**社群媒體分析第九組第一次讀書會作業**

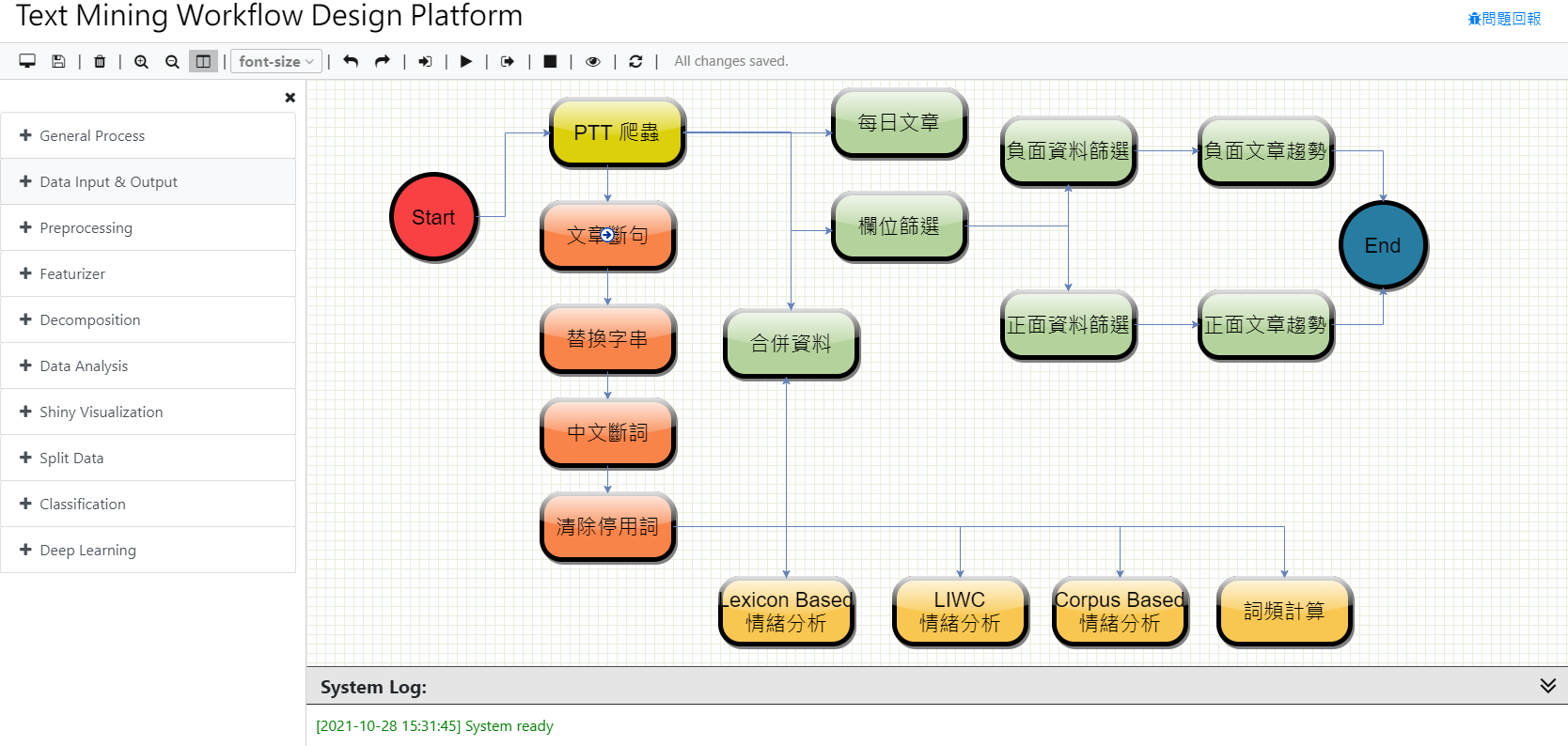
1. 分析熱門主題:魷魚遊戲
2. 使用平台:中山大學工作流程平台
3. 所使用的工作流程名稱:魷魚遊戲、建立時間2021/10/23 11:09:40

一張含有 桌 的圖片

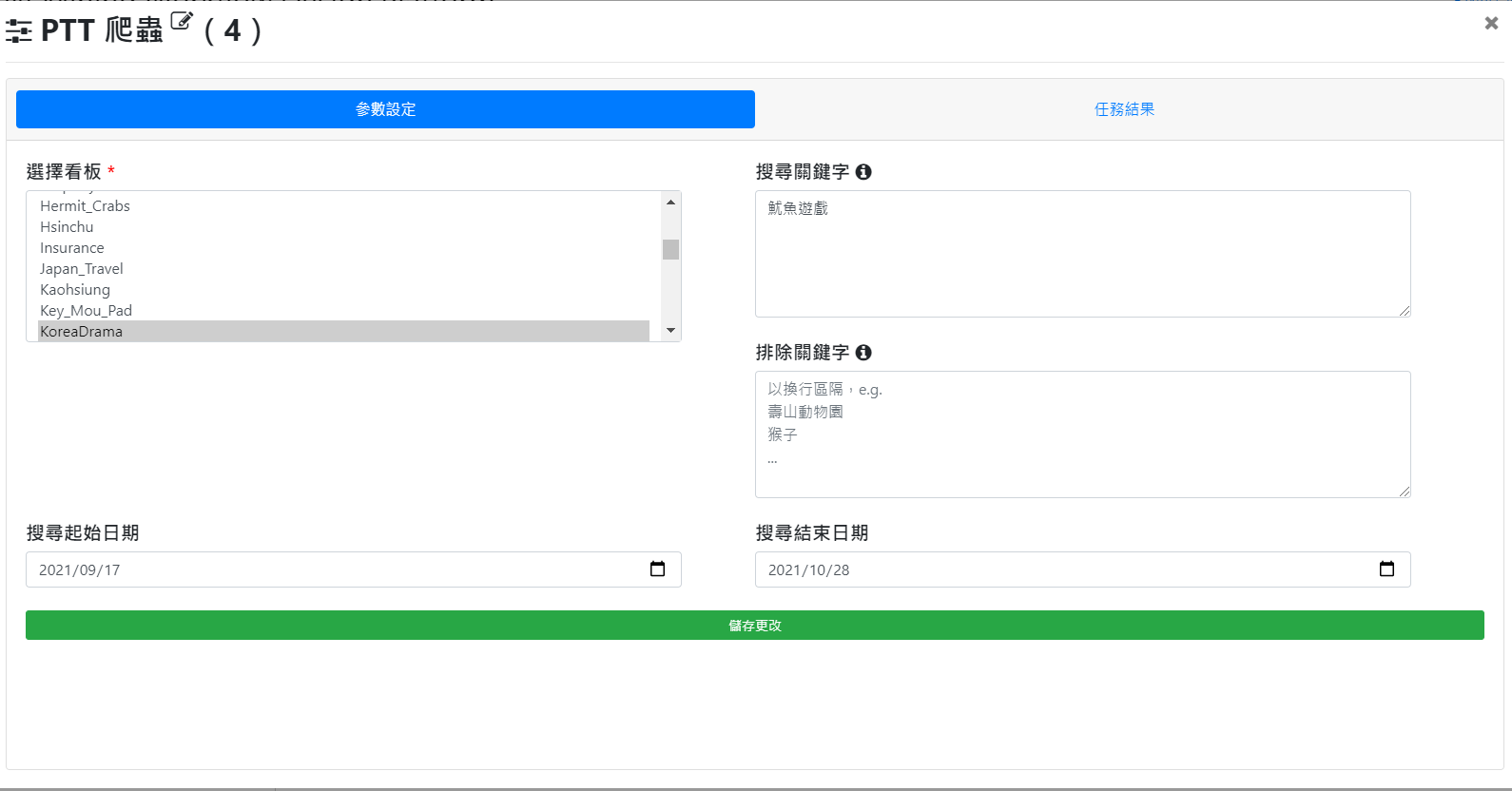
自動產生的描述

使用此工作流程

1. 魷魚遊戲工作流程圖:



1. 資料來源:使用工作流程平台的PTT爬蟲元件:
   1. 參數設定:選擇PTT八卦版及韓劇版，搜尋的關鍵字為:「魷魚遊戲」，搜尋資料起訖日期為:2021/09/17~2021/10/28。



