結果說明

	accuracy	onlyGPT	onlyGPTfull	onlyHKBDB	onlyHKBDBfull
score	0.9016041	3.9355556	0.7653061	0.8978676	1.0544218

- 所有匹配到的欄位其準確率 (相似度) 為 90.16 %
- 平均每個表單 GPT 能多擷取 3.93 筆欄位
- 平均每個表單 GPT 能多擷取 0.76 筆 "事件" (也就是二維資料的筆數)
- 平均每個表單 HKBDB 多 0.89 筆欄位
- 平均每個表單 HKBDB 多 1.05 筆 "事件"

已知

- onlyHKBDBfull 較高主要集中在 "publication" 表單, 因 HKBDB 包含 "hasEditor" 的資料，若此些資料不納入計算，可降低
- accuracy 只算計匹配到的資料 (GPT 及 HKBDB 都有，且欄位名稱相同)
- onlyGPT 計算 GPT 多的欄位 (GPT 有, HKBDB 沒有 (既定欄位情況下)) (full 代表多的整筆資料)
- onlyHKBDB 計算 HKBDB 多的欄位 (GPT 沒有, HKBDB 沒有 (既定欄位情況下)) (full 代表多的整筆資料)
- 某些表單 GPT 沒有, HKBDB 沒有，此情況 (代表 GPT 無亂擷取 (一維資料)) 無納入計算

計算公式 (accuracy)

- 完全匹配: 一模一樣才給分
  - (台灣教育局, 台灣教育所) = 0
  - (台灣教育局, 台灣教育所) = 100
- 是否包含: 只要有一方包含在內，即得滿分，否則獲得相似分數
  - (金融學系, 金融系) = 100
- 無序匹配: 以空格為斷行, 應用在如筆名 (多個答案放在 list, 順序不重要)
  - ([查理, 姚馥蘭], [姚馥蘭, 查理]) = 100

計算方式 (onlyGPT & onlyHKBDB)

- 一維資料，一欄位算一分
- 二維資料分兩種計分
    - 同一筆事件中多填的欄位，一欄位加一分
    - 若是直接多一整筆事件，加一分，記在 full 表單

compare flow

- 刪除空白的 sheet
- 刪除無 "關鍵欄位" 欄位的二維資料，如 "article" 表單有大量無 "文章名稱" 的資料，會將其刪除
- 判斷目前計算的 sheet 屬於一維還二維資料
    - 一維: 從 data&ans 取得 key 對應的值,
        - 若兩種都取到，計算相似度 (accuracy)
        - data 有, ans 沒有, 計算 onlyGPT
        - data 沒有, ans 有, 計算 onlyHKBDB
    - 二維: 因二維是由多個事件構成，彼此沒有順序, 直接比對可能會比錯筆資料如
GPT	HKBDB
Data 1	Data 2
Data 2	Data 1

        - 先找到 data 的關鍵欄位 (可能像是學校名稱、公司名稱)
        - 比對 ans 所有的關鍵欄位，以找到相關事件
        - 接著比對方法如一維資料
- 若有某一 excel 少對方 sheet, 走 onlyGPT or onlyHKBDB 計算