

Text Mining 實習課

這次的內容主要是針對 Text 內容進行處理，所需要的套件有：

- pandas
- numpy
- re
- jieba
- nltk
- genism

後面三個套件是需要另外安裝的，在 Anaconda 內是沒有預設安裝，安裝方式可以參考以下說明。

額外套件安裝說明

● jieba 繁體版

<https://github.com/APCLab/jieba-tw>

- 1.開啟終端機
- 2.在提示符號下鍵入 `pip install git+https://github.com/APCLab/jieba-tw.git`
3. 安裝後，可在 Jupyter Notebook 中輸入 `import jieba`，看是否有錯誤訊息，如果沒有表示安裝成功，如果有就看錯誤內容進行調整

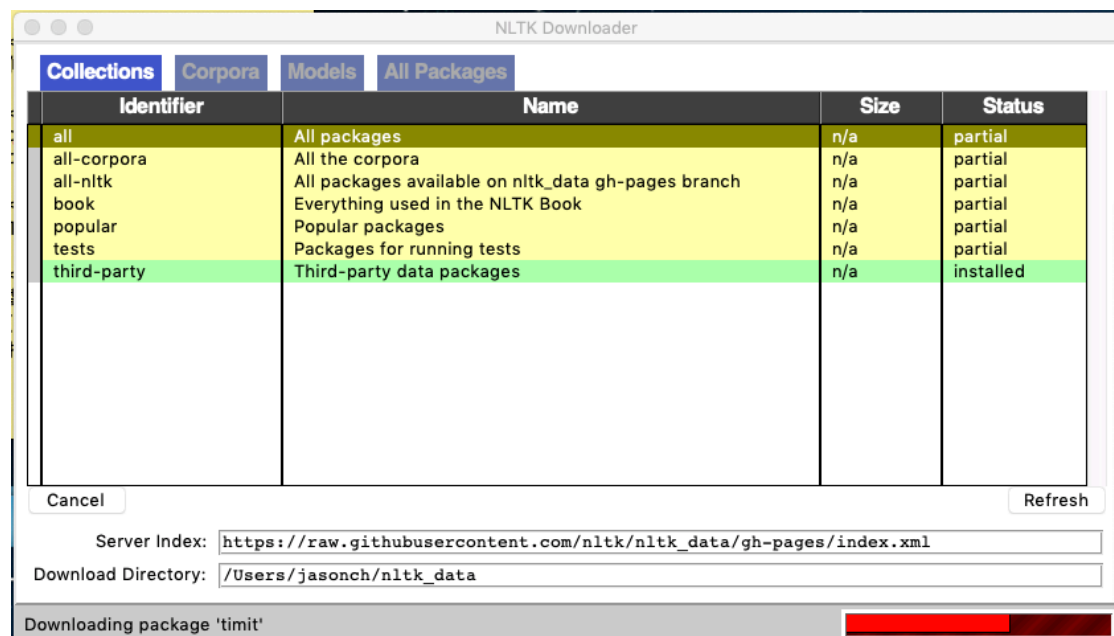
● nltk

<https://www.nltk.org>

nlTK 全名是 Natural Language Tool Kit，是一套基於 Python 的自然語言處理工具箱。官方的文件算是相當完備，可以透過該內容了解。

安裝方式：

1. 開啟終端機
2. 在提示符號下鍵入 `pip install nltk`
3. 安裝後，可在 Jupyter Notebook 中輸入 `import nltk`，看是否有錯誤訊息，如果沒有表示安裝成功，如果有就看錯誤內容進行調整
4. 載入 `nltk` 後，可先輸入 `nltk.download()`，在 Python Launcher 會顯示下面視窗內容，可選擇安裝所有的 Packages，以免去之後還要再進行安裝



● gensim

<https://radimrehurek.com/gensim/>

gensim 是一個免費 python package，能夠從文件中有效地自動抽取語義主題。支援多個演算法，可以處理原始的，非結構化的文字。

安裝方式：

1.開啟終端機

2.在提示符號下鍵入 `pip install --upgrade gensim` 或是 `conda install -c conda-forge gensim`

以上兩種擇一使用

3.安裝後，可在 Jupyter Notebook 中輸入 `import gensim`，看是否有錯誤訊息，如果沒有表示安裝成功，如果有就看錯誤內容進行調整