

國立臺南大學資訊工程學系105學年度專題 結合引文網絡之論文推薦系統

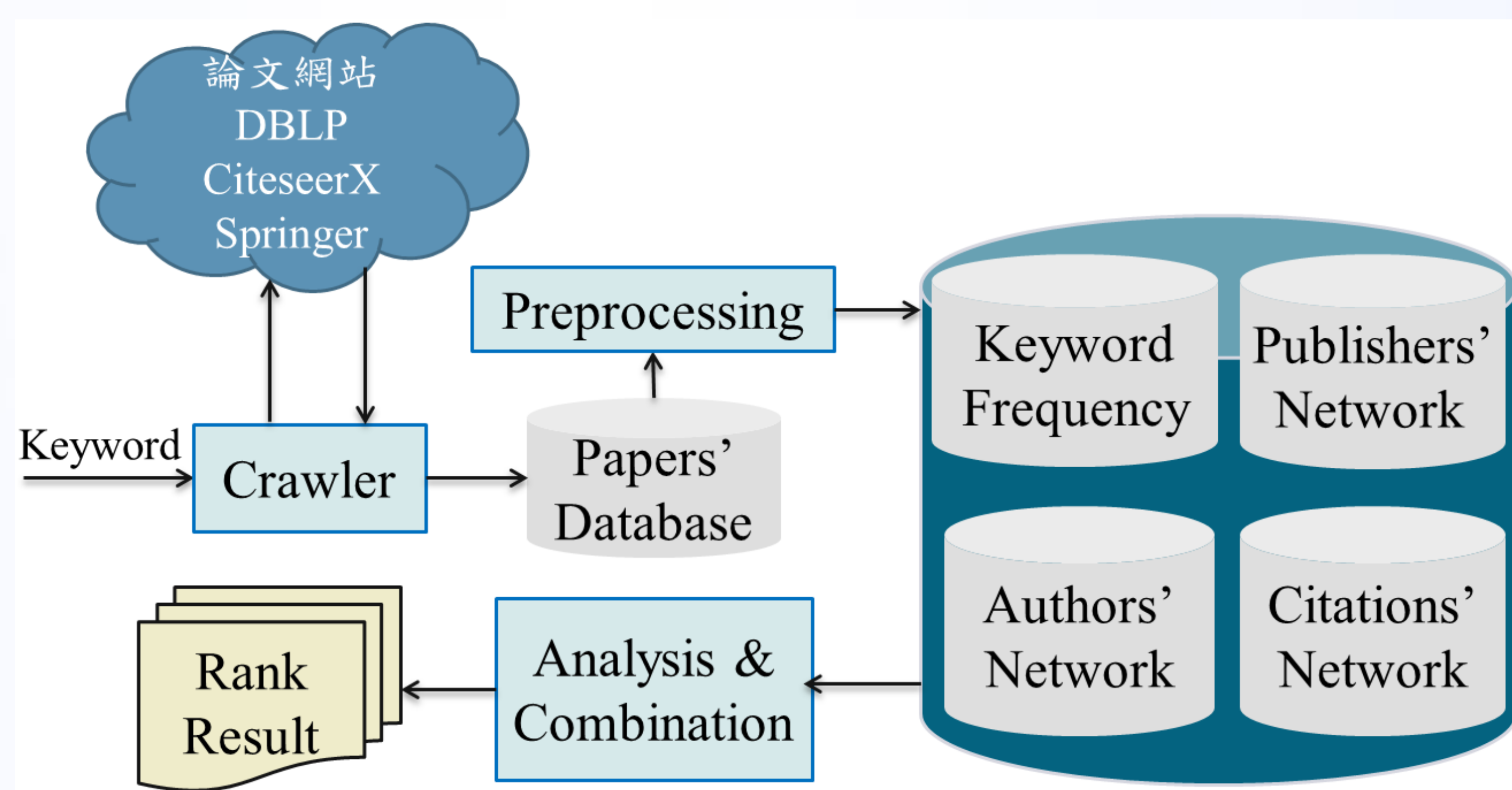
成員：劉軒佑

組別：第17組

摘要

面對龐大的論文資料庫，是否能廣泛收集使用者喜好資訊，以及能否即時呈現相關性高的推薦資訊是目前問題所在，本研究將分析論文間參考及引用的關聯，嘗試建構出論文間的網絡架構，並透過網頁排名以及推薦系統的技術，將其應用在計算論文網絡中各論文的權重和論文之關鍵字出現頻率，進一步設計整合排序方法，透過簡單清楚的使用者介面呈現計算結果，以提供解決推薦資訊缺乏及檢索效率不彰之問題。

