

網路智慧型測驗資料庫之分析與改進

蕭經武

國立暨南國際大學資訊管理學系
臺灣省南投縣桃米里埔里鎮大學路 1 號
g7213508@ncnu.edu.tw

黃國禎

國立暨南國際大學資訊管理學系
臺灣省南投縣桃米里埔里鎮大學路 1 號
gjhwan@ncnu.edu.tw

摘要

在遠距教學之中，題庫的建立往往是一項費時的工作，為了加速測驗資料庫的建立，由多位出題教師透過網路來共同建立題庫，已經成為現今發展的趨勢。然而眾多專家在未經協調的情形之下，出題方向難免失之於偏頗，無法完整的包含測驗的重點和觀念，甚至於出現雷同的題目，這些情形都會使得評估的結果大打折扣。關於這些問題，國內學者開始在中文環境下進行研究工作，試圖提供解決的方式，冀望不僅能使得問題發生的機率降低，且又兼顧學習過程之中的各項重點。而本研究的目的，則在於改進之前所提出的方法，期望能夠讓題庫在建構時更加符合學習評估測量的精神，同時也可以減輕出題者的負擔。

關鍵詞：網路測驗資料庫、題意式比對、概念關聯表格重複性、一致性、

1.簡介

在[9]中，作者詳細的討論了在網路測驗資料庫中一般常見的問題，並將其歸納為「重複性」、「一致性」、「完整性」和「關聯性」這幾個方面來探討。其中：

1-1 重複性分析

是指題目與題目之間，可能會發生題意相近或相同的重複出現現象，造成題目的存在失去意義或測驗效果的減低。為了避免這種情形的發生，作者利用下列的流程 (Figure 1) 來對題目進行檢查與篩選，以下將對每一步驟作說明：

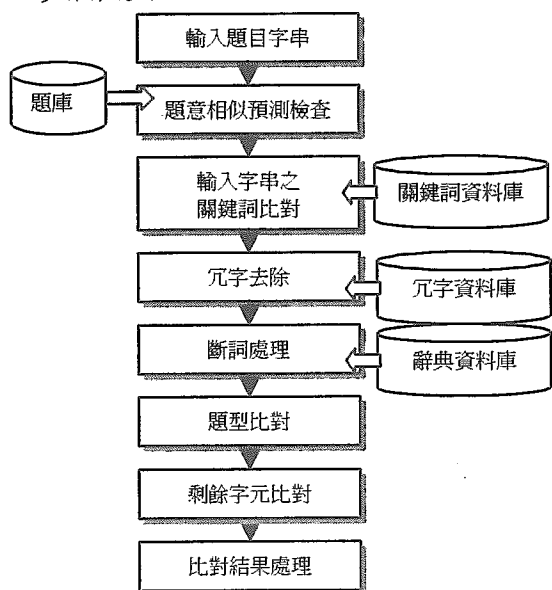


Figure 1 重複性流程檢查圖

1-1-1 題意相似預測檢查

所謂相似或相同的題目，除了在題目的字串內容及排列會很類似之外，和各章節的相關性必定也十分相近。為了能讓斷字及做字串比對時節省系統搜尋和比對的時間，可以先利用題意相似預測來篩選出一些可能會造成題意相似的候補，再從其中來做比對。由於在定義題目內容時，為了確切了解每個題目和相關概念之間的關係，以進行學習結果評估，作者使用評估平衡表 (Evaluation Balance Table) 來表示題目與重點概念關係 e_{ij} [19]

- 5：測驗問題對於此重點項目非常重要
- 4：測驗問題對於此重點項目具重要性
- 3：測驗問題對於此重點項目略微重要
- 2：測驗問題與此重點項目有相關性
- 1：測驗問題與此重點項目有少許關聯
- 0：測驗問題與此重點項目無任何關聯

這 0~5 的值可以轉換為以 100 為區間的數值：0, 20, 40, 60, 80, 100。使得對於每個題目與概念間會形成如下圖 (Figure 2) 的關係：

	概念 1(K ₁)	概念 2(K ₂)	概念 m
[Q ₁]	e_{11}	e_{12}	e_{1m}
[Q ₂]	e_{21}	e_{22}	e_{2m}
⋮	⋮	⋮	⋮	⋮
[Q _m]	e_{m1}	e_{m2}	e_{mm}
	E_1	E_2	E_m

