修正後提交的檔案清單如下：

1. 執行程序 - sdg\_process\_v2.py

2. 配合調整項目的資料庫 Schema - sdgs\_2020-05-13.sql (Mysql)

3. csv - 資料庫表單輸出，共 10 個檔案

4. raw\_data - 使用的原始資料檔，兩份課程大綱 + 兩份關鍵字庫清單

---

建置環境及操作步驟如下 ：

1. 建立 mysql 資料庫 (參考https://dev.mysql.com/downloads/installer/)

2. 在指定資料庫中建立表單：執行 sdgs\_2020-05-13.sql 建立表單

3. 建置 python 環境 (v3.8.2)

4. 將 sdg\_process\_v2.py 放在指定位置，修改下列參數：

(1) 關鍵字詞庫檔案路徑 (line 23)

(2) 課程大綱檔案路徑 (line 43)

(3) 資料庫連線資訊 - host/user/password/db (line 120)

(4) 開啟terminal(cmd)畫面執行程式碼

(a-1) python sdg\_process\_v2.py 1:CH => 建立新的中文SDG關鍵詞庫(會自動把該語言的舊詞庫清空，建立新的)

(a-2) python sdg\_process\_v2.py 1:EN => 建立新的英文SDG關鍵詞庫(會自動把該語言的舊詞庫清空，建立新的)

(b) python sdg\_process\_v2.py 2 => 將課綱資料存放在資料庫

(c-1) python sdg\_process\_v2.py 3:CH => 針對所有存在資料庫的input進行中文詞庫的分析

(c-2) python sdg\_process\_v2.py 3:EN => 針對所有存在資料庫的input進行英文詞庫的分析

＊執行(a)的指令時，請確保檔案符合格式，可參考附件raw\_data/關鍵字詞庫 lin.xlsx

每一個類別的第一列為表頭：「SDG N」與「WEIGHT-N」（N為類別編號）

＊將資料存放在 input 的功能僅針對這次提供的課綱檔案格式，若有其他格式的檔案

則無法使用 python sdg\_process\_v2.py 2 這個指令將之存放到資料庫

---

# 透過資料庫的 view 查看結果：

1. v\_output : 查看每個課綱的所有類別分析結果

2. v\_process\_detail : 可透過 text\_code 細看每一筆 text\_content 的所有關鍵字出現頻率

3. v\_term\_frequency：所有關鍵字出現的頻率列表（頻率由高到低&照類別排序）

4. v\_term\_length\_less\_than\_1：單詞長度為一個字的列表

5. v\_term\_not\_used：沒有被使用到的關鍵詞列表（照類別排序）

# input 欄位解析：

id - 編號

text\_code - 課綱的「學年度-學期」\_「課程代號」\_「班次」(如果有班次)

class\_name - 課程名稱

teacher\_name - 授課教師名字

text\_content - 「課程大綱課程概述」+「課程目標」

text\_length - text\_content 長度

# output 欄位解析：

id - 編號

text\_code - 對應到 input 的 text\_code

category\_id - 對應到 sdgs\_category 的 id

score - 該文章在 category\_id 這個類別的 關鍵字\*出現頻率\*權重 總和

language - 用以分析的關鍵字語言

create\_time - 分析時間，若有多次分析結果可由時間做區別

# process\_detail 欄位解析：

id - 編號

text\_code - 對應到 input 的 text\_code

term\_id - 對應到 sdgs\_term 的 id

frequency - 出現次數

language - 用以分析的關鍵字語言

create\_time - 分析時間

# sdgs\_term 欄位解析：

id - 編號

category\_id - 對應到 sdgs\_category 的 id

term - 關鍵字內容

weight - 權重

language - 語言

# sdgs\_category 欄位解析：

id - 編號

name - 類別名稱