external hostname

Backgrounded

Pause until all child processes have exited.

Shorthand for “stdout and stderr together”

More common commands (1)

- **diff**: Compare two files

- Example use: How does this config file differ from the known-good backup?

```
$ diff config config-backup
```

```
2d1
```

```
< evil=true
```



Second line, first column



Left file ('<') has this extra line

- **md5sum/sha*sum**: Hash files

- Example use: Hash all static files, compare hashes later (e.g. using diff)

```
$ find /path -exec sha256sum '{}' ';' > SHA256SUM.orig
```

... (time passes) ...

```
$ find /path -exec sha256sum '{}' ';' > SHA256SUM.now
```

```
$ diff SHA256SUM.orig SHA256SUM.now
```

- **dd**: Do block IO with fine degree of control

- Example use: Overwrite the first 1MB of a hard drive (destroys filesystem, but data is still intact – insecure but fast drive erasure)

```
$ dd if=/dev/zero of=/dev/sda bs=1k count=1k
```

More common commands (2)

- **hd/hexdump/od**: Hex dump (comes in a few variants)
 - Example use: Examine a config file for non-printable or Unicode characters that may be triggering a parser bug.

```
$ hd config1
00000000  73 65 74 74 69 6e 67 31  3d 79 65 73 ff 0a 73 65  |setting1=yes..se|
00000010  74 74 69 6e 67 32 3d 6f  6b 0a                                |tting2=ok.|
0000001a
```

- **strings**: Scan an otherwise binary file for printable strings
 - Example use: Quickly assess an unknown binary file for clues as to its nature

```
$ strings setup.exe | less
```

(scroll through lots of content quickly)

```
<assemblyIdentity version="1.0.0.0" processorArchitecture="X86"
name="DS.SolidWorks.setup" type="win32"></assemblyIdentity><description>This file
will allow SolidWorks to take advantage of xp themes.</description>
```

Conclusion: this is an installer for SolidWorks.

More common commands (3)

- **file**: Identify what kind of file you have by its format

- Example use: Attacker pulled down an opaque file, what is it?

```
$ file hax.dat
```

```
hax.dat: gzip compressed data, last modified: Thu Aug  9 16:50:37 2018, from Unix
```

```
$ gzip -cd hax.dat | file -
```

Most programs that take a filename can take '-' to mean stdin.

```
/dev/stdin: PE32+ executable (console) x86-64, for MS Windows
```

Conclusion: It's a gzip'd Windows executable

- **wget/curl**: Fetch internet stuff via HTTP (and other protocols)

- **wget** downloads to file by default, **curl** writes to stdout by default (but either can do the other with options)

- Example use 1: Download a file

```
$ wget http://150.2.3.5/attacker-kit.tgz
```

- Example use 2: Hit a web API (the URL below tells you your external IP)

```
$ curl http://dsss.be/ip/
```

```
152.3.64.179
```

```
vcm-292.vm.duke.edu
```

Examples (1)

- Quick SSH banner recon:

It's like echo, but it's printf.

```
$ for H in `cat hostlist` ; do printf "%-30s" "$H" ; echo hi | nc $H 22 | head -n1 ; done
```

```
remote.eos.ncsu.edu      SSH-2.0-OpenSSH_7.4
x.dsss.be                SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.4
dsss.be                  SSH-2.0-OpenSSH_6.6.1p1 Ubuntu-2ubuntu2.10
reliant.colab.duke.edu  SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.4
davros.egr.duke.edu     SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.4
esa00.egr.duke.edu      SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.4
esa01.egr.duke.edu      SSH-2.0-OpenSSH_7.6p1 Ubuntu-4
storemaster.egr.duke.edu SSH-2.0-OpenSSH_7.2p2 Ubuntu-4ubuntu2.4
```

Examples (2)

- Download all the course notes (well, all linked PDFs):

```
$ wget -r -ll -A pdf http://people.duke.edu/~tkb13/courses/ece560/
```

```
$ find  Default behavior prints everything below here in the directory tree – a quick way to check what we got.
```

```
.  
./people.duke.edu  
./people.duke.edu/~tkb13  
./people.duke.edu/~tkb13/courses  
./people.duke.edu/~tkb13/courses/ece560  
./people.duke.edu/~tkb13/courses/ece560/slides  
./people.duke.edu/~tkb13/courses/ece560/slides/01-intro.pdf  
./people.duke.edu/~tkb13/courses/ece560/slides/02-overview.pdf  
./people.duke.edu/~tkb13/courses/ece560/slides/03-networking.pdf  
./people.duke.edu/~tkb13/courses/ece560/slides/04-crypto.pdf  
./people.duke.edu/~tkb13/courses/ece560/resources  
./people.duke.edu/~tkb13/courses/ece560/resources/appx  
./people.duke.edu/~tkb13/courses/ece560/resources/appx/C-Standards.pdf  
./people.duke.edu/~tkb13/courses/ece560/resources/appx/F-TCP-IP.pdf  
./people.duke.edu/~tkb13/courses/ece560/resources/appx/I-DomainNameSystem.pdf  
./people.duke.edu/~tkb13/courses/ece560/homework  
./people.duke.edu/~tkb13/courses/ece560/homework/homework0.pdf  
./people.duke.edu/~tkb13/courses/ece560/homework/Ethics Pledge.pdf
```

Examples (3)

Search a big directory tree for a file in old dBase format

- Using find's -exec option:

```
$ find -exec file '{}' ';' | grep -i dbase
```

```
./server01-back/dat/cust20150501/dbase_03.dbf: FoxBase+/dBase III DBF, 14 records * 590,  
update-date 05-7-13, at offset 1025 1st record "0507121      CMP      circular      12"
```

-exec will run a command for each file found, with {} as the filename, terminating the command with ';'.

- Using xargs for efficiency (run fewer discrete processes):

```
$ find | xargs file | grep -i dbase
```

```
./server01-back/dat/cust20150501/dbase_03.dbf: FoxBase+/dBase III DBF, 14 records * 590,  
update-date 05-7-13, at offset 1025 1st record "0507121      CMP      circular      12"
```

xargs takes files in stdin and runs the given command on many of them at a time

- Using xargs with null delimiters to deal with filenames with spaces:

```
$ find -print0 | xargs -0 file | grep -i dbase
```

```
./server01-back/dat/cust20150501/spacey filename.dbf: FoxBase+/dBase III DBF, 14 records *  
590, update-date 05-7-13, at offset 1025 1st record "0507121      CMP      circular      12"
```

Both find's output and xargs's input are set to null-delimited instead of whitespace delimited.

Advanced uses of SSH

Advanced SSH: Tunnels

- Secure Shell (SSH): We know it logs you into stuff and is encrypted. *It does WAY MORE.*
- **SSH tunnels:** Direct TCP traffic through the SSH connection
 - `ssh -L <bindport>:<farhost>:<farport> <server>`
 - **Local forward:** Opens port `bindport` on the local machine; any connection to it will tunnel through the SSH connection and cause `server` to connect to `farhost` on `farport`.
 - `ssh -R <bindport>:<nearhost>:<nearport> <server>`
 - **Remote forward:** Opens port `bindport` on `server`; any connection to it will tunnel back through the SSH connection and the local machine will connect to `nearhost` on `nearport`.
 - `ssh -D <bindport> <server>`
 - **Dynamic proxy:** Opens port `bindport` on the local machine. This port acts as a SOCKS proxy (a protocol allowing clients to open TCP connections to arbitrary hosts/ports); the proxy will exit on the server side. Browsers and other apps support SOCKS proxy protocol.
 - Easy way to punch into or out of a restricted network environment.

Advanced SSH: Tunnel examples

- Example **local forward**:
 - You want to connect to an app over the network, but it doesn't support encryption and/or you don't trust its security.
 - Solution:
 - Set app daemon to only listen on loopback connections (127.0.0.1) port 8888
 - SSH to server with local forward enabled:
`ssh -L 8888:localhost:8888 myserver.com`
 - Connect your client to localhost:8888 instead of myserver.com:8888.
All traffic is tunneled through encryption; access requires SSH creds.
- Example **remote forward**:
 - You're an attacker with SSH credentials to a machine behind a NAT. You have an exploit that lets you run a command on another machine behind the NAT.
 - Solution: SSH to a server you control with a reverse SSH forwarder:
`ssh -R 2222:victim:22 hackerserver.com`
 - Can then connect to hackerserver.com's loopback port 2222 to get to victim.
- Example **dynamic proxy**: Turn it on. Set browser to use it. Surf via server.
 - Bypass censorship, do web-admin on a restricted network, tunnel through a NAT, etc.

Advanced SSH: Keys

- You're used to using passwords to login. That's...decent.
- **Alternative: SSH supports public/private key pairs!**
 - Pro: Allows passwordless login (or you can protect the key with a passphrase)
 - Pro: Key file is random and way longer than password (kills dictionary attack)
 - Pro: Can distribute your public key to any server you want easy access to
 - Con: *Private key must be kept secure!* It allows login!!

Advanced SSH: Key generation

- Create key pair:

```
$ ssh-keygen      (can provide various options, but default are ok)
```

```
Generating public/private rsa key pair.
```

```
Enter file in which to save the key (/home/tkbletsch/.ssh/id_rsa): mykey
```

```
Enter passphrase (empty for no passphrase):
```

```
Enter same passphrase again:
```

```
Your identification has been saved in mykey.
```

```
Your public key has been saved in mykey.pub.
```

```
The key fingerprint is:
```

```
SHA256:kywUn3nyI+LHOnsOYND5+FY7qIaTS+Ta0bXVjGTVY3Y tkbletsch@FREEMAN
```

```
The key's randomart image is:
```

```
+----[RSA 2048]-----+
```

```
|      .   ..      |
```

```
|    . . o +  = E  |
```

```
|  . o . B .o o   |
```

```
|    . + + O      |
```

```
|  . + = S =      |
```

```
| o o = O + .     |
```

```
| +o. B =         |
```

```
| ++..o.+..      |
```

```
|. o+. o=.       |
```

```
+-----[SHA256]-----+
```

Advanced SSH: Key files

- Examining the keys:

```
$ cat mykey
```

```
-----BEGIN RSA PRIVATE KEY-----
```

```
MIIEpAIBAAKCAQEAq6vZKqVSLfZoiXd6yEgu3ZdLO/gv8mBaepWvJbISe5YKQw63  
dBqnLAZc0rJcoqzHgwBjddWUyzDh7g7+MZYgf+n+xE+3QDchqdrktPxj96TMfWUZ  
tH1tpY1UNdbIStAhMbGr/L6aKFs/Ouk5RhWw+GPA7N1diATD0SYibTqdG5+JQqGn
```

```
...
```

```
/4zTb3GDIXFIY9+raaFZ1XLJKBzfhi3ED4ga3nqmeKK60CDTvx8QbA==
```

```
-----END RSA PRIVATE KEY-----
```

```
$ cat mykey.pub
```

```
ssh-rsa
```

```
AAAAB3NzaC1yc2EAAAADAQABAAQCrq9kqpVIt9miJd3rISC7dl0s7+C/yYFp6la8l  
shJ7lgpDDrd0GqcsBlzSslyirMeDAGN11ZTLMOHuDv4xliB/6fuJK0D4BCFbhD8Y2eGh  
TZ/l/g9uIwIv7merL+UQduCSKvqLo1X4JYsI5VSkNKCjcLo7lJoCOUazqmttkX2EBSGd  
3VYp97Eu3XC3rqDAa/FnUe3E4w8nHLk9mB6/qbyr tkblets@FREEMAN
```

Informational comment, defaults to
username@hostname, could be anything.

Advanced SSH: Key usage

- Authorizing a key:
 - Copy only `mykey.pub` to the remote machine you want to establish access to
 - On remote machine, add it to `~/.ssh/authorized_keys`:

```
$ cat mykey.pub >> ~/.ssh/authorized_keys
```
- Using a key to login:
 - Provide the *identity file* (private key) with `-i`:

```
$ ssh -i mykey remotehost.com
```
 - SSH will use `~/.ssh/id_rsa` by default – can use this “default key” without extra options.
- Remember: **keep your private key safe!!!**

Advanced SSH: Commands

- Can give a command with ssh to only do that command (no interactive session). Stdin/stdout/stderr are tunneled appropriately!
 - Really works great with passwordless keys!
 - Find out uptime of server quickly:

```
$ ssh myserver uptime
```
 - Reboot nodes in a cluster:

```
$ for A in node{0..7} ; do ssh root@$A reboot ; done
```
 - Back up remote physical disk image:

```
$ ssh root@server bash -c "gzip -c < /dev/sda" > server.img.gz
```

Advanced SSH: SCP, SFTP, and Rsync

- Almost any SSH server is also a file server using the SFTP protocol. The `scp` command is one way to use this.

