

以中文斷詞系統及 TF-IDF 演算法為基礎 之網路優質文章搜尋推薦系統

黃仁鵬¹ 李文雄²

¹南台科技大學資訊管理研究所

jehuang@mail.stust.edu.tw

²南台科技大學資訊管理研究所

m9990301@stust.edu.tw

摘要

搜尋引擎(Search Engine)是人們使用網際網路時，最常使用的網路服務之一。據調查結果顯示，現在使用網路的人口中，最常使用搜尋引擎，從網際網路搜尋想要的專業知識或其它如生活等各式資訊。然而，因為每個使用者網路知識高低不一、關鍵字使用技巧深淺不同，以及每次搜尋時的意向並不盡相同，就算如 Google 這已使用搜尋效果優異的 PageRank 演算法之全球最大之搜尋引擎，其搜尋的結果亦只能部分準確預測與滿足每個使用者的搜尋需求。幾乎每一個使用者依然必須浪費不少時間來檢視與過濾搜尋結果，才能找到想要的資訊。本研究有鑑於此，提出一套網路優質文章搜尋與推薦的系統架構，並依此架構開發一套網路優質文章搜尋與推薦系統雛形，主要是利用現有搜尋引擎之搜尋結果，一一加以過濾及評鑑，最後萃取與篩選出符合使用者搜尋關鍵字之現有網路上的優質文章，並加以儲存及系統性的歸檔供使用者查閱，可大幅減少使用者檢視搜尋結果的時間，提昇資料搜集的效率及成效。

關鍵字：網路文章、推薦系統、TF-IDF、斷詞系統、搜尋引擎