

## 十一、研究計畫中英文摘要：請就本計畫要點作一概述，並依本計畫性質自訂關鍵詞。

### (一) 計畫中文摘要。(五百字以內)

在過去十幾年中，為了達到協助使用者擷取有效的訊息，有許多資料探勘 (Data Mining) 的技術紛紛被提出來以完成各式各樣的知識探勘 (Knowledge Discovery) 任務。在這些技術的使用下，各種不同形式的樣式 (Pattern) 也相繼產生出來，例如：序列樣式 (Sequential Pattern)、頻繁項目組 (Frequent Itemset) 以及封閉式樣式 (Closed Pattern) 與最大樣式 (Maximum Pattern) 等等。然而，在資料探勘的研究領域裡，如何有效地使用這些發現的樣式，則仍是一個懸而未決的議題。在大部份文件探勘的技術中，均採取關鍵字的方法以建造由單一文字 (Word) 或單一項目 (Term) 所形成的文字內容表示法，然而其他的研究技術則相信，詞組片語所攜有的資訊比單一文字來得多的假說，而捨棄了關鍵字的方法，改選擇以詞組片語來建造文字內容的表示法。令人遺憾的是，這些以詞組片語為基礎的技術方法並未帶來明顯的效果。推究其原因，則應是高頻率的詞組(通常是較短的詞組)通常擁有較高量的涵蓋性 (Exhaustivity)，但卻也含有較低量的具體性 (Specificity)，於是那些具描述性的詞組便會遭逢所謂低頻次的問題。樣式分類法模型 (Pattern Taxonomy Model, PTM) 是一個以樣式為基礎的技術方法，其採用了序列樣式探勘法並以封閉式樣式作為文字代表法的元素。PTM 針對較長的具體性樣式，運用樣式映射的策略，試圖解決上述低頻次的問題。然而，在 PTM 系統的內容學習階段中，負向資料 (Negative Example) 仍被忽略而未被妥善的使用，而系統所發現的樣式則需要這些資訊來做重新的評估。因此，本計畫將以發展具有效能及效率的樣式進化 (Pattern Evolution) 方法為目標，以期能夠解決上述的問題。所提的方法將會以實際的知識探勘任務來做測試，實驗的結果也將會和現有的方法來做比較，以評估系統的效能。