Figure 2 概念關係圖

對於第 i 題，這種關係可以一個概念向量 Q_i 來表示：

$$Q_i = (e_{i1}, e_{i2}, \dots, e_{im})$$

對於輸入題目之概念向量 Q_{input} 與題庫內題目之概念向量 Q_j ，可以取其概念值間的差之總合作平均：

$$error_distance(Q_{input}, Q_j) = \sum_{r=1}^m |e_{input\ r} - e_{jr}| \div m$$

其中 m 為重點的個數，而誤差距離 (error_distance) 則表示這兩題所包含重點概念的差距，誤差距離越小表示兩題目間所包含的重點觀念越相近，同時也越有可能產生題目重複的情形。

1-1-2 關鍵詞比對、冗字去除與斷詞處理

由於題目與課程內容有相關，所以可以將題目所包含的概念之關鍵詞，利用建置另一關鍵詞資料庫來先做比對，以縮小斷詞時所需的搜尋範圍，並減少斷詞時的錯誤率。此外，為了節省斷詞所需花費的時間，可先將題目中較常出現的一些冗字 (不具特殊意義的詞)，如：“請”、“問”、“的”、“有”、“個”、“和”、“，”、“？”、“。”...等加以去除，因其不具備重要的意義，並且會增加詞語匹配的時間。

1-1-3 題型比對和結果處理

經過以上步驟之後，將會產生三個字串集合，分別是關鍵詞字串、冗字字串、和剩餘字串。接下來再結合這些字串來和原來題庫中已有的題目作比對，求出 Q_i 的 match 百分比為：

$$\frac{Q_i \text{與題目匹配到的字數}}{Q_i \text{原輸入字串總數}} \times 100\%$$

最後設定一門閥值，決定 match 百分比的下限，以作為是否將題目視為「重複」之列出與否的參考依據，亦即只要超過此一門閥的話，則我們會將兩個題目列出，供專家參考是否加入題庫或設為互斥。

1-2 一致性分析

一致性的檢查，其目的是為了要避免在同一份試題中，產生某題答案出現在另一題的題目中的情況，而必須在出題時所做的一些控制與篩選。由於這類題目本身的相似性並不一定很高，所以可將 match 百分比的門閥值略微調低，最後再於選出的題目中做選項與題目的交叉比對，檢查是否有答案出現在題目之中。若有的話，則列出給出題老師作為是否將此兩題設為互斥之參考，作法及流程如下圖 Figure 3 所示

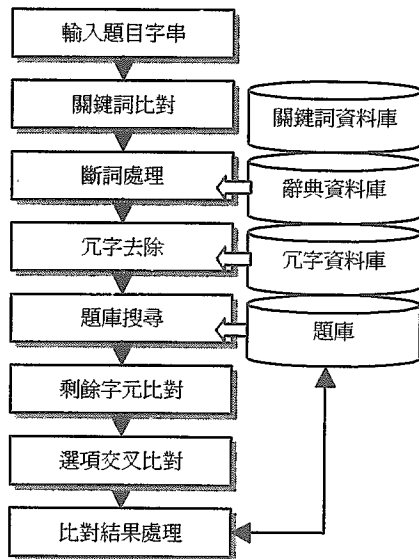


Figure 3 一致性檢查流程

1-3 完整性分析

為了避免題庫的問題都偏重於某一方面，作者利用評估平衡表，做各題目重點的題數百分比統計圖，提供教師做為目前題庫中的題目是否有分佈不均的參考。此外，對於某重點題數若介於離平均題數 3 個標準差之外的話，亦將特別列出給專家做參考，也就是

$$\text{if } (\mu \pm K_j > 1.5 \times S) \\ \text{then} \\ \{ \text{pruntln}(K_j); \}$$

其中

$$\mu \text{ 為平均題數} \\ K_j \text{ 為某重點題數} \\ S = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

