

## 第四章 研究步驟及實驗結果

本研究以中華民國電腦教育發展協會所開發的 Word 文書處理系統認證 PreMOUS (MOCC 標準級) 及 MOCC 專業級的試題為素材 (有關於 PreMOUS 及 MOCC 專業級的試題內容, 請參考附錄 1、2)。而針對系統的需求, 規劃三大作業, 字詞剖析作業、回歸測試分析作業及新增試題分析作業規劃系統架構如圖 4-1 ;

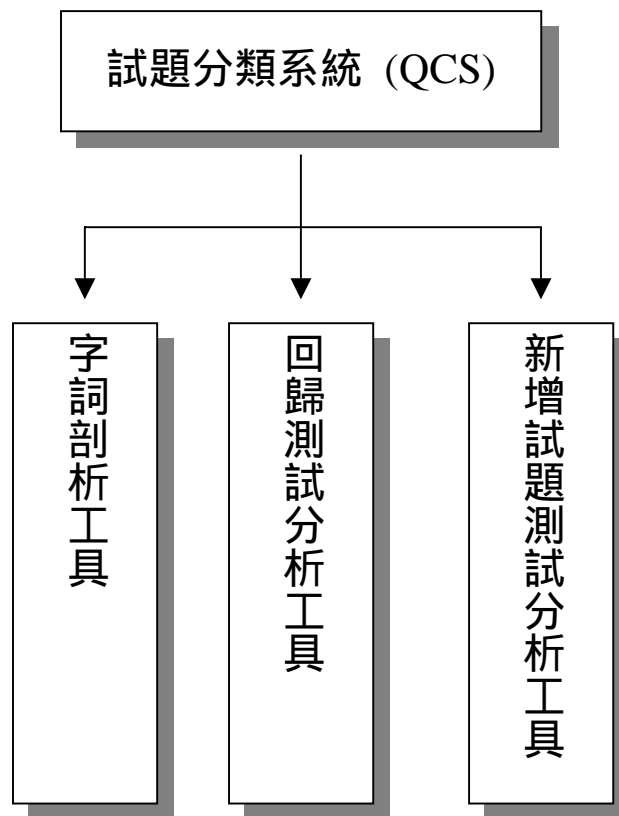


圖 4-1 試題分類系統 (Question Categorization System, QCS)架構圖

- 一、 字詞剖析工具：針對 PreMOUS 每一道試題, 剖析出特徵關鍵字詞, 並將特徵關鍵字詞, 存放於「特徵關鍵字詞庫」的資料表中, 並且記錄每一關鍵字在每一道試題的出現頻率

(Term Frequency)。而每一道試題都有其對應的能力項目，也可以由「特徵關鍵字詞」資料表中透過加總聚合 (Aggregate) 運算，得到關鍵字詞  $t$  在各能力項目  $a$  的出現頻率 (Term Frequency)。

二、 回歸 (Recall) 測試分析工具：當「特徵關鍵字詞庫」完成之後，此一詞庫，在內容上已經具備有 PreMOUS 試題特徵意義。接著配合上 TF、IDF、TFIDF、TFIDF<sup>2</sup>、TF/WIDF、WIDF 等字詞權重函數，測試所有樣本試題回歸到原有能力項目的比例，以判別各種公式優劣。

三、 新增試題分類及分析：為了解系統是否可以正確判讀新增試題的能力項目分類，採用各種字詞權重函數中，表現比較好的公式。將已經被專家分類的 MOCC 專業級的試題作為新增試題，藉由系統分類結果與專家分類結果相互比較的結果，評鑑分類的品質，並分析分類錯誤的因素，作為題庫品質管控的參考。

## 第一節 系統資料結構及規劃

本詞庫採用 PreMOUS (MOCC 專業級) 作為建立「特徵關鍵字詞庫」的素材。PreMOUS 每一道試題內容都具有獨立的能力項目，因此所建立的「特徵關鍵字詞庫」會比較單純化。而詞庫的儲存系統應用檔案型資料庫系統 Access 2000，共有五大資料表：