- Copy a file to remote home directory:

```
$ scp file1.txt username@myserver:
```

- Copy a directory to remote server's web root:

```
$ scp -r dir1/ webadmin@myserver:/var/www/
```

- Can also use a tool called `rsync` to copy *only changes* to files

- Here's the script I use to update the course site:

```
echo COLLECTING COURSE SITES
```

```
rsync -a --delete-delay ./ECE560/website/  
./www/courses/ece560/
```

```
rsync -a --delete-delay ./ECE250/website/  
./www/courses/ece250/
```

```
rsync -va --progress --delete-delay --no-perms www/*  
tkb13@login.oit.duke.edu:public_html/
```

Understanding and controlling the terminal

Brief terminal history

- Original terminal: **the teletype machine**
 - Based on typewriter technology
 - It's why we say “carriage return” and “line feed”
- Then came: **the serial terminal**
 - CRT display with basic logic to speak serial protocol
 - Many hooked up to one mainframe
 - Needed new *codes* to do new things like “clear screen” and “underline” without breaking compatibility
- Now we have: **terminal emulators** like **xterm**
 - Even more *codes* to do color, cursor movement, non-scrolling regions, etc.
- In Linux, the physical display is a “TTY” (teletype), e.g. `/dev/tty1`.
- Logical terminals like SSH sessions are “pseudoterminals”, e.g. `/dev/pts/0`

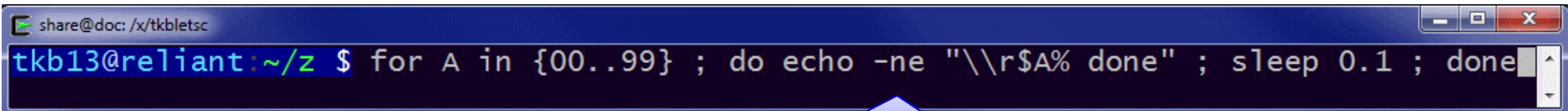


```
share@doc: ~/kikletsic
Tasks: 137 total, 1 running, 136 sleeping, 0 stopped, 0 zombie
  CPU:  0.0 us, 0.0 sy, 0.0 ni, 100.0 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
Mem:   202832 total, 332760 free, 191156 used, 884576 buff/cache
k18 Swap: 2097148 total, 2097148 free, 0 used, 1986160 avail mem

PID USER   PR  NI  VIRT  RES  SHR  SPM  TIME+ P COMMAND
2354 tkbl1  20  0  41856 3996 R  1 0 0.2 0:00.08 1 top
1 root   20  0  37600 5656 S  0 0 0.3 0:10.21 0 /sbin/init
2 root   20  0  0 0 0 0 0 0:00.35 0 [kthreadd]
3 root   20  0  0 0 0 0 0 0:00.56 0 [ksoftirqd/0]
5 root   0 -20  0 0 0 0 0 0:00.00 0 [worker/0-0]
7 root   20  0  0 0 0 0 0 4:26.71 1 [rcu_sched]
8 root   20  0  0 0 0 0 0 0:00.00 0 [rcu_bh]
9 root   rt  0  0 0 0 0 0 0:00.39 0 [migration/0]
10 root  rt  0  0 0 0 0 0 0:09.12 0 [watchdog/0]
11 root  rt  0  0 0 0 0 0 0:05.22 1 [watchdog/1]
12 root  rt  0  0 0 0 0 0 0:00.29 1 [migration/1]
13 root  20  0  0 0 0 0 0 0:03.55 1 [ksoftirqd/1]
15 root  20  0  0 0 0 0 0 0:00.00 0 [kdevfsd]
17 root  20  0  0 0 0 0 0 0:00.00 0 [netns]
18 root  0 -20  0 0 0 0 0 0:00.00 0 [perf]
19 root  20  0  0 0 0 0 0 0:01.03 0 [khungtaskd]
20 root  0 -20  0 0 0 0 0 0:00.00 1 [writeback]
21 root  25  5  0 0 0 0 0 0:00.00 1 [ksmd]
22 root  39 19  0 0 0 0 0 0:27.54 0 [khugepaged]
23 root  0 -20  0 0 0 0 0 0:00.00 1 [cryptd]
24 root  0 -20  0 0 0 0 0 0:00.00 1 [kintegrityd]
25 root  0 -20  0 0 0 0 0 0:00.00 1 [biset]
```

Terminal control sequences: Basic idea

- Some ASCII values are special:
 - 0x0A = Linefeed (move cursor to next line), written as `\n`
 - 0x0D = Carriage return (move cursor to left), written as `\r`
 - 0x07 = Bell (beep), written as `\a`
 - 0x1B = Escape – indicates a special multi-byte sequence, written as `\e`
 - MANY sequences exist. [Full documentation here](#).
- Example: Show a progress line without scrolling:



```
share@doc: /x/tkblets  
tkb13@reliant:~/z $ for A in {00..99} ; do echo -ne "\\r$A% done" ; sleep 0.1 ; done
```

-n means “no newline”, -e means “allow escape characters”.

- Example: How does the ‘clear’ command work?



```
share@doc: /x/tkblets  
tkb13@reliant:~ $ clear | hd  
00000000 1b 5b 48 1b 5b 32 4a | . [H. [2]  
00000007 7 bytes  
tkb13@reliant:~ $
```


Terminal control sequences: Color!

Terminal Escape Code Table 1.6 by Tyler Bletsch

Color control codes are of the form:

`\e[<Code>m` OR `\e[<Code>;<Code>;...m`

where:

* `\e` is the escape character (ASCII 27, `\033`, `\x1B`)

* `<Code>` is a style or color code below, or none to indicate a reset.

For example:

`\e[36;44m`Cyan on blue. `\e[96m`Hi-cyan. `\e[1m`**Bold**, `\e[4m`underline. `\e[m` Reset.

Yields:

Cyan on blue. Hi-cyan. Bold, underline. Reset.


style codes: 0=Reset, 1=Bold, 2=Faint, 4=Underline, 7=Reverse, 9=Strikeout


Colors:

40	41	42	43	44	45	46	47
90 ;1	30 90 ;1	30 90 ;1	30 90 ;1	30 90 ;1	30 90 ;1	30 90 ;1	30 90 ;1
31 91 ;1	31 91 ;1	31 91 ;1	31 91 ;1	31 91 ;1	31 91 ;1	31 91 ;1	31 91 ;1
32 92 ;1	32 92 ;1	32 92 ;1	32 92 ;1	32 92 ;1	32 92 ;1	32 92 ;1	32 92 ;1
33 93 ;1	33 93 ;1	33 93 ;1	93 ;1	33 93 ;1	33 93 ;1	33 93 ;1	33 93 ;1
34 94 ;1	34 94 ;1	34 94 ;1	34 94 ;1	94 ;1	34 94 ;1	34 94 ;1	34 94 ;1
35 95 ;1	35 95 ;1	35 95 ;1	35 95 ;1	35 95 ;1	95 ;1	35 95 ;1	35 95 ;1
36 96 ;1	36 96 ;1	36 96 ;1	36 96 ;1	36 96 ;1	36 96 ;1	96 ;1	36 96 ;1
37 97 ;1	37 97 ;1	37 97 ;1	37 97 ;1	37 97 ;1	37 97 ;1	37 97 ;1	97 ;1

The xterm 256-color gamut:

`\e[38;5;Nm`=Foreground to code N. `\e[48;5;Nm`=Background to code N.

N in 0..15 standard colors above: []

N in 232..255 Grayscale: []

N in 16..231 (N-16) is the RGB in base 6, $36r+6g+b+16$ for r,g,b in 0..5:

16	[]
88	[]
160	[]

^ Re-run with the `-e` option for extended color list.

Why bother?

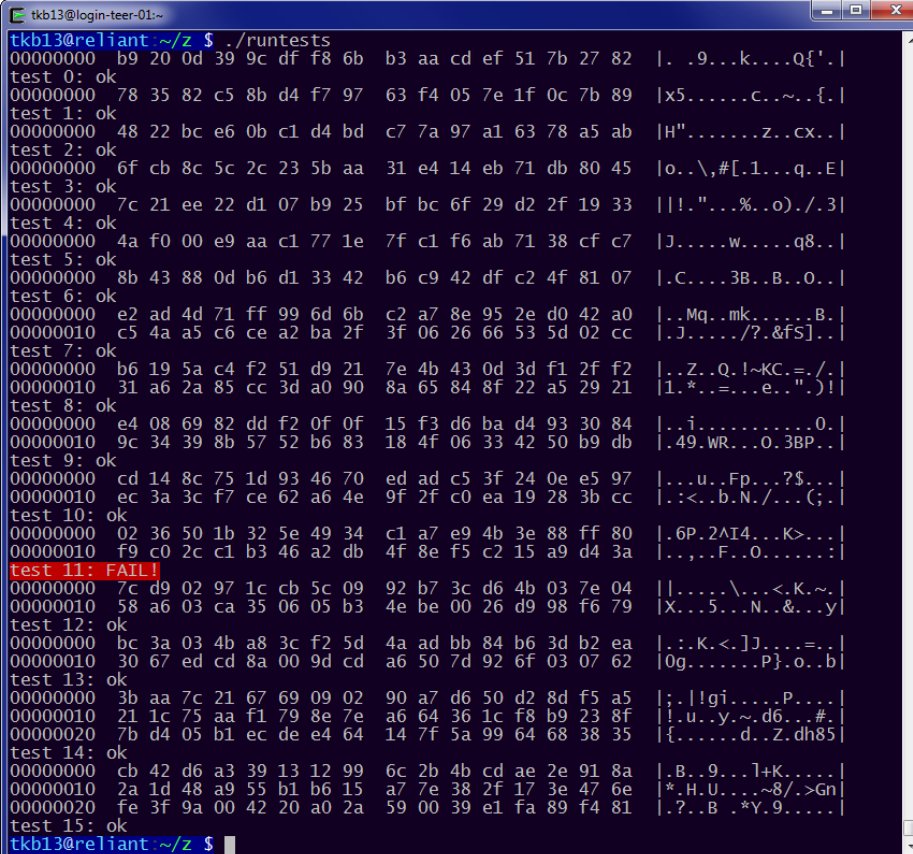
- Making output visually distinctive can *greatly* accelerate a task!
- Tester for ECE650 root kit: which would you rather use?

```
tkb13@login-teer-01:~  
See <file:///usr/share/doc/gcc-5/README.Bugs> for instructions.  
Makefile:12: recipe for target 'sneaky_process' failed  
make: *** [sneaky_process] Error 1  
Removing sneaky_mod in case one is already loaded from a past run.  
rmmod: ERROR: Module sneaky_mod is in use  
Pre-test fault: sneaky is already in /etc/passwd, trying to auto-fix...  
Auto-fix successful, proceeding...  
Running sneaky_process...  
bash: /home/tkb13/h5/sneaky_process: No such file or directory  
The PID is '4476'  
*** TESTING BEGINS ***  
  
Test 1  
test$ ls  
Test 1 (ls): pass  
  
Test 2  
test$ ls -l  
Test 2 (ls -l): pass  
  
Test 3  
test$ ls -l /home/tkb13/h5  
Test 3 (ls -l /home/tkb13/h5): pass  
  
Test 4  
test$ find . -name sneaky_process  
Test 4 (find . -name sneaky_process): pass  
  
Test 5  
test$ find /home/tkb13/h5 -name sneaky_process  
Test 5 (find /home/tkb13/h5 -name sneaky_process): pass  
  
Test 6  
test$ lsmod  
Test 6 (lsmod): FAIL  
  
Test 7  
test$ ls /proc/  
Test 7 (ls /proc/): pass  
  
Test 8  
test$ ps -au  
Test 8 (ps -au): pass  
  
Test 9  
test$ cat /etc/passwd  
Test 9 (cat /etc/passwd): pass  
  
Test 10  
test$ grep -E -B 1 -A 4 '/etc/passwd\sneakyuser' strace.log  
** For this test, examine the strace highlights below and verify that /etc/passwd was  
modified as expected...  
Test 10 (grep -E -B 1 -A 4 '/etc/passwd\sneakyuser' strace.log): (manual inspection)  
^^ Are there syscalls above that would have modified /etc/passwd as expected?  
*** TESTING ENDS ***  
Sending 'q' to exit gracefully  
tkb13@reliant ~/h5 $
```

```
tkb13@login-teer-01:~  
See <file:///usr/share/doc/gcc-5/README.Bugs> for instructions.  
Makefile:12: recipe for target 'sneaky_process' failed  
make: *** [sneaky_process] Error 1  
Removing sneaky_mod in case one is already loaded from a past run.  
rmmod: ERROR: Module sneaky_mod is in use  
Pre-test fault: sneaky is already in /etc/passwd, trying to auto-fix...  
Auto-fix successful, proceeding...  
Running sneaky_process...  
bash: /home/tkb13/h5/sneaky_process: No such file or directory  
The PID is '4476'  
*** TESTING BEGINS ***  
  
Test 1  
test$ ls  
Test 1 (ls): pass  
  
Test 2  
test$ ls -l  
Test 2 (ls -l): pass  
  
Test 3  
test$ ls -l /home/tkb13/h5  
Test 3 (ls -l /home/tkb13/h5): pass  
  
Test 4  
test$ find . -name sneaky_process  
Test 4 (find . -name sneaky_process): pass  
  
Test 5  
test$ find /home/tkb13/h5 -name sneaky_process  
Test 5 (find /home/tkb13/h5 -name sneaky_process): pass  
  
Test 6  
test$ lsmod  
Test 6 (lsmod): FAIL  
  
Test 7  
test$ ls /proc/  
Test 7 (ls /proc/): pass  
  
Test 8  
test$ ps -au  
Test 8 (ps -au): pass  
  
Test 9  
test$ cat /etc/passwd  
Test 9 (cat /etc/passwd): pass  
  
Test 10  
test$ grep -E -B 1 -A 4 '/etc/passwd\sneakyuser' strace.log  
** For this test, examine the strace highlights below and verify that /etc/passwd was  
modified as expected...  
Test 10 (grep -E -B 1 -A 4 '/etc/passwd\sneakyuser' strace.log): (manual inspection)  
^^ Are there syscalls above that would have modified /etc/passwd as expected?  
*** TESTING ENDS ***  
Sending 'q' to exit gracefully  
tkb13@reliant ~/h5 $
```