1-4 關聯性分析

完整性分析是為了要讓各重點都能夠均勻分配到題目，但若是題目之間無法突顯出重點間的相異程度，可能會使得評量時無法清楚的分辨學生的學習瓶頸與障礙在於何處；為了使出題者瞭解重點間的關聯性，作者使用類似收集表格[25]的觀念來作分析。其重點間相似程度則用以下公式測量：

$$D_{ij} = \frac{\sum_k |A_i^k - A_j^k|}{MN}$$

D_{ij} = 重點 i 與重點 j 之相似距離

A_i^k = 重點 i 之第 k 個題目之值

A_j^k = 重點 j 之地 k 個題目之值

M = 關聯度最大值

N = 題目之總數

當某組 D_{ij} 之值太小時（相似距離很接近），表示這兩個重點之間的區隔並不明顯，此時可以加入一些題目，針對這兩個重點，使得其差異性增加，避免讓評估的結果在每次考到這兩個重點時，總是會產生學生認知程度相等的情形。

2. 概念關聯題庫題意式比對分析策略

雖然上述的方法可以改進題庫在重複性、一致性及完整性方面可能造成的問題，但仍然有些無法處理的狀況，例如下面兩個題目：

- Q1. 請問引起白血病的
Q2. 什麼是造成的血癌重要因素？

由題意看來，我們可以很明顯的看出這是兩個完全相同的題目，但是由先前所提出的方法來做重複性的比對，我們卻發現：

Q1. 關鍵詞：白血病
一般詞：引起、主要、原因
贅詞：請、問、的、是、？

Q2. 關鍵詞：血癌
一般詞：造成、重要、因素
贅詞：什、麼、是、的、？

代入公式後： $(3(\text{匹配到的字數}) \div 13(\text{原輸入字串字數})) \times 100\% = 23.08\%$ （在此，我們暫訂門閥值為 50%）如此低的相似度，顯然不會被挑選出來。

那麼，問題出在哪裡？首先是同義詞的部份。我們知道，其實「血癌」和「白血病」的意義是相同的，「主要」和「重要」這兩個詞的相似度則很高，「原因」和「因素」則可以說是同義詞，但是這些在比較時卻不會被察覺，因為在之前的研究方式裡，並沒有做相似詞之間的關聯比較。

為了解決這個問題，我們認為系統中應該建立一個關鍵詞與一般詞的概念關聯表格（Conceptual Relation Table），在這個表格中所存的，是詞與詞之間的關聯性。由於一般詞與一般詞的關聯性是固定的，在找尋專家並建立表格數據後便不會再更改，而關鍵詞則由於課程可能會有變動，所以可由專家來新增或修改和刪除其相似度，於是我們將關聯性分為六個階級，如下圖（Figure 4、Figure 5）所示：

	白血病	血友病	血癌
:	:	:	:	:	:
白血病			0	5	:
血友病	0		0	:
血癌	5	0		:
:	:

Figure 4 關鍵詞概念關聯表格

	...	引起	因素	造成	主要	重要	原因	...
:	:	:	:	:	:	:	:	:
引起	...		2	5	0	0	2	:
因素	...	2		2	0	0	5	:
造成	...	5	2		0	0	2	:
主要	...	0	0	0		4	0	:
重要	...	0	0	0	4		0	:
原因	...	2	5	2	0	0		:
:	:

Figure 5 一般詞概念關聯表格

其中，數字 1~5 所代表的意義是

- 5: 意義完全相同 (同義詞)
- 4: 意義非常接近
- 3: 意義接近
- 2: 有一點關聯
- 1: 不太有關聯
- 0: 完全沒有關聯

若轉換成百分比數字，則可以由 100、80、60、40、20、0 等來代表。

另外，在[9]一文中所提的計算門檻值公式裏，冗字也被列為比較的對象，且其重要性和關鍵詞與一般詞相同，這是非常不合理的現象。所謂「冗字」，顧名思義就是不會對題意造成決定性影響的字與符號，如介係詞、標點符號和一般題目常見的詞語等(例：請問、何者.....)。此外，從另一方面來看，除了關鍵詞以外的一般詞以及冗字即使相似的程度再高，卻並不保證題目的題意會相同，我們可以看看以下的例子：

