

106 學年度大四工工專題摘要

第 7 組	新聞留言篩選器-以內容特徵值為基礎
指導教授	侯建良 教授
參與學生	103034035 連鈞瀚 103034062 陳奕廷 103034063 邱家堯
摘 要	
近年來，隨著科技的日新月異以及網路世界的發達與普及，人們有了更多元的方式可以透過不同媒介一窺世界各地發生的大小事，而各家網路新聞平台也猶如雨後春筍般建立，藉由隨時隨地的更新報導，供讀者得以透過電腦或手機等 3C 用品觀看第一手的新聞，而不需如傳統般必須守在電視機前才能獲知新聞消息。除了觀看新聞的即時性與便利性之外，網路新聞同時也能提供讀者在每則新聞底下發表自身的評論並與其他讀者交流，讓讀者也能自由針對新聞表達立場與意見，藉此促進了新聞之於大眾的討論性與可讀性，尤其某些讀者可能由於時間等因素而無法完整閱讀一則新聞，抑或是對新聞本身的議題不了解或者想有更深入的探討，藉由觀看新聞下方其他讀者的留言，也能增進自我對新聞的理解程度。但也因為這些言論平台的設立，導致越來越多的言論氾濫，舉凡無關新聞的留言、過度偏激的情緒性言論等，都導致讀者在觀看新聞評論的困難以及不便，進而減少讀者閱讀新聞評論甚至是觀看新聞的意願。 本專題的研究目的在於透過建立一套完整的系統模式，藉由 jieba 斷詞法進行分詞，以及 TF-IDF 法以及 TextRank 算法尋找新聞的關鍵字，最後透過演算法與詞庫的比對，計算出詞向量並篩選出與新聞本體相關性較高的留言而加以保留，並刪除部分不相關的留言，藉此提升新聞與留言的相關性，以期節省讀者閱讀時間，並提供讀者較為客觀而正確的言論，而研究過程也會蒐集多項樣本，並將主觀的留言相關性判斷結果與客觀的系統實驗結果進行比對，以增加系統的可靠度與可信度。	