

Audience Issue 2

RD2 黃彥鈞

2022-03-30

大綱

- Audience-API v 2.3
 - 改版目標
 - 改版項目
 - 操作流程
 - 實際測試
- Audience Django Toolkit (Site)
 - 修改目標
 - 修改項目
 - 操作流程
- 注意事項
 - 貼標測試須知

Audience-API v 2.3

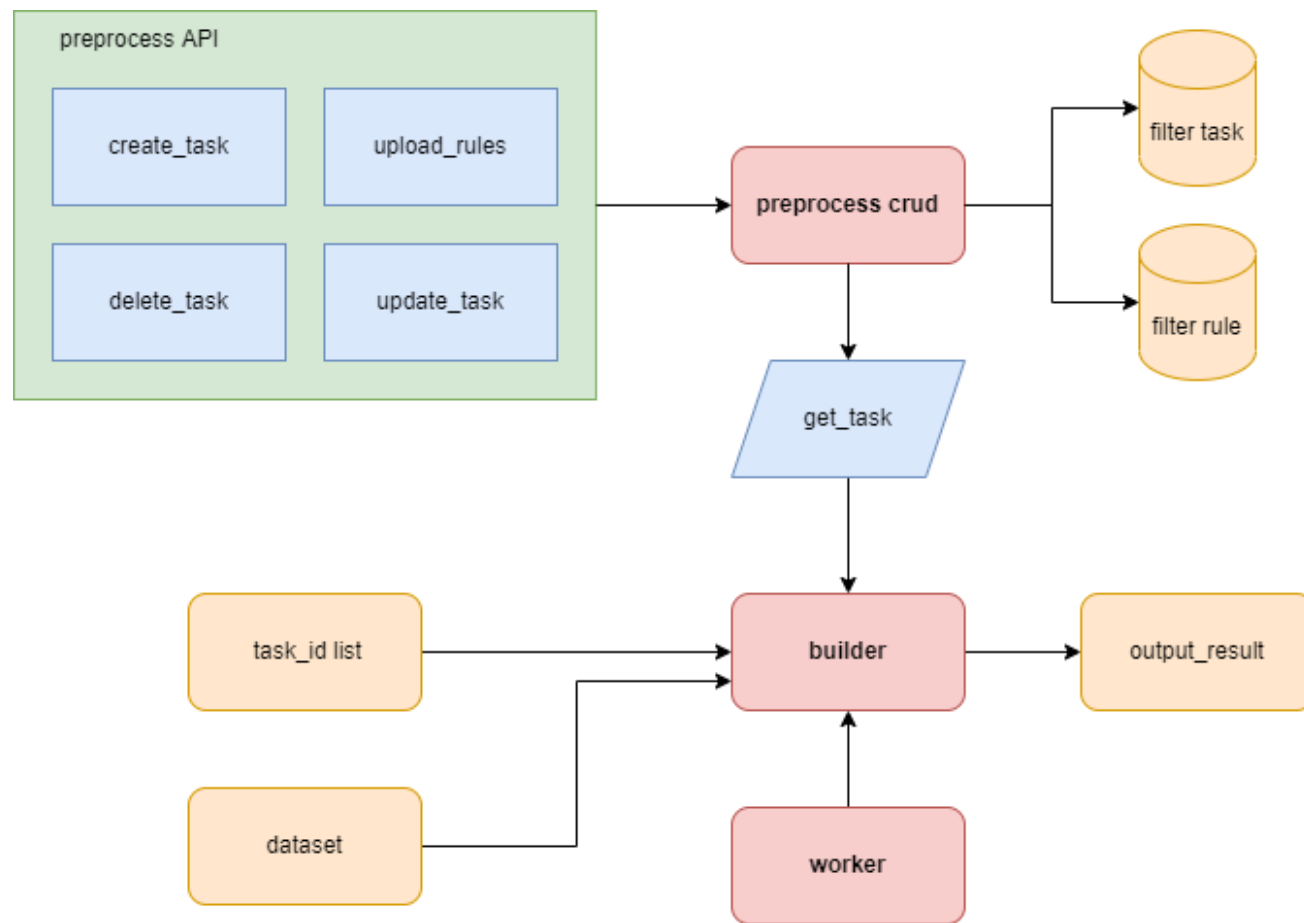
改版目標

- 因應 AS1 2020年提出的需求，預期新增前處理模組，使用關鍵字與正則表達式模型，貼標或是訓練模型前，排除或修改特定內容資料。

改版項目

- 新增
 - PreprocessCRUD 負責前處理模組任務和規則資料庫操作
 - DataFilterWorker 執行前處理模組任務
 - DataFilterBuilder 串接前處理模組任務與模型任務與貼標任務
- 修改
 - PrepressingWorker 加入前處理步驟
 - PredictWorker, get_batch_by_timedelta 加入前處理步驟

操作流程1



操作流程2

- AS1: 提供 RD2 希望排除的規則資料表 (.csv)
- 資料表須包含以下欄位:
 - content: 規則內容
 - rule_type: 規則類型 (請 關鍵字 或 正則表達式 擇一)
 - label: 標籤 (請 刪除 或 修改 擇一)
 - match_type: 比對方式 (請參考 站台 > 資料整備任務 > 任務列表 > 關鍵字規則)
- 請將規則拆成四個檔案傳給 RD2，檔案 rule_type - label 獨立，如:
 - A 檔案包含所有 關鍵字 - 刪除 的規則資料
 - B 檔案包含所有 關鍵字 - 修改 的規則資料
 - C 檔案包含所有 正則表達式 - 刪除 的規則資料
 - D 檔案包含所有 正則表達式 - 修改 的規則資料

操作流程3

- RD2 透過 preprocess API 建立前處理任務，匯入前處理規則
