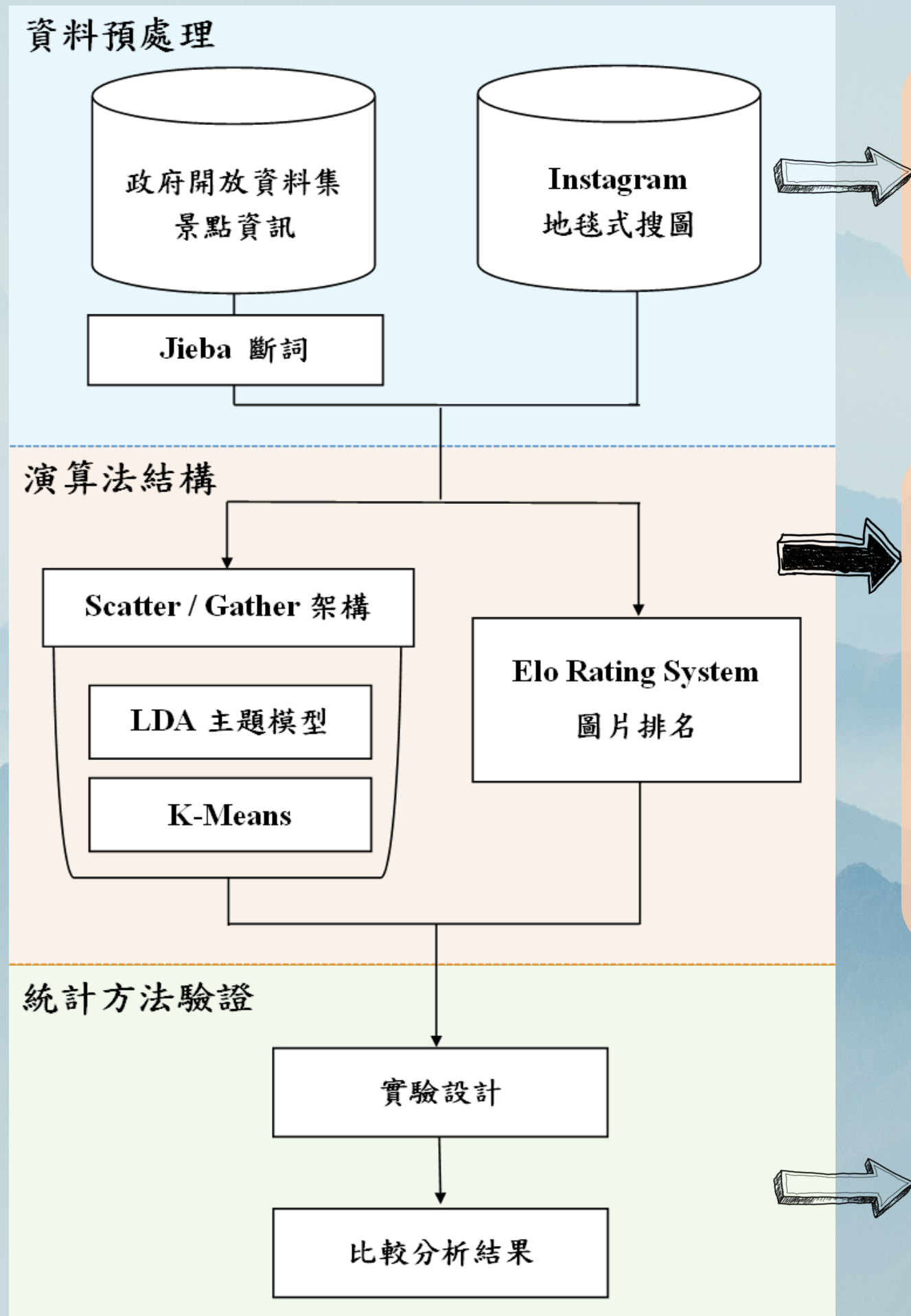


PICTURE YOUR 圖窮 WAY 景獻

- 景點推薦系統

- LDA
- KMEANS
- SCATTER / GATTER
- ELO RATING SYSTEM



系統特色

1. 推薦才貌兼併的全能型圖片

Jieba斷詞與LDA主題分布可以挑選出有內涵的圖片，而Elo Rating System則是挑出美麗的圖片，將兩者結合後才貌兼備的圖片就此產生。

2. 高隱私，低操作

透過各種演算的結合與正確的使用，得以讓使用者不需提供任何私人訊息，只需動動手指即可得到推薦結果。

3. 減少使用者搜索時間

不需要閱讀冗長的景點敘述，只需使用本研究的網頁系統，就能簡單且快速的提供使用者所感興趣的景點。

研究流程-1

在預處理的部分，本研究地毯式蒐集 **政府開放資料集** 中的景點於Instagram上的圖片，同時利用 **Jieba斷詞** 演算法，對資料集中的景點簡述進行預處理以利於後續的建模。

研究流程-2

第二部分則是將一般用於大量文件分群的 **Scatter/Gather概念** 運用在景點分群，並融合 **Latent Dirichlet Allocation主題模型** (簡稱LDA)與 **K-Means分群法** 作為其分群過程使用的演算法。從LDA主題模型挖掘各景點的潛在主題，並萃取各景點的LDA主題機率分布作為K-Means分群的變數，以取代過往文件分群使用的稀疏矩陣。

考量到除了景點類型外，景點美觀亦為社會大眾選擇景點時的重要因素，因此透過**Elo Rating System**反覆訓練所有使用的景點圖片並賦予景點圖片權重，以優化景點挑選之結果。

研究流程-3

最後為了證明研究方法的可行性以及研究結果的可信度，使用F-Measure方法加以驗證，亦設計實驗蒐集問卷，將搜集得來結果進行**Spearman相關分析**，用以證明Elo Rating System相對於Instagram中圖片本身的按讚數量來得可行。

核心演算法的運用

1. Scatter/Gather：

Scatter/Gather的概念為快速收斂縮減資料筆數，將資料反覆收斂與發散，能使其筆數以群數為底數的對數函數快速縮減，有效的找到使用者理想的景點類型。

2. LDA主題模型分析：

將LDA主題模型作為Scatter/Gather概念中的第一次分群，為預先建模的分群用以提高準確度。與TF-IDF的差別在於其分群依據不只考慮詞頻，亦考慮了詞與詞之間的相對分布。同時可透過各群中的詞語分布判斷該群景點所屬的景點類型。

3. K-Means分群法：

將K-Means分群法用於Scatter/Gather概念中後續的分群，透過統計方法驗證後，萃取各景點的LDA主題機率分布作為K-Means分群中計算歐式距離的變數，取代過往文件分群使用的稀疏矩陣，得以解決線上即時大量數據計算耗時的問題。

4. Elo Rating System：

透過Elo Rating System訓練所有圖片後，將圖片賦予美觀程度的排名再結合前項所述的分群結果，找出既切合使用者內心又美麗的景點，隨著越多人使用，本系統將會集合群眾智慧(crowd intelligence)，增加網頁系統推薦美麗景點圖片的客觀性。