HW4\_第11組

工作流程名稱(讀書會last\_朱)

選擇「PTT爬蟲」

看板=政黑版

期間=2024/05/20-2024/05/23

一張含有 螢幕擷取畫面, 文字, 圖表 的圖片

自動產生的描述

一張含有 文字, 軟體, 數字, 行 的圖片

自動產生的描述

* 替換字串

整理PTT常用格式進行一張含有 文字, 數字, 行, 字型 的圖片

自動產生的描述

* 中文斷詞

定義詞彙包含有：

台灣民眾黨 500

民眾黨 500

國民黨 500

民進黨 500

黃國昌 500

徐巧芯 500

韓國瑜 500

柯文哲 500

一張含有 文字, 軟體, 螢幕擷取畫面, 數字 的圖片

自動產生的描述

* 清除停用詞

清除英文字母、停止詞包含網址/轉錄/現在/今天/記者/台北/台中/希望等

一張含有 文字, 螢幕擷取畫面, 數字, 行 的圖片

自動產生的描述

* 轉為DTM

一張含有 螢幕擷取畫面, 文字, 軟體, 電腦圖示 的圖片

自動產生的描述

* 字詞網路圖

取不同的詞彙關聯度範圍

一張含有 文字, 地圖, 螢幕擷取畫面, 圖表 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 地圖, 圖表 的圖片

自動產生的描述

* 單中心網路圖

這邊我們有匯入關鍵字字典

一張含有 文字, 軟體, 數字, 電腦圖示 的圖片

自動產生的描述

* 透過DTM匯出CSV檔取得詞彙篩選前200個，設定自定義字典並匯入

一張含有 文字, 行, 數字, 平行 的圖片

自動產生的描述

* 區分幾個大類：中共、台灣、國民黨、民眾黨、民進黨、立法院、學運等關鍵類別

一張含有 圖表, 螢幕擷取畫面, 行, 設計 的圖片

自動產生的描述

民眾黨的次層詞類包含有國昌、黃國昌、柯文哲、民眾黨、阿北等

一張含有 圖表, 螢幕擷取畫面, 行, 設計 的圖片

自動產生的描述

國民黨的次層詞類包含有傅崑萁、國民黨、藍營、韓國瑜、徐巧芯等

一張含有 螢幕擷取畫面, 圖表, 行, 設計 的圖片

自動產生的描述

民進黨的次層詞類包含有民進黨、柯建銘、綠營、綠粉、蔡英文、賴清德等

一張含有 圖表, 螢幕擷取畫面, 行, 設計 的圖片

自動產生的描述

* 留言萃取

一張含有 文字, 軟體, 螢幕擷取畫面, 網頁 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

* 欄位篩選

選出文章回覆者

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

得知每篇文回覆者帳號

一張含有 文字, 螢幕擷取畫面, 數字, 軟體 的圖片

自動產生的描述

* 欄位篩選

取得發文者帳號

一張含有 文字, 螢幕擷取畫面, 軟體, 行 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 字型 的圖片

自動產生的描述

* 合併資料

透過system\_ID把欄位篩選合併

一張含有 螢幕擷取畫面, 文字, 軟體, 數字 的圖片

自動產生的描述

可以取得每一篇發文者、對應的回覆帳號與回覆數量

一張含有 文字, 螢幕擷取畫面, 數字, 軟體 的圖片

自動產生的描述

* 分群彙總

計算每一個帳號在期間內回覆對象與回覆次數

一張含有 文字, 軟體, 螢幕擷取畫面, 電腦圖示 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 數字, 軟體 的圖片

自動產生的描述

* 社會網路圖

**一張含有 文字, 數字, 行, 字型 的圖片

自動產生的描述**

**一張含有 文字, 螢幕擷取畫面, 地圖, 圖表 的圖片

自動產生的描述**

**一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述**