## Network Administration Practice Homework 1: Python Scripts

weicc & blzhuang

### Requirements

- $\Box$  1-1 Web crawler (45%+5%)
  - Entry point: `nahw1-1\_{student\_ID}.py`
  - Login options
    - a) NCTU Portal (5% Bonus)
    - b) Course Registration System
- □ 1-2 Auth Log parser Count ssh login failed (45%)
  - Entry point: `nahw1-2\_{student\_ID}.py`
- ☐ Readme file (Readme.md) (10%)
  - List used Package / library for each part
  - Login by NCTU Portal or Course Selection System
- ☐ Due data: 2018/03/22 23:59
  - Upload `nahw1-{\$student\_ID}-{\$YYYYMMDD-HHMM}.tar` on New E3 (https://e3new.nctu.edu.tw)

### Readme File: Sample

```
# Network Adminstration - Homework 1
     ## Part I - Web crawler
     Login to NCTU portal and get schedule on course selection system.
     ### Dependency
     - Pillow
     - pytesseract
     BeautifulSoup
10
     - PrettyTable
11
12
13
     ## Part II - Auth log parser
14
15
     Parse auth.log file.
16
17
     ## Dependency
18
```

## 1-1: Web Crawler – Requirements (1/2)

- ☐ Input format
  - python script\_name username`
- ☐ Argument parser 5%
  - -h show usage
- ☐ Captcha Recognize 20%
  - Login by course selection system
  - Login by NCTU Portal
- ☐ Parse HTML & Print Table (one of following)
  - Can parse schedule as list from HTML 10%
  - Can parse schedule as list from HTML and print as table 20%
- ☐ Bonus 5% Login by NCTU Portal
  - Solve all kinds of captcha
  - Relay login session from NCTU Portal to course selection system

## 1-1: Web Crawler – Requirements (2/2)

- ☐ Only Python script allowed
  - Call shell script is not allowed
  - Call NodeJS script is not allowed
  - subprocess not allowed
  - `os.system` not allowed
- ☐ Python 3 only

:20~~14:10	           數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117		I Sup in TrowSoup C115 I.join(f	+	Dock	個別面   個別面   系統語	研究 EC015	上 上 上 上 上 上 上 上 上 上 上 上 上 上 上 上 上 上 上
:00~~07:50 :00~~08:50 :00~~09:50 :10~~11:00 :10~~12:00 :20~~13:10 :20~~14:10	           數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117	     網路安全(英文授課)E   網路安全(英文授課)E   網路安全(英文授課)E	I Sup in TrowSoup C115 I.join(f		Dock	個別面 素統 Nission (ontrol	研究 EC015	<ul><li>●性與陽私權</li></ul>
:00~~08:50 :00~~09:50 :10~~11:00 :10~~12:00 :20~~13:10 :20~~14:10	         數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117	     網路安全(英文授課)E   網路安全(英文授課)E   網路安全(英文授課)E	I Sup in TrowSoup C115 I.join(f		Dock	系统。 Mission (ontrol	福言與地區 安全	性與隱私權
:00~~09:50 :10~~11:00 :10~~12:00 :20~~13:10 :20~~14:10	       數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117	   網路安全(英文授課)E   網路安全(英文授課)E   網路安全(英文授課)E	C115 / I		Dock	Mission Ontrol		全性與隱私權
:10~~12:00 :20~~13:10 :20~~14:10	   數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117	│ 網路安全(英文授課)E │ 網路安全(英文授課)E	C115 / I		Dock	Mission (ontrol		全性與隱私權
:20~~13:10 :20~~14:10	   數位遊戲與學習ED117   數位遊戲與學習ED117   數位遊戲與學習ED117	│ 網路安全(英文授課)E │ 網路安全(英文授課)E	C115 / I		Dock	Mission Control		全性與隱私權
:20~~14:10	│ 數 位 遊 戲 與 學 習 ED117 │ 數 位 遊 戲 與 學 習 ED117		115		Dock	Mission Control		生性與隱私權
	數位遊戲與學習ED117	77			5 ĬÏ	Control		E IT >< MOUTH
:20~~15:10		I 78						
				Ⅰ 軟體測試CS10	5 I	1		1
:30~~16:20		79 pTable = Pretty	Table(table[0	Ⅰ 軟體測試CS10	5 I			1
:30~~17:20		80						
:30~~18:20		81 for x in range	1, ler(table)		671.60	38.63		n 主 総 的
		82 pTable.add_	row(table[x])			海讯 :		掃描器
		83				!		!
		84 print(pTable)		Ⅰ 網路管理實務EC	122	!		!
:3	30~~20:20 30~~21:20	00~~19:20   00~~20:20   00~~21:20   00~~22:20	80~~20:20	10~~20:20	60~~20:20	Max	80~~20:20	60~~20:20

