## 01418235 UNIX OS & Shell Programming ปฏิบัติการที่ 6 สำหรับอาทิตย์ที่ 1-3 สิงหาคม 2566 Regular Expressions

วัตถุประสงค์ เพื่อให้นิสิตเข้าใจ RE และทดสอบ RE ด้วยคำสั่ง grep และ/หรือ vim ได้

## กิจกรรม

- 1. คำสั่ง grep มี options ที่ใช้งานบ่อย ๆ คือ
  - 1.1 -b prints the byte offset (O-based) before each output line
  - 1.2 -c prints only a count of the number of lines matching the pattern
  - 1.3 -i ignores upper/lower case in matching text
  - 1.4 -1 prints a list of files that contain at least one line matching the pattern
  - 1.5 -n shows line number of each line before the line
  - 1.6 -o prints only the matched parts of a matching line
  - 1.7 -s suppresses error messages
  - 1.8  $-\mathbf{v}$  (nverse output prints lines that do not match pattern
  - 1.9 x prints only lines that entirely match pattern
- 2. ตรวจสอบว่า regular expression ต่อไปนี้ match กับบรรทัดใดบ้าง และในบรรทัดที่ match นั้น string ที่ match คืออะไร

( Files ที่เตรียมไว้ให้ใช้ใน lab นี้ ถูก tar รวมกันอยู่ใน file "Lab06-RE.tgz" ให้ copy file นี้จาก sukreep ลงใน working directory ของตนเอง location ของ file ใน sukreep คือ /var/235/Lab06-RE.tgz แล้วเปิด terminal เพื่อรันคำสั่ง tar zxvf Lab06-RE.tgz จะเป็นการ extract files ที่ต้องใช้ ลงใน working directory โดยอยู่ใน directory ย่อยที่ชื่อ Lab06-RE)

```
2.1 [ABC]
   2.1.1 A
   2.1.2 AB
   2.1.3 ADBC
   2.1.4 D
2.2 [^A-G]
   2.2.1 A
   2.2.2 AB
   2.2.3 MBC
   2.2.4 D
2.3 [ABC] [^AB] ใช้ text ในข้อ 2.1 และ 2.2
2.4 ^ [ABC] [^C] *$ ใช้ text ในข้อ 2.2
2.5 [ABC] [^C] *$ ใช้ text ในข้อ 2.2
2.6 \..\{3,\}$
   2.6.1 rs.ef$tt
   2.6.2 abc.ab
   2.6.3 abc.$$$$
   2.6.4 abc
```

2.7 ตรวจสอบความถูกต้องโดยสร้าง file ที่เก็บ text ในแต่ละข้อเก็บไว้ แล้วใช้คำสั่ง grep -b -o <pattern> <file>

01418235 Lab 06 2023/01

3. พิจารณาผลของ RE ในแต่ละข้อต่อไปนี้กับไฟล์ตัวอย่าง หลังจากนั้น<del>จ</del>ึงทดลองว่าได้ผลตามที่คิดไว้หรือไม่

```
Linux and unix
  Linux is modeled on Unix, but developed independently,
  and is free, open sourced software. Unix, in its many
  flavors is propriatery, although I think Sun might have
  open sourced their version of Unix,
  but I could be very wrong.
  Linux/Unix are very powerful, multi-user capable systems,
  and can be run either from command line or through GUIs.
  I am not sure what limitations Unix has in terms of using
  the numerous GUIs that run on Linux.
  UNIX
  I think Unix is more hardware demanding than Linux,
  but I could be wrong.
  File systems on unix and Linux are similar. File information
  is kept inside an inode.
  DOS
  DOS is a command line OS, single user, much less
  powerful than Linux or Unix. Many commands for DOS were
  also present in (and copied from?) CP/M.
  Many Linux/UNIX commands are different from those of DOS.
   3.1 grep '...' testFile
   3.2 grep '^...$' testFile
   3.3 grep -vn '....' testFile
   3.4 grep -c '^$' testFile
   3.5 grep -c '.' testFile
   3.6 grep 'unix' testFile // ลองใช้ -i
   3.7 grep -iv 'UNIX' testFile
   3.8 grep -i 'UNIX' testFile
   3.9 grep -i '^UNIX' testFile
   3.10 grep -i 'file.*file' testFile
   3.11 grep -v '^$' testFile // ลองใช้ -c
   3.12 grep -n '^ *D' testFile
   3.13 grep -n '^[A-G]' testFile
   3.14 grep -n '^[^A-G]' testFile
4. ใช้ไฟล์ตัวอย่างเช่นเดียวกับข้อที่แล้ว
   4.1 ใช้คำสั่ง grep ร่วมกับ regular expression ที่เหมาะสมในการหาบรรทัดที่มีคำซ้ำกันสอง (duplicate) คำยื่นไป
         คำสั่ง grep หรือ RE ที่ใช้ คือ \(\<.\\>\).*\<\1\>
   4.2 หาบรรทัดที่มีคำซ้ำกันสาม (duplicate) คำปืนไป
        คำสั่ง grep หรือ RE ที่ใช้ คือ
```

01418235 Lab 06 2023/01 2

5. จากข้อมูลตัวอย่างต่อไปนี้ ให้หา RE ร่วมกับคำสั่ง grep ที่ได้ผลลัพธ์ตามที่ระบุ

```
* This is a line with only one asterisk.
* This is a line that starts with an asterisk, but has more.*
*** This is a line with three asterisks in a row.
This line has no asterisk.
* This is another line with only one asterisk.
This line has two asterisks * separated by other characters. *
*** The start and end have three asterisks. ***
This line has two asterisks with a character between them.
*** The start has three, but the end has four asterisks. ****
*The start and end have one asterisk.*
5.1 หาบรรทัดที่ขึ้นต้นด้วย * เพียงตัวเดียว ยึ่งในบรรทัดดังกล่าวอาจมี * อยู่ที่อื่นได้อีก
       คำสั่ง grep หรือ RE ที่ใช้ คือ _____'\^*[^*]*$'
5.2 หาบรรทัดที่มี * สองตัวยิงถูกคั่นด้วยตัวอักขระอื่น 1 ตัว
       คำสั่ง grep หรือ RE ที่ใช้ คือ <u>\*[^*]*</u>'
 5.3 หาบรรทัดที่มีตัว * ติดกันไม่เกิน 3 ตัว (ใช้ grep pipe ต่อกันได้หลายครั้ง)
       คำสั่ง grep หรือ R6 ที่ใช้ คือ <u>grep \*\{1,3\}' asterisks.txt l grep --color -nv \*\{4,\}'</u>
5.4 หาบรรทัดที่มีกลุ่มของ * ที่อยู่ต้นและท้ายบรรทัด ยึ่งมีจำนวนของ * เท่ากัน
       คำสั่ง grep หรือ RE ที่ใช้ คือ '^\(*\{1,\}\).*[\^*]\1$'
```

## สิ่งที่ต้องส่ง

ให้พิมพ์คำตอบ ข้อ 4 และ 5 ลงใน text file พร้อมรหัสนิสิตและชื่อ แล้ว turn in ภายในวันจันทร์ที่ 14 สิงหาคม 2566

01418235 Lab 06 2023/01 3