

## Aims

This week we develop your Perl programming skills.

## Assessment

**Submission:** give cs2041 lab08 courses.pl lectures0.pl lectures1.pl lectures2.pl [tags.py shortest\_path.py]

**Deadline:** either during the lab, or Monday October 19 11:59pm (midnight)

**Assessment:** Make sure that you are familiar with the [lab assessment criteria](#).

## Background

We have covered only a small amount of Perl in lectures. In fact, to cover the whole language in detail would take a whole semester, so we're going to rely on you finding out about the language yourself in tutes, labs and assignments. A good place to start is the Perl documentation & tutorial links on the class home page For example you might find these useful:

- [Perl language syntax](#)
- [Perl functions](#)
- [Perl operators](#)

As usual **lab08** sub-directory of your git repo:

```
cd
cd 2041-labs/lab08
gedit courses.pl &
```

## Exercise: Scraping Course Codes

Write a Perl script `courses.pl` which prints the course codes with a given prefix of all UNSW courses with lectures on the Kensington Campus this session. For example:

```
./courses.pl VISN
VISN1101
VISN1111
VISN1221
VISN2111
VISN2211
VISN2231
VISN3111
VISN3211
VISN4003
```

**Hints:** The course codes with prefix `VISN` can be found in this web page: <http://www.timetable.unsw.edu.au/current/VISNKENS.html> . You can assume this is the case for all prefixes. You saw how to retrieve a web page using `wget` in a previous lab. So you can perform this task in less than 10 lines of Perl.

As usual:

```
~cs2041/bin/autotest lab08 courses.pl
git add courses.pl
git commit -a -m "courses.pl was easy - Perl is cool"
```

## Exercise: Scraping Lecture Times

Write a Perl script `lectures0.pl` which given course codes as arguments prints details of their lectures.

```
./lectures0.pl COMP2041
COMP2041: S2 Tue 13:00 - 15:00 (Weeks:1-9,10-12), Thu
./lectures0.pl PSYC1011 COMP9024 COMP3231
PSYC1011: S2 Mon 11:00 - 12:00 (Weeks:1-9,10-12), Wed
PSYC1011: S2 Mon 18:00 - 19:00 (Weeks:1-9,10-12), Wed
COMP9024: S1 Tue 18:00 - 21:00 (Weeks:1-4,5-12)
COMP9024: S2 Thu 18:00 - 21:00 (Weeks:1-9,10-12)
COMP3231: S1 Tue 14:00 - 16:00 (Weeks:1-4,5-12), Wed
```

**Hint:** You can assume that a course's lecture times will be found in a web page equivalent to: <http://timetable.unsw.edu.au/current/COMP2041.html>. Its difficult to use a regexp to match the line containing the lecture description but you match a previous line, then skip a certain number of lines.

You can also get the teaching period this way. Don't panic if you can't get this quite right your tutor will be generous with hints.

**Hint:** a hash can be easily used to avoid repeated output.

**Hint:** make sure you have the URL exactly as above - e.g. don't have repeated slashes (the timetable website uses fragile rewriting rules).

As usual:

```
~cs2041/bin/autotest lab08 lectures0.pl
git add lectures0.pl
git commit -a -m "lectures0.pl passes autotest"
```

## Semi-Challenge Exercise: Lecture Times as Tuples

The output from `lectures0.pl` is (more or less) human readable but is less convenient for other uses. Copy `lectures0.pl` to `lectures1.pl` and modify it so that if a `-d` option is specified it prints the hourly details of lectures in the format shown in the examples below:

./lectures1.pl COMP2041

COMP2041: S2 Tue 13:00 - 15:00 (Weeks:1-9,10-12), Thu

./lectures1.pl -d COMP2041

S2 COMP2041 Tue 13

S2 COMP2041 Tue 14

S2 COMP2041 Thu 17

./lectures1.pl -d COMP4121

S2 COMP4121 Mon 11

S2 COMP4121 Mon 12

S2 COMP4121 Thu 9

S2 COMP4121 Thu 10

./lectures1.pl -d COMP1927

X1 COMP1927 Wed 9

X1 COMP1927 Wed 10

X1 COMP1927 Wed 11

X1 COMP1927 Fri 9

X1 COMP1927 Fri 10

X1 COMP1927 Fri 11

S1 COMP1927 Tue 16

S1 COMP1927 Tue 17

S1 COMP1927 Wed 12

S2 COMP1927 Mon 10

S2 COMP1927 Mon 11

S2 COMP1927 Wed 15

S2 COMP1927 Wed 16

./lectures1.pl -d PSYC1011 COMP9024 COMP3231

S2 PSYC1011 Mon 11

S2 PSYC1011 Wed 14

S2 PSYC1011 Thu 12

S2 PSYC1011 Mon 18

S2 PSYC1011 Wed 18

S2 PSYC1011 Thu 18

S1 COMP9024 Tue 18

S1 COMP9024 Tue 19

S1 COMP9024 Tue 20

S2 COMP9024 Thu 18

S2 COMP9024 Thu 19

S2 COMP9024 Thu 20

S1 COMP3231 Tue 14

```
S1 COMP3231 Tue 15
S1 COMP3231 Wed 16
```

Getting this exercise completely correct is difficult - your tutor will be generous with nearly correct attempts.

As usual:

```
~cs2041/bin/autotest lab08 lectures1.pl
```

```
git add lectures1.pl
```

```
git commit -a -m "lectures1.pl working for 1 hour lectures"
```

## Semi-Challenge Exercise: Lecture Times as a Table

Copy `lectures1.pl` to `lectures2.pl` and modify it so that when a `-t` option is specified it prints a count of many lectures occur at each day/time in an ASCII table in the format shown in the example below.

```
./lectures2.pl COMP9020 COMP9021 COMP9024
```

```
COMP9020: S1 Fri 14:00 - 17:00 (Weeks:1-4,5-12)
COMP9020: S2 Thu 12:00 - 15:00 (Weeks:1-9,10-12)
COMP9021: S1 Tue 18:00 - 21:00 (Weeks:1-4,5-12)
COMP9021: S2 Thu 18:00 - 21:00 (Weeks:1-9,10-12)
COMP9024: S1 Tue 18:00 - 21:00 (Weeks:1-4,5-12)
COMP9024: S2 Thu 18:00 - 21:00 (Weeks:1-9,10-12)
```

```
./lectures2.pl -d COMP9020 COMP9021 COMP9024
```

```
S1 COMP9020 Fri 14
S1 COMP9020 Fri 15
S1 COMP9020 Fri 16
S2 COMP9020 Thu 12
S2 COMP9020 Thu 13
S2 COMP9020 Thu 14
S1 COMP9021 Tue 18
S1 COMP9021 Tue 19
S1 COMP9021 Tue 20
S2 COMP9021 Thu 18
S2 COMP9021 Thu 19
S2 COMP9021 Thu 20
S1 COMP9024 Tue 18
S1 COMP9024 Tue 19
S1 COMP9024 Tue 20
S2 COMP9024 Thu 18
S2 COMP9024 Thu 19
```

S2 COMP9024 Thu 20

./lectures2.pl -t COMP9020 COMP9021 COMP9024

S1	Mon	Tue	Wed	Thu	Fri
09:00					
10:00					
11:00					
12:00					
13:00					
14:00					1
15:00					1
16:00					1
17:00					
18:00		2			
19:00		2			
20:00		2			

S2	Mon	Tue	Wed	Thu	Fri
09:00					
10:00					
11:00					
12:00				1	
13:00				1	
14:00				1	
15:00					
16:00					
17:00					
18:00				2	
19:00				2	
20:00				2	

./lectures2.pl -t COMP1911 COMP2041 COMP2121 COMP3121 COMP3311 COM

S1	Mon	Tue	Wed	Thu	Fri
09:00	1				
10:00	1				
11:00	2		1		
12:00	1		1		
13:00		1	1	1	
14:00		1	1		1
15:00			1	1	1
16:00			1		1
17:00			1		
18:00		2			

19:00		2			
20:00		2			
S2	Mon	Tue	Wed	Thu	Fri
09:00				1	
10:00				1	
11:00	1				
12:00	1			1	
13:00		1		1	
14:00		1	1	2	
15:00		1		2	
16:00	1	1		1	
17:00	1			1	
18:00				2	
19:00				2	
20:00				2	

Getting this exercise completely correct is difficult - your tutor will be generous with nearly correct attempts.

