LESSON 3

X-PROGRAM

在終端機裡執行以下指令 (TERMINAL / 模擬終端機 ... 等等)

- 安裝 Poltergeist
  - 輸入: gem install poltergeist
- 執行 which phantomjs
  - 執行結果會指出系統內的 phantomjs 在什麼位置
  - 例如 /usr/local/bin/phantomjs
  - 如果沒有 phantomjs 請進行下一頁的操作

安裝 PHANTOMJS

- Mac (三選一即可)
  - Homebrew: brew install phantomjs
  - MacPorts: sudo port install phantomjs
  - Manual install: Download this
- Linux
  - Download the 32 bit or 64 bit binary.
  - Extract the tarball and copy bin/phantomjs into your PATH

新增一個 RUBY 檔案

- 新增一個副檔名為 .rb 的文字檔案
  - 可使用 Sublime Text 內的 File -> New File
- 或在終端機中輸入 touch gogo.rb
  - 然後輸入 subl gogo.rb 使用 Sublime Text 開啟該檔

在剛剛新增的.RB 檔案內輸入程式碼以引用套件

- require 'capybara'
- require 'capybara/dsl'
- require 'capybara/poltergeist'
- include Capybara::DSL

CAPYBARA 設定

- options = {
- :window_size => [1280, 800],
- :js_errors => false
- }
- Capybara.javascript_driver = :poltergeist
- Capybara.current_driver = :poltergeist
- Capybara.register_driver :poltergeist do |app|
- Capybara::Poltergeist::Driver.new(app, options)
- end

按照慣例，我們先決定要打誰

- 輸入程式碼
  - url = "http://140.118.31.215/querycourse/ChCourseQuery/QueryCondition.aspx"
- 用瀏覽器看一下這是什麼網站，開啟開發者工具的 Network 分頁，看看每一個操作發送了哪些東西給 Server

思考時間

- 我們要進行什麼操作？
- 會遇到什麼困難？
- 解決辦法是什麼？

解決問題

- 我們要進行什麼操作？
  - 我們要送出搜尋，然後把結果一一印出
- 會遇到什麼困難？
  - 參數好多好煩，而且還有 ASPX Server 端的驗證問題
- 解決辦法是什麼？
  - ？？？

設定必要參數

- 學年
  - year = 105
- 學期
  - semester = 2
- 對應網頁上 Select 元素，文字部份是 xxxy (一百零zzzzzz)
  - 怎麼選到我們指定的學期？

選擇指定選項

- all('select#semester_list option').each do |option|
- if option.text.include?("#{year}#{semester}")
- option.select_option
- end
- end
- 然後 save_screenshot 觀察看看是不是選到我們要的選項了

送出查詢

- find('input#QuerySend').trigger('click')
- 這個查詢動作會讓目標網頁讀取資料庫內容，由於不可抗力的關係，它很慢
- 所以我們讓程式在這邊睡個兩百秒
- 然後輸入
- save_screenshot('screenshot.jpeg', full: true)
- 確認兩百秒之後是不是在頁面上有我們要的資料

過濾結果

- 觀察網頁上，我們需要的資料有哪些？
- 我們需要的資料，包含在什麼元素之內？
- 我們要怎麼拿到 poltergeist 正在開啟的網頁HTML結構？

過濾結果

- 觀察網頁上，我們需要的資料有哪些？
  - 課程代碼、課程名稱、學分、上課星期節次等等
- 我們需要的資料，包含在什麼元素之內？
  - 包在 table#my_dg 裡面，每一個 tr 元素代表一門課
  - 每一個 td 元素代表該堂課的一種資料
- 我們要怎麼拿到 poltergeist 正在開啟的網頁HTML結構？
  - 使用 page.html 或 page.body 方法即可

中場休息，如何有效率的除錯？

- 使用 poltergeist 的時候，每次都要用 save_screenshot 來看網頁擷圖，其實滿麻煩的
- 這邊介紹一個好東西 "pry-byebug"
- 在終端機內輸入 gem install pry-byebug
- 在程式碼頂端加入 require 'pry-byebug'
- 在想要暫停除錯的程式碼輸入 binding.pry

中場休息，如何有效率的除錯？

- 當程式執行到 binding.pry 的時候就會出現像這樣的畫面

```
From: /home/xprogram/桌面/gogo.rb @ line 20 :

    15: end
    16:
    17: url = "http://140.118.31.215/querycourse/ChCourseQuery/QueryCondition.as
px"
    18: visit(url)
    19: binding.pry
 => 20: year = '105'
    21: semester = '2'
    22:
    23: all('select#semester_list option').each do |option|
    24:   if option.text.include?("#{year}#{semester}")
    25:     option.select_option

[1] pry(main)>
```

- 輸入 url 並按下 enter

中場休息，如何有效率的除錯？

- next - 執行下一行程式
- continue - 離開 pry-byebug 並繼續執行之後的程式

回到分析資料

- courses_list = Nokogiri::HTML(page.body)
- courses_list_trs = courses_list.css('table#my_dg tr:not(:first-child)')
- courses_list_trs_count = courses_list_trs.count
- puts courses_list_trs_count #印出開課總數

分析每一門課的資料和元素長相

- courses_list_trs.each do |row|
  - table_data = row.css('td')
  - puts table_data[2] #也可以印出其它位置觀察
- end