### 一、基礎字詞庫

由 Windows 作業系統內附的注音輸入法的相關字詞庫轉置而成，共有 44,579 筆字詞。資料表格式如表 4-1 所示：

表 4-1 「基礎字詞庫」資料表格式

欄位名稱	資料格式	資料長度
編號	自動編號	4 Bytes
關鍵字	文字	50 Bytes

原相關字詞內容如圖 4-2(a)所示，經由轉置程序，變成 4-2(b)的格式。

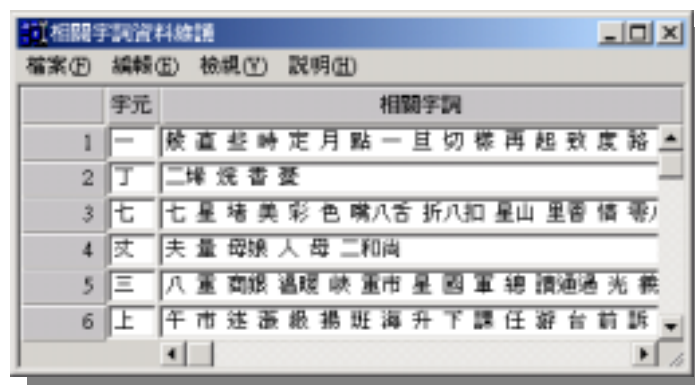


圖 4-2 (a) 原相關字詞庫格式

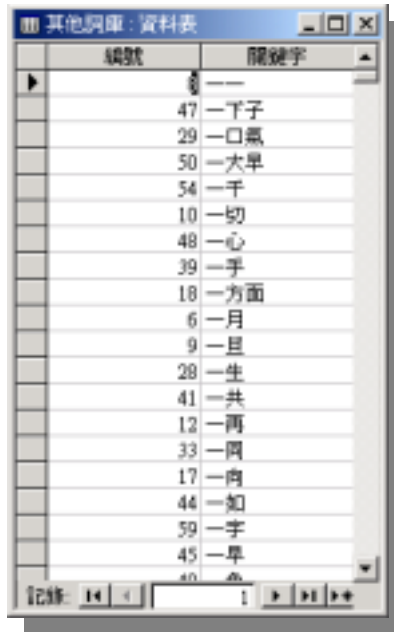


圖 4-2(b)處理後之「基礎字詞庫」

## 二、能力項目

PreMOUS 共分為 24 種能力項目，為了方便資料表間的關聯，每個能力項目都會對應到一個能力項目碼，「能力項目」資料表的資料表格式，如表 4-2 所示。

表 4-2 「能力項目」資料表格式

欄位名稱	資料格式	資料長度
能力項目碼	數字	2 Bytes
能力項目	文字	50 Bytes

24 組能力項目碼，分別使用 1~24 來代表，以簡化運算的程序，並方便作資料表間的關聯。

其內容如圖 4-3 所示。

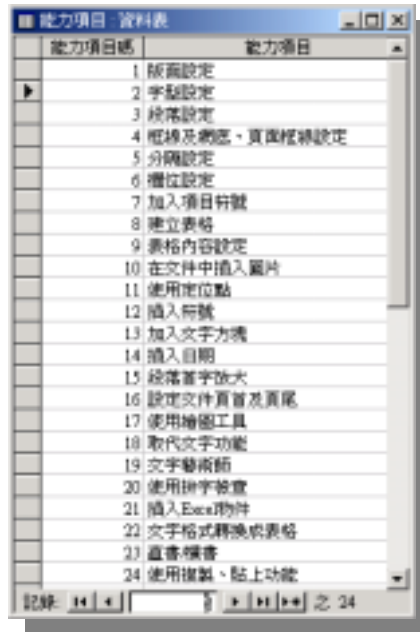


圖 4-3 「能力項目」資料表的內容

### 三、 特徵關鍵字詞庫

PreMOUS 的試題內容，經過剖析工具後，便可以得到該試題的特徵關鍵字詞，及該特徵關鍵字詞的出現頻率。資料表格式規劃如 4-3 所示。

表 4-3 「特徵關鍵字詞庫」資料表格式

欄位名稱	資料格式	資料長度
編號	自動編號	4 Bytes
關鍵字	文字	50 Bytes
試題	文字	50 Bytes
出現頻率	數字	2 Bytes
能力項目碼	數字	2 Bytes

