人工智慧系統 學習單 #7

AI/ Python/自然語言處理

班級： 學號： 姓名：

學習目標： 1. 瞭解安裝套件操作 2. 瞭解網頁爬蟲指令

1. 課堂斷詞使用哪一套? 如何安裝斷詞？

2. 上網搜尋其他可以斷詞的套件?

3. 文字雲使用的套件? 如何安裝套件?

4. 文字雲的遮罩須符合什麼要件?

[課堂活動#1] 課堂活動 AI 斷詞技術與文字雲應用

以數位人文活動為範圍，以課堂教授的 jieba & 文字雲技術。

上網以四書五經或其他數位人文議題，下載 原文 txt，並製作成斷詞後的文字雲。

要求：

1. 能夠用結巴斷詞
2. 新增專業字典 / 停用詞字典。
3. 自定義文字雲形狀&設定字體為黑體。
4. 找出一個應用。

學號： 姓名：

主題：論語文字雲分析

應用：找出論語中最常見的詞，發現子貢曰 出現很多。君子、小人、夫子出現很多。

其他應用：可以用於分析數位學習論壇，分析學生常見問題。

原文下載網址：

https://www.chineseclassic.com/content/2

子
曰
吾
君子
孔子
夫子
小人
言
不可
邦
亦
已
路
人
無
道
如
好学
仁者
子貢曰
不知
可謂
使
欲
知
顏淵
問
未
夫
三年
10
11
12
13
14
15
16
17
18
19
20
23

子
曰
吾
君子
孔子
言
夫子
子貢曰
道
人
邦
亦
路
可謂
不可
小人
已
天下
知
欲
死
問
10
11
12
13
14
15
16
18
20
23
24

子
曰
君子
吾
孔子
人
道
言
子貢曰
何如
可謂
不知
亦
邦
夫
不可
天下
知
已
問
使
求
子夏
三年
斯
如之何
未
欲
死
樂
好