* 任務結果:10個欄位，資料筆數1136筆

一張含有 文字 的圖片

自動產生的描述

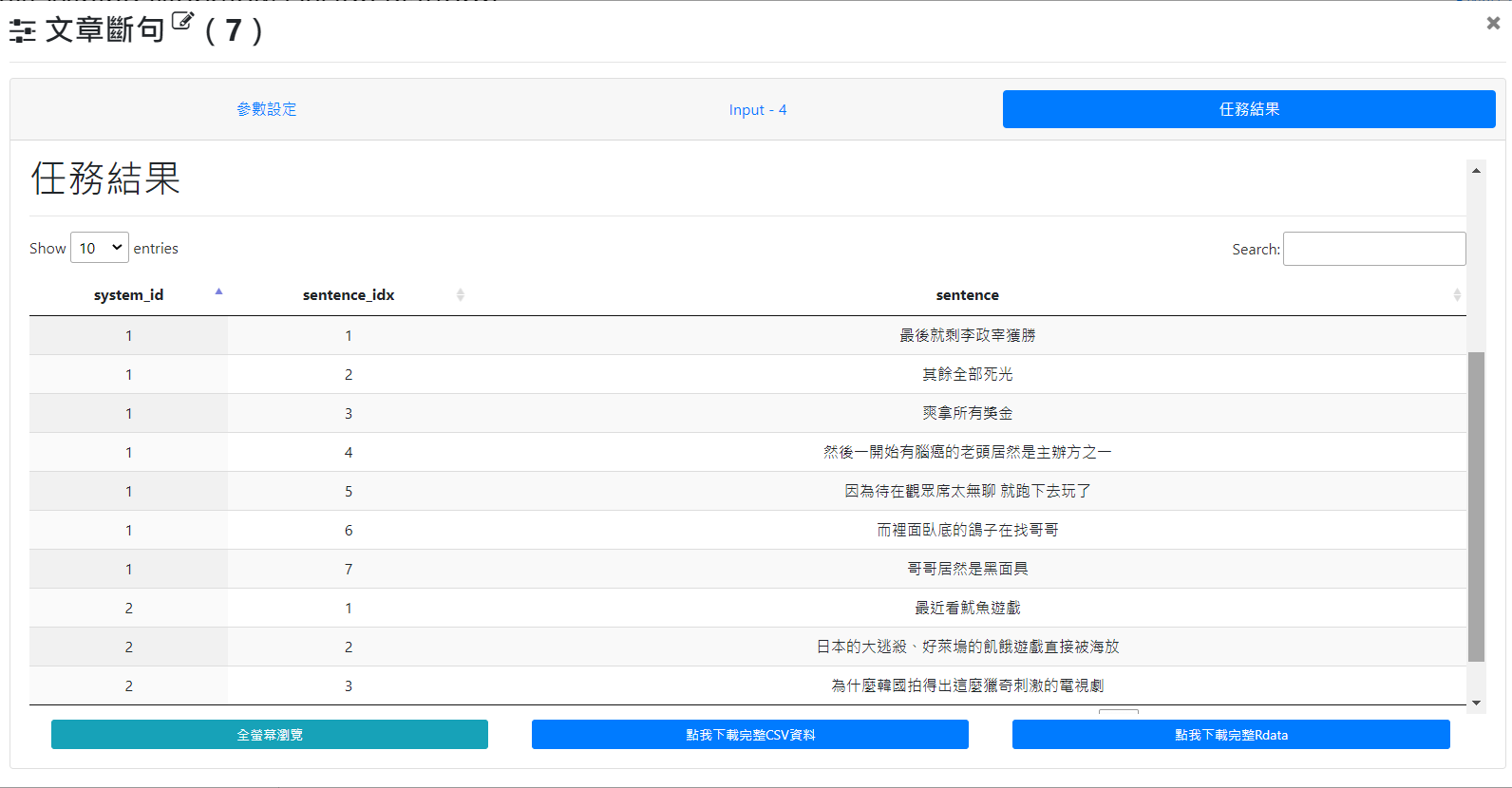
1. 資料清理:文章斷句、替換字串、中文斷詞、清除停用字
   1. 文章斷句:
      * 參數設定:



* + - 任務結果:

一張含有 文字 的圖片

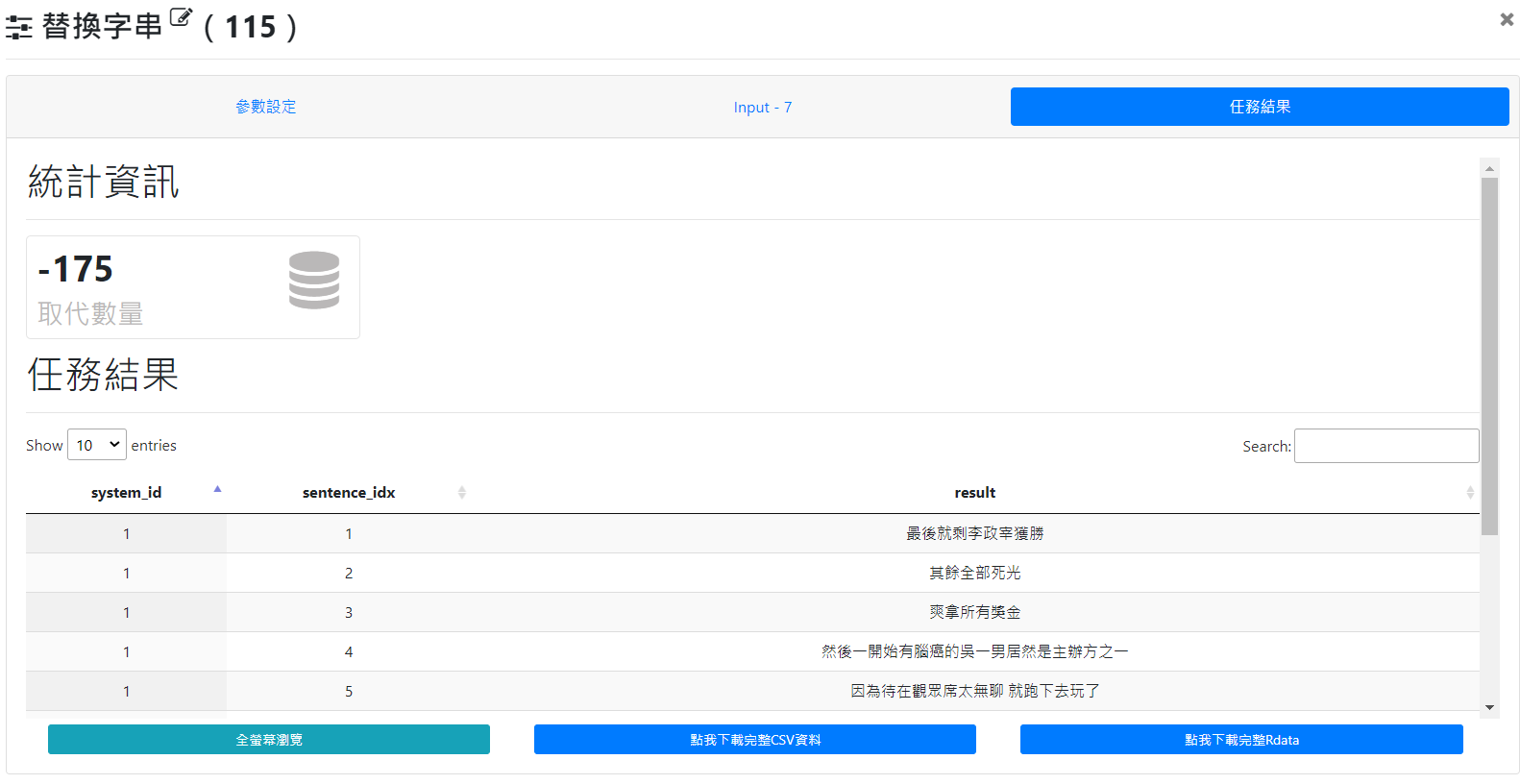
自動產生的描述



* 1. 替換字串:
     + 參數設定: 老頭>>吳一男、裝b仔>>自卑、BOSS>>老闆、YT>>Youtube、fb>>facebook



* + - 任務結果:共取代175個字詞





* 1. 中文斷詞:
     + 參數設定: 「主辦方」會斷詞為「主辦」及「方」、；「魷魚遊戲」會斷詞為「魷魚」及、「遊戲」；「吳一男」會斷詞為「吳一」及「男」….等等，所以在定義詞彙的參數設定，加重主辦方、魷魚遊戲及吳一男的權重(設為1000)