資料表內容參考如圖 4-4 所示（該圖編號為流水號，在分類上不具任何意義。

編號	關鍵字	題目	出現頻率	能力項目碼
5811	紙張	WD280101a.bt	30	1
5812	邊界	WD280101a.bt	30	1
5813	設定	WD280101a.bt	30	1
5814	縱向	WD280101a.bt	24	1
5810	公分	WD280101a.bt	48	1
5815	方向	WD280103a.bt	8	1
5816	列印	WD280103a.bt	8	1
5817	右邊	WD280104a.bt	3	1
5818	文件	WD280104a.bt	3	1
5819	大小	WD280104a.bt	3	1
5820	整份	WD280104a.bt	3	1
5823	頁緣	WD280105a.bt	6	1
5824	頁首	WD280105a.bt	6	1
5821	距離	WD280105a.bt	6	1

圖 4-4 「特徵關鍵字詞庫」資料表內容

透過樞紐分析後，可知道各關鍵字在各能力項目分佈的出現頻率

TF。

關鍵字	合計	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
設定	117	30	18	9	18	2	2	2	24	4	6	2	2			2	2			4					
公分	114	48	18			2	2	2	26	4	6													4	4
文字	102	23	4					2	16	10	6	18				2	2	2	6		1	4	2	4	
標題	54	22	4	6	2	2	2		6			2	2	2						2			2		
表格	46								8	30	2												2	4	
大小	40	2	23					2	2		2	4							2	2					
距離	40			16	2	2				8	2		4			2	2							2	
對齊	34			16					12	2	2													2	
位置	33	6								12	6			4		2	2			1					
字號	33	13							2			6	2	8		2									
格式	32	10	2	4		2	4					2		6	2										
邊界	31	30		1																					
方式	30		8	2			2		6	2						4	2		4						
紙張	30	30																							
圖案	28			4						20															4
色彩	27	15							2	2					2	4	2								
儲存	26								16	2	2														6
縱向	26	24																						2	
字體	26	18						2	2		2									2					

圖 4-5 關鍵字詞分佈於各能力項目的頻率

然而上述的頻率值，可能會因為某一個字詞總數特別高，而造成分配不平均的效應，因此出現頻率  $TF$ ，必須做過歸一化 (Uniform) 的處理，公式如下：

$$TF = tf_n / tf_{total} \quad (4-1)$$

其中

$TF$ ：歸一化後關鍵字詞的出現頻率

$tf_n$ ：單一關鍵字詞在第  $n$  個能力項目的出現頻率

$tf_{total}$ ：單一關鍵字詞所有能力項目的出現頻率總和

經歸一化處理後，其內容如圖 4-6 所示。

圖 4-6 採用歸一化 (Uniform) 表示的 TF 值

#### 四、 單題

每當讀入一道試題，會透過字詞剖析程序擷取該題關鍵字詞，最後存放在「單題」的資料表中。單題資料表，會與「特徵關鍵字詞」庫作關聯，並運用加總聚合 (Aggregate) 運算得到該試題在各能力項目的比重為何。找出比重最高的能力項目，即為該題的能力項目。「單題」資料表的如表 4-4 所示。

表 4-4 「單題」資料表格式

欄位名稱	資料格式	資料長度
編號	自動編號	4 Bytes
關鍵字	文字	50 Bytes
試題	文字	50 Bytes

資料表內容參考如圖 4-7 所示（編號為流水號，分析上不具任何意義）。

編號	關鍵字	題目
162836	公分	WD2k0101a.txt
162837	設定	WD2k0101a.txt
162838	紙張	WD2k0101a.txt
162839	邊界	WD2k0101a.txt
162840	縱向	WD2k0101a.txt

圖 4-7 單題資料表內容

## 五、 試題

紀錄試題的檔名及所對應的能力項目編號。資料表格式如下表。

表 4-5 「試題」資料表格式

欄位名稱	資料格式	資料長度
編號	自動編號	4 Bytes
試題	文字	50 Bytes
能力項目碼	數字	2 Bytes
第一順位	是/否	1 bit
第二順位	是/否	1 bit
第三順位	是/否	1 bit
判讀類別 1	數字	2 Bytes
判讀類別 2	數字	2 Bytes
判讀類別 3	數字	2 Bytes



除此之外，使用第一順位、第二順位及第三順位的欄位紀錄該題作回歸測試的時候，在第幾順位能夠正確判讀，1 代表正確判讀，如果第一順位正確判讀，第二及第三順位也視為正確判讀。而判讀類別 1~3，紀錄該題作回歸測試時，能力項目分數最高的前 1 至 3 名。