Q3. 關於腎上腺的敘述何者正確？

Q4. 關於腦垂腺的敘述何者正確？

雖然兩題除了關鍵詞之外的敘述都完全相同，但是問題的重點卻完全不一樣。由此處我們可以清楚的瞭解到，一個問題往往存在著一個重心，若不能抓住重心作比較，做再多的功夫也是白費；然而題目的重心常常就在關鍵詞上面，接下來才是一般詞，於是我們決定採用加權的方式來做處理，可以寫成以下的公式：

$$\text{關鍵詞} \times W_1 + \text{一般詞} \times W_2 + \text{贅詞 (含冗字)} \times W_3$$

其中 $W_1 > W_2 > W_3$ 。

(例如說我們可以取 $W_1=2$ 、 $W_2=1$ 、 $W_3=0$)

這裡值得注意的是，我們採用「詞」而並非「字」為單位，這樣在分析上才具有其價值，因為中文的詞一旦被拆開來，就會失去其原有的含意，而採用這種方式，才符合以「題意」作為重複性比較的精神。然而當一個詞與另一題目中的許多詞都有關聯時，我們將只取關聯性最高的詞與其匹配；此外，若兩題的題目中都有關鍵詞的存在，但其重複性為零，則不可能有相同之題意。於是改良後的流程將變為如下圖(Figure 6)所示

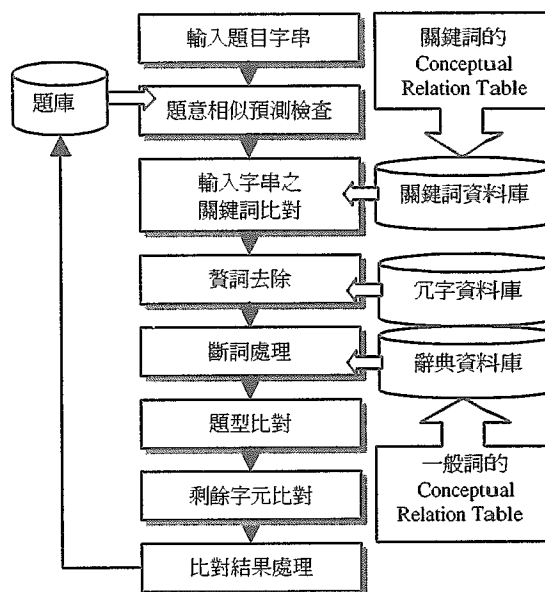


Figure 6 重複性檢查流程

公式也修正而變成：

$$Q_i \text{ match 百分比} = \left(\sum Q_i \text{ 與題庫題目比對後匹配到的詞} \times \text{權數} \right) \div \left(\sum Q_i \times \text{權數} \right) \times 100\%$$

※ Q_i 為去除贅詞後的詞串

我們定「引起」與「造成」的關聯性為 100%、「主要」與「重要」的關聯性為 80%、「原因」與「因素」的關聯性為 100%而「血癌」又等於「白血病」；於是之前的 Q_1 與 Q_2 經計算之後變成：

$$(1 \times 2 + 2 \times 1 + 1 \times 0.8) \div (1 \times 2 + 3 \times 1) \times 100\% = 96\%$$

很顯然的將高過門檻值(50%)，於是將被列出以作為參考。

接下來讓我們先來分析一下，之所以會造成「這一題的答案出現在另一題的題目中」的原因有哪些？(一致性分析)

第一種，答案在另一題的題目中，如：

Q5. 請問人類是屬於哪一類？(1)哺乳類(2).....

Q6. 除了人類是哺乳類之外，下列哪一個選項中的動物也是屬於哺乳類？

這一類型的題目出現的機率較低，但是由於题目的相似性不一定很高，只能先降低相似度的門檻值，再對選項作交叉比對，找出是否有在彼此的題目中出現，以作為專家是否將其設為互斥或重新出題的參考(也就是先前學者研究所提供的方式)

第二種，答案在另一題的答案中

Q7. 下列敘述有關遺傳的敘述，哪一項是正確的？(1)血友病和色盲均是性聯遺傳(2).....

Q8. 血友病和色盲是屬於(1)基因突變(2)性聯遺傳(3).....