一張含有 文字 的圖片

自動產生的描述

* + - 任務結果:



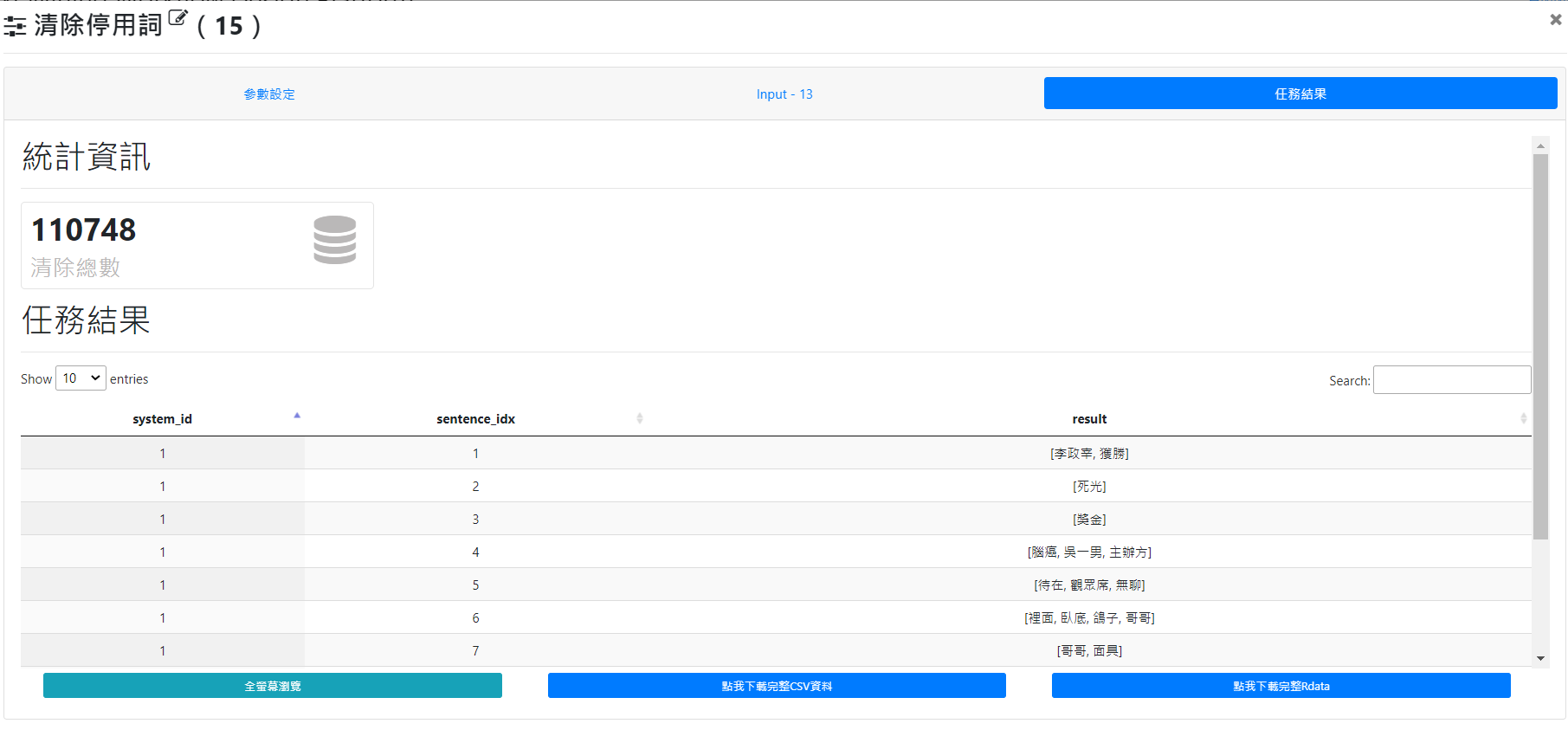
* 1. 清除停用字
     + 參數設定: 自定義停用詞1897字

一張含有 文字 的圖片

自動產生的描述

共1897字

* + - 任務結果:共清除110748個停用字



1. 情緒分析: 分別使用Lexicon Based、LIWC Based及Corpus Base三個元件進行情緒分析
   1. Lexicon Based情緒分析
      * 參數設定:使用預設情緒字典