「試題」資料表內容如圖 4-8 所示。

題號	題名	能力項目碼	第一順位	第二順位	第三順位	判讀類別1	判讀類別2	判讀類別3
1	WQGM001a.tst	1	1	1	1	1	23	3
2	WQGM002a.tst	1	1	1	1	1	23	3
3	WQGM003a.tst	1	1	1	1	1	23	3
4	WQGM004a.tst	1	1	1	1	1	23	8
5	WQGM005a.tst	1	1	1	1	1	3	16
6	WQGM006a.tst	1	1	1	1	1	3	16
7	WQGM007a.tst	2	1	1	1	2	15	12
8	WQGM008a.tst	2	1	1	1	2	19	12
9	WQGM009a.tst	2	1	1	1	2	19	12
10	WQGM004a.tst	2	1	1	1	2	5	22
11	WQGM005a.tst	2	1	1	1	2	19	17
12	WQGM006a.tst	2	1	1	1	2	19	3
13	WQGM007a.tst	2	8	1	1	19	3	13
14	WQGM008a.tst	2	1	1	1	2	13	19
15	WQGM009a.tst	2	1	1	1	2	6	13
16	WQGM010a.tst	2	8	8	1	23	12	3
17	WQGM011a.tst	2	1	1	1	2	18	13
18	WQGM001a.tst	2	1	1	1	2	1	7

圖 4-8 「試題」資料表的內容

## 第二節 關鍵字詞剖析處理

關鍵字剖析處理，主要的目的在於作試題內容的斷句，進而取得各試題所內含的關鍵字詞。字詞剖析的層次上，以圖 4-9 為例，共有三個層次。

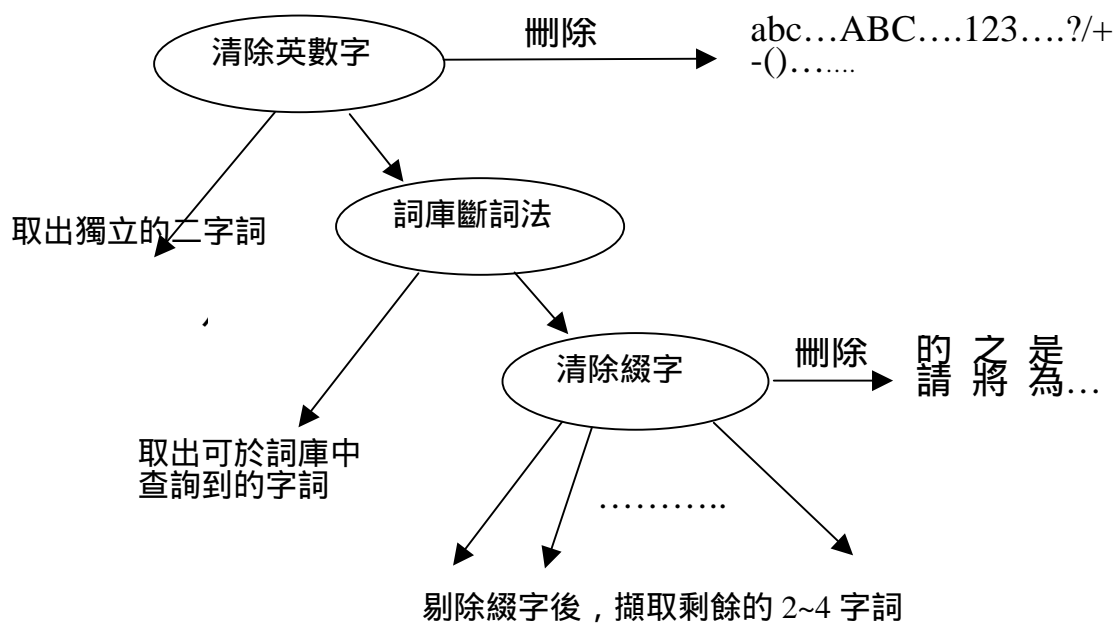


圖 4-9 字詞剖析樹 (Parse Tree)

中文是由幾個獨立的單字組成一個詞，而英文屬於拼音式的語言，其單字可能就等於中文的一個詞，因此在斷詞的方法上，有極大的差異。如英文斷詞處理，只要以空白或標點符號來分段，然後將代名詞（如；I, He, She, There, Here....）、介詞（如；of, in, on, at ...）及助動詞（如；is, am, are, was, were....）予以刪除，便可以留下比較具有特徵意義的單字，作為關鍵字詞 (Term)。

