

讓電腦讀懂海量財經訊息-財經文詞解析引擎 及企業價值與風險評鑑(2/3)

核心詞彙特徵詞知識庫 說明文件

103 年 9 月

銘傳大學財務金融中心

111 台北市士林區中山北路五段 250 號 (02)28824564#2514

一、 特徵詞篩選模組簡介

在資訊爆炸的時代，數以萬計的公開新聞訊息流通於世界各地，為能快速剖析海量資訊，唯有透過機器將資料結構化的擷取並進行系統化管理，才能達到更有效率的決策與加值應用。主題分類之特徵詞更可應用於公司新產品之開發，準確且有效捕捉消費者的市場偏好及反應；國家政策之訂定亦可有賴於特徵詞篩選模組將各來源之民意反映進行分析處理，深入了解民眾之需求，促進公共建設之開發及改進。

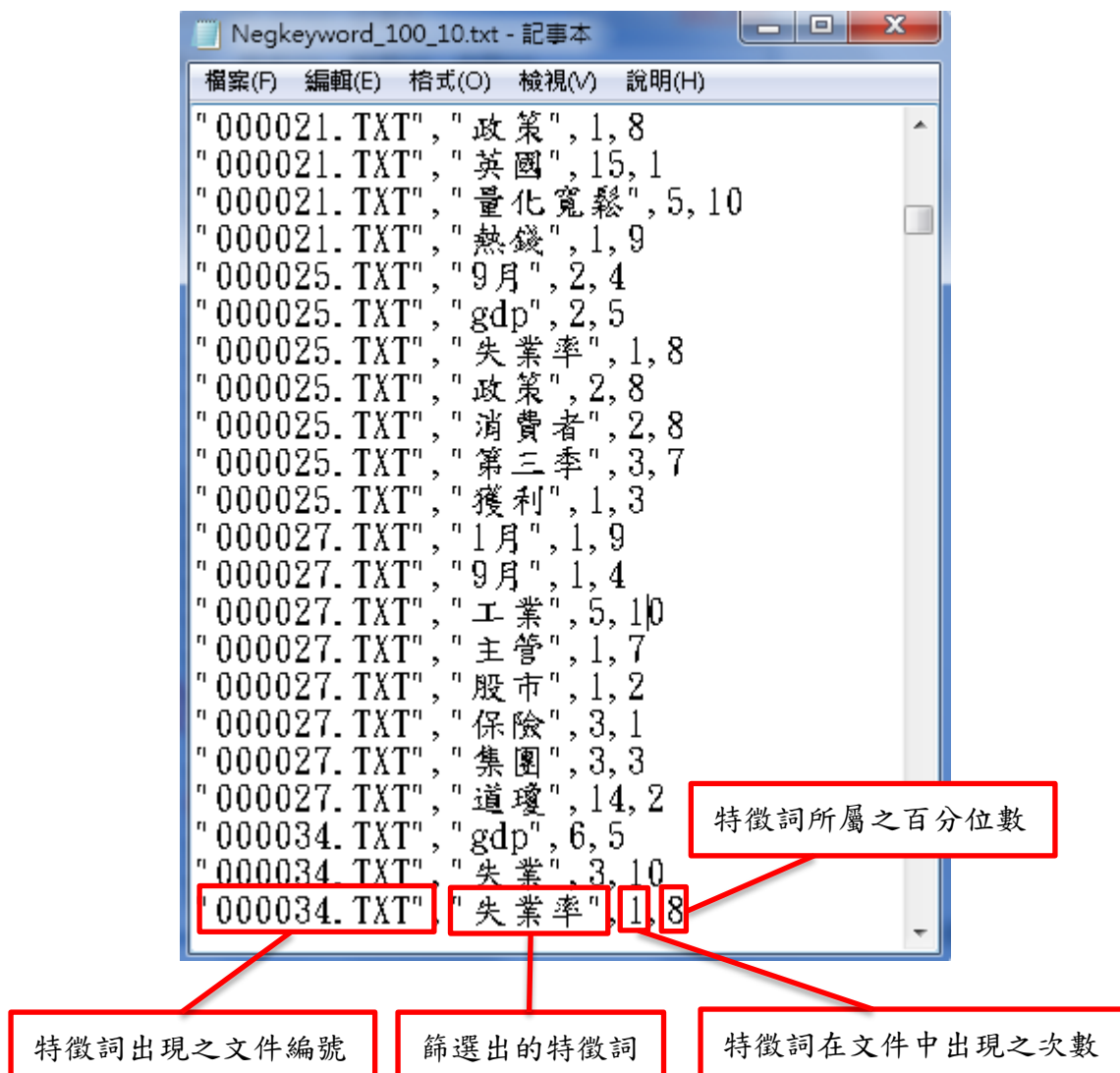
使用者蒐集原始新聞後，可依需求建立特定領域或單一新聞事件主題特徵詞，所建構之特徵詞主要可提供後續相關新聞或不同來源之語料分類，減少人工判讀之時間。以下以財經特徵詞為例，由圖 1 特徵詞篩選流程圖可見，利用特徵詞篩選模組及機器學習方式，由大量財經新聞過濾出財經專業用語，建立財經相關用語之參數檔。各類語料篩選特徵詞建構一涵蓋各主題的特徵詞詞庫，後續語料分類可直接利用已建構之財經相關用語之參數檔，透過特徵詞篩選模組並自訂篩選比例，篩選出現比例較高之財經詞彙，即可利用模組篩選之結果進行後續的分析應用，有助於提升使用效益及檢索資料之正確性，亦可更有效地剖析語義探勘技術文件之資訊內涵。



