

組別資料

- 組別：第一組
- 成員(組長未定)：劉語萱 09170101、陳柏尹 09170109、廖曉琇 09170131

專案目的

- 目標：匯出檔名為論文名稱&內容為僅包含該資訊的 PDF 檔

完整流程	舉例
匯入/選取「檔名非論文名稱」的 PDF 檔	匯入/選取 123.pdf
用 Python 篩選出與「主要」論文相關的論文資料	第 1 頁：論文 A 的最末頁 第 2~16 頁：論文 B 的內容 第 17 頁：論文 C 的起始頁
刪除其他雜訊	刪除第 1 頁、第 17 頁的內容 保留第 2~16 頁
匯出 PDF 檔案（檔名為論文名稱）	匯出 B.pdf

專案時間

- 4/20—8/10，共 16 週

預計進度

- 4/27~5/11：進行 EDA 分析，制定初步的處理方向
- **5/11：進行進度報告（需有畫面分享）**

本次任務

項目	說明
EDA 分析－了解 PDF 的格式	了解 PDF 的基本格式、資料特性(如：Meta)
EDA 分析－了解論文格式	了解各論文格式(如：關鍵字、引用)
Python－適合處理論文的方式	找適合處理論文的 PY 開源套件

※以上皆尚未確認執行人

第二次會議 | 時間待定

- 參與者：全體組員(3)
- 議程：

選組長&建立專題的 Github 專案(需含全體組員&講師)
確認每週固定會議的相關資訊
完成第一次的任務分派

其他

1. 建議可用 PDF 的 Meta Data 抓出論文名稱
2. 由於論文格式無統一，故格式(如：關鍵字前為標題)僅能作為判斷之一，不能僅以此為判斷方式
3. 最後編輯時間：2022/4/27 21:20