【robots.txt】

* 是一種存於網站根目錄（head）下的 ASCII 文字檔，表示此網站中的哪些內容是不應被取得的。
* 通常為「限制」（但非強制，僅為自訂規則）
* 爬取前可先看看robots.txt歡不歡迎爬蟲=w=
* User-agent: \* ：適用哪些使用者代理(瀏覽器)、爬蟲、快取……（\*為所有）
* Disallow: / ：禁止的範圍及工具（/為根目錄，即整個網頁內容）
* Allow: / ：允許的範圍及工具（/為根目錄，即整個網頁內容）

【ubuntu & 資料庫】

* sudo apt-get update ：更新套件清單
* sudo apt-get upgrade：比對清單進行安裝更新
* 在資料庫中，curd表示create、update、read、delete統稱。

【在linux中移除已安裝套件】

* sudo apt-get xxxxxx / sudo apt-get --remove xxxxxx（以purge方式移除）
* sudo apt-get --purge autoremove xxxxxx （移除無用套件，多為移除套件後才殘留下來的無用套件，使用autoremove+ purge徹底移除）
* dpkg -l （列出已安裝套件）

【python & 資料庫】

* 在python檔案中登入SQL並執行指令操作，需先安裝──

sudo apt-get install python3-pip （python執行套件）

pip3 install pymysql（python及mysql連結使用套件）

sudo service mysql status （查看SQL的執行狀態）

sudo netstat -ntlp | grep mysql （查看SQL IP及port狀態）

* pymysql.connect ( host=”A“, port=xxxx, user=”B“, password=”yyy”, database=”C”, charset=”編碼方式” )：

1. host：連結方式，可為IP、localhost……
2. user：登入SQL的使用者名稱
3. port：SQL的所在PORT號，若為localhost可不填
4. password：使用者的登入密碼
5. database：登入時直接進入的指定資料庫
6. charset：設定編碼方式