Simple example – make errors obvious

```
for testnum in {0..15} ; do
    if ./dotest $testnum ; then
        echo "test $testnum: ok"
    else
        echo -e "\e[41mtest $testnum: FAIL!\e[m"
    fi
done
```



```
tkb13@reliant:~/z $ ./runtests
00000000 b9 20 0d 39 9c df f8 6b b3 aa cd ef 51 7b 27 82 |. .9...k...Q{'.|
test 0: ok
00000000 78 35 82 c5 8b d4 f7 97 63 f4 05 7e 1f 0c 7b 89 |x5.....c...~..{|
test 1: ok
00000000 48 22 bc e6 0b c1 d4 bd c7 7a 97 a1 63 78 a5 ab |H".....z..cx..|
test 2: ok
00000000 6f cb 8c 5c 2c 23 5b aa 31 e4 14 eb 71 db 80 45 |o..\.# [.1...q..E|
test 3: ok
00000000 7c 21 ee 22 d1 07 b9 25 bf bc 6f 29 d2 2f 19 33 |!|!"...%..o)/.3|
test 4: ok
00000000 4a f0 00 e9 aa c1 77 1e 7f c1 f6 ab 71 38 cf c7 |J.....w.....q8..|
test 5: ok
00000000 8b 43 88 0d b6 d1 33 42 b6 c9 42 df c2 4f 81 07 |.C....3B..B..O..|
test 6: ok
00000000 e2 ad 4d 71 ff 99 6d 6b c2 a7 8e 95 2e d0 42 a0 |..Mq..mk.....B.|
00000010 c5 4a a5 c6 ce a2 ba 2f 3f 06 26 66 53 5d 02 cc |.J...../?.&fS]..|
test 7: ok
00000000 b6 19 5a c4 f2 51 d9 21 7e 4b 43 0d 3d f1 2f f2 |..Z..Q.!~KC.=./.|
00000010 31 a6 2a 85 cc 3d a0 90 8a 65 84 8f 22 a5 29 21 |1.*...=...e..")!|
test 8: ok
00000000 e4 08 69 82 dd f2 0f 0f 15 f3 d6 ba d4 93 30 84 |..i.....0..|
00000010 9c 34 39 8b 57 52 b6 83 18 4f 06 33 42 50 b9 db |.49.WR...O.3BP..|
test 9: ok
00000000 cd 14 8c 75 1d 93 46 70 ed ad c5 3f 24 0e e5 97 |...u..Fp...?$.|
00000010 ec 3a 3c f7 ce 62 a6 4e 9f 2f c0 ea 19 28 3b cc |.:< .b.N./...(;.|
test 10: ok
00000000 02 36 50 1b 32 5e 49 34 c1 a7 e9 4b 3e 88 ff 80 |.6P.2^I4...K>...|
00000010 f9 c0 2c c1 b3 46 a2 db 4f 8e f5 c2 15 a9 d4 3a |...F..O.....:|
test 11: FAIL!
00000000 7c d9 02 97 1c cb 5c 09 92 b7 3c d6 4b 03 7e 04 ||. . . . \ . . . < . K . ~ . |
00000010 58 a6 03 ca 35 06 05 b3 4e be 00 26 d9 98 f6 79 |X...5...N.&...y|
test 12: ok
00000000 bc 3a 03 4b a8 3c f2 5d 4a ad bb 84 b6 3d b2 ea |.:.K.<.]J.....=..|
00000010 30 67 ed cd 8a 00 9d cd a6 50 7d 92 6f 03 07 62 |0g.....P}.o..b|
test 13: ok
00000000 3b aa 7c 21 67 69 09 02 90 a7 d6 50 d2 8d f5 a5 |;.!gi....P....|
00000010 21 1c 75 aa f1 79 8e 7e a6 64 36 1c f8 b9 23 8f |!u..y..~.d6...#.|
00000020 7b d4 05 b1 ec de e4 64 14 7f 5a 99 64 68 38 35 |{.....d..Z.dh85|
test 14: ok
00000000 cb 42 d6 a3 39 13 12 99 6c 2b 4b cd ae 2e 91 8a |.B..9...]+K.....|
00000010 2a 1d 48 a9 55 b1 b6 15 a7 7e 38 2f 17 3e 47 6e |*.H.U....~8/./>Gn|
00000020 fe 3f 9a 00 42 20 a0 2a 59 00 39 e1 fa 89 f4 81 |.?.B.*Y.9.....|
test 15: ok
tkb13@reliant:~/z $
```

Also you can do cool crap



A terminal window titled "tkb13@login-teer-07:~" showing a command prompt. The prompt is "tkb1etsc@doc: ~/motd2015 \$". The command "cat mario1.ansi" has been executed, resulting in a pixelated ASCII art representation of Mario's head and shoulders. The character is rendered using a limited color palette of red, blue, orange, and black on a dark background. The terminal window includes standard window controls (minimize, maximize, close) in the top right corner and a scrollbar on the right side.

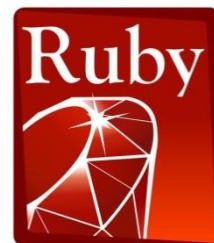
```
tkb13@login-teer-07:~
tkb1etsc@doc: ~/motd2015 $ cat mario1.ansi
tkb1etsc@doc: ~/motd2015 $
```

Scripting languages and regular expressions

Regular expression material is adapted from “Regular Expressions” in “Python for Informatics: Exploring Information” by Charles Severance at Univ. Michigan and “Regular Expressions” by Ian Paterson at Rochester Institute of Technology

Higher-level scripting languages

- Key languages categories commonly used:
 - **Application:** Java, C#, maybe C++
 - **Systems programming:** C, maybe Rust
 - **Shell:** bash (or ksh, tcsh, etc.)
 - **Scripting:** Python, Perl, or Ruby
- You *can* do everything in bash, but it gets ugly. Things bash is awkward at:
 - Math
 - Arrays
 - Hash/dictionary data structures
 - Really any data structures...
- Turn to **scripting languages:** dynamic, interpreted, compact



Scripting language key insight: three fundamental types

- Most data manipulation tasks can be phrased as simple algorithms against these three types:
 - **Scalar:** simple value, numeric or string
 - **Array:** list of values (can nest)
 - **Hash/dictionary/map:** relationship between keys and values (can nest)

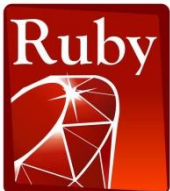


```
my %pairs = ( "hello" => 13,  
             "world" => 31,  
             "!" => 71 );  
foreach my $key ( keys %pairs ) {  
    print "key = $key, value = $pairs{$key}\n";  
}
```



```
myDict = { "hello": 13,  
           "world": 31,  
           "!" : 71 }  
for key, value in myDict.items():  
    print ("key = %s, value = %s" % (key, value))
```

[Examples from here.](#)



```
my_dict = { "hello" => 13,  
           "world" => 31,  
           "!" => 71 }  
my_dict.each {|key, value| puts "key = #{key}, value = #{value}"}
```

One-liners



- Scripting languages support one-liners (typed from shell as a single command).
- Perl is the king of one-liners.
 - `-e` to provide code
 - `-n` automatically wraps code in “for each line of input from stdin or files”
 - `-i` replaces the content of given files with stdout of program (can provide filename extension to back up original data to)
- Quick Perl intro
 - Scalar variables start with a dollar sign, e.g. `$var`
 - Most functions, if you don't specify, affect a variable called `$_`
 - Reference an array element value with `$array[$i]`, whole array is `@array`
 - Reference a hash element value with `$hash{$k}`, whole hash is `%hash`
 - Variables you make reference to are automatically created if they don't exist (including arrays and hashes)
 - One-line comments with `#`

Perl one-liner example



- Remove duplicate lines from a file while preserving original order

Long-winded Perl:

```
while (<>) { # for each line
    if (!$hash{$_}) {
        print;
    }
    $hash{$_}=1;
}
```

Run it:

```
$ perl dedupe.pl in.txt
```

in.txt

```
alpha
delta
alpha
bravo
bravo
alpha
charlie
alpha
bravo
alpha
```

One-liner:

```
$ perl -ne 'if (!$h{$_}){print} $h{$_}=1;' in.txt
alpha
delta
bravo
charlie
```

Crazy dense one-liner:

```
$ perl -ne '$h{$_}++||print;' in.txt
```

Manipulating text

- Task: extract the hostname part of a URL, e.g. <http://google.com/images>
- Thought process:
 - Idea: Start at character 7, capture until you find a slash
 - Problem: what about https?
 - Idea: Go until you see two slashes in a row, then capture until you find a slash
 - Problem: can have more than two slashes at start
 - Idea: Go until you see two or more slashes in a row, then capture until you find a slash
 - Problem: What about username specifier (user@) and port number (:80)?
 - ugh nevermind just give up 😞
- Solution: We need a **language** to describe string processing!

THE LANGUAGE OF STRING PROCESSING



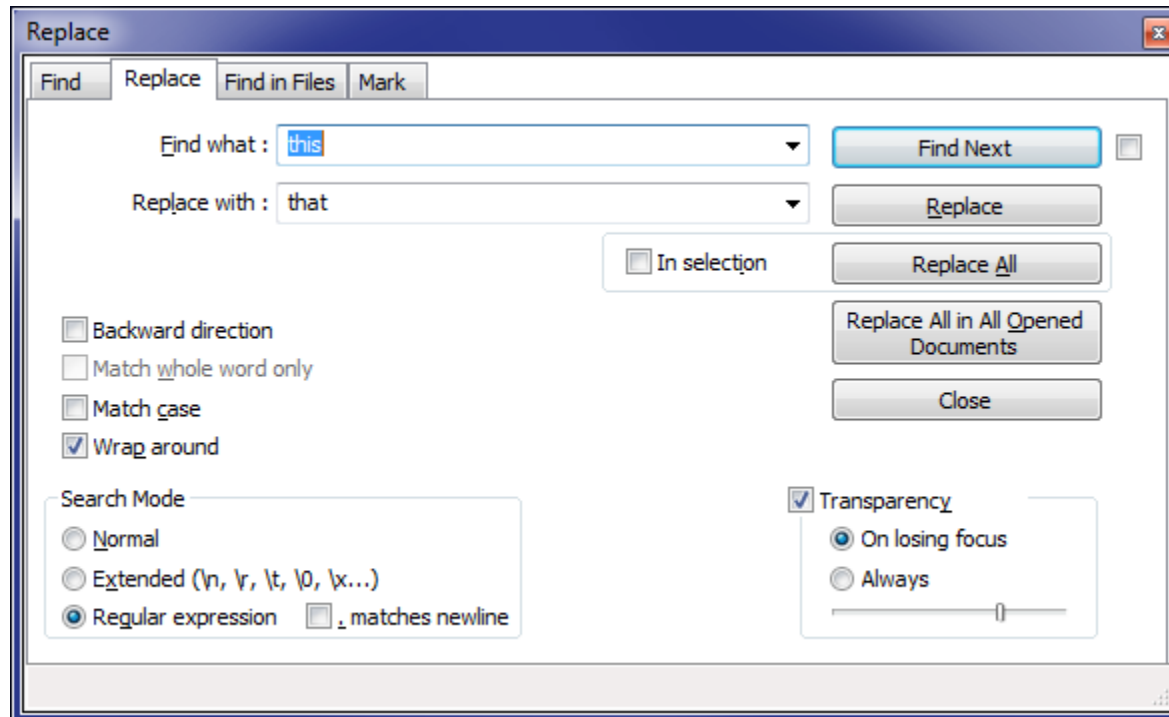
REGULAR

EXPRESSIONS



Regular Expressions

- Regular expressions are expressive rules for walking a string
 - May capture parts of the string (parsing) or modify it (substitution)
 - Like a fancy find-and-replace



Understanding Regular Expressions

- Very powerful, cryptic, and fun
- Regular expressions are a language:
 - Based on "marker characters" - programming with characters
- The “gold standard” variant is from Perl:
Perl-Compatible Regular Expressions (PCRE)
- Many languages support Perl-Compatible Regular Expressions:
Perl, grep (with -P), **sed** (mostly), **Python, Ruby, Java,**
most text editors, basically any language/tool worth using.
- **Common among languages:** Actual syntax *inside* regex
- **Differs between languages:** Syntax to call a regex, get feedback from it, provide options, etc.
- We’ll use both Perl and Python examples – easy to port to others

Introduction to Regular Expressions



- Basic syntax

- In Perl and sed, RegEx statements begin and end with /
(This is language syntax, not the case for Python and others)

- `/something/`

- Escaping reserved characters is crucial

- `/(i.e. /` is invalid because `(` must be closed

- However, `/\(i\.e\.` is valid for finding '(i.e.'

