



System Administration Practice Homework 2: Shell Programming

wlliou

Requirements

- ❑ 2-1: Filesystem Statistics (20%)
- ❑ 2-2: Web Crawler + Dialog(60%)
 - 用 shell 爬課表，並且用 column 排版輸出。 ❤️
- ❑ Modify code by yourself at demo (20%)
- ❑ Please write the scripts in Bourne Shell (sh)
 - No score if you use csh, bash or other languages.
- ❑ Due: 2017/10/12 upload $\${student_ID}.tar$ on E3
- ❑ Demo at next week. 2017/10/16~19 19:00~22:00

2-1: Filesystem Statistics

```
$ wget https://github.com/Thomas-Tsai/partclone/archive/0.2.89.tar.gz -O - | tar jxf -
--2016-09-29 10:32:25-- https://github.com/Thomas-Tsai/partclone/archive/0.2.89.tar.gz
正在查找主機 github.com (github.com)... 192.30.253.113
正在連接 github.com (github.com)|192.30.253.113|:443... 連上了。
已送出 HTTP 要求，正在等候回應... 302 Found
位置: https://codeload.github.com/Thomas-Tsai/partclone/tar.gz/0.2.89 [跟隨至新的 URL]
--2016-09-29 10:32:26-- https://codeload.github.com/Thomas-Tsai/partclone/tar.gz/0.2.89
正在查找主機 codeload.github.com (codeload.github.com)... 192.30.253.120
正在連接 codeload.github.com (codeload.github.com)|192.30.253.120|:443... 連上了。
已送出 HTTP 要求，正在等候回應... 200 OK
長度: 1051296 (1.0M) [application/x-gzip]
Saving to: 'STDOUT'

-                               100%[=====]

2016-09-29 10:32:30 (362 KB/s) - written to stdout [1051296/1051296]

$ cd partclone-0.2.89/
$ ../../sahw2-1.sh
1:402607 Makefile.in
2:312642 configure
3:173953 xfs_bmap.c
4:118770 ChangeLog
5:111736 extent-tree.c
Dir num: 24
File num:428
Total: 4992643
```

2-1: Filesystem Statistics - Requirement

- Inspect the current directory(“.”) and all sub-directory.
- Calculate the number of directories.
- Do not include ‘.’ and ‘..’
- Calculate the number of files.
- Calculate the sum of all file size.
- List the top 5 biggest files.
- Only consider the regular file. Do not count in the link, FIFO, block device... etc.

2-1: Filesystem Statistics - Requirement

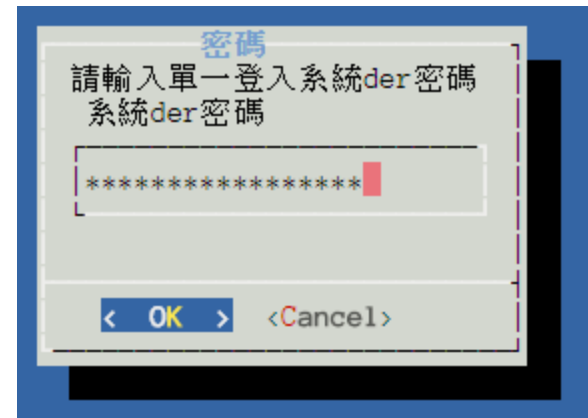
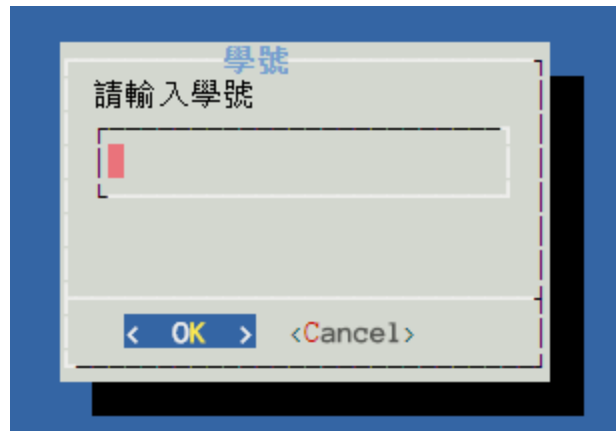
- Use **one-line** command
- No temporary file or shell variables.
- No “&&” “||” “>” “>>” “<” “;” “&”, but you can use them in the awk command. Actually, you don't need them to finish this homework.
- Only pipes are allowed.
- Hint: ls(1) with -A and -R

2-1: Filesystem Statistics - Requirement

❑ Grade

- File is executable. (4%)
- List top 5 file size and name. (4%)
- Dir num is correct. (4%)
- File num is correct. (4%)
- Total size is correct. (4%)

2-2: Web Crawler + Dialog



Mon.	Tue.	Wed.	Thu.	Fri.	Sat.	Sun.
.	.	.	雲端計算系統與實務	.	個別研究	.
.	.	.	.	演算法	.	.
.	.	.	.	演算法	.	.
雲端計算系統與實務
雲端計算系統與實務	演算法
.
.	科技計畫與專利創意提案	.	系統管理實務	.	.	.
.	科技計畫與專利創意提案	.	系統管理實務	.	.	.
.	科技計畫與專利創意提案	.	系統管理實務	.	.	.