As usual:

```
~cs2041/bin/autotest lab08 lectures2.pl
git add lectures2.pl
git commit -a -m "lectures2.pl half working"
```

## Challenge Exercise: Extracting Tags in Python

The introduction to Python in lectures will come later.

The [example Python scripts](#) and links to external Python resources should help - but you may need more info - Google is your friend. Write a Python program, `tags.py` which given the URL of a web page fetches it by running `wget` and prints the HTML tags it uses.

Don't count closing tags (e.g. `</a>` )

The tag should be converted to lower case and printed in sorted order with a count of often each is used.

You are expected (not required) to use regexes but there are problems processing HTML this way - don't be surprised if it's hard to get right.

**Hint:** make sure you don't print tags within HTML comments.

For example:

```
./tags.py http://www.cse.unsw.edu.au/~cs2041/intro.html
a 20
b 8
body 1
```

br 17  
center 1  
div 2  
em 2  
font 15  
h1 1  
h3 18  
head 1  
hr 2  
html 1  
li 52  
link 1  
ol 2  
p 32  
pre 1  
script 1  
small 15  
table 3  
tbody 1  
td 20  
th 5  
title 1  
tr 9  
ul 11  
[./tags.py http://www.cse.unsw.edu.au/~cs2041/15s2/](http://www.cse.unsw.edu.au/~cs2041/15s2/)  
a 138  
b 3  
body 1  
br 27  
code 1  
div 9  
em 36  
head 1  
html 1  
link 1  
p 1  
script 1  
small 3  
table 3  
tbody 3

```
td 162
th 11
thead 3
title 1
tr 29
./tags.py http://www.cse.unsw.edu.au
a 159
body 1
br 2
div 445
fieldset 1
form 1
h2 27
h3 1
head 1
html 1
img 28
input 6
label 2
legend 1
li 85
link 10
meta 11
p 15
script 9
span 39
title 1
ul 14
```

Add an `-f` option to `tags.py` which indicates the tags are to be printed in order of frequency. For example:

```
./tags.py -f http://www.cse.unsw.edu.au/~cs2041/intro.html
pre 1
h1 1
title 1
tbody 1
html 1
body 1
head 1
link 1
```



```
center 1
script 1
em 2
hr 2
ol 2
div 2
table 3
th 5
b 8
tr 9
ul 11
font 15
small 15
br 17
h3 18
td 20
a 20
p 32
li 52
```

**Hint:** see last week's tute for a sample solution in Perl. As usual:

```
~cs2041/bin/autotest lab08 tags.py
```

```
git add tags.py
```

```
git commit -a -m "tags.py almost working"
```

## Challenge Exercise: Finding the Shortest Journey in Python

Write a Python program `shortest_path.py` that given the road distances between a number of towns (on standard input) calculates the shortest journey between two towns specified as arguments. Here is an example of how your program should behave.

```
./shortest_path.py Parkes Gilgandra
```

Bourke Broken-Hill	217
Bourke Dubbo	23
Bourke Gilgandra	62
Bourke Parkes	71
Canowindra Dubbo	35
Canowindra Gilgandra	13
Canowindra Parkes	112
Dubbo Gilgandra	91

```
<ctrl-d>
```

```
Shortest route is length = 105: Parkes Dubbo Canowind
```

Hints: Python's strings have a [split method](#) which can break up input lines. Its easy to implement [Sets in python](#).

**Hint:** see last week's tute for a sample solution in Perl. As usual:

```
~cs2041/bin/autotest lab08 shortest_path.py
```

```
git add shortest_path.py
```

```
git commit -a -m "my shortest_path.py rocks!"
```

## Testing

You will need to do your own testing but to assist you as usual some autotest tests are available for this lab.

To run all tests:

```
~cs2041/bin/autotest lab08
```

You can run a single test if you also pass the test label as the second argument to autotest. For example, to run just test courses\_2 type:

```
~cs2041/bin/autotest lab08 courses_2
```

## Finalising

You must show your solutions to your tutor and be able to explain how they work. Once your tutor has discussed your answers with you, you should submit them using:

```
give cs2041 lab08 courses.pl lectures0.pl lectures1.pl lectures2.pl [tags.py sho
```

Whether you discuss your solutions with your tutor this week or next week, you must submit them before the above deadline.