- Reserved characters include:

`. * ? + () [] { } / \ |`

- Also some characters have special meanings based on their position in the statement

Regular Expression Matching



- Text Matching
 - A RegEx can match plain text
 - ex. `if ($name =~ /Dan/) { print "match"; }`
 - But this will match Dan, Danny, McDaniel, etc...
- Full Text Matching with Anchors
 - Might want to match a whole line (or string)
 - ex. `if ($name =~ /^Dan$/) { print "match"; }`
 - This will only match Dan
 - `^` anchors to the front of the line
 - `$` anchors to the end of the line

In Python: The Regular Expression Module

- Before you can use regular expressions in your program, you must import the library using `import re`
- Use `re.search()` to see if a string matches a regex
- Use `re.findall()` extract parts of a string that match your regex
- Use `re.sub()` to replace a regex match with another string
- Use `re.split()` to separate a string by a regex separator

- Example:

- `if re.search(r'Dan', name): print "match"`

In Python, r-quotes mean “raw string”, i.e. “don’t interpret escapes in this string”, which makes it convenient to write Regexes which use all sorts of weird punctuation

General operation

- Engine searches string from the beginning
 - Plain text is treated literally
 - Special characters allow more flexible matching
- A regex is just a way to write a finite state machine (FSM)
 - FSM proceeds through states as matching characters are encountered; if a full regex is walked, that's a match.
- Every character matters!
 - `/ s/` is not the same as `/ s/`

Regular Expression Char Classes

- Allows specification of only certain allowable chars
 - `[dofZ]` matches only the letters d, o, f, and Z
 - If you have a string 'dog' then `/[dofZ]/` would match 'd' only even though 'o' is also in the class
 - So this expression can be stated "match one of either d, o, f, or Z."
 - `[A-Za-z]` matches any letter
 - `[a-fA-F0-9]` matches any hexadecimal character
 - `[^*/\]` matches anything BUT *, \$, /, or \
 - The `^` in the *front* of the char class specifies 'not'
 - In a char class, you only need to escape: `\ (] - ^`

Regular Expression Char Classes

- Special character classes match specific characters
 - `\d` matches a single digit
 - `\w` matches a word character: `[A-Za-z0-9_]`
 - `\b` matches a word boundary,
e.g. `/\bword\b/` catches “`my word!`” but not “`mr. wordly`”
 - `\s` matches a whitespace character (space, tab, newline)
 - `.` wildcard matches everything but newlines (can make it include newlines)
 - Use very carefully, you could get anything!
- To match “anything but...” capitalize the char class
 - i.e. `\D` matches anything that isn’t a digit

Regular Expression Char Classes



- Character Class Examples
 - `/e\w\w/`
 - Matches ear, eye, etc
 - `$thing = '1, 2, 3 strikes!'; $thing =~ /\s\d/;`
 - Matches '2'
 - `$thing = '1, 2, 3 strikes!'; $thing =~ /[\s\d]/;`
 - Matches '1'
 - Not always useful to match single characters
 - `$phone =~ /\d\d\d-\d\d\d-\d\d\d\d/;`
 - There's got to be a better way...

Regular Expression Repetition



- Repetition allows for flexibility
 - Range of occurrences
 - `$weight =~ /\d{2,3}/;`
 - Matches any two- or three-digit integer
 - `$name =~ /\w{5,}/;`
 - Matches any name longer than 5 letters
 - `if ($SSN =~ /\d{9}/) { print "valid SSN!"; }`
 - Matches exactly 9 digits

Regular Expression Repetition



- General Quantifiers

- Some more special characters

*** matches *zero or more* of the preceding, so `/\d*/` matches any size number *or no number at all*

+ matches one or more of the preceding, so `/\w+/` matches one or more characters

? causes the preceding to be optional (match 0 or 1 times), so `/\d?/` means “an optional digit”

Regular Expression Repetition



- Greedy vs Non-greedy matching
 - Greedy matching gets the longest results possible
 - Non-greedy matching gets the shortest possible
 - Let's say `$robot = 'The12thRobotIs2ndInLine'`
 - `$robot =~ /\w*\d+/`; (greedy)
 - Matches `The12thRobotIs2`
 - Maximizes the length of `\w`
 - `$robot =~ /\w*?\d+/`; (non-greedy)
 - *Add a '?' to a repetition to make it non-greedy!*
 - Matches `The12`
 - Minimizes the length of `\w`

Regular Expression Repetition



- Greedy vs Non-greedy matching
 - Suppose `$txt = 'something is so cool'`
 - `$txt =~ /something/;`
 - Matches 'something'
 - `$txt =~ /so (something) ?/;`
 - Matches 'something' and the second 'so'
 - Parenthesis can be used for **grouping** (e.g. being modified by '?') and **capture** (covered later)

Regular Expression Real Life Examples



- Using what you've learned so far, you can...
- Validate an email address (note: regex below is a little oversimplified)
 - `$email =~ /^[\\w\\.\\-]+@(\\w+\\.)* (\\w+) $/`
- Determine if log entry includes an IPv4 address
 - `/\\d{1,3}\\.\\d{1,3}\\.\\d{1,3}\\.\\d{1,3}/`
- Regular expressions can be hard to write and even harder to read
- Two techniques can help:
 - Languages have various 'verbose' or 'extended' modes so that a regex can be multiple lines, include comments, etc.
 - Can use an interactive regex editor such as <http://regex101.com/>

Regex101 example

- The IP address example

regular expressions 101 @regex101 donate contact bug reports & feedback wiki

REGULAR EXPRESSION v1 5 matches (~0ms)

`/\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}/ig`

TEST STRING SWITCH TO UNIT TESTS

```
Log start|
Initializing kernel
Updating software to version 21.3.1p6
Flushing memcache
Attempted login as root from 152.34.6.3
Attempted login as root from 152.34.6.3
Attempted login as root from 152.34.6.3
Attempted login as root from 152.34.6.3
Login accepted as root from 152.34.6.3
Rebooting
```

EXPLANATION

- `/\d{1,3}\.\d{1,3}\.\d{1,3}\.\d{1,3}/ig`
 - `\d{1,3}` matches a digit (equal to `[0-9]`)
{1,3} Quantifier — Matches between 1 and 3 times, as many times as possible, giving back as needed (*greedy*)
 - `\.` matches the character `.` literally (case insensitive)
 - `\d{1,3}` matches a digit (equal to `[0-9]`)
{1,3} Quantifier — Matches between 1 and 3 times, as many times as possible, giving back as needed (*greedy*)
 - `\.` matches the character `.` literally (case insensitive)
 - `\d{1,3}` matches a digit (equal to `[0-9]`)
{1,3} Quantifier — Matches between 1 and 3 times, as many times as possible, giving back as needed (*greedy*)
 - `\.` matches the character `.` literally (case insensitive)
 - `\d{1,3}` matches a digit (equal to `[0-9]`)
{1,3} Quantifier — Matches between 1 and 3 times, as many times as possible, giving back as needed (*greedy*)
 - Global pattern flags**
 - i modifier:** insensitive. Case insensitive match (ignores case of `[a-zA-Z]`)
 - g modifier:** global. All matches (don't return after first match)

Alternation



- Alternation allows multiple possibilities

- Let `$story = 'He went to get his mother'`

```
$story =~ /^ (He|She) \b.*?\b (his|her) \b.*? (mother|father|brother|sister|dog) /;
```

- Also matches 'She punched her fat brother'

- Make sure the grouping is correct!

```
$ans =~ /^(true|false)$/
```

- Matches only 'true' or 'false'

```
$ans =~ /^true|false$/ (same as /^(^true|false$)/)
```

- Matches 'true never' or 'not really false'

Grouping for Backreferences

- **Backreferences** (also known as **capture groups**)
 - We want to know what the expression finally ended up matching
 - Parenthesis give you **backreferences** let you see what was matched
 - Can be used *after* the expression has evaluated or even *inside* the expression itself!
 - Handled differently in different languages
 - Numbered from left to right, starting at 1

Grouping for Backreferences

- Perl backreferences

- Used inside the expression

- `$txt =~ /\b(\w+)\s+\1\b/`

- Finds any duplicated word, must use `\1` here (true in most languages)

- Used after the expression

- `$class =~ /(.+?) - (\d+) /`

- The first word between hyphens is stored in the Perl variable `$1` (not `\1`) and the number goes in `$2`. (This part varies between languages)

- `print "I am in class $1, section $2";`

- Equivalent Python:

```
import re
cls = "ECE560-02"
m = re.match(r'(.+?) - (\d+)', cls)
print "I'm in class "+m.group(1)+" , section "+m.group(2)
```

Example: Email Headers

- Here are some email headers.

Date: Sep 15, 2018, 5:15 PM

X-Sieve: CMU Sieve 2.3

X-DSPAM-Result: Innocent

X-DSPAM-Confidence: 0.8475

X-Content-Type-Message-Body: text/plain

- Let's write a regex to just match just the X- ones:

`/X-.*:.* /`

Using [Regex101.com](https://www.regex101.com) to understand this

The screenshot shows the Regex101.com interface. At the top, the site name "regular expressions 101" is on the left, and navigation links for "@regex101", "donate", "contact", "bug reports & feedback", and "wiki" are on the right. The main area is divided into three sections: "REGULAR EXPRESSION", "TEST STRING", and "EXPLANATION".

REGULAR EXPRESSION: The input field contains the regex `/ X-.*: .* / gm`. A status bar indicates "4 matches, 73 steps (~0ms)".

TEST STRING: The test string is a block of text with several lines highlighted in blue, corresponding to the matches found. The text is: "Date: Sep 15, 2018, 5:15 PM", "X-Sieve: CMU Sieve 2.3", "X-DSPAM-Result: Innocent", "X-DSPAM-Confidence: 0.8475", and "X-Content-Type-Message-Body: text/plain". A "SWITCH TO UNIT TESTS" button is visible.

EXPLANATION: This section provides a detailed breakdown of the regex components. It lists the following:

- `X-` matches the characters `X-` literally (case sensitive).
- `.*` matches any character (except for line terminators) **Quantifier** — Matches between **zero** and **unlimited** times, as many times as possible, giving back as needed (**greedy**).
- `:` matches the characters `:` literally (case sensitive).
- `.*` matches any character (except for line terminators) **Quantifier** — Matches between **zero** and **unlimited** times, as many times as possible, giving back as needed (**greedy**).
- Global pattern flags:**
 - `g` **modifier:** global. All matches (don't return after first match)
 - `m` **modifier:** multi line. Causes `^` and `$` to match the

MATCH INFORMATION: This section lists the four matches found in the test string. Each match is shown with its "Full match" and the corresponding text from the test string:

- Match 1: Full match `28-50 `X-Sieve: CMU Sieve 2.3``
- Match 2: Full match `51-75 `X-DSPAM-Result: Innocent``
- Match 3: Full match `76-102 `X-DSPAM-Confidence: 0.8475``
- Match 4: Full match `103-142 `X-Content-Type-Message-Body: text/pl ain``

QUICK REFERENCE: This section is currently empty.

Example: Email Headers

Capturing name and value

- We still have these email headers

```
Date: Sep 15, 2018, 5:15 PM
X-Sieve: CMU Sieve 2.3
X-DSPAM-Result: Innocent
X-DSPAM-Confidence: 0.8475
X-Content-Type-Message-Body: text/plain
```
- Let's amend our regex to capture the NAME and VALUE.

`/(X-.*): (.*)/`

What if we want to PARSE those headers?

- Parenthesis used for **capture** of part of a match

The screenshot shows a web-based regular expression testing tool. The interface is divided into several sections:

- REGULAR EXPRESSION:** The pattern `/ (X-.*): (.*) / gm` is entered in the input field.
- TEST STRING:** A sample email header is provided:

```
Date: Sep 15, 2018, 5:15 PM
X-Sieve: CMU Sieve 2.3
X-DSPAM-Result: Innocent
X-DSPAM-Confidence: 0.8475
X-Content-Type-Message-Body: text/plain
```
- EXPLANATION:** A detailed breakdown of the regex components:
 - `/`: matches the characters `/` literally (case sensitive)
 - `(X-.*):`: 1st Capturing Group
 - `X-`: matches the characters `X-` literally (case sensitive)
 - `.*`: matches any character (except for line terminators)
 - `*`: Quantifier — Matches between zero and unlimited times, as many times as possible, giving back as needed (greedy)
 - `:`: matches the characters `:` literally (case sensitive)
 - `(.*)`: 2nd Capturing Group
 - `.*`: matches any character (except for line terminators)
 - `*`: Quantifier — Matches between zero and unlimited times, as many times as possible, giving back as needed (greedy)
- MATCH INFORMATION:** A table showing the results of the matches:

Match	Full match	Group 1.	Group 2.
Match 1	28-50 `X-Sieve: CMU Sieve 2.3`	28-35 `X-Sieve`	37-50 `CMU Sieve 2.3`
Match 2	51-75 `X-DSPAM-Result: Innocent`	51-65 `X-DSPAM-Result`	67-75 `Innocent`
Match 3	76-102 `X-DSPAM-Confidence: 0.8475`	76-94 `X-DSPAM-Confidence`	96-102 `0.8475`
- QUICK REFERENCE:** A section for additional resources.