### 1-1: Web Crawler – Hint: Argument Parser Sample

#### 1-1: Web Crawler – Hint

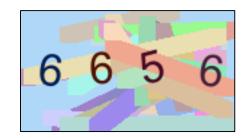
□ 善用Browser Development tool → Network

```
chkpas.php
                                          AV OTC NOI DSP COR"
                                          Pragma: no-cache
login.php
                                          Server: Microsoft-IIS/10.0
jquery-1.8.3.min.js
                                          Strict-Transport-Security: max-age=157680000
is.portal.login.LoadSystemName.js
                                          X-Powered-By: PHP/5.2.9
pic.php
                                      ▼ Request Headers
                                                            view source
                                          Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/web
m pic.php
                                          p,image/apng,*/*;q=0.8
refresh.gif
                                          Accept-Encoding: gzip, deflate, br
ga.js
                                          Accept-Language: zh-TW,zh;q=0.9,en-US;q=0.8,en;q=0.7
                                          Cache-Control: no-cache
nctuloginbg.jpg
                                          Connection: keep-alive
data:application/ja...
                                          Content-Length: 99
semi-transparent.png
                                          Content-Type: application/x-www-form-urlencoded
fun.inc.relay_conf.js.portal.login.LoadSy...
                                          Cookie: PHPSESSID=91lhhubs09uurv2r01vg4v8ie5; PHPSESSID=86qe492jce9v4fts
                                          fk3kbjimj6
                                          DNT: 1
                                          Host: portal.nctu.edu.tw
                                          Origin: https://portal.nctu.edu.tw
                                          Pragma: no-cache
                                          Referer: https://portal.nctu.edu.tw/portal/login.php
                                          Upgrade-Insecure-Requests: 1
                                          User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_13_3) AppleWebKit/5
                                          37.36 (KHTML, like Gecko) Chrome/64.0.3282.186 Safari/537.36
                                       ▼ Form Data
                                                                    view URL encoded
                                                      view source
                                          username: 0656087
                                          Submit2: 登入(Login)
                                          pwdtvpe: static
                                          password: testing
                                          seccode: 0795
12 requests | 102 KB transferred | Finish: 29...
```

## 1-1: Web Crawler – Hint: Captcha Recognize Login Option (a): NCTU Portal

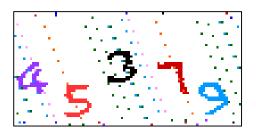






- ☐ Preprocess
  - Convert image to grayscale
  - Adjust brightness and contrast
  - Packages may be used
    - > Pillow
- ☐ Recognize
  - Pytesseract (Python interface for Tesseract OCR By google)

## 1-1: Web Crawler – Hint: Captcha Recognize Login Option (b): Course Registration System



#### ☐ Preprocess

- Convert image to grayscale
- Remove salt and pepper noise
- Packages may be used
  - > Pillow
  - OpenCV

#### ☐ Recognize

Pytesseract (Python interface for Tesseract OCR By google)

### 1-2: Auth log parser – Requirements (1/6)

- ☐ With following option
  - -h show usage 5%
  - -u sort by user 5%
  - -after filter log after special date 10%
  - -before filter log before special date 10%
  - -n show only the user of most #-th times 5%
  - -t show only the user of attacking equal or more than # times 5%
  - -r sort in reverse order 5%
- ☐ Only Python script allowed
  - Call shell script is not allowed
  - Call NodeJS script is not allowed
  - `subprocess` not allowed
  - `os.system` not allowed
- ☐ Python 3 only

