第五章　結論與建議

本研究旨在應用資訊檢索技術中潛在語意分析的方法，分析題庫中試題相似性的情形。本章將依據研究過程、研究目的和研究成果，呈現具體之研究結論與建議。

第一節　結論

在本研究中，研究者利用資訊檢索技術中潛在語意分析的方法，探討此方法能否分析出題庫中的相似試題，以提供給題庫管理者作參考，具體結論如下：

一、有去除冗詞在判斷完全相同、非常相似、部分相似與些微相似試題的情況時，其效果優於無去除冗詞者，但相似試題的數量取的愈多，其差異愈不明顯。

二、在調整詞彙和試題關係矩陣權重的方式上，因為每一試題包含的詞彙較少，因此調整 local 的權重對效果的影響較有限，而在 global 權重的部份，使用 idf2 與 entropy 的效果優於 idf 與 normal，而 idf2 與 entropy 兩者效果上之差異較不明顯。local 權重和 global 權重的組合，判斷完全相同試題的情況時，使用 binary-entropy 的效果較佳，判斷非常相似、部分相似與些微相似試題情況時，使用 log-entropy 的效果較佳。

三、在維度約化的最佳維度方面，判斷完全相同試題的情況時，保留的維度愈高，效果愈佳，但保留的維度大於 5 時，精確率的增加會趨緩；而在判斷非常相似、部分相似與些微相似試題的情況時，保留的維度到達某一臨界值後，精確率的增加會趨緩甚至降低，因此保留的維度愈高效果不一定愈好，判斷非常相似試題情況時最佳的保留維度

為 30，判斷部分相似試題情況時最佳的保留維度為 15，判斷部分相似試題情況時最佳的保留維度為 14。

四、系統判斷正確方面

對於「電腦軟體應用技能檢定丙級學科」92 年與 93 年兩個年度共 1000 題選擇題，下列四種情況的相似試題，系統皆能正確的判斷出來：

（一）用詞完全相同的試題。

（二）部分辭彙不同的相似試題。

（三）敘述方式不同，但題意相同的試題。

（四）辭彙不同，但意義相同的試題。

五、系統判斷錯誤方面

對於系統判斷為相似試題，但實際上並非相似的情況，分析有以下之原因：

（一）斷詞的錯誤。

（二）兩試題只有少部分辭彙不同，但不同之辭彙對於題意佔有較大的影響。

第二節　建議

依據本研究對於題庫相似度比對之結果，以下提出未來可能發展方向與後續研究之建議：

一、本研究使用之題庫為「電腦軟體應用技能檢定丙級學科」92 年與 93 年兩個年度的試題，未來可嘗試在不同科目進行研究，評估本研究所使用之方法在其他科目是否有效。

二、利用潛在語意分析作試題相似度比對時，並未考量詞性（即詞彙是名詞、動詞、介詞等）與試題意義的關係，建議後續研究可對於不同的詞性，給予不同的權重，評估是否能解決試題只有少部分辭彙不同，但不同之辭彙對於題意佔有較大的影響的情形。

三、本研究之斷詞方法採用實作較容易的反向最大匹配法，未來可嘗試利用其他的斷詞方法，如法則式斷詞、統計式斷詞等，以評估是否能增加系統判斷時的正確性。

四、資訊檢索技術的發展日新月異，後續研究可利用其他資訊檢索的技術，如 fuzzy set model、HMM/N-gram-based model 等，評估其是否能有效找出題庫中試題相似的情形。

五、在電腦化適性測驗（comprterized adaptive testing ,CAT）中，因為選題策略的關係，會依受試者做答反應否給予難度較高或較低的試題，但如此能力較高或較低的受試者，可能會一直做到某類型的試題，例如測驗數學的四則運算時，因為除法運算

較加法困難，所以能力值高的受試者可能會只做到除法的試題，能力值低的受試者可能會只做到加法的試題。因此在電腦化適性測驗的選題策略方面，可探討加入試題相似度這項參數時，題庫中不同概念的試題是否較能平均出現，與其對測驗長度和能力估計的影響。

六、在電腦化適性測驗中，受試者的作答反應會受猜測度與題意了解與否的影響，即答對者有可能是猜測正確，答錯者有可能只是題意不了解或誤解題意，因此可探討選題時加入相似試題，是否能降低作答反應受猜測度與題意了解與否的影響。