Q: 目標網站 && 目標內容?

- 網站：https://www.alexanderwang.com/tw-zh/women-t-by-alexanderwang (https://www.alexanderwang.com/tw-zh/women-t-by-alexanderwang)
- 欲獲取之內容
  - 圖片
  - 產品名稱

爬動態網站兩法子：

- 手動分析
- selenium

安裝

```
$ pip install selenium
```

下載

- 請根據電腦 Chrome 瀏覽器的版本 & 作業系統安裝
- Chrome Driver (http://chromedriver.chromium.org/downloads)

In [46]:
```
# selenium v3.141
# 查看版本號: $ pip show selenium
# $ pip search selenium

from selenium import webdriver

# 訪問頁面
browser = webdriver.Chrome()

from selenium.webdriver.support.ui import WebDriverWait
# 頁面加載問題 - 隱性, 還有別的可能: 顯性 & 暴力處理法 (下面有連結)
browser.implicitly_wait(10)

# 取得網頁原始碼
browser.get('https://www.alexanderwang.com/tw-zh/women-t-by-alexanderwang')


from selenium.webdriver.support import expected_conditions as EC
from selenium.common.exceptions import NoSuchElementException

from bs4 import BeautifulSoup
import lxml

# 提取想要的目標內容
# selenium提取方法、正則表達式 (難度比較高)
soup = BeautifulSoup(browser.page_source, 'lxml')

for title in soup.find_all('picture', attrs = {"title": True}):  # True / N
    print(title.get_text().replace("\n", ""))  # 為什麼 html 的`屬性`被竄改了?

    # your turn: 觀察兩三個網站, 了解為什麼 selenium 會竄改`屬性`?

    # your turn: 請擷取連結 & 商品名稱

# 關閉瀏覽器
browser.close()
```

```
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
38cf18e5/hi-res/4W194008815F.jpg?sw=100" alt="水洗尼龙裤">
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
851f453a/hi-res/4W194008815R.jpg?sw=100" alt="水洗尼龙裤">
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
c93d2c95/hi-res/4W194008815E.jpg?sw=100" alt="水洗尼龙裤">
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
9915cdc8/hi-res/4W194008815D.jpg?sw=100" alt="水洗尼龙裤">
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
803b45e3/hi-res/4W194008815B.jpg?sw=100" alt="水洗尼龙裤">
        <img class="image image--fallback" src="https://www.alexanderwan
g.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw
c2197b03/hi-res/4W194009001F.jpg?sw=100" alt="水洗尼龙短裤">
        <img class="image image--fallback" src="https://www.alexanderwan
```

In [13]:
```
from IPython.display import Image
Image(filename = "1-1.png")
```

Out[13]:
```
<picture title="水洗尼龙裤" class="lazyload product-tile_image "> == $0
  <source media="(max-width: 600px) and (-webkit-min-device-pixel-ratio: 2), (max-width: 600px) and (min-resolution: 192dpi)" data-srcset="https://www.alexanderwang.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw38cf18e5/hi-res/4W194008815F.jpg?sw=1200">
  <source media="(max-width: 600px)" data-srcset="https://www.alexanderwang.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dw38cf18e5/hi-res/4W194008815F.jpg?sw=600">
```

In [12]:
```
Image(filename = "1-2.png")
```

Out[12]:
```
<picture title="水洗尼龙短裤" class="lazyload product-tile_image "> == $0
  <source media="(max-width: 600px) and (-webkit-min-device-pixel-ratio: 2), (max-width: 600px) and (min-resolution: 192dpi)" data-srcset="https://www.alexanderwang.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dwc2197b03/hi-res/4W194009001F.jpg?sw=1200">
  <source media="(max-width: 600px)" data-srcset="https://www.alexanderwang.com/dw/image/v2/BCCC_PRD/on/demandware.static/-/Sites-master/default/dwc2197b03/hi-res/4W194009001F.jpg?sw=600">
```

selenium 文檔

https://selenium-python-docs-zh.readthedocs.io/zh_CN/latest/index.html (https://selenium-python-docs-zh.readthedocs.io/zh_CN/latest/index.html)

加載頁面等待

https://huilansame.github.io/huilansame.github.io/archivers/sleep-implicitlywait-wait (https://huilansame.github.io/huilansame.github.io/archivers/sleep-implicitlywait-wait)

正則表達式

http://www.runoob.com/regexp/regexp-syntax.html (http://www.runoob.com/regexp/regexp-syntax.html)