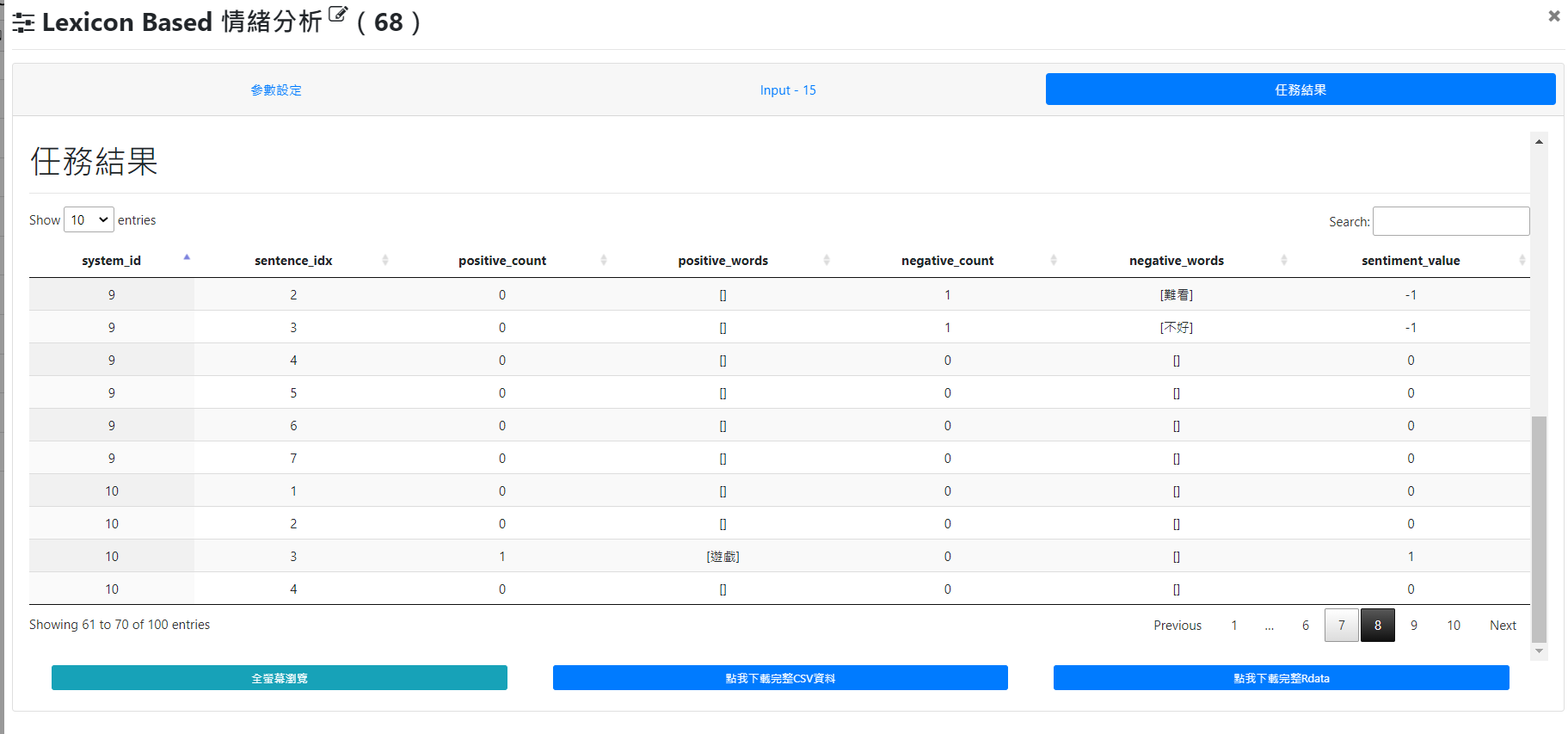
一張含有 文字 的圖片

自動產生的描述

* 任務結果:5個最大正向情緒、5個最大負向情緒

一張含有 文字, 螢幕擷取畫面, 室內 的圖片

自動產生的描述

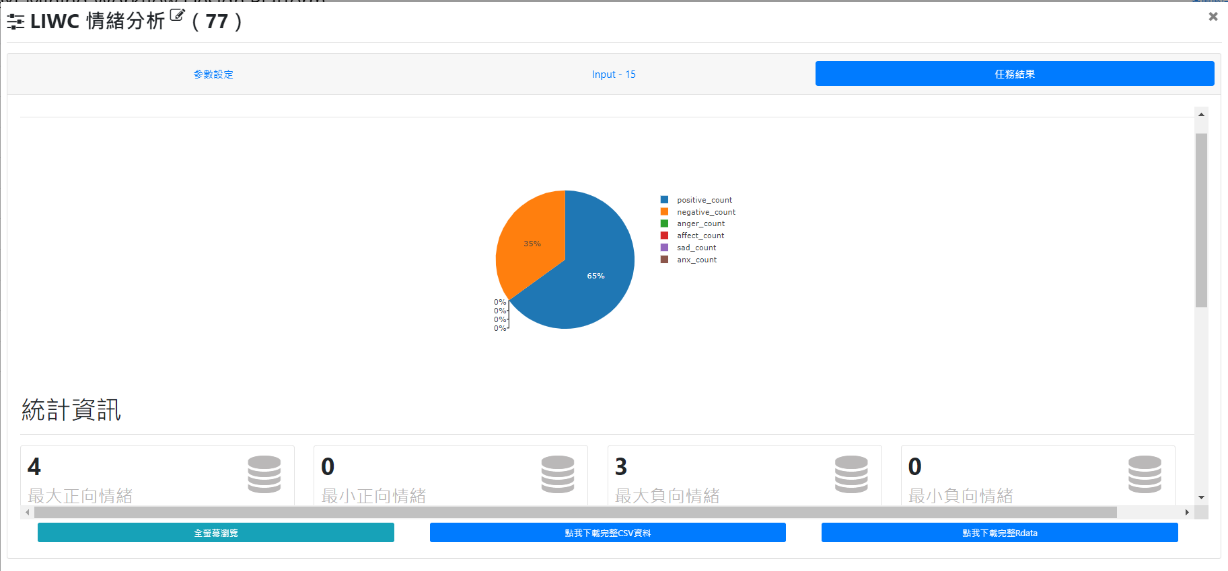


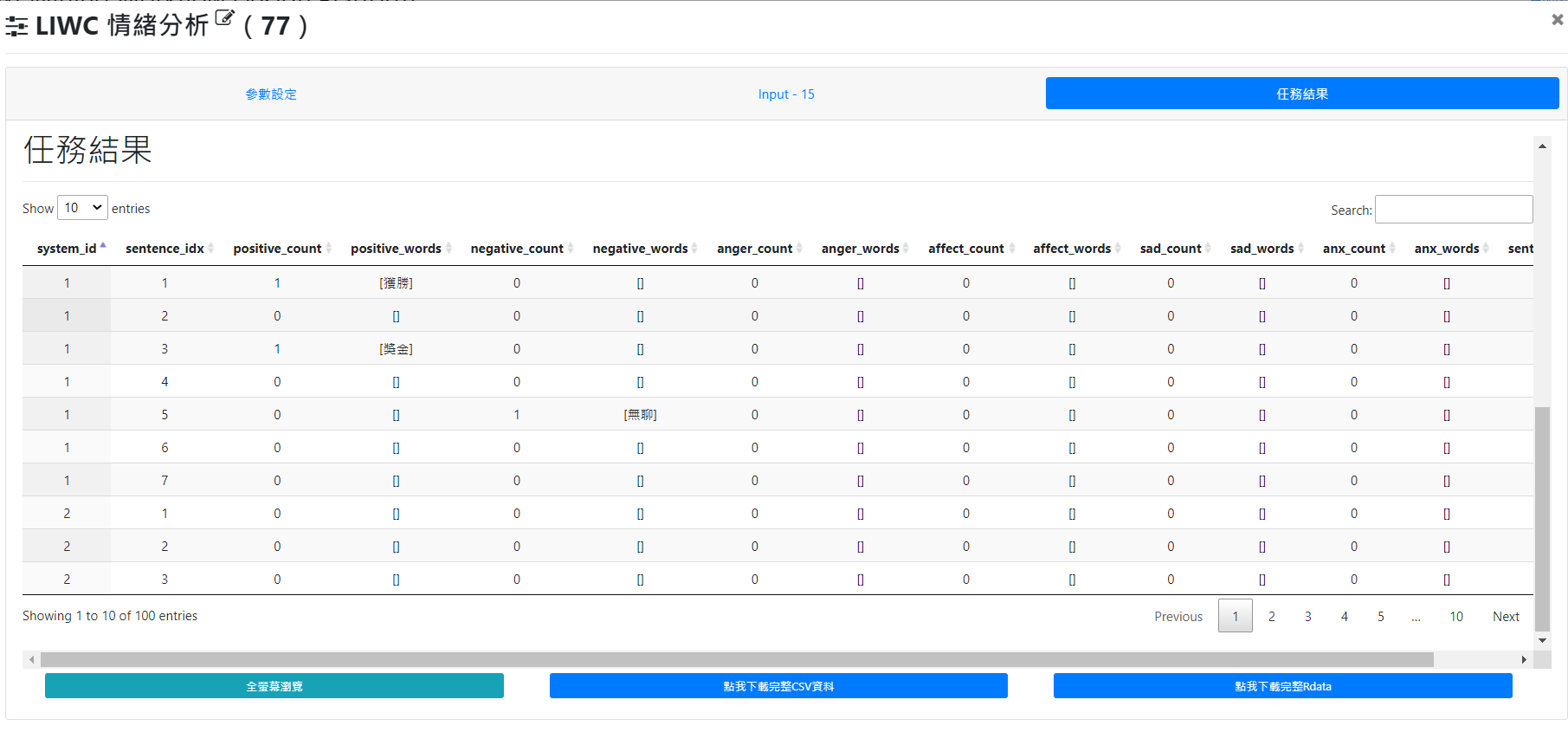
* 1. LIWC情緒分析
     + 參數設定:

一張含有 文字 的圖片

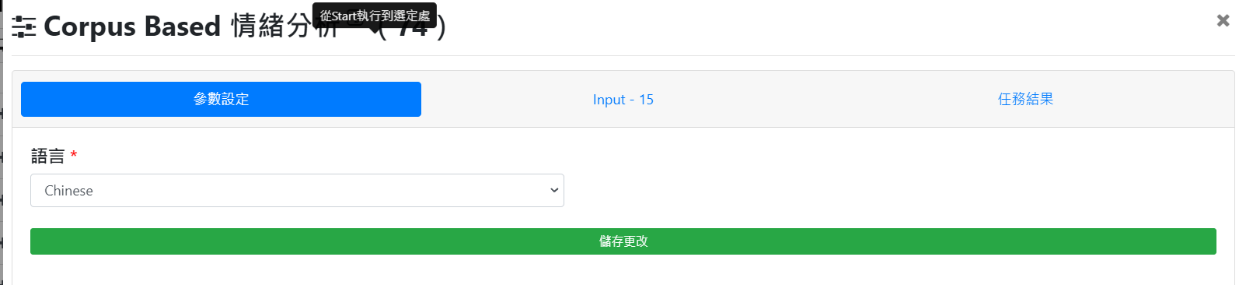
自動產生的描述

* + - 任務結果:4個最大正向情緒、3個最大負向情緒





* 1. Corpus Based情緒分析
     + 參數設定:



* + - 任務結果:



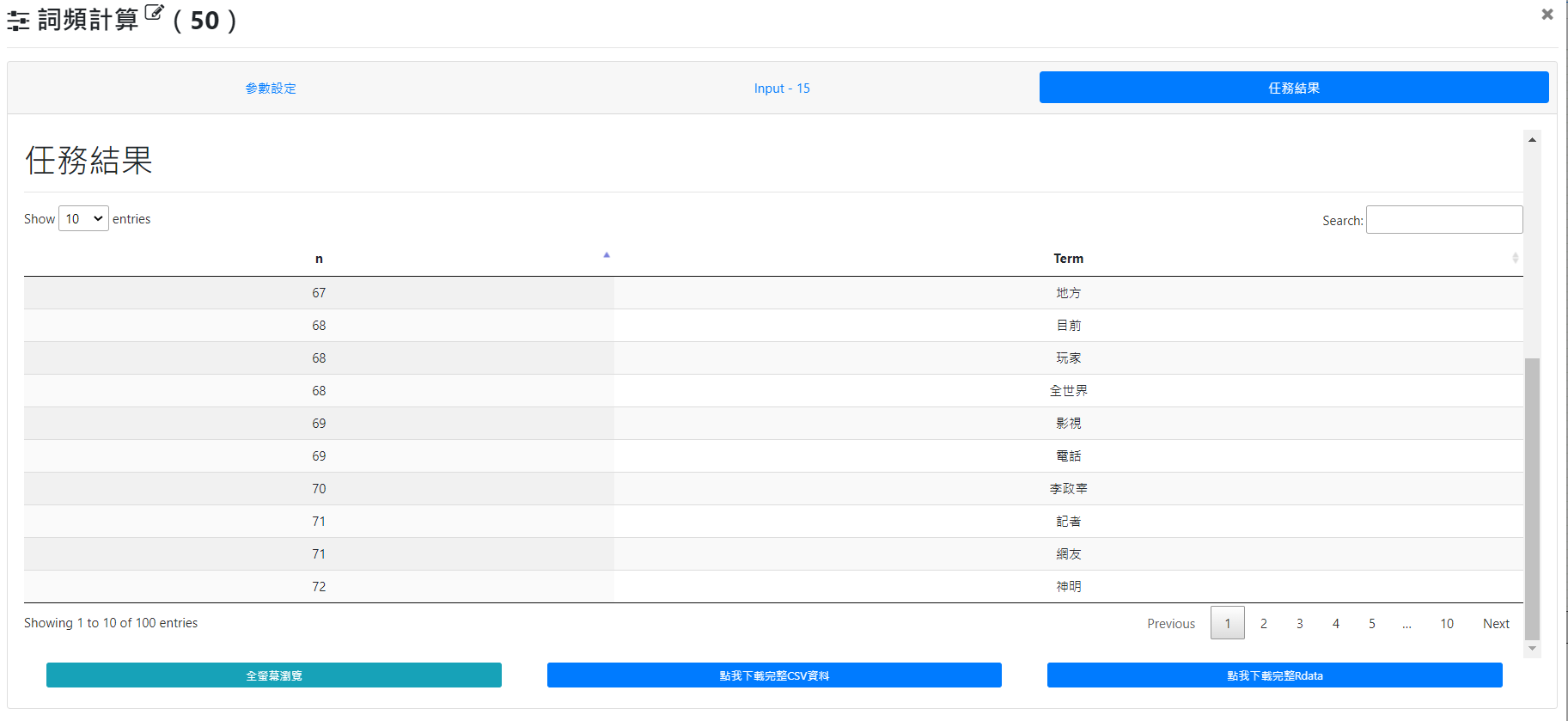
1. 詞頻計算:自定義停用詞有好像、目前、這部、剩下、討論、這種、有沒有、都是、參加…
   1. 參數設定:

一張含有 文字 的圖片

自動產生的描述

* 任務結果





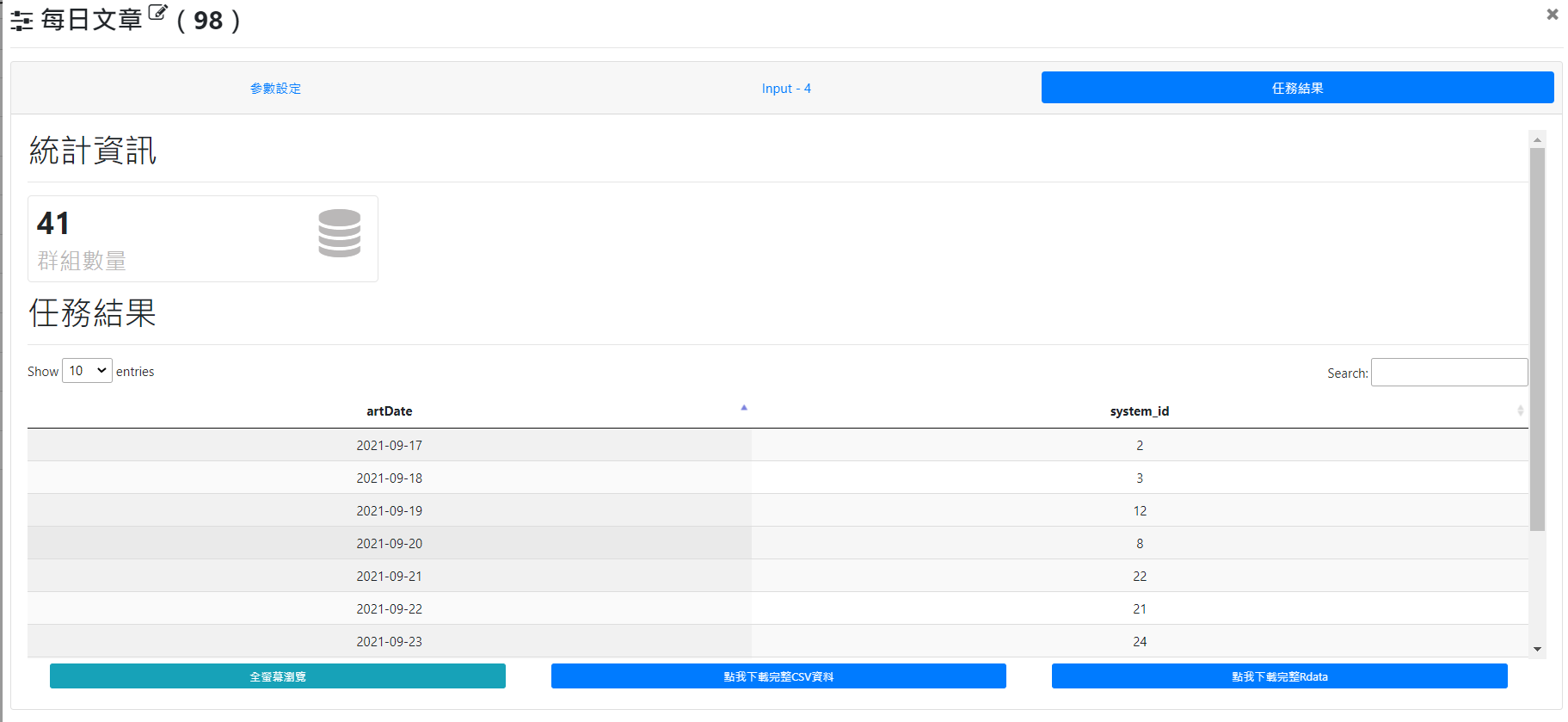
1. 每日文章

* 參數設定

一張含有 文字 的圖片

自動產生的描述

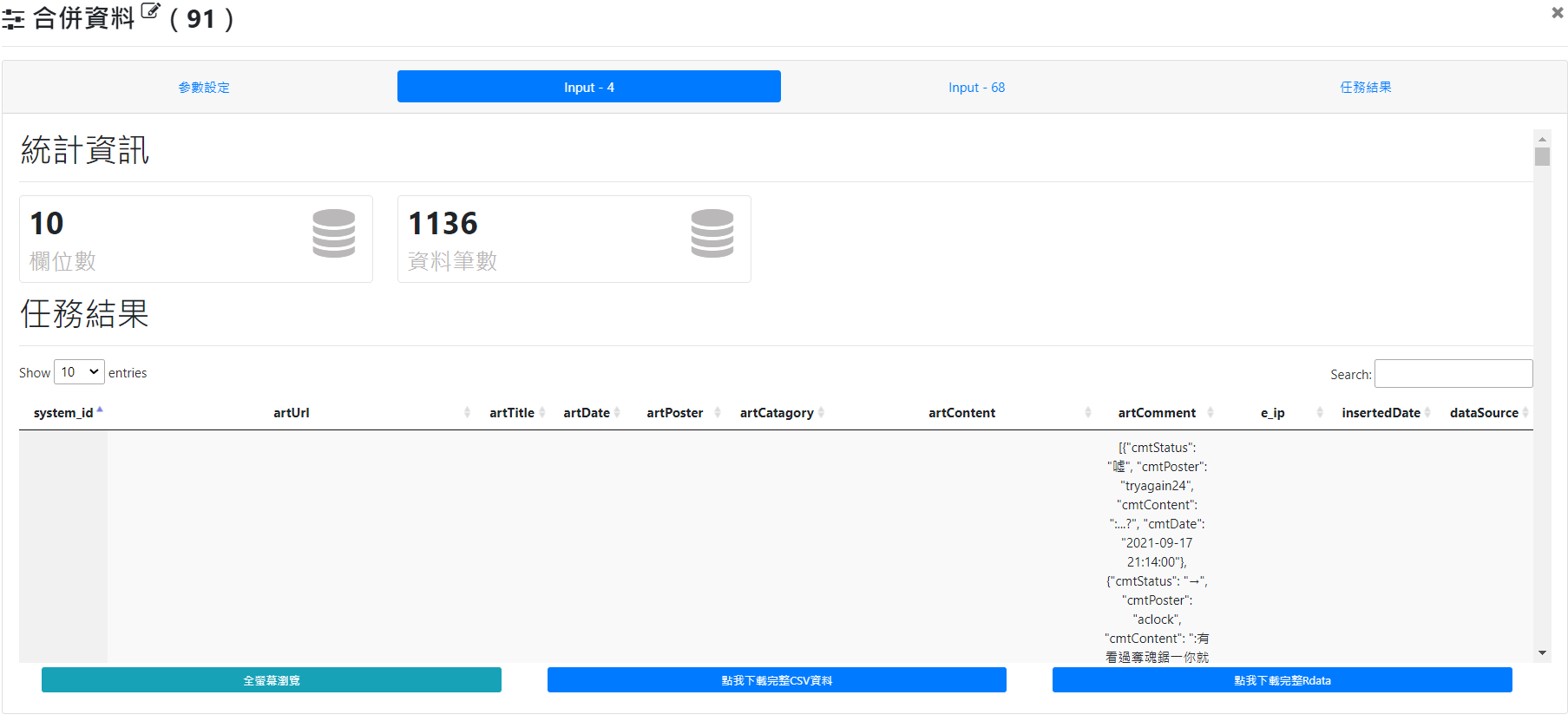
* 任務結果



1. 分群彙總
   1. 合併資料:
      * 參數設定:

一張含有 文字 的圖片

自動產生的描述



一張含有 文字, 螢幕擷取畫面, 室內 的圖片

自動產生的描述

* + - 任務結果:

一張含有 文字 的圖片

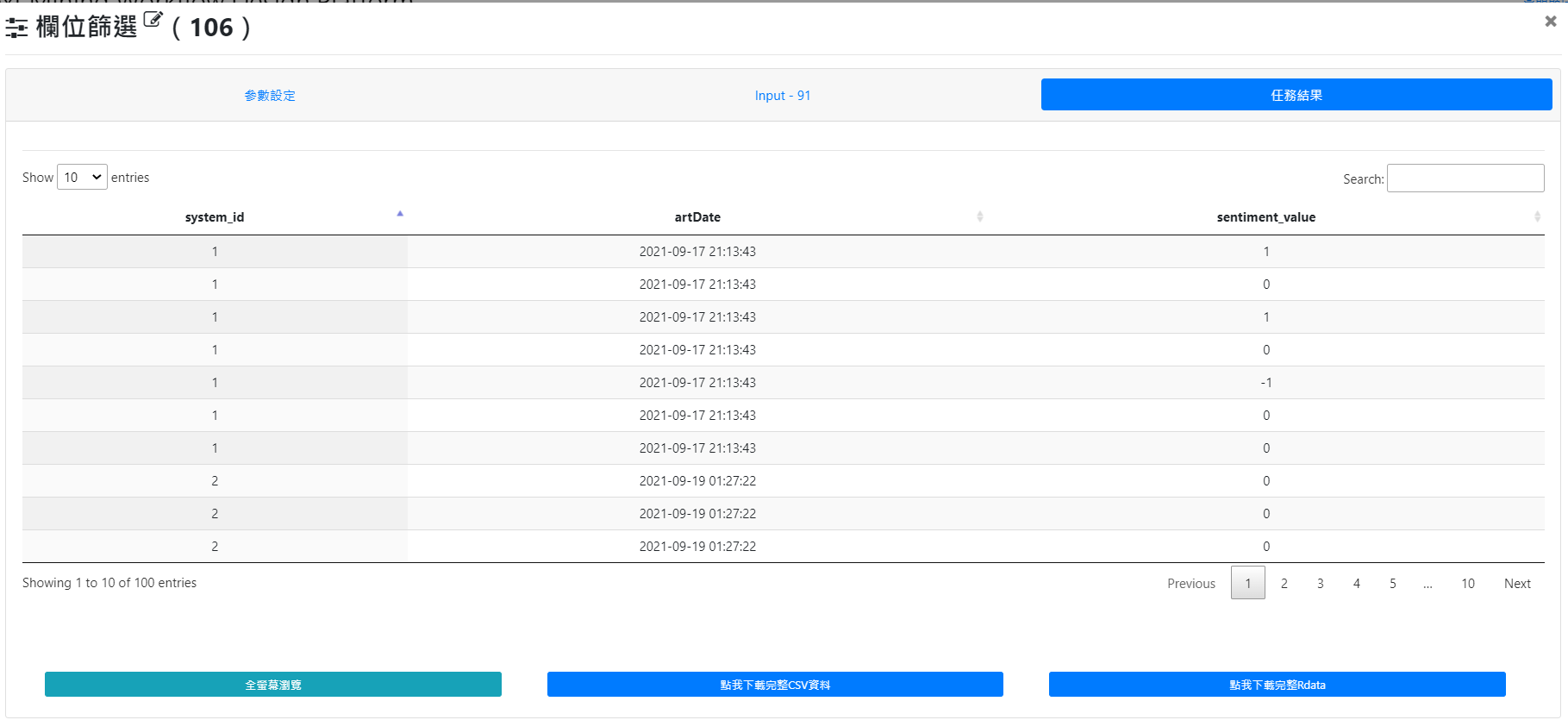
自動產生的描述

* 1. 欄位篩選:
     + 參數設定: 選取artDate和 sentiment\_value

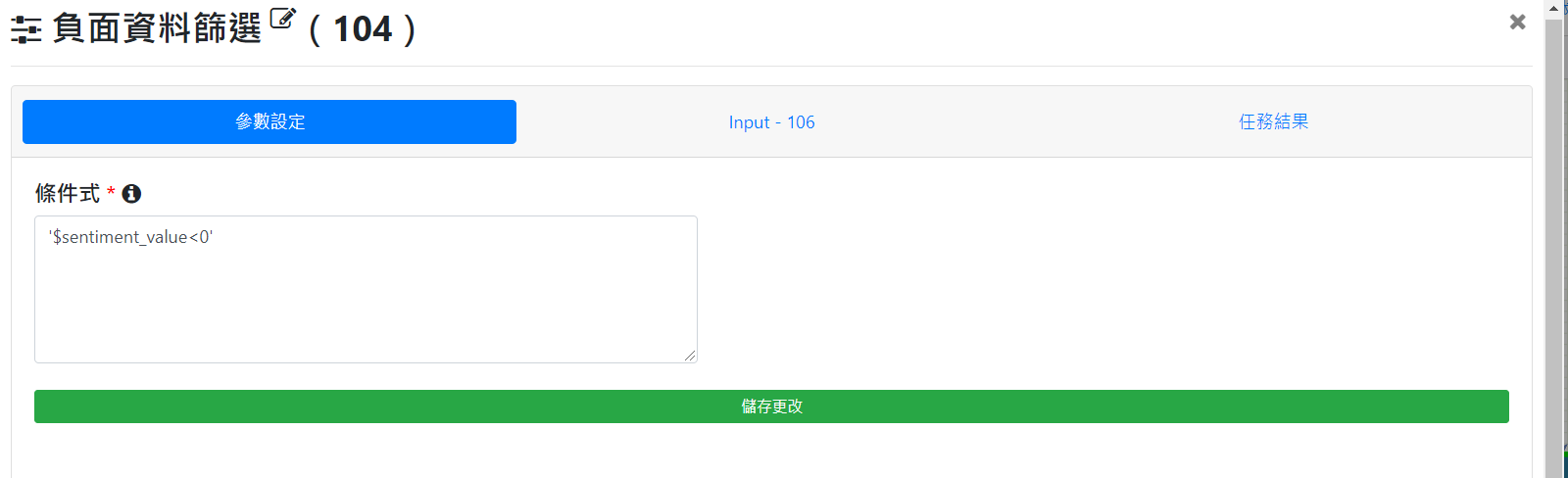
一張含有 文字 的圖片

自動產生的描述

* + - 任務結果:



* 1. 負面資料篩選
     + 參數設定: $sentiment\_value < 0



* + - 任務結果

一張含有 桌 的圖片

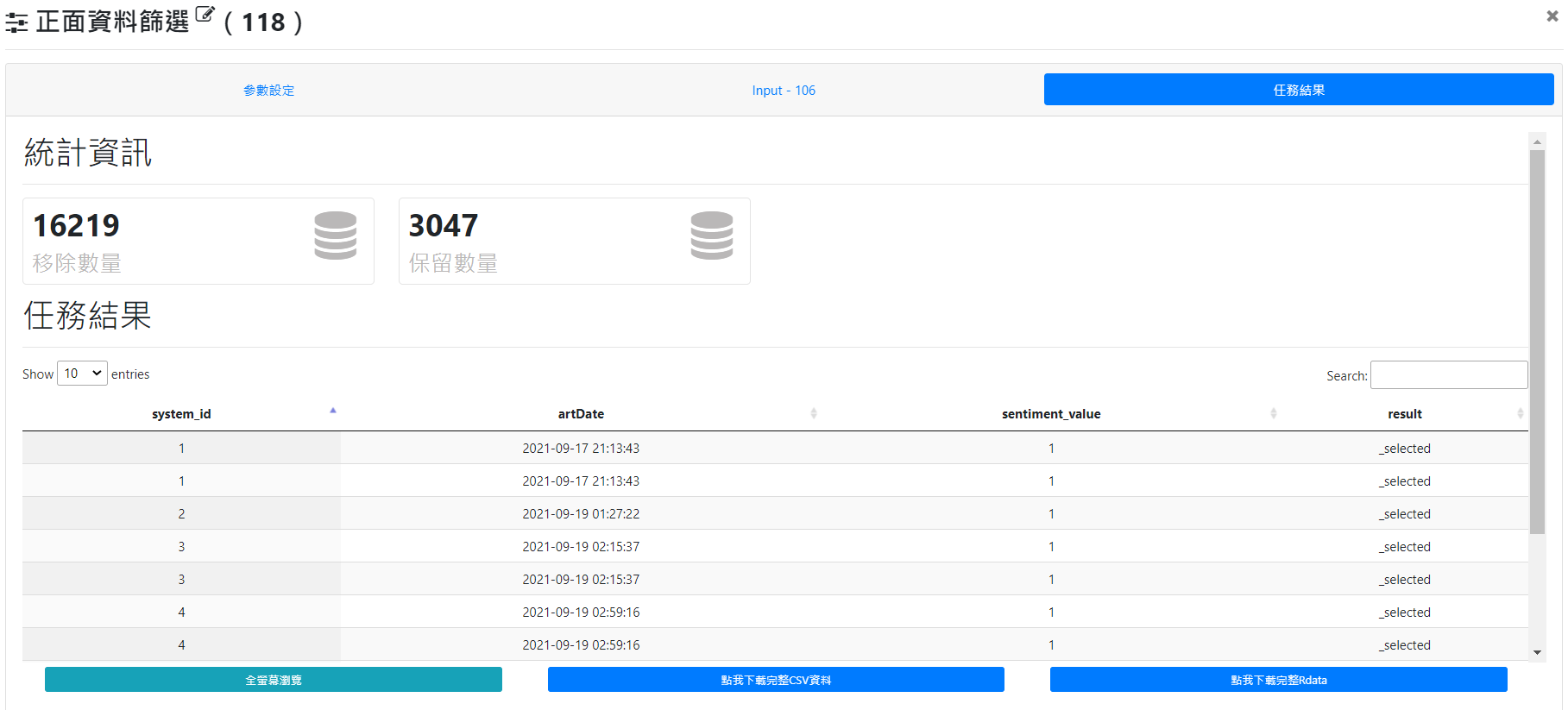
自動產生的描述

* 1. 正面資料篩選:
     + 參數設定: $sentiment\_value > 0

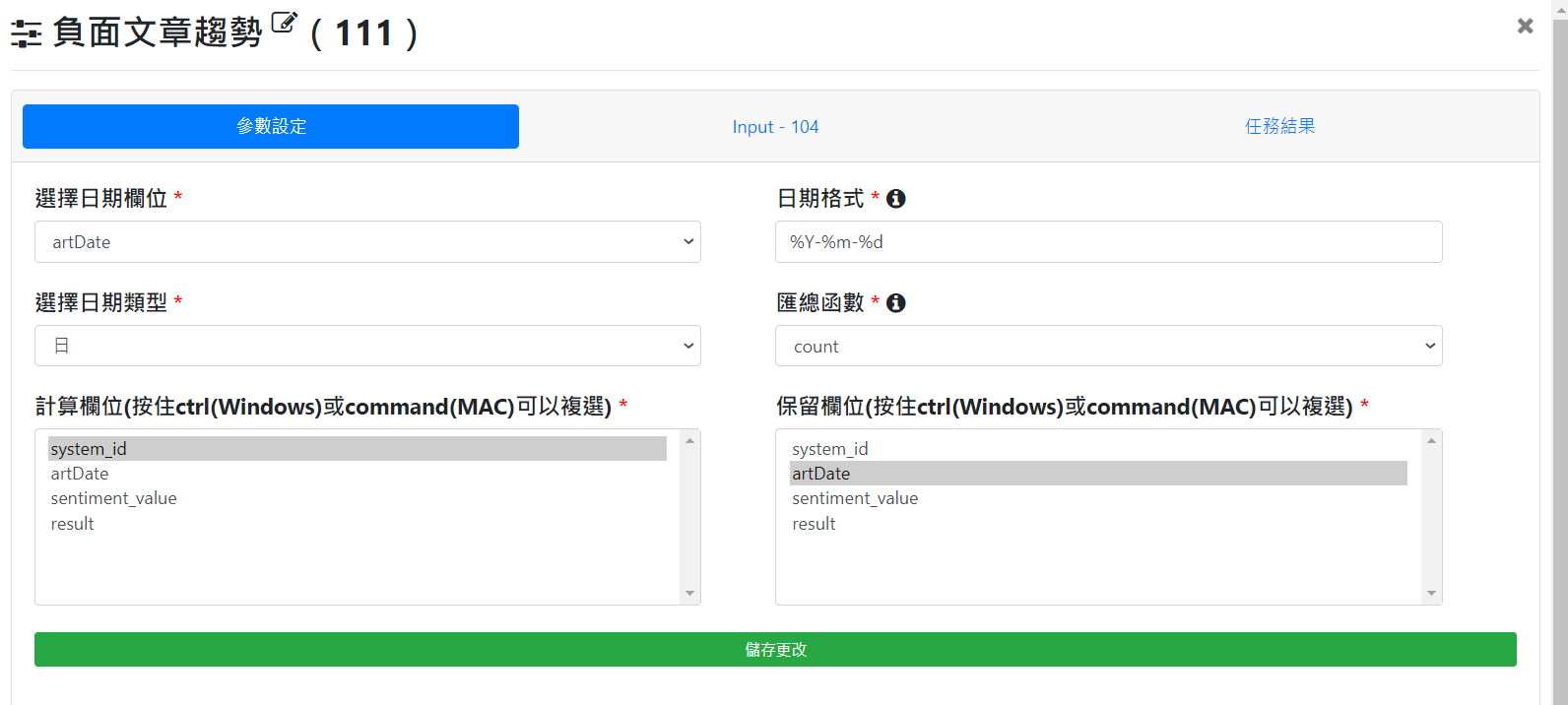
一張含有 文字 的圖片

自動產生的描述

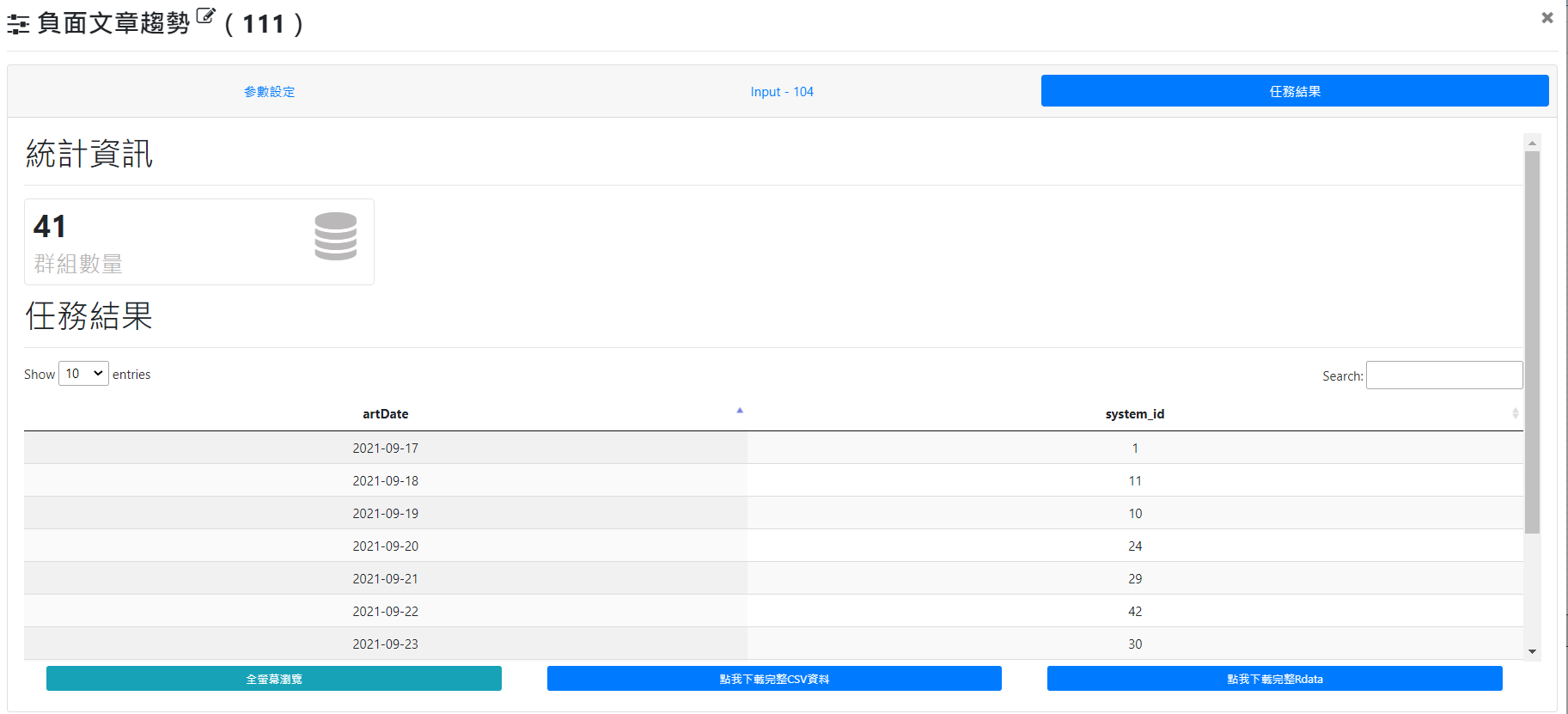
* + - 任務結果



* 1. 負面文章趨勢
     + 參數設定:



* 任務結果

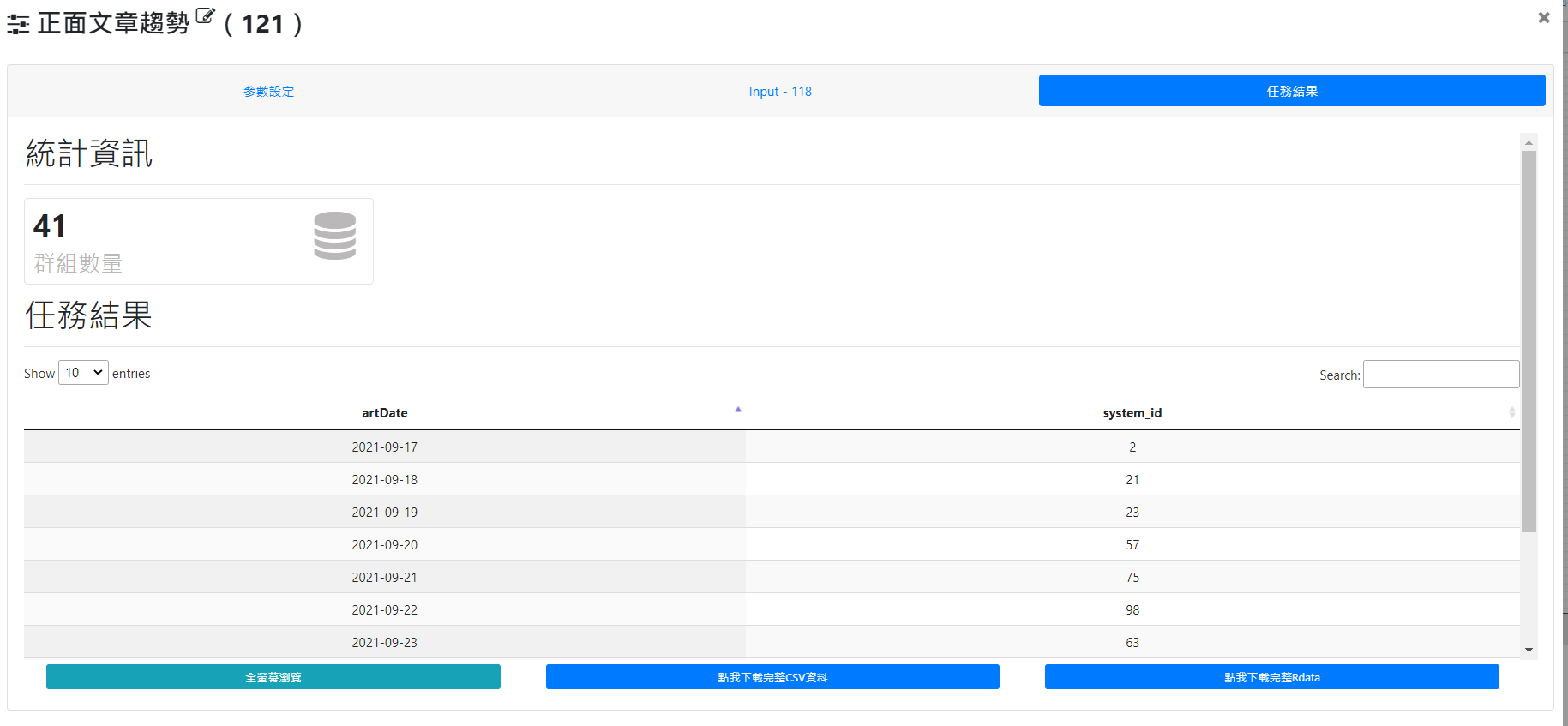


* 1. 正面文章趨勢
* 參數設定:

一張含有 文字 的圖片

自動產生的描述

* 任務結果:



1. 視覺化圖表
   1. 每日文章統計圖



* 1. 負面文章趨勢圖

一張含有 文字, 室內, 螢幕擷取畫面 的圖片

自動產生的描述

* 1. 正面文章趨勢圖



* 1. 詞頻計算文字雲



十二、結論

1. 由每日文章統計圖可發現，魷魚遊戲自9/17上映後，每日討論熱度持續攀升，至9/28到達頂點(共85篇)，之後每日聲量有所起伏，但仍維持每日20篇以上數量。