類似這種情形出現的機率相當的高，但很不幸的，通常這類型的題目與題目之間的相似度卻非常的低，因為對 Q7 而言，事實上真正的題目是出在選項當中，因此若只是單純的比較題目，很可能因為相似度太低而沒有發現潛在的一致性問題。

幸而這類的問題大都有類似的題型，如：「試問以下關於○○的敘述，哪一項是正確的」或者如「下列敘述何者為非」……等，我們便可以利用此一特性，將符合此類題型的題目另做處理，把它的每一個選項都視為一個完整的題目來和其他題目作比對，這樣雖然要花費較長的時間，但是卻可以有效找出題目與題目之間一致性問題的存在。

修改之後的流程如下圖(Figure 7)所示：

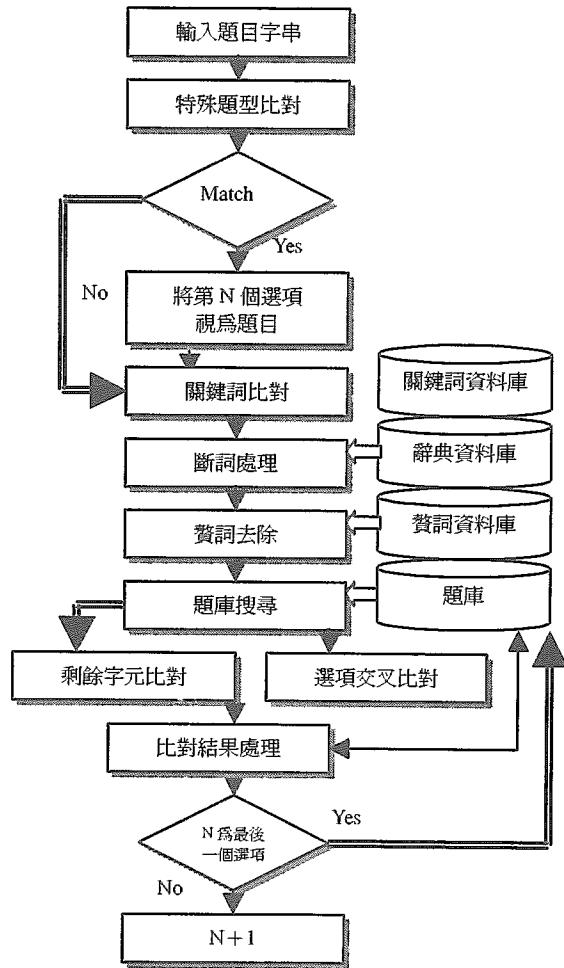


Figure 7 一致性檢查流程

特殊題型比對和一般題型比對不同的地方在於，關鍵詞在此處的出現反而變得不是如此的重要（對題型沒有影響），所以並不需要做加權。

3.系統架構及實作

本系統在架構設計上如右圖(Figure 8)所示，包含了一個以Java設計的題庫建構介面單元、查詢單元及一致性、完整性及重複性建構單元，以及辭典資料庫、關鍵字資料庫（均包含概念關聯表格）、冗字資料庫、教師資料庫和測驗資料庫。在本系統中，出題的學者透過瀏覽器註冊之後，經由題庫建構介面可以進行題庫的題目新增、修改及刪除等功能，也可以瀏覽資料庫內試題或查詢試題的重點分佈狀況，並且可以進入討論區參與討論或進入意見公布區發表意見。同時針對此一題目作線上題庫的重複性、一致性與完整性的檢查，來加以判

斷是否新增題目至題庫中，或與題庫中的其他題目設為互斥，相關的演算法在第三節中有詳細的說明。

在贅詞資料庫中存放了一些題目常見的非關鍵性詞句，不僅可以減少斷詞的時間，亦可增加比對的正確率；辭典資料庫存放的是一般常見的詞語，以提供斷詞時使用；關鍵詞資料庫則存放了測驗範圍內的觀念或比較重要的詞句，這個資料庫由出題專家和教師們所共同提供，有必要時也能夠做出修正（如課程內容的變更）。關於辭典與關鍵詞資料庫 Conceptual Relation Table 的建立，將採用 Linking List 的方式，如此不僅能加快比對時的效率，也可以大量減少空間的浪費；測驗資料庫中記錄了题目的基本資料如出題者之帳號、難易度、鑑別度、互斥題號……等，以供出題時之參考；教師資料庫則存放了教師的基本資料和出過的題數。想要建構一個完善的網路測驗與評量系統，需要多方面的考量加以整合，方能達到其真正的功效而對學習者有所幫助。學習與測驗評量系統除了教學策略、出題策略之外，題庫的建構與管理更是影響測驗評量好壞的重要因素。