系統介紹



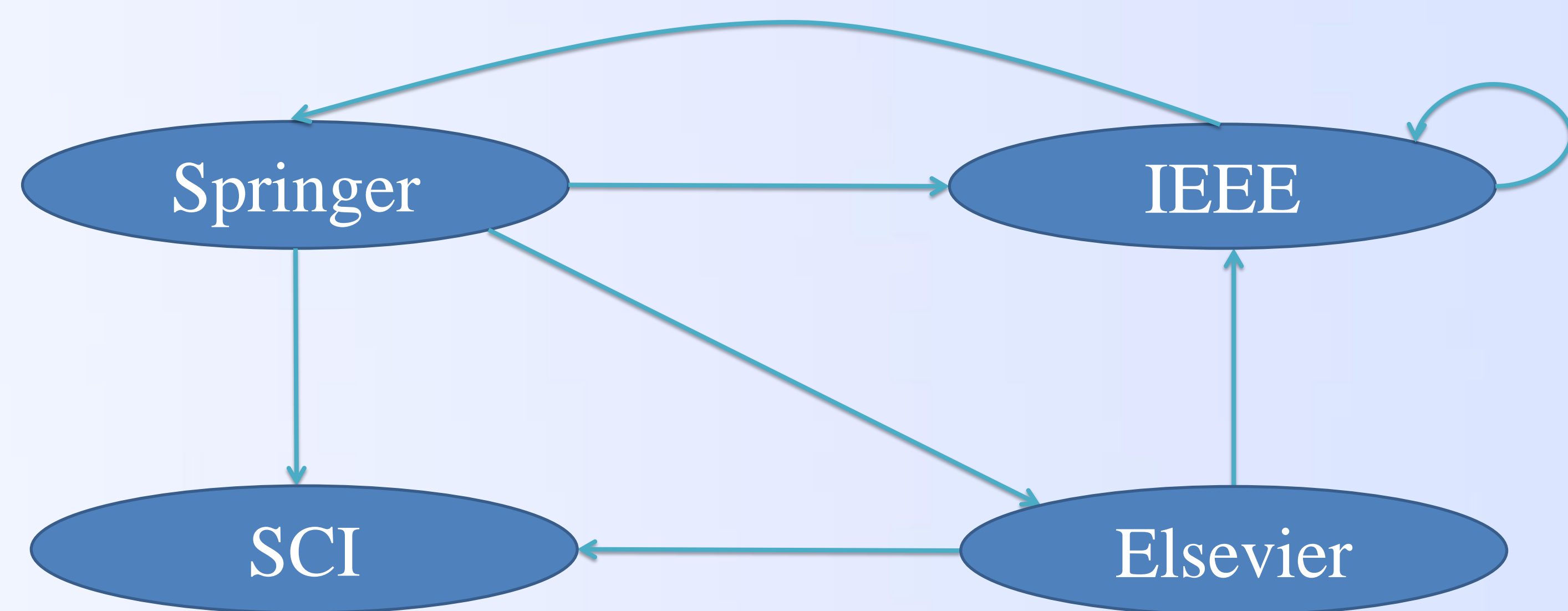
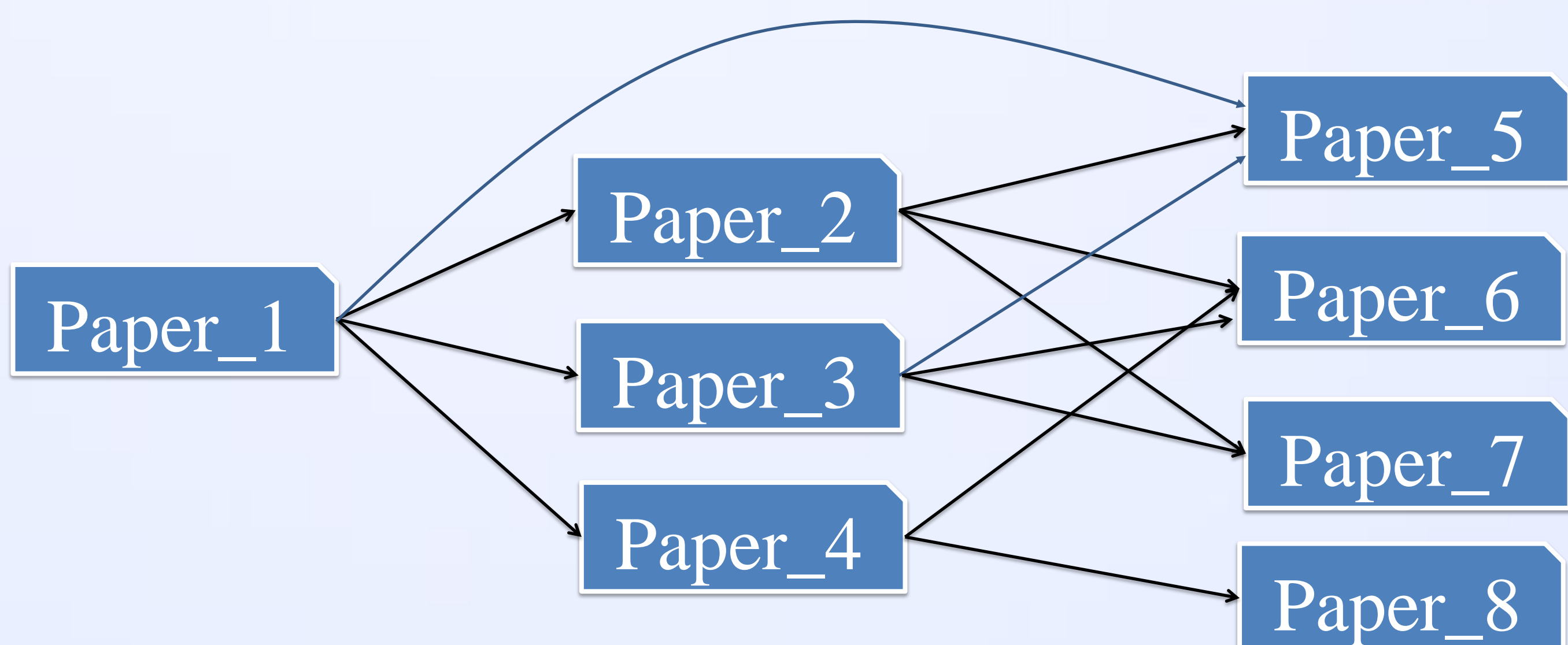
Crawler：論文自動搜尋引擎，蒐尋對象包含DBLP、CiteseerX、Springer。