2. 從篩選之區間分析，每日正面情緒的討論度大於負面情緒，亦提升該影集被關注度，連帶影響不追劇用戶加入觀看及討論。
3. 透過文字雲分析，魷魚遊戲上映後，分別登上Netflix台灣、日本、韓國收視排行榜的冠軍寶座，也可以發現此議題對於上述等國家討論度較高，與另從talkwalker上分析討論熱度是一致的。

**本次作業心得**

* **資料來源的選擇:**

以PPT資料為例，如果以PPT所有版下去篩選關鍵字，不同版別抓出來資料筆數相差很多，如八卦版的資料遠比韓劇版的資料多、及以關鍵字會蒐集到文章內容提及「魷魚遊戲」，但整體內文較不相關(如food版店家推出碰糖商品) ，故可從資料分析的目的考量選取要抓哪個版的資料。

* **關鍵字定義:**

為避免干擾資料蒐集，如有熱門議題或流行詞彙，可從反覆觀察後調整或增加關鍵字，並以相關性的版面為主(如gossiping版、EAseries版、koreadrama版)，以同篇文章能出現多個關鍵字詞將能提高相關度，若資料來源及類型太多時只藉由增加關鍵字反而無法提高

* **資料清洗:**

初始資料經反覆調整得到有意義之資料時，須進行斷句、斷詞、清除停用字等作業，把無關字詞或調整適當權重，得到結果會比較正確，對比 talkwalker分析之文字雲，本小組呈現結果會跟主題比較相關。