

Lab 11

擷取網路資料 (II)

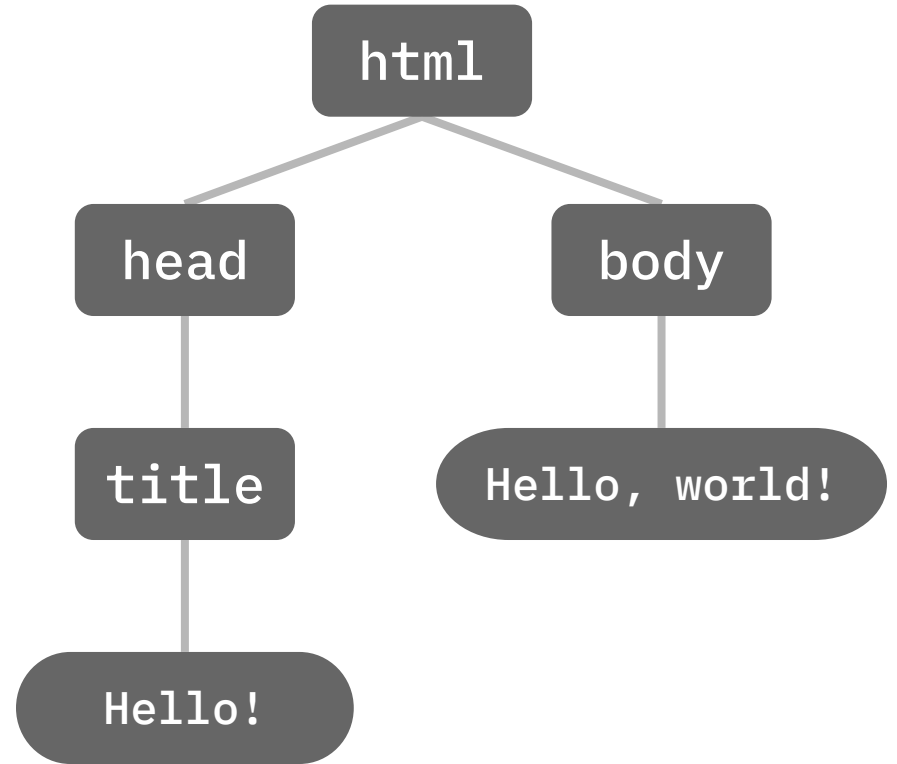
HTML, CSS, 網頁剖析

HTML

(HyperText Markup Language)

Document Object Model

```
<!DOCTYPE html>
<html>
  <head>
    <title>Hello!</title>
  </head>
  <body>
    Hello, world!
  </body>
</html>
```



CSS

(cascading style sheets)

CSS Selector

selector

h1 {

color: red;

text-align: center;

}

← style 1

← style 2



模仿學院

Jessica

1. 校長
2. jessica@ntu.edu.tw
3. Room 3004
4. 專長: 法國人
5. [Web site](#)

Claire

1. 模仿學院學員
2. claire@ntu.edu.tw
3. Room 3003
4. 專長: ?
5. [Web site](#)

Andrea

1. 模仿學院練習生
2. andrea@ntu.edu.tw
3. Room 3002
4. 專長: ?
5. [Web site](#)

Puff

1. 模仿學院練習生
2. puff@ntu.edu.tw
3. Room 3001
4. 專長: ?
5. [Web site](#)

Amber

1. 模仿學院榮譽會員
2. amber@ntu.edu.tw
3. Room 1111
4. 專長: IG 排版
5. [Web site](#)

mcku

1. 模仿學院榮譽會員
2. mcku@ntu.edu.tw
3. Room 9999
4. 專長: 麻將
5. [Web site](#)

GitHub Pages

rvest

瀏覽器開發者工具 (Web Developer Tools)

➤ 開啟 Chrome / Safari 開發者工具

- **Ctrl + Shift + I**
- **⌥ (Option) + ⌘ (Command) + I (Mac)**

不必凡事都自己來

(先 survey 別人辛苦的結晶)

➤ 如果你要 twitter (API) 資料 ...

- [rtweet](#)

➤ 如果你要爬 PTT ...

- [PTTmineR](#) (推, 作者很有熱情)

➤ 如果你要爬 reddit ...

- [reddit](#)
- [RedditExtractoR](#)

進階爬蟲 (Browser simulation)

<http://theautomatic.net/2019/07/23/beautifulsoup-vs-rvest/>