在中文方面由於介詞、主詞或助動詞也可能是其他字詞的一部份。如：「我的電腦」在 Windows 軟體中，屬於一個特有的名詞，然而如果採用英文的斷詞法，則會造成「我」、「的」兩的字被析離後，而只剩下「電腦」兩個字。這樣的斷詞方式，顯然會讓字詞的分類屬性產生很大的差距。

因此為了可以完整的擷取具有意義的關鍵字詞，規劃步驟及方法說明如下：

步驟一：將英數字刪除。由於使用的中文系統，因此將英文字視為較不具特徵的部分予以刪除。英數字刪除後，若有兩個中文字相連，則可以視為一個關鍵字詞。

步驟二：採用適合中文的「詞庫式斷詞法」，先擷取試題中既有的關鍵字詞。如此可以避免介詞、助動詞及主詞夾雜在某個字詞中而造成斷詞後，特徵意義消失的困擾。詞庫式斷詞法本身便需要一個豐富的「基礎詞庫」，本實驗的「基礎詞庫」來源採用 Windows 98/2000 作業系統所內涵的注音輸入法的相關字詞庫，由於該詞庫，採用文字檔格式，為方便採用 SQL 語法查詢，將該詞庫調整成資料表的格式。

有關詞庫斷詞的演算法敘述如下：

1. 試題內容中拆成單一中文字所組成的字列 WL。

如：『利用[直書/橫書]功能將文字方向改為縱向』這句話。在步驟一，已經因為英數字被刪除而變成『利用直書橫書功能將文字方向改為縱向』，擷取『利用』、『直書』、『橫書』等二字詞之後，剩下『功能將文字方向改為縱向』，拆成單一中文字的數列之後則該字列 WL 變成如下的樣子：

```
WL = '功','能','將','文','字','方','向','改','為','縱','向'
```

2. 使用 SQL 語法的 Select 語法查詢「基礎字詞」資料表，將字詞字首含有 WL 字列的中文字查詢出來。語法如下：

```
select 關鍵字 from 關鍵字詞
where Left(關鍵字,1)
      in('功','能','將','文','字','方','向','改','為','縱','向')
order by len(關鍵字) desc
```

所有查詢結果依照關鍵字的長度降冪排列，目的是要讓關鍵字可以依照字詞的長度由大到小的先

後次序擷取，以免長度小的字詞先擷取，而造成長度大的關鍵字不完整，影響特徵的完整性。

3. 將所查詢到的字詞依照長度由大到小，去比對試題內容，如果在內容中與所查到的字詞比對吻合，即可將該字詞列到「特徵關鍵字詞」資料表。

步驟三：將綴字予以刪除。經詞庫式斷詞法之後，介詞、主詞及助動詞便可以放心的刪除。

步驟四：然而刪除後不可避免的會殘留一些字詞，這些字詞可能有特徵意義。因此擷取二到四個連續排列的中文字，捨棄獨立的中文字，並放入特徵關鍵字詞庫。

經由上述的四個步驟之後，比較具有特徵性的字詞，已經擷取並放入特徵關鍵字詞庫，而該資料表也詳細紀錄某個關鍵字出現於那道試題出現的次數及其所歸屬的能力項目分類。其內容如圖 4-10 所示：

關鍵字	題目	次數	能力項目碼
方塊	WD2k1301a.txt	8	13
日期	WD2k0903a.txt	4	9
日期	WD2k1401a.txt	6	14
日期	WD2k1604a.txt	6	16
比例	WD2k0907a.txt	2	9
水平	WD2k0903a.txt	2	9
水平	WD2k1701a.txt	2	17
水果	WD2k0206a.txt	2	2
水果	WD2k0302a.txt	4	3
水果	WD2k0403a.txt	2	4
以外	WD2k0201a.txt	3	2
以後	WD2k0210a.txt	1	3

圖 4-10 「特徵關鍵字詞庫」資料表內容

由圖 4-10 可以看出『方塊』這個關鍵字在Wd2k1301a.txt 的試題中共出現 8 次，座落於 13 的能力項目；而『日期』這個關鍵字詞分別在WD2k903.txt、WD2k14101a.txt 及 WD2k1604a.txt 三道試題出現過，而且分別座落在 9、14 及 16 三個能力項目。

