<https://hackmd.io/@johnsonsy/Hkl5t9EiF>

<https://hackmd.io/@johnsonsy/NKUST_20211130>

<https://medium.com/pierceshih/%E7%AD%86%E8%A8%98-%E4%BD%95%E8%AC%82-http-%E5%82%B3%E8%BC%B8%E5%8D%94%E5%AE%9A-1d9b5be3fd24>

目前只能先爬”SSR”的，”CSR”之後再說

1. jsoup 是由java 寫的一個包，讓,maven可以調用

<https://jsoup.org/cookbook/>

Maven則類似python的pip，可以把外面的包給download進來

Jsoup對應到Python的beautiful soup

1. Domain Name System(DNS): (<https://yakimhsu.com/project/project_w4_Network_http.html>)

當瀏覽器發送 request 的時候，其實是發送一個網址，但要怎麼知道網址的 IP 位置呢？就是靠 DNS 啦

DNS可以將網址轉成IP

1. 瀏覽器只是一個程式:

(<https://yakimhsu.com/project/project_w4_Network_http.html>)

前面有提過，其實瀏覽器幫我們做的就是：發送一堆 request、接收一堆 response，根據 html 裡面的內容再進行資源下載，而更詳細的 request 發送順序可以參考以下：

一開始先發送 request 到副檔名為 .html 的 url

DNS 解析成 IP 位置

Server 接收到 request、 回傳 response

瀏覽器解析 html 檔案

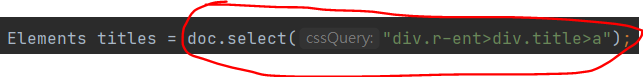
根據 html 內容，一但發現有 CSS、JS、圖片檔案，再發送各檔案的 request（ 如同步驟 1.2.3 ）

發送完包含在 html 所有資源的 request

開始下載資源 ( CSS、JS、圖片檔 )

進行渲染網頁

1. CSS selector:



<https://www.w3schools.com/cssref/css_selectors.asp>

1. Tag、class:

tag就是<>，大致有head、body

class: 每個tag可以至多有一個class是用來”渲染”

href: 網址