Preprocessing：論文前處理步驟，分為兩個部分：

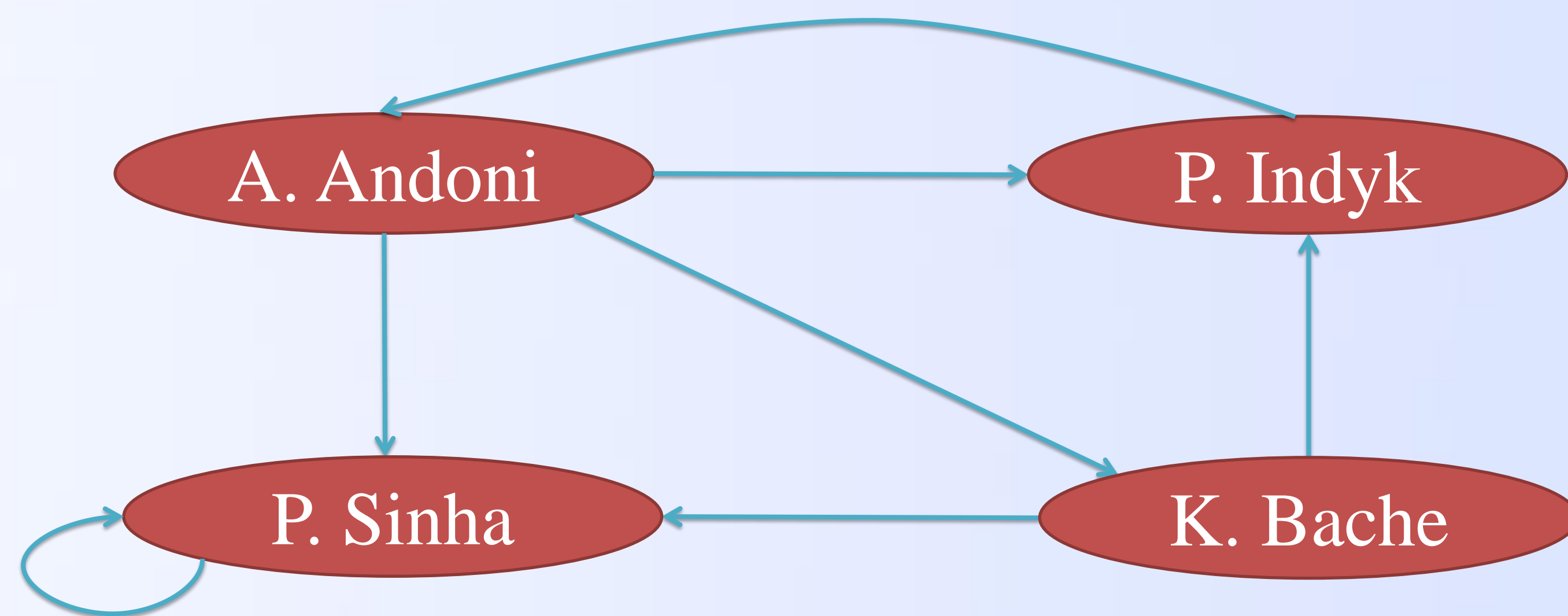
- **文字前處理**：執行移除標點符號、去除冗字(Stop Words)、字根還原(Stemming)以取得重要關鍵字並統計詞頻。
- **論文網絡前處理**：從論文資料庫中依作者(Author)、出版者(Publisher)、參考(Citation)分類並記錄其關聯。

Analysis and Combination：分析處理後的論文資訊：

- **Keyword Frequency**：將每篇論文的詞頻資訊進行相似度比對。
- **Publishers' Network**：出版者網絡為一有向有環圖(DCG)，利用權重遞移演算法(Transitive weight algorithm)計算各出版者在出版者網絡中的重要程度。
- **Authors' Network**：作者網絡為一有向有環圖，利用權重遞移演算法計算各作者在作者網絡中的重要程度。
- **Citations' Network**：參考及引用網絡為一有向無環圖(DAG)，利用PageRank演算法計算每篇論文在論文網絡中的重要程度。



有向有環圖(DCG)_出版者



有向有環圖(DCG)_作者

貢獻

節省時間

- 相較以往需仰賴人力逐篇審閱，論文推薦較為省時。

提升研究效率

- 能於短時間內取得較多與目標研究相關的論文。

易於操作

- 簡單且直覺的使用者介面，省去繁雜的檢索步驟。

使用者介面