2-2: Requirements

- ❑ 爬到自己的課表(HTML), 用任何入口都可以。(40%)
- ❑ 用 column 排版輸出。(5%)
- ❑ 需要用戶輸入的地方使用 dialog。(10%)
- ❑ 登入或辨識驗證碼時如果失敗要能自動重試(retry)。(5%)
- ❑ 輸入密碼時要用星號遮罩。(-10%)
- ❑ 不能自建伺服器。(-40%) 除了 cookie jar 和驗證碼圖片以外, 寫一個檔案-10分, 請善用變數。(-10%/次)

2-2: Dialog

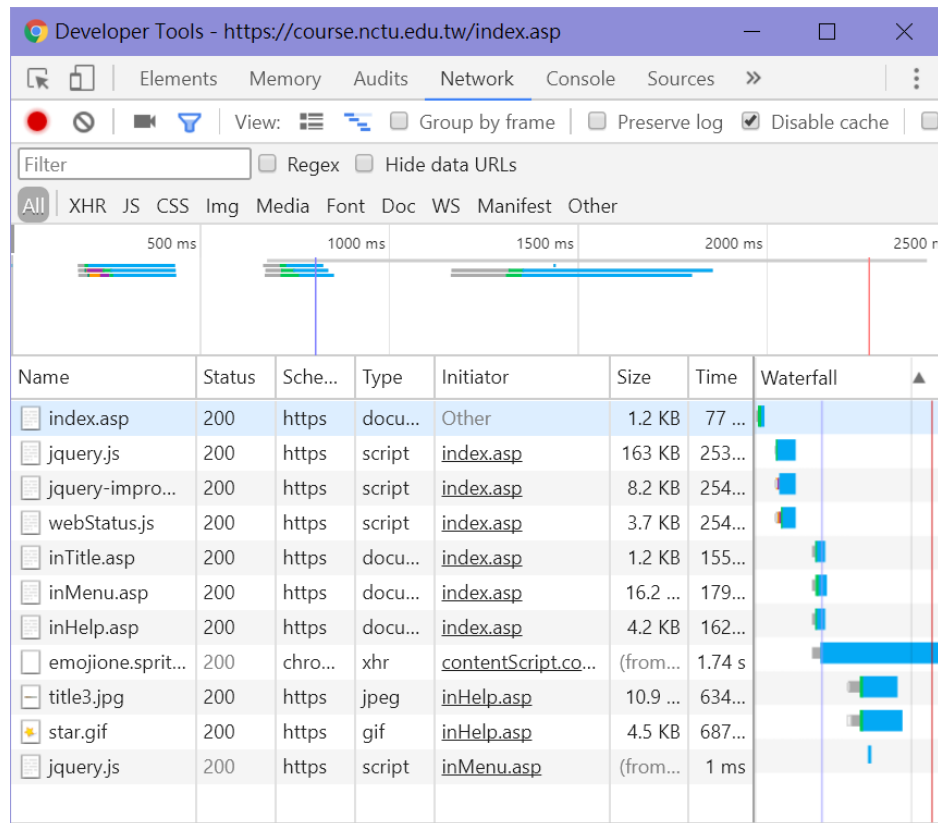
Dialog is a program that will let you to present a variety of questions or display messages using dialog boxes from a shell script.

These types of dialog boxes are implemented (though not all are necessarily compiled into dialog):

buildlist, calendar, checklist, dselect, editbox, form, fselect, gauge, infobox, inputbox, inputmenu, menu, mixedform, mixedgauge, msgbox (message), passwordbox, pause, prgbox, programbox, progressbox, radiolist, rangebox, passwordform, tailbox, tailboxbg, textbox, timebox, treeview, and yesno (yes/no).

2-2: Hint

- ❑ 課表在一般選課系統裡有。
- ❑ 可以用校園系統單一入口來登入一般選課系統。
- ❑ 善用Browser Devtool (F12) / Network



2-2: Hint (Keywords)

- dialog
- function (retry)
- sed
- awk
- grep
- curl
 - Cookie jar
- iconv (轉編碼)
- while
- if
- HTML
- HTTP form-data

2-2: Recommend Workflow

辨識驗證碼

- 先存取 <https://portal.nctu.edu.tw/captcha/pic.php> 讓登入系統建立 session 與驗證碼
- 再使用 <https://portal.nctu.edu.tw/captcha/pitctest/pic.php> 取得驗證圖片
- 可以 call 提供的 API, 請參照 <https://nasa.cs.nctu.edu.tw/sap/2017/hw2/captcha-solver/README.md>
- Bonus(+5%): 使用自己的方式辨識, 可使用外部服務或工具。

登入「校務系統單一入口」

- POST <https://portal.nctu.edu.tw/portal/chkpas.php>?
- body: username=<帳號>&password=<密碼>&seccode=<驗證碼>&pwdtype=static&Submit2=登入(Login)

登入「一般選課系統」

- <https://portal.nctu.edu.tw/portal/relay.php?D=cos>
- 可使用提供的 tool 處理 HTML: <https://github.com/nctu-cscc/sa2017-hw2/tree/master/extractFormdata>
- <https://course.nctu.edu.tw/index.asp>
- 注意: 系統限制10分鐘內僅能登入5次, 超過必須等待時間解鎖。

爬+剖析課表

- <https://course.nctu.edu.tw/adSchedule.asp>

2-2: Bonus

□ 3-2

- 使用自己的方式辨識驗證碼，可使用外部服務或工具。(+5%)
- 自己處理登入一般選課系統的 Form (將 HTML 的 form 轉為 url 格式)。(+5%)

2-2: Help

- ❑ FB Group: nctu nasa 課程討論區
<https://www.facebook.com/groups/NCTUCSNASA/>
- ❑ Email ta@nasa.cs.nctu.edu.tw or E3

- ❑ Q為何出這個作業？
A覺得好玩😁