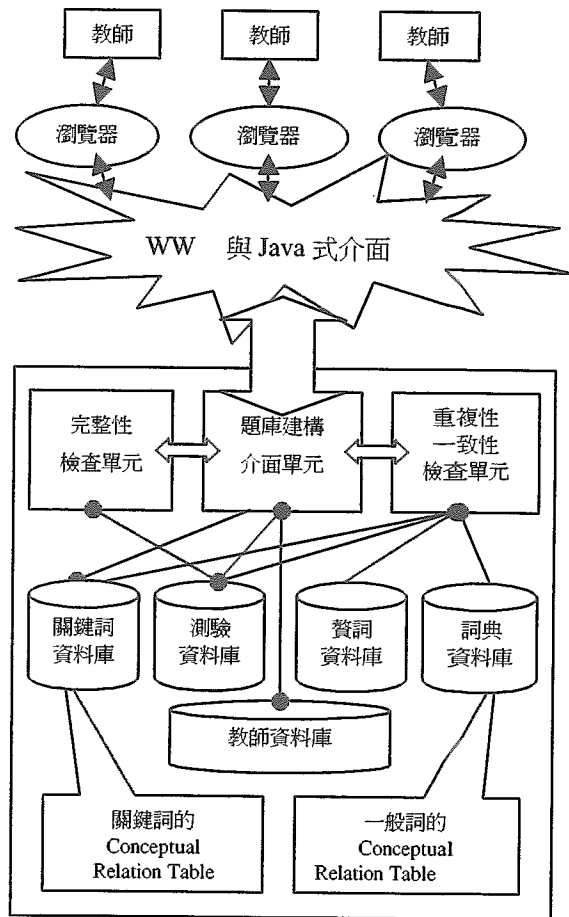


Figure 8 網路教學智慧型測驗資料庫建構圖

4.結論

為了改善網路合作多人出題時經常會發生的現象，如：重複出題及某題答案出現在另一題題目中的情形，以及各重點的題目分配不均和重點間的差異沒有凸顯的問題，使學習者能得到更佳的測驗試題，前人提出了重複性、一致性、完整性和關聯性的分析。而本研究的目的則是將其研究再做加強，在一致性方面，導入了以題意為主的分析比對方式，而不再是傳統的字串比

對，找出問題的重心，才能作更精確的比較。在一致性的方面，則是加強了其比對的方式，讓一些常用的題型，也能夠接受一致性比對的考驗，而不致成為漏網之魚。除了上述所提到的改進，我們更藉由相似表格的建立，導入了相似詞的比對，讓比對的能力大幅提升，擺脫單兵作戰的桎梏，使得比對的範圍更加的廣泛而能夠更完善的處理重複性以及一致性的問題。

在網際網路發達的今日，資訊檢索技術也跟著日新月異，未來對於資訊檢索技術是否能更應用於本研究之字串比對及斷詞處理並縮短搜尋的時間是很值得注意的研究方向。此外，現在出題的趨勢已逐漸向更活潑與更多樣化的方式，目前本系統只能對文字型態的選擇和是非的出題方式做處理，希望將來能導入問答題的形式，甚至能對具有圖形的題目作影像比對處理，讓題庫的組成更加的具備全面性，並且為教師共同建構試題提供一個良好的環境。