- 在 settings 配置前處理任務 id 列表

實際測試

- **Wh_backpackers** 資料庫，執行 6 個(正則式內文與關鍵字作者)為期一年範圍的貼標任務，最長執行時間 27 分鐘。
- 其他測試須知請詳閱**貼標測試須知**，若有任何疑問請聯絡 RD2 負責人。

Audience Django Toolkit (Site)

改版目標

- 修改以下 BUG:
 - Labeling_jobs 規則列表 翻頁 BUG
 - Rd2demo 跨域問題
 - Modeling_jobs 部分內容無法正常顯示

修改項目

- 修改項目
 - Modeling_jobs tasks 新增發送請求任務
 - Modeling_jobs views 新增串接 tasks class
 - Modeling_jobs urls 新增 term-weight url
 - Modeling_jobs job_details 修改 AJAX url 內容
 - Base 模板 新增 Content-Security-Policy tag
 - Settings 新增 SECURE_PROXY_SSL_HEADER 等設定

操作流程

- 站台流程無修改，請參考 Audience 站台前兩份 Slides

注意事項

貼標測試須知

- 請 AS1 夥伴先試貼 YT 等大型資料庫，範圍 **3 個月**，一次執行 **3 個**貼標任務。觀察貼標時間，與是否成功，若有狀況請通知 RD2 負責人。
- 以上無問題，同上述步驟，再行試貼 YT 等大型資料庫，範圍 **6 個月**，一次執行 **5 個**貼標任務。
- 若有單個任務執行超過 4 個小時，且站台無任何錯誤訊息，請通知 RD2 負責人，協助檢查後台。
- 若任務失敗 (貼標進度 **FAILURE**)，請將以下任務資訊提供給 RD2 負責人：
 - 任務編號、任務 ID、檢查點、錯誤訊息
- 現行前處理流程，AS1 夥伴若有更新前處理規則需求，請聯絡 RD2 負責人。
- 感謝夥伴配合，若產品有其他需求，請發送需求書給 RD2，或是來信與 RD2 負責人討論。