Refining a regex (1)

- What if our content includes some confusing non-headers mixed in?

The screenshot shows a web-based regular expression testing tool. The interface is divided into several sections:

- REGULAR EXPRESSION:** The pattern `(X-.*): (.*)` is entered. It shows 3 matches and 130 steps (~0ms).
- TEST STRING:** A text block containing:

```
Date: Sep 15, 2018, 5:15 PM
X-Sieve: CMU Sieve 2.3
X-Content-Type-Message-Body: text/plain

The X-Plane status is: behind schedule :-(
```
- EXPLANATION:** A detailed breakdown of the regex components:
 - 1st Capturing Group (X-.*):**
 - `X-` matches the characters `X-` literally (case sensitive).
 - `.*` matches any character (except for line terminators) between zero and unlimited times, as many times as possible, giving back as needed (greedy).
 - `:` matches the character `:` literally (case sensitive).
 - 2nd Capturing Group (.*):**
 - `.*` matches any character (except for line terminators) between zero and unlimited times, as many times as possible, giving back as needed (greedy).
- MATCH INFORMATION:** A table showing the results of the three matches:

Match	Full match	Group 1	Group 2
Match 1	28-50 `X-Sieve: CMU Sieve 2.3`	28-35 `X-Sieve`	37-50 `CMU Sieve 2.3`
Match 2	51-90 `X-Content-Type-Message-Body: text/plain`	51-78 `X-Content-Type-Message-Body`	80-90 `text/plain`
Match 3	96-134 `X-Plane status is: behind schedule :-(`	96-113 `X-Plane status is`	
- QUICK REFERENCE:** A section for additional reference information.

Refining a regex (2)

- Make regex more specific so it *just* matches what we want:

$\wedge(X-\backslash S^*): (.*)$

Must be start of line \wedge Non-whitespace characters only $\backslash S^*$

The screenshot shows a regex testing interface. The 'REGULAR EXPRESSION' field contains `^(X-\S*): (.*)` with flags `/gm`. The 'TEST STRING' field contains the following text:

```
Date: Sep 15, 2018, 5:15 PM
X-Sieve: CMU Sieve 2.3
X-Content-Type-Message-Body: text/plain

The X-Plane status is: behind schedule :-(
```

The 'EXPLANATION' section details the components of the regex:

- `^`: asserts position at start of a line
- 1st Capturing Group** `(X-\S*)`:
 - `X`: matches the characters `X` literally (case sensitive)
 - `\S*`: matches any non-whitespace character (equal to `[\r\n\t\f\v]`)
 - Quantifier**: Matches between **zero** and **unlimited** times, as many times as possible, giving back as needed (*greedy*)
- `:`: matches the characters `:` literally (case sensitive)
- 2nd Capturing Group** `(.*)`:
 - `.`: matches any character (except for line terminators)
 - Quantifier**: Matches between **zero** and **unlimited** times, as many times as possible, giving

The 'MATCH INFORMATION' section shows the following results:

Match	Full match	Group 1.	Group 2.
Match 1	28-50 `X-Sieve: CMU Sieve 2.3`	28-35 `X-Sieve`	37-50 `CMU Sieve 2.3`
Match 2	51-90 `X-Content-Type-Message-Body: text/plain`	51-78 `X-Content-Type-Message-Body`	80-90 `text/plain`

Grouping *without* Backreferences



- Sometimes you just need to make a group
 - If important groups must be backreferenced, disable backreferencing for any unimportant groups
 - `$sentence =~ /(?:He|She) likes (\w+)\./;`
 - I don't care if it's a he or she
 - All I want to know is what he/she likes
 - Therefore I use `(?:)` to forgo the backreference
 - `$1` will contain that thing that he/she likes

Matching Modes

- Matching has different functional modes
 - In Perl, these are specified as letters after the regex.
 - `$name =~ /[a-z]+/i;`
 - `i` turns off case sensitivity
 - `$xml =~ /title="([\w]*)" . *keywords="([\w]*)"/s;`
 - `s` enables `.` to match newlines
 - `$report =~ /^ \s* Name: [\s\S]* ? The End. \s* $/m;`
 - `m` allows newlines between `^` and `$`
 - In Python, you pass an additional optional argument with [named constants](#) (either short like the above or with full names), e.g.:
 - `re.search(r' [a-z]+', name, re.I) # or re.IGNORECASE`

Regular Expression Substitution

- **Substitutions** simplify complex data modification
 - First part is a regex of what to find, second part is text to replace it
 - Backreferences can be included in replacement
 - For sophisticated work, most languages let you give a callback function so that the replacement can be programmatically generated for each match
- Perl replacement syntax
 - `$phone =~ s/\D//;`
 - Removes the first non-digit character in a phone number
(Leaving the replacement blank means “replace with nothing”, i.e. “delete”)
 - `$html =~ s/^(\\s*)/$1\\t/;`
 - Adds a tab to a line of HTML using backreference `$1`
- Python uses `re.sub()`

Substitutions Modes

- Substitutions have modes like matches (ignore case, multiline, etc.)
- Important one: Substitutions can be performed singly or **globally**
 - In Perl, use the **g** flag to force the expression to scan the entire string
 - `$phone =~ s/\D//g;`
 - Removes all non-digits in the phone number
 - In Python's `re.sub()` function, specify a **count** parameter to limit replacements (e.g. `count=1` for traditional “first match only” behavior)

Combining one-liners and regexes

- Remember this slide when I compared color output to plain?

Why bother?

- Making output visually distinctive can *greatly* accelerate a task!
- Tester for ECE650 root kit: which would you rather use?



```
use of file (/usr/share/doc/gcc-12/FAQs_help) for instructions...
loadfile [1]: use file for target 'asm2process' failed
load: *** [asm2process] [Error 1]
Warning: asm2proc is deprecated and is already loaded from a past run.
Warning: asm2proc will be removed in a future version.
load: Error: Module 'asm2proc' and its use
Auto-Fix successful: processing...
Warning: asm2process: no such file or directory
File #1 is: /etc/passwd
***
TESTING BEGINS
***
Test 1
test1 in: /etc/passwd
test1 1 (/etc/passwd): pass
Test 2
test2 in: /etc/passwd
test2 1 (/etc/passwd): pass
Test 3
test3 in: /etc/passwd
test3 1 (/etc/passwd): pass
Test 4
test4 in: /etc/passwd
test4 1 (/etc/passwd): pass
Test 5
test5 in: /etc/passwd
test5 1 (/etc/passwd): pass
Test 6
test6 in: /etc/passwd
test6 1 (/etc/passwd): pass
Test 7
test7 in: /etc/passwd
test7 1 (/etc/passwd): pass
Test 8
test8 in: /etc/passwd
test8 1 (/etc/passwd): pass
Test 9
test9 in: /etc/passwd
test9 1 (/etc/passwd): pass
Test 10
test10 in: /etc/passwd
test10 1 (/etc/passwd): pass
Test 11
test11 in: /etc/passwd
test11 1 (/etc/passwd): pass
Test 12
test12 in: /etc/passwd
test12 1 (/etc/passwd): pass
Test 13
test13 in: /etc/passwd
test13 1 (/etc/passwd): pass
Test 14
test14 in: /etc/passwd
test14 1 (/etc/passwd): pass
Test 15
test15 in: /etc/passwd
test15 1 (/etc/passwd): pass
Test 16
test16 in: /etc/passwd
test16 1 (/etc/passwd): pass
Test 17
test17 in: /etc/passwd
test17 1 (/etc/passwd): pass
Test 18
test18 in: /etc/passwd
test18 1 (/etc/passwd): pass
Test 19
test19 in: /etc/passwd
test19 1 (/etc/passwd): pass
Test 20
test20 in: /etc/passwd
test20 1 (/etc/passwd): pass
Test 21
test21 in: /etc/passwd
test21 1 (/etc/passwd): pass
Test 22
test22 in: /etc/passwd
test22 1 (/etc/passwd): pass
Test 23
test23 in: /etc/passwd
test23 1 (/etc/passwd): pass
Test 24
test24 in: /etc/passwd
test24 1 (/etc/passwd): pass
Test 25
test25 in: /etc/passwd
test25 1 (/etc/passwd): pass
Test 26
test26 in: /etc/passwd
test26 1 (/etc/passwd): pass
Test 27
test27 in: /etc/passwd
test27 1 (/etc/passwd): pass
Test 28
test28 in: /etc/passwd
test28 1 (/etc/passwd): pass
Test 29
test29 in: /etc/passwd
test29 1 (/etc/passwd): pass
Test 30
test30 in: /etc/passwd
test30 1 (/etc/passwd): pass
Test 31
test31 in: /etc/passwd
test31 1 (/etc/passwd): pass
Test 32
test32 in: /etc/passwd
test32 1 (/etc/passwd): pass
Test 33
test33 in: /etc/passwd
test33 1 (/etc/passwd): pass
Test 34
test34 in: /etc/passwd
test34 1 (/etc/passwd): pass
Test 35
test35 in: /etc/passwd
test35 1 (/etc/passwd): pass
Test 36
test36 in: /etc/passwd
test36 1 (/etc/passwd): pass
Test 37
test37 in: /etc/passwd
test37 1 (/etc/passwd): pass
Test 38
test38 in: /etc/passwd
test38 1 (/etc/passwd): pass
Test 39
test39 in: /etc/passwd
test39 1 (/etc/passwd): pass
Test 40
test40 in: /etc/passwd
test40 1 (/etc/passwd): pass
Test 41
test41 in: /etc/passwd
test41 1 (/etc/passwd): pass
Test 42
test42 in: /etc/passwd
test42 1 (/etc/passwd): pass
Test 43
test43 in: /etc/passwd
test43 1 (/etc/passwd): pass
Test 44
test44 in: /etc/passwd
test44 1 (/etc/passwd): pass
Test 45
test45 in: /etc/passwd
test45 1 (/etc/passwd): pass
Test 46
test46 in: /etc/passwd
test46 1 (/etc/passwd): pass
Test 47
test47 in: /etc/passwd
test47 1 (/etc/passwd): pass
Test 48
test48 in: /etc/passwd
test48 1 (/etc/passwd): pass
Test 49
test49 in: /etc/passwd
test49 1 (/etc/passwd): pass
Test 50
test50 in: /etc/passwd
test50 1 (/etc/passwd): pass
Test 51
test51 in: /etc/passwd
test51 1 (/etc/passwd): pass
Test 52
test52 in: /etc/passwd
test52 1 (/etc/passwd): pass
Test 53
test53 in: /etc/passwd
test53 1 (/etc/passwd): pass
Test 54
test54 in: /etc/passwd
test54 1 (/etc/passwd): pass
Test 55
test55 in: /etc/passwd
test55 1 (/etc/passwd): pass
Test 56
test56 in: /etc/passwd
test56 1 (/etc/passwd): pass
Test 57
test57 in: /etc/passwd
test57 1 (/etc/passwd): pass
Test 58
test58 in: /etc/passwd
test58 1 (/etc/passwd): pass
Test 59
test59 in: /etc/passwd
test59 1 (/etc/passwd): pass
Test 60
test60 in: /etc/passwd
test60 1 (/etc/passwd): pass
Test 61
test61 in: /etc/passwd
test61 1 (/etc/passwd): pass
Test 62
test62 in: /etc/passwd
test62 1 (/etc/passwd): pass
Test 63
test63 in: /etc/passwd
test63 1 (/etc/passwd): pass
Test 64
test64 in: /etc/passwd
test64 1 (/etc/passwd): pass
Test 65
test65 in: /etc/passwd
test65 1 (/etc/passwd): pass
Test 66
test66 in: /etc/passwd
test66 1 (/etc/passwd): pass
Test 67
test67 in: /etc/passwd
test67 1 (/etc/passwd): pass
Test 68
test68 in: /etc/passwd
test68 1 (/etc/passwd): pass
Test 69
test69 in: /etc/passwd
test69 1 (/etc/passwd): pass
Test 70
test70 in: /etc/passwd
test70 1 (/etc/passwd): pass
Test 71
test71 in: /etc/passwd
test71 1 (/etc/passwd): pass
Test 72
test72 in: /etc/passwd
test72 1 (/etc/passwd): pass
Test 73
test73 in: /etc/passwd
test73 1 (/etc/passwd): pass
Test 74
test74 in: /etc/passwd
test74 1 (/etc/passwd): pass
Test 75
test75 in: /etc/passwd
test75 1 (/etc/passwd): pass
Test 76
test76 in: /etc/passwd
test76 1 (/etc/passwd): pass
Test 77
test77 in: /etc/passwd
test77 1 (/etc/passwd): pass
Test 78
test78 in: /etc/passwd
test78 1 (/etc/passwd): pass
Test 79
test79 in: /etc/passwd
test79 1 (/etc/passwd): pass
Test 80
test80 in: /etc/passwd
test80 1 (/etc/passwd): pass
Test 81
test81 in: /etc/passwd
test81 1 (/etc/passwd): pass
Test 82
test82 in: /etc/passwd
test82 1 (/etc/passwd): pass
Test 83
test83 in: /etc/passwd
test83 1 (/etc/passwd): pass
Test 84
test84 in: /etc/passwd
test84 1 (/etc/passwd): pass
Test 85
test85 in: /etc/passwd
test85 1 (/etc/passwd): pass
Test 86
test86 in: /etc/passwd
test86 1 (/etc/passwd): pass
Test 87
test87 in: /etc/passwd
test87 1 (/etc/passwd): pass
Test 88
test88 in: /etc/passwd
test88 1 (/etc/passwd): pass
Test 89
test89 in: /etc/passwd
test89 1 (/etc/passwd): pass
Test 90
test90 in: /etc/passwd
test90 1 (/etc/passwd): pass
Test 91
test91 in: /etc/passwd
test91 1 (/etc/passwd): pass
Test 92
test92 in: /etc/passwd
test92 1 (/etc/passwd): pass
Test 93
test93 in: /etc/passwd
test93 1 (/etc/passwd): pass
Test 94
test94 in: /etc/passwd
test94 1 (/etc/passwd): pass
Test 95
test95 in: /etc/passwd
test95 1 (/etc/passwd): pass
Test 96
test96 in: /etc/passwd
test96 1 (/etc/passwd): pass
Test 97
test97 in: /etc/passwd
test97 1 (/etc/passwd): pass
Test 98
test98 in: /etc/passwd
test98 1 (/etc/passwd): pass
Test 99
test99 in: /etc/passwd
test99 1 (/etc/passwd): pass
Test 100
test100 in: /etc/passwd
test100 1 (/etc/passwd): pass
```