國科會計畫編號：NSC 89-2520-S-260-001

5. 參考文獻

- [1]何榮桂(民 83). 電腦化題庫概述, 測驗與輔導, 126 期, 2576-2577 頁。
- [2]何榮桂&蘇建誠&郭再興(民 85). 遠距適性測驗系統架構, 資訊與教育雜誌, 42 期, 29-35 頁。
- [3]何榮桂(民 86). 網路環境題庫與測驗之整合系統, 八十六年度電腦輔助學習及遠距教學專題研究計畫成果討論會摘要, 144-162 頁。
- [4]何榮桂&陳麗如(民 87). 電腦化適性測驗題庫品質管理策略之研究, 第七屆國際電腦輔助教學研討會論文集, 409-416 頁。
- [5]余民寧(民 84). 成就測驗的編製原理。
- [6]杜淑芬&黃國禎(民 87). 網路測驗及評估系統試題動態配置最佳化之研究, 第七屆國際電腦輔助教學研討會論文集, 337-343 頁。
- [7]周倩&簡榮宏(民 86.6). 網路評量系統之發展與研究, 遠距教育, 4 期, 12-15 頁。
- [8]周倩(民 86.10). 智慧型遠距合作學習環境中評量管理之整合研究—電腦輔助測驗與評估模式之研究, 八十六年度電腦輔助學習及遠距教學專題研究計畫成果討論會摘要, 469-474 頁。
- [9]祝鈞毅(民 87). 網路智慧型測驗資料庫重複性、一致性及完整性分析, 國立暨南國際大學資訊管理研究所碩士論文。
- [10]陳正佳(民 84). 一套中文語法分析系統的研究與設計, 國立台灣大學資訊工程學研究所碩士論文。
- [11]陳克健&張麗麗&張莉萍&謝清俊(民 86). 國語中的複合詞和語言剖析, 中華民國七十六年全國計算機會議論文集, 415-422 頁。
- [12]張百慈(民 85). 以卡瑞爾學習模組為基礎建立智慧型電腦輔助教學之架構, 資訊與教育雜誌, 52 期。
- [13]許慶昇&杜淑芬&黃國禎(民 87). 概念繼承關係在網路智慧型學習診斷系統之應用, 第七屆國際電腦輔助教學研討會論文集, 602-609 頁。
- [14]黃國禎&祝鈞毅&許慶昇&杜淑芬&葉士毅(民 86). 遠距學習環境中智慧型學習測驗及評估系統之研製, 中華民國八十六年全國計算機會議論文集 第二冊, D-129 - D-134 頁。
- [15]葉士毅&黃國禎(民 87). 遠距學習環境中學生學習行為即時診斷之研究, 第七屆國際電腦輔助教學研討會論文集, 71 頁。
- [16]曾元顯(民 86). 關鍵詞自動擷取技術與相關詞回饋, 中國圖書館學會會報, 59 期。
- [17]孫春在(民 86). 遠距教學策略:「遠距合作設計」簡介, 遠距教育, 3 期。
- [18]趙元任(民 81). 語言成份裡意義有關的程度問題, 中國現代語文學的開拓與發展:趙元任語言學論文集, 袁毓林主編, 清華大學出版社, 北京。
- [19]蕭漢威(民 83). 智慧型電腦輔助教學之評估模式, 國立交通大學資訊科學研究所碩士論文。
- [20]Fan, Joshua Poh-Onn & Mak Tina Kwai-Lan & Shue Li-Yen(1996). "Development of a knowledge based computer assisted instruction system," Proceedings 1996 International Conference Software Engineering: education and Practice, Dnnedin, New Zealand.
- [21]Gaines, B. R. & Shaw M. L. G.(1997). "Knowledge acquisition, modeling and inference through the World Wide Web," Int. J. Human-Computer Studies, 46, 729-759.
- [22]Hwang, G. J.(1995). "Knowledge acquisition for fuzzy expert systems," International Journal of Intelligent Systems, 10, pp.541-590.
- [23]Hwang, G. J.(1996a), "Development of an intelligent distance cooperative learning environment Sub-project 5 - A decision support system for tutoring strategies (II)," 行政院國家科學委員會專題研究計畫成果報告.
- [24]Hwang, G. J.(1996b), "A fuzzy expert system t support tutoring strategies for an intelligent distant cooperative CAL," Proceedings of The 5th Inter-national Conference on Computer -Assiste Instruction, Taipei, pp.137-147.
- [25]Kelly, G.A.(1995), The Psychology of Personal Construct, New Yourk: Norton.
- [26]Lewis, D. D. & Spark Jones K.(1996). "Natural language processing for information retrieval," Com-munications of the ACM, 39(1), pp.92-101.
- [27]Lorraine, Sherry (1995). "Issues in distance learning," International Journal of Educational Tele-communications, 1(4), pp.337-365.
- [28]Richard Sproat & Chilin Shih & William Gale Nancy Chang (1996). "A stochastic finite-state word-segmentation algorithm for Chinese," Computational Linguistics, 22(3), pp.376-404.
- [29]Sun, C. T. & Chou C.(1996). "Experiencing CORAL: Design and implementation of distant cooperative learning," IEEE Transactions on Education, 39(3), pp.357-366.