

## Week8

1.學習方向：該領域內隨機踏步 or 專心找到想要的工具。

2.Crawler

=>BeautifulSoup: pip install beautifulsoup4

=>02-pptHot: pttHot.py

爬特定的版面以及繞過一些條件

=>sql:視窗化

==>python qt(缺點：太大)

==>python Tk

3.preprocess

=>排序就是一種前處理

=>排序

==>index sort

優點：

缺點：每次加一個就要全部排一次(連插入最糟情況也需要  $O(n)$  時間)

==>Hash sort // 要盡量留空間，否則碰撞機率很高

優點：計算只要常數時間

缺點：如果全部元素都雜湊到同一行，機率非常低

==>時間複雜度

| 插入方式  | O           | Average     |
|-------|-------------|-------------|
| Hash  | $O(n)$      | $O(1)$      |
| Index | $O(n)$      | $O(n)$      |
| 紅黑樹   | $O(\log n)$ | $O(\log n)$ |

4.正規表達式

5.壓縮檔

6.KMP 演算法

7.Trie

=>Radix Trie：更節省空間

=>Patricia Tries(compressed Tries)

8.NCBI