- Did I write a whole separate script that omitted colors? NO!

```
$ perl -pe 's/\e.*?m//g' orig > plain
```

- p means “read one line at a time like -n, but print the line afterwards”
- \e is the escape character.

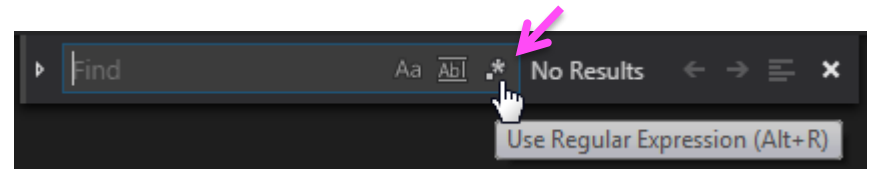
Regular Expression Quick Guide

<code>^</code>	Matches the beginning of a line
<code>\$</code>	Matches the end of the line
<code>.</code>	Matches any character (except newline, unless you give an option)
<code>\s</code>	Matches whitespace
<code>\S</code>	Matches any non-whitespace character
<code>\w</code>	Matches a “word-like” character (letters/numbers/underscore)
<code>\d</code>	Matches a decimal digit (0-9)
<code>\b</code>	Matches a word boundary
<code>?</code>	Makes a character or group <i>optional</i> (appears zero or one times)
<code>*</code>	Repeats a character or group zero or more times
<code>*?</code>	Repeats a character zero or more times (non-greedy)
<code>+</code>	Repeats a character one or more times
<code>+?</code>	Repeats a character one or more times (non-greedy)
<code> </code>	Alternation – allows either/or. Usually used with parens: <code>(this) (that)</code>
<code>[aeiou]</code>	Matches a single character in the listed set
<code>[^XYZ]</code>	Matches a single character not in the listed set
<code>[a-z0-9]</code>	The set of characters can include a range
<code>(and)</code>	Indicates a group (used to capture part of a match <i>or</i> group stuff for modifiers)

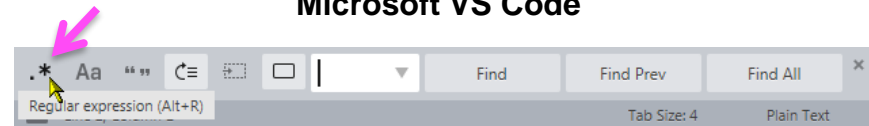
A [more complete quick-ref guide is here](#) and linked on the course site. See also the [Python re module docs](#).

Where can I use regexes?

- Obviously, in Perl and Python
- Also: Javascript, Java, .NET, PHP, R, C/C++, PowerShell, Ruby
- Also: your text editor



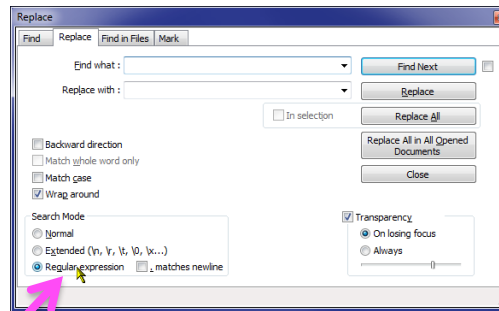
Microsoft VS Code



Sublime Text

- Tons of shell tools:

- **grep** *-P*
- **sed** *(just has em)*
- **awk** *(just has em)*
- **less** *(Press /)*



Notepad++

No screenshot because I ain't launching that thing but you can type "C-M-s" for regex search (whatever that means)

emacs

`/(regex) goes [here]`

vi and vim (press /)

- Everyone cool is using regexes! Don't get left behind!!!!

References for learning more about regexs

- Regex editor, code generator, and community database of regexs

- <http://regex101.com/>



It's-a good site!

- Tutorials for various programming languages

- <http://www.regular-expressions.info/>

- Python in-depth docs

- <https://docs.python.org/3/library/re.html>

- Perl in-depth docs

- <https://perldoc.perl.org/perlreref.html>

Manipulating tabular data

Hey, how about Excel? That thing's cool, right?

- Terminals are nice, but did you know: **GUIs exist?**
- Some tasks benefit from non-terminal interface
- Example: tabular data wants to be in a **spreadsheet**

- Let's cover some quick tips on (ab)using Excel (or Google Sheets)

Data in/out

- File format for getting in/out of Excel:
Comma-Separated Values (CSV)
 - Trivial for simple data: bob,2,19
 - If you have commas in data, enclose in quotes: "Jimmy, PhD",4,50
 - If you have quotes, double them up: "This is a ""quote""",7,94
 - Save with “.csv” extension and Excel loads it right up
 - Can generate well enough with simple commands
 - Can use common libraries to do everything “right” (quoting, etc.); e.g. Python has a built-in **csv** module
- For fast stuff, can just use the **clipboard**
 - Often quick just to copy/paste instead of making actual files
 - Format for spreadsheet <-> plaintext via clipboard is **tab-separated**
 - For a single column of data, there’s no tabs – it’s just in lines!

Formulas

- Spreadsheet formulas are outside of our scope – if you aren't familiar, you **need** to learn them
- One thing to add: you can do string manipulation as well as math
 - & is the concatenation operator
 - TEXT() can format numbers in arbitrary formats

Auto-filter

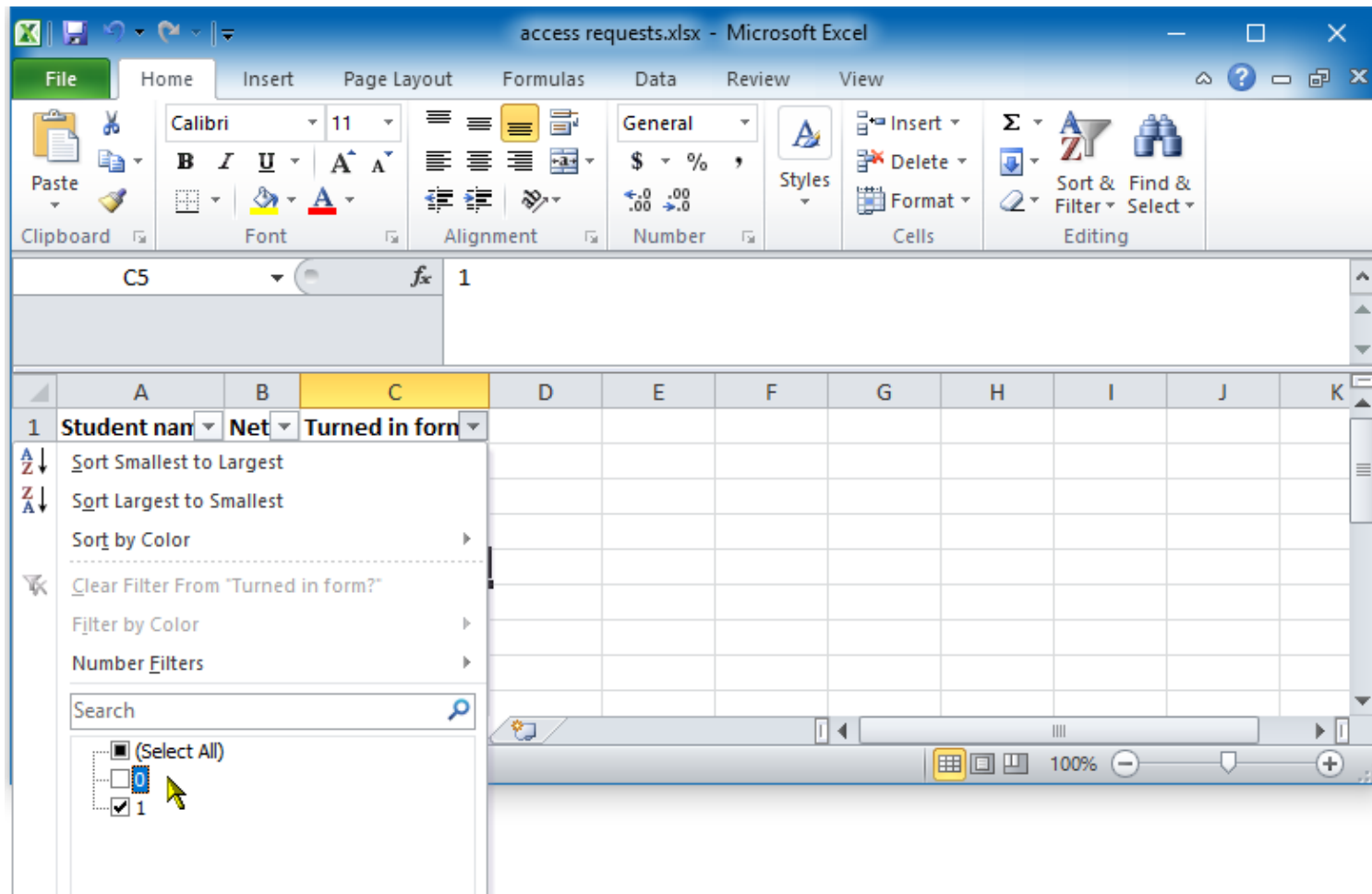
- Take a sheet, make sure it has headers, highlight your data, turn on **auto filter** → Bam! **instant sort/filter controls**.
 - Example: Requesting badge access for some students.

The screenshot shows the Microsoft Excel interface with the 'Sort & Filter' dropdown menu open. The menu options are: Sort Smallest to Largest, Sort Largest to Smallest, Custom Sort..., Filter (highlighted), Clear, and Reapply. The data table below is:

	A	B	C	D	E	F	G	H
1	Student name	NetID	Turned in form?					
2	Alice	ag11	1					
3	Bob	bc22	1					
4	Charlie	cz33	0					
5	Denise	dp44	1					
6	Erwin	ej55	0					
7	Frank	fo66	1					

Auto-filter

- Take a sheet, make sure it has headers, highlight your data, turn on **auto filter** → Bam! **instant sort/filter controls**.
 - Example: Requesting badge access for some students.



Auto-filter

- Take a sheet, make sure it has headers, highlight your data, turn on **auto filter** → Bam! **instant sort/filter controls**.
 - Example: Requesting badge access for some students.

The screenshot shows the Microsoft Excel interface with the 'access requests.xlsx' file open. The ribbon is set to 'Home', and the 'Filter' button is visible in the 'Editing' group. The data table is filtered to show only rows where the 'Net' column contains the values 'ag11', 'bc22', 'dp44', and 'fo66'. A red callout arrow points to the filtered data with the text: 'I paste this NetID list into my email to IT to get students badge access.'

