# 摘要

論文名稱：基於Word2Vec之熱門主題偵測 頁數：1頁

校所別：國立臺北科技大學 資訊工程 研究所

畢業時間：一百零四學年度 第二學期 學位：碩士

研究生：謝宗廷 指導教授：王正豪 博士

關鍵詞： Topic Detection、Word2Vec、Keyword Extraction、Chinese Segmentation、Topic Modeling

批踢踢實業坊是目前台灣規模前幾大的論壇，討論的話題常與社會輿論相互連動，影響力非同一般。本論文提出一個熱門主題偵測方法，讓一般大眾即使處於一個資訊爆炸的時代，也能夠快速了解當今網路討論的話題趨勢。

現有的主題偵測方法無論是基於統計模型或是自然語言處理，大多以文章內容擷取而來的關鍵字作為文章特徵。然而在文章相似度比較上並未考量特徵彼此的語意關係，對於在網路上用語多元且平均篇幅較短的文章，效果相當有限。

本論文使用Latent Dirichlet Allocation搭配斷詞工具擷取文章的關鍵字，並利用 Word2Vector Model進行文章特徵的語意相似度分析，接著使用改良過後的 Hierarchical, Centroid Based的分群演算法來進行話題分群，最後計算偵測各話題熱度來選出熱門主題。

實驗使用批踢踢實業坊八卦版 2014/04/04 ~ 2016/05/10的文章，手動標記53個熱門主題約2700篇文章，並與系統偵測的主題做比較，結果顯示在Adjusted Rand index, Adjusted Mutual Information, V-measure等各項指標上均有不錯的結果。