# 1-2: Auth log parser – Requirements (2/6) show help

```
→ 2018NA=homework=1 git:(master) 1× python3 nahw1=2_0656087.pyr=h doc
usage: nahw1-2_0656087.py [-h] [-u] [-after AFTER] [-before BEFORE] [-n N]
                         [-t T] [-r]
                         filename
Auth log parser.
positional arguments:
 filename Log file path.
optional arguments:
 -h, --help
                 show this help message and exit
                 Summary failed login log and sort log by user.
 -u
 -after AFTER
                 Filter log after date. format YYYY-MM-DD-HH:MM:SS
 -before BEFORE
                 Filter log before date. format YYYY-MM-DD-HH:MM:SS
                 Show only the user of most N-th times
 -n N
 -t T
                 Show only the user of attacking equal or more than T times
                 Sort in reverse order
 -r
```

# 1-2: Auth log parser – Requirements (3/6) default output

```
2018NA-homework-1 git:(master) x python3 nahw1-2_0656087.py auth-sample.log
           count
 user
 admin
            104
   pi
             61
service
             14
 ubnt
 ethos
 squid
skyhertz
usuario
 user
 mother
```

## 1-2: Auth log parser – Requirements (4/6) with -r output

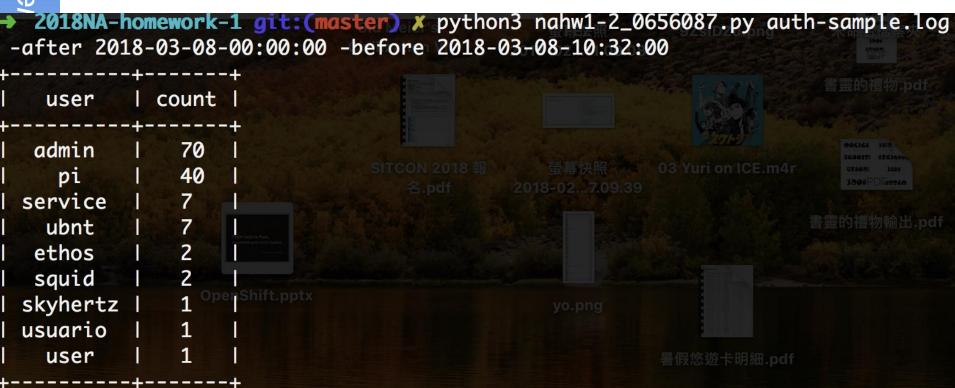
```
2018NA-homework-1 git:(master) x python3 nahw1-2_0656087.py auth-sample.log -r
           count
  user
skyhertz
usuario
  user
mother
 ethos
 squid
  ubnt
service
             14
             61
   pi
 admin
            104
```

## 1-2: Auth log parser – Requirements (5/6) with -u output

```
2018NA-homework-1 git:(master) x python3 nahw1-2_0656087.py auth-sample.log -u
           count
 user
admin
            104
ethos
mother
   pi
             61
service
             14
skyhertz |
squid
 ubnt
 user
usuario
```

## 1-2: Auth log parser – Requirements (6/6)

with -after -before output



### 1-2: Auth log parser – Hint: log sample

- ☐ Log format (a)
  - Mar 8 00:48:38 <auth.info> bsd4 sshd[85194]: Invalid user Chiangmj840306 from 36.230.109.223
- □ Log format (b) without <{facility}.{severity level}>
  - Mar 8 09:50:21 linux7 sshd[15899]: Invalid user ubnt from 188.187.121.68 port 47664
- ☐ Set log message year as 2018, if year is not defined on log message.
- □ No matter how many times the password is tried, each connection is only counted once.
  - Hint: just counting `Invalid user...`
- ☐ There is different that include facility and severity level of log message or not between them.
- ☐ Your script must can parse both (a) and (b).
- ☐ You can get raw sample on <a href="https://goo.gl/siimub">https://goo.gl/siimub</a>

## Help

- ☐ Email to <u>ta@nasa.cs.nctu.edu.tw</u>
- □ New E3 <a href="https://e3new.nctu.edu.tw">https://e3new.nctu.edu.tw</a>