	A	B	C	D	E	F	G	H	I	J	K
1	Student name	Net	Turned in form								
2	Alice	ag11	1								
3	Bob	bc22	1								
5	Denise	dp44	1								
7	Frank	fo66	1								
8											
9											
10											
11											

Ready 4 of 6 records found Count: 4 100%

Putting it all together

Example data manipulation task: Planning Homework 1

- **What I have:** The Homework 1 draft writeup
- **My goal:** Plan out point allocation for questions
- **What I want:** Table of question number, topic, and points

answering the questions, including classmates and the textbook. Please use authoritative sources like RFCs, ISOs, NIST SPs, man pages, etc. for your references.

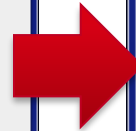
This assignment is adapted from material by Samuel Carter (NCSU).

Question 1: Internet Standards (3 points)

In Chapter 0 and Appendix C of the course textbook, we begin to look at technology standards and standard-setting organizations. Various organizations are involved in the development of standards related to data and computer communications. It is important to understand who the major organizations are and the standards they are responsible for. These standards bodies will be heavily referenced throughout the course and can be useful references when trying to understand different security technologies. Give a short description of each organization, its key primary responsibilities around standards, and an example a security-related standard that it has developed.

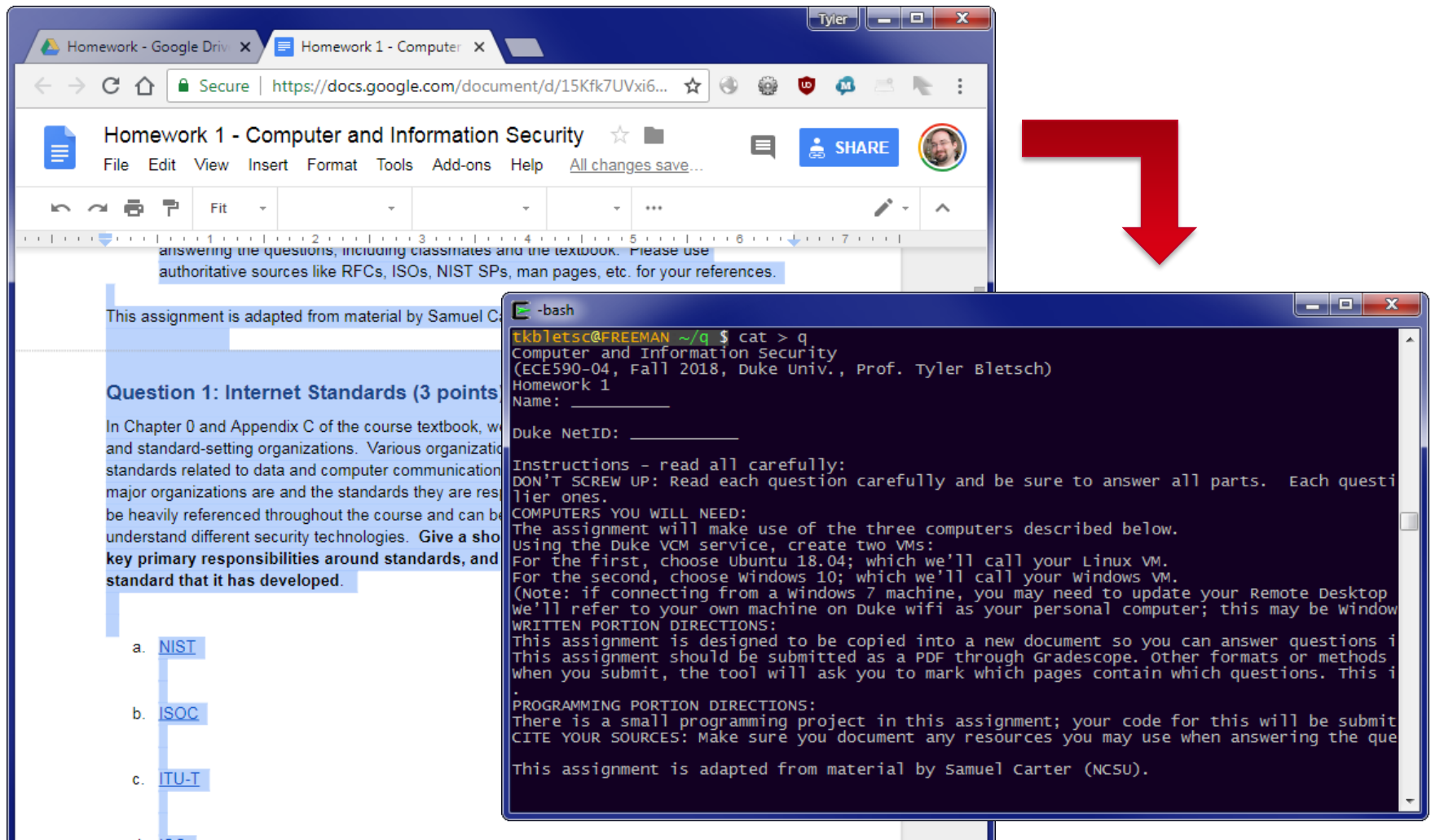
- [NIST](#)
- [ISOC](#)
- [ITU-T](#)
- [ISO](#)

	A	B	C	D
1	1	Internet Standards	3	
2	2	A Model for Computer Security	7	
3	3	Threats and Attacks	5	
4	4	IP Addressing	12	
5	5	Physical Addresses	4	
6	6	Networking Protocols	4	
7	7	Ports	5	
8	8	DNS	3	
9	9	Network Traffic Analysis with Wireshark	7	
10	10	Network Traffic Analysis with TCPDump	2	
11	11	Network Mapping	5	
12	12	Ncat, Telnet, Netstat, and Sockets	2	
13	13	Banner Grabbing: Services Spilling Their Guts	5	
14	14	Networking Tools	6	
15	15	Tor Project: Anonymity Online	4	
16	16	AWK Programming Language	2	
17	17	MS05-30 Attack Script	3	
18	18	Cryptography Theory	3	
19	19	Analysis of LM Hashing Algorithm	2	
20	20	Bitlocker, FileVault, and LUKS	3	
21				



Planning Homework 1: Acquire source data

- Select all, copy
- In shell, run “`cat > q`” and paste (middle-click), then **Ctrl+D** for EOF



The image shows a Google Docs document titled "Homework 1 - Computer and Information Security" and a terminal window. A red arrow points from the document to the terminal. The terminal shows the command `cat > q` being executed, followed by the document's content being pasted into the file `q`.

Google Docs Content:

answering the questions, including classmates and the textbook. Please use authoritative sources like RFCs, ISOs, NIST SPs, man pages, etc. for your references.

This assignment is adapted from material by Samuel Carter (NCSU).

Question 1: Internet Standards (3 points)

In Chapter 0 and Appendix C of the course textbook, we discuss standard-setting organizations. Various organizational standards related to data and computer communication major organizations are and the standards they are responsible for. These standards are heavily referenced throughout the course and can be used to understand different security technologies. Give a short answer to the following questions, citing the key primary responsibilities around standards, and the standard that it has developed.

- [NIST](#)
- [ISOC](#)
- [ITU-T](#)

Planning Homework 1: Develop regex

```
-bash
tkblets@FREEMAN ~/q $ grep -E '^Question \d+' q
tkblets@FREEMAN ~/q $ grep -P '^Question \d+' q
Question 1: Internet Standards (3 points)
Question 2: A Model for Computer Security (7 points)
Question 3: Threats and Attacks (5 points)
Question 4: IP Addressing (12 points)
Question 5: Physical Addresses (4 points)
Question 6: Networking Protocols (4 points)
Question 7: Ports (5 points)
Question 8: DNS (3 points)
Question 9: Network Traffic Analysis with Wireshark (7 points)
Question 10: Network Traffic Analysis with TCPDump (2 points)
Question 11: Network Mapping (5 points)
Question 12: Ncat, Telnet, Netstat, and Sockets (2 points)
Question 13: Banner Grabbing: Services Spilling Their Guts (5 points)
Question 14: Networking Tools (6 points)
Question 15: Tor Project: Anonymity Online (4 points)
Question 16: AWK Programming Language (2 points)
Question 17: MS05-30 Attack Script (3 points)
Question 18: Cryptography Theory (3 points)
Question 19: Analysis of LM Hashing Algorithm (2 points)
Question 20: Bitlocker, FileVault, and LUKS (3 points)
tkblets@FREEMAN ~/q $ perl -e '/^Question \d+:.*\(\d+ point/' / and print "$1\t$2" q
> ^C
tkblets@FREEMAN ~/q $ perl -e '/^Question \d+:.*\(\d+ point/' / and print "$1\t$2" q
tkblets@FREEMAN ~/q $ echo perl -e '/^Question \d+:.*\(\d+ point/' / and print "$1\t$2" q
perl -e '/^Question \d+:.*\(\d+ point/' / and print "$1\t$2" q
tkblets@FREEMAN ~/q $ perl -ne '/^Question \d+:.*\(\d+ point/' / and print "$1\t$2" q
```

Oh, I must need a perl-level regex. Switch to -P

Looks good, let's switch to perl to capture fields.

Mismatched quotes so shell waited for more input, Ctrl+C.

These last commands didn't match anything - I realize I forgot to turn on 'read a file line by line' mode (-n)

Planning Homework 1: Debug regex

```
-bash
tkblets@FREEMAN ~/q $ perl -ne '/^Question \d+:.*(\d+ point/ and print "$1\t$2\n"' q

tkblets@FREEMAN ~/q $ perl -ne '/^Question (\d+):.*\((\d+) point/ and print "$1\t$2\n"' q
1      3
2      7
3      5
4     12
5      4
6      4
7      5
8      3
9      7
10     2
11     5
12     2
13     5
14     6
15     4
16     2
17     3
18     3
19     2
20     3

tkblets@FREEMAN ~/q $ perl -ne '/^Question (\d+): (.*) \((\d+) point/ and print "$1\t$2\n"' q
1  Internet Standards
2  A Model for Computer Security
3  Threats and Attacks
4  IP Addressing
5  Physical Addresses
6  Networking Protocols
7  Ports
```

So it's matching but no output? Oh, I forgot to capture the fields I wanted! Add capture parens.

Working, but I realize I want the topics too. Add those parens.

Forgot to include \$3 in output.

Planning Homework 1: Clean output

```
-bash
tkblets@FREEMAN ~/q $ perl -ne '/^Question (\d+): (.*) \\((\d+) point/ and print "$1\t$2\t$3\n"' q
1   Internet Standards      3
2   A Model for Computer Security  7
3   Threats and Attacks      5
4   IP Addressing           12
5   Physical Addresses       4
6   Networking Protocols     4
7   Ports                    5
8   DNS                      3
9   Network Traffic Analysis with Wireshark 7
10  Network Traffic Analysis with TCPDump  2
11  Network Mapping          5
12  Ncat, Telnet, Netstat, and Sockets      2
13  Banner Grabbing: Services Spilling Their Guts  5
14  Networking Tools         6
15  Tor Project: Anonymity online  4
16  AWK Programming Language      2
17  MS05-30 Attack Script        3
18  Cryptography Theory          3
19  Analysis of LM Hashing Algorithm      2
20  Bitlocker, Filevault, and LUKS        3
tkblets@FREEMAN ~/q $ perl -ne '/^Question (\d+): (.*) \\((\d+) point/ and print "$1\t$2\t$3\n"' q > out
tkblets@FREEMAN ~/q $ notepad++ out &
[1] 10080
```