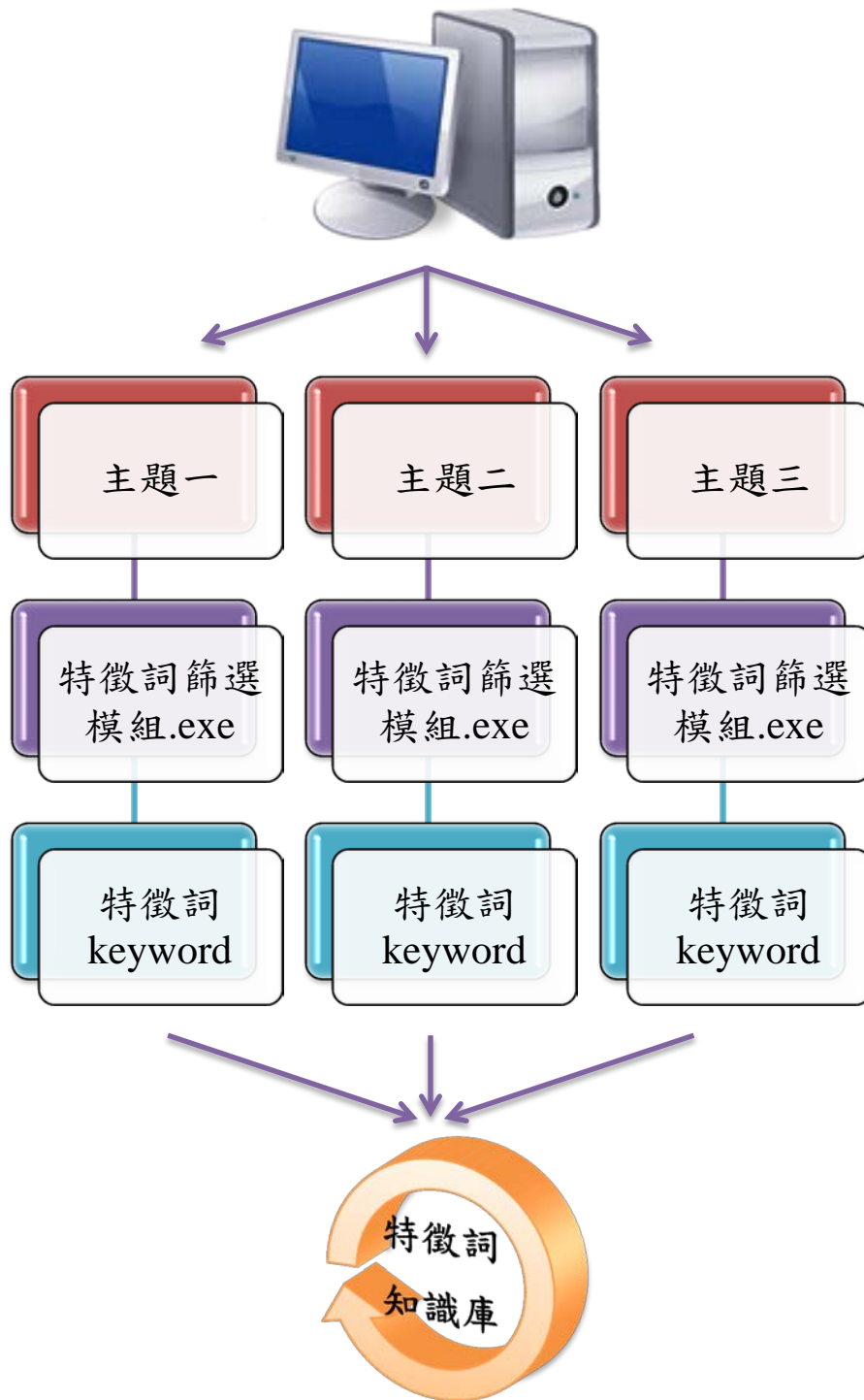
■ 圖 1、財經特徵詞篩選流程示意圖

二、 範例說明

透過特徵詞篩選模組進行原始新聞特徵詞篩選，評估所有詞彙於文章分布狀況及出現次數，評估其權重，設定篩選比例，進而篩選出與主題相符且具有意義之特徵詞，篩選結果如下圖 2 所示，並建立特徵詞資料庫，其流程請參閱圖 3，將有利於文件標記、議題分類、關鍵詞組搜尋等，並有助於管理文件之統整性及統一性。



■ 圖 2、特徵詞篩選模組結果說明



■ 圖 3、特徵詞篩選模組流程圖