

第一章 緒論

本研究旨在利用資訊檢索技術中潛在語意分析 (latent semantic analysis, LSA) 的方法，分析題庫中的試題是否有相同或相似的情形。本章將針對研究背景與動機、研究目的、研究範圍與限制依序說明之。

第一節 研究背景與動機

在人類生活中，測驗佔十分重要的地位，舉凡教育上的學習評量、診斷、輔導等應用，就業方面的人員甄試、升等、證照授與等，皆需使用到測驗。而在資訊社會中，利用電腦科技的輔助，提昇測驗的精確性及便利性，已是不可避免的發展趨勢(何榮桂，1997)。

題庫(item bank)是測驗的核心，廣義的題庫是指一大群試題的集合 (Millman & Arter, 1984)，「一大群」指題庫內試題的數量，必須是任何使用此題庫之測驗其題數的好幾倍，「易於使用」是指題庫內的試題有利用索引 (indexed)、結構化 (structured) 等方式，使編製測驗時易於選題；狹義的題庫是指把多數的試題依其類別 (如單元或科目) 或統計特質 (如難度、鑑別度等) 加以整理、分類，並以數位化方式儲存，採電腦來管理及操作的試題組合 (何榮桂，1994)。

一個大型的題庫，可能會有意義相似甚至相同的試題，或出現暗示他題正確答案的情況，而新增試題到題庫時，也須考慮新增的試題是否已有相似者。要解

決上述的問題，直覺的方法是比對兩試題間相同的詞彙，再依相同詞彙之多寡決定其相似程度，但此方法無法解決兩試題有同義詞或相似詞的情形。在資訊檢索技術中的潛在語意分析，只要兩個句子間有相同的概念，即可分析出其相似程度 (Furnas et al., 1988)，如此即可克服同義詞或相似詞的情形。因此，研究者嘗試用 LSA 的技術對現用題庫做分析，找出意義相同或相似的試題，以提供給題庫管理者做參考。

第二節 研究目的

本研究旨在應用資訊檢索技術中潛在語意分析的方法，分析題庫中的試題是否有相同或相似的情形，以提供給題庫管理者做參考。而在使用潛在語意分析時，有許多增進其效果的方法，如冗詞去除、權重的調整與維度約化 (dimension reduction) 之最佳維度等，本研究亦將探討之，因此本研究之目的可具體臚列如下：

- 一、探討潛在語意分析是否能有效找出題庫中相同或相似的試題？
- 二、探討使用潛在語意分析時，冗詞去除與否、何種調整權重方式與約化的維度，在分析試題相似度時效果較佳？

第三節 研究範圍與限制

本研究依據前述之研究目的，考量研究環境、背景等因素，茲將本研究之範圍與限制敘述如下：

- 一、本研究著重於利用資訊檢索技術中潛在語意分析的方法分析測驗題庫，其他資訊檢索的技術不在研究範圍內。
- 二、本研究使用之題庫為行政院勞工委員會 (<http://www.cla.gov.tw/>) 所編製的「電腦軟體應用技能檢定丙級學科」試題，92 年與 93 年共 1000 題選擇題，試題範例如附錄一。