Perfect, let's save it to a file now

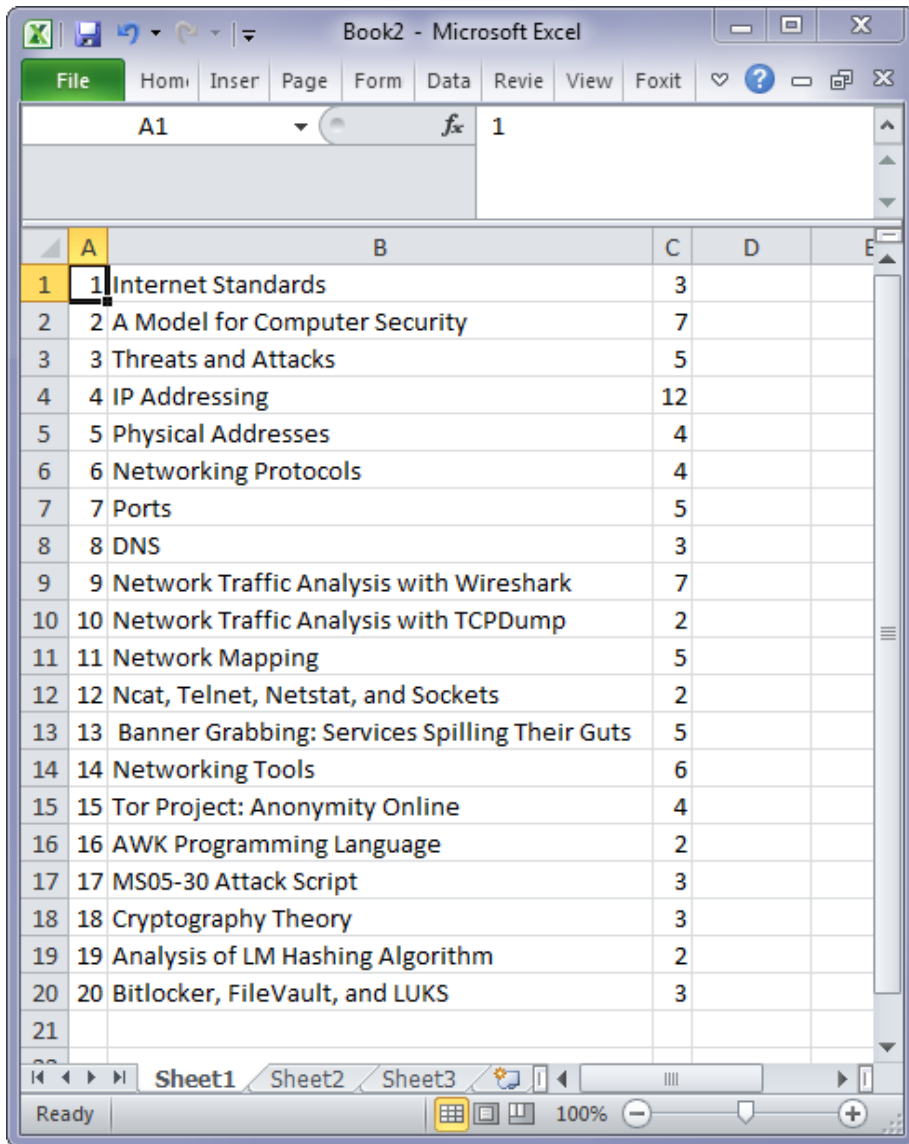
Open in my GUI editor

Appended >out

```
C:\cygwin64\home\tkblets\q\out - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
SCRATCH.TXT x shell notes.txt x out x
1 1 Internet Standards 3
2 2 A Model for Computer Security 7
3 3 Threats and Attacks 5
4 4 IP Addressing 12
5 5 Physical Addresses 4
6 6 Networking Protocols 4
7 7 Ports 5
8 8 DNS 3
9 9 Network Traffic Analysis with Wireshark 7
10 10 Network Traffic Analysis with TCPDump 2
11 11 Network Mapping 5
```

Why copy from text editor instead of shell? Shell will *render* those tabs as spaces for clipboard purposes; editor preserves them.

Planning Homework 1: Check results



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E
1	1	Internet Standards	3		
2	2	A Model for Computer Security	7		
3	3	Threats and Attacks	5		
4	4	IP Addressing	12		
5	5	Physical Addresses	4		
6	6	Networking Protocols	4		
7	7	Ports	5		
8	8	DNS	3		
9	9	Network Traffic Analysis with Wireshark	7		
10	10	Network Traffic Analysis with TCPDump	2		
11	11	Network Mapping	5		
12	12	Ncat, Telnet, Netstat, and Sockets	2		
13	13	Banner Grabbing: Services Spilling Their Guts	5		
14	14	Networking Tools	6		
15	15	Tor Project: Anonymity Online	4		
16	16	AWK Programming Language	2		
17	17	MS05-30 Attack Script	3		
18	18	Cryptography Theory	3		
19	19	Analysis of LM Hashing Algorithm	2		
20	20	Bitlocker, FileVault, and LUKS	3		
21					

- Paste into excel
- Resize columns
- Compare a few rows against document to confirm (never forget to check that your work actually got what you think it did!)
- I can immediately see Q13 has an extra space so I fix that in the doc
- Can consider and assign points accordingly.



Example task: Organizing PPTs

Gathering info

- Reviewing ECE651 course content, need to organize slides
- Have syllabus and downloaded content, want to put in order
- Can't fully automate: matching syllabus to filenames is fuzzy
- Excel can help:

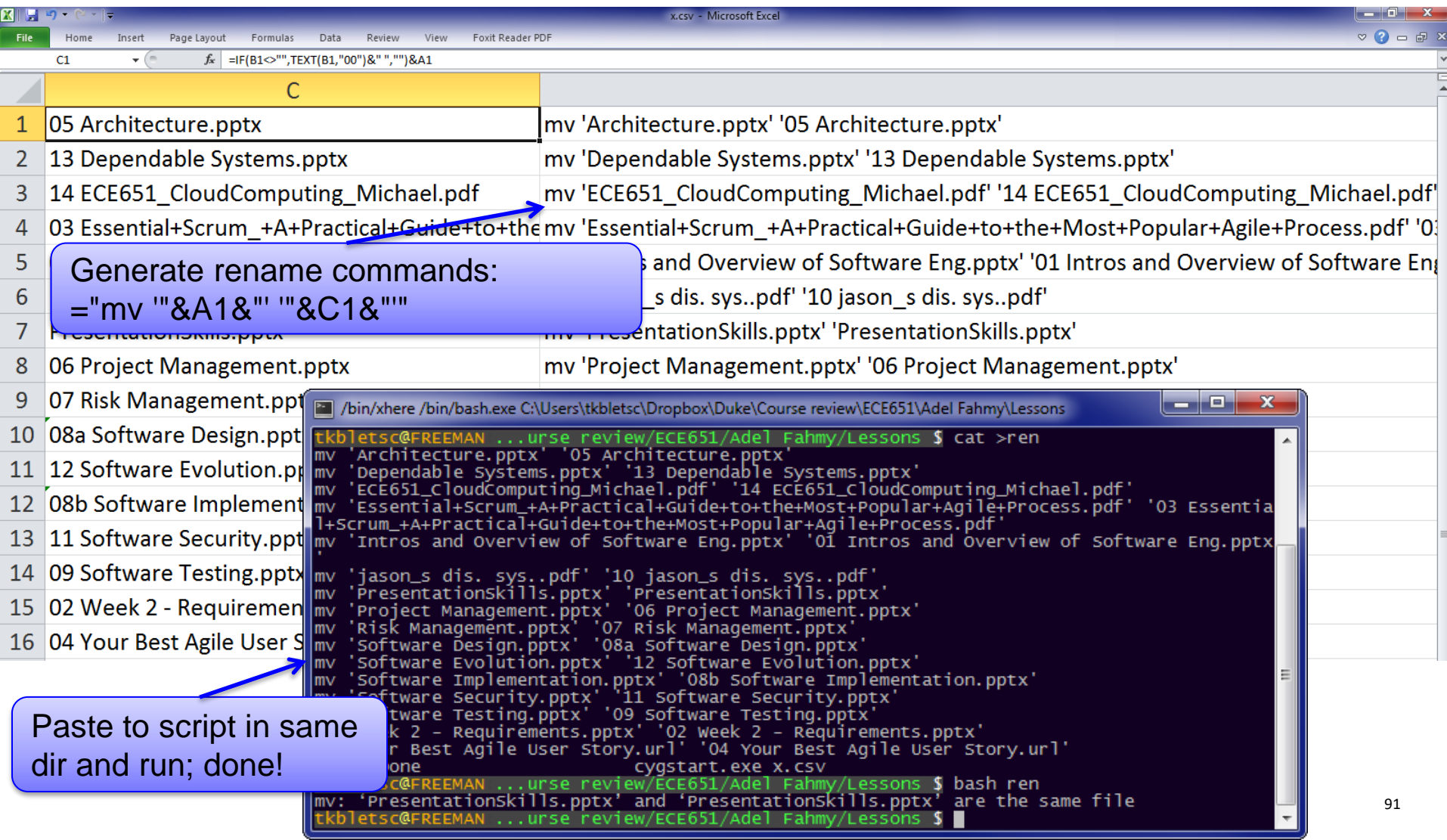
```
$ ls > x.csv
```

	A	B	
1	Architecture.pptx	5	05 Architecture.pptx
2	Dependable System	13	13 Dependable Systems.pptx
3	ECE651_CloudComp	14	14 ECE651_CloudComputing_Michael.p
4	Essential+Scrum_+A+Practical+Guide+to+the+most+popular+Agile+Process.pdf	3	03 Essential+Scrum_+A+Practical+Guid
5	Intros and Overview of Software Eng.pptx	1	01 Intros and Overview of Software Eng
6	jason_s dis. sys..pdf		10 jason_s dis. sys..pdf
7	PresentationSkills.pptx		PresentationSkills.pptx
8	Project Management.pptx	6	06 Project Management.pptx
9	Risk Management.pptx	7	07 Risk Management.pptx
10	Software Design.pptx	8a	08a Software Design.pptx
11	Software Evolution.pptx		re Evolution.pptx
12	Software Implementation.pptx		are Implementation.pptx
13	Software Security.pptx	11	11 Software Security.pptx

Example task: Organizing PPTs

Generating script and renaming

- We can even have Excel make our renaming shell script:



The screenshot shows a Microsoft Excel spreadsheet with a CSV file open. The spreadsheet has two columns: the first column contains file names, and the second column contains the corresponding rename commands. A blue callout box points to the formula in cell C1, which is `=IF(B1<>"",TEXT(B1,"00")&" " &"")&A1`. This formula concatenates the file name from column A with a zero-padded number from column B, separated by a space, to create the new filename for the rename command.

	C	
1	05 Architecture.pptx	mv 'Architecture.pptx' '05 Architecture.pptx'
2	13 Dependable Systems.pptx	mv 'Dependable Systems.pptx' '13 Dependable Systems.pptx'
3	14 ECE651_CloudComputing_Michael.pdf	mv 'ECE651_CloudComputing_Michael.pdf' '14 ECE651_CloudComputing_Michael.pdf'
4	03 Essential+Scrum_+A+Practical+Guide+to+the	mv 'Essential+Scrum_+A+Practical+Guide+to+the+Most+Popular+Agile+Process.pdf' '03 Essential+Scrum_+A+Practical+Guide+to+the+Most+Popular+Agile+Process.pdf'
5		and Overview of Software Eng.pptx' '01 Intros and Overview of Software Eng.pptx'
6		s dis. sys..pdf' '10 jason_s dis. sys..pdf'
7	PresentationSkills.pptx	mv 'PresentationSkills.pptx' 'PresentationSkills.pptx'
8	06 Project Management.pptx	mv 'Project Management.pptx' '06 Project Management.pptx'
9	07 Risk Management.pptx	
10	08a Software Design.pptx	
11	12 Software Evolution.pptx	
12	08b Software Implementation.pptx	
13	11 Software Security.pptx	
14	09 Software Testing.pptx	
15	02 Week 2 - Requirements.pptx	
16	04 Your Best Agile User Story.url	

Generate rename commands:
`= "mv "&A1&" "&C1&" "`

```
/bin/xhere /bin/bash.exe C:\Users\tkblets\Dropbox\Duke\Course review\ECE651\Adel Fahmy\Lessons
tkblets@FREEMAN ...urse review/ECE651/Adel Fahmy/Lessons $ cat >ren
mv 'Architecture.pptx' '05 Architecture.pptx'
mv 'Dependable Systems.pptx' '13 Dependable Systems.pptx'
mv 'ECE651_CloudComputing_Michael.pdf' '14 ECE651_CloudComputing_Michael.pdf'
mv 'Essential+Scrum_+A+Practical+Guide+to+the+Most+Popular+Agile+Process.pdf' '03 Essential+Scrum_+A+Practical+Guide+to+the+Most+Popular+Agile+Process.pdf'
mv 'Intros and Overview of Software Eng.pptx' '01 Intros and Overview of Software Eng.pptx'
mv 'jason_s dis. sys..pdf' '10 jason_s dis. sys..pdf'
mv 'PresentationSkills.pptx' 'PresentationSkills.pptx'
mv 'Project Management.pptx' '06 Project Management.pptx'
mv 'Risk Management.pptx' '07 Risk Management.pptx'
mv 'Software Design.pptx' '08a Software Design.pptx'
mv 'Software Evolution.pptx' '12 Software Evolution.pptx'
mv 'Software Implementation.pptx' '08b Software Implementation.pptx'
mv 'Software Security.pptx' '11 Software Security.pptx'
mv 'Software Testing.pptx' '09 Software Testing.pptx'
mv 'Week 2 - Requirements.pptx' '02 week 2 - Requirements.pptx'
mv 'Your Best Agile User Story.url' '04 Your Best Agile User Story.url'
one
cygstart.exe x.csv
tkblets@FREEMAN ...urse review/ECE651/Adel Fahmy/Lessons $ bash ren
mv: 'PresentationSkills.pptx' and 'PresentationSkills.pptx' are the same file
tkblets@FREEMAN ...urse review/ECE651/Adel Fahmy/Lessons $
```

Paste to script in same dir and run; done!

Conclusion

- Time it took to do this: 3 minutes
- Time it would have taken to do this manually: 6 minutes?
 - Odds of a typo or transcription error: way higher than automated way
- Time it would take to do this if I were learning the skills for the first time: way more than 6 minutes.

- So is it worth it to learn to automate?
 - **Yes: you learn once, benefit many times.** 😊
 - One of the sources of developer or sysadmin productivity!
- Keep doing things the dumb manual way because it's faster?
 - **You never improve!** 😞

Conclusion: PRACTICE THIS STUFF!!