「特徵關鍵字詞」資料表完成了以後，便可以配合字詞權重函數產生權重向量矩陣，以供後續能力項目分類之用。

### 第三節 關鍵字剖析工具 (Term Parser Tools, TPT)

#### 一、處理流程

針對第一節中文斷詞工具的功能需求，系統規劃的流程圖如圖

4-11 所示：

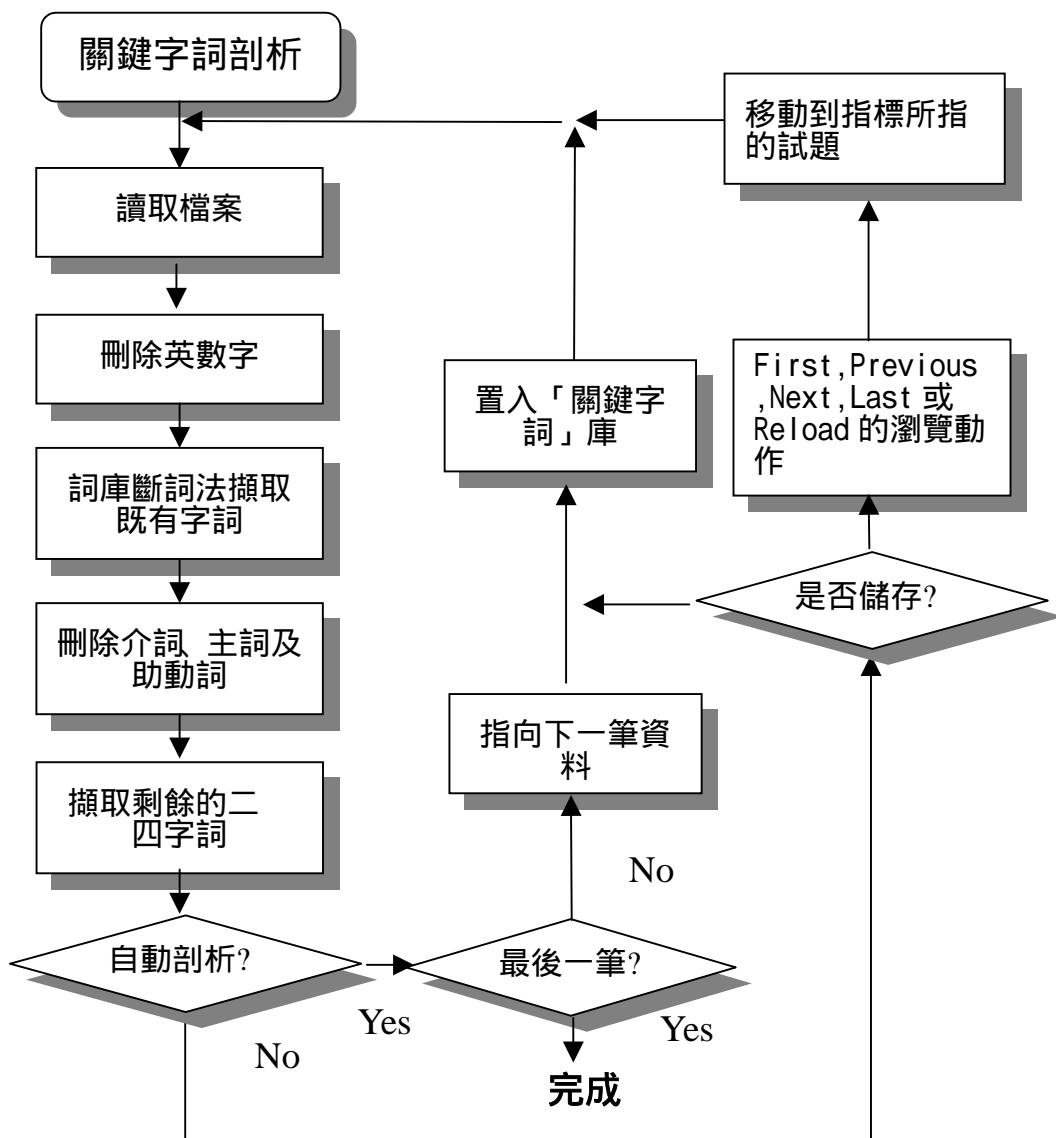


圖 4-11 關鍵字詞剖析(TPT)流程

## 二、執行及成效

經設計完成工具程式如圖 4-13 所示，程式上方為瀏覽介面，分別具有第一筆、上一筆、下一筆、最後一筆及重新載入，此五個按鈕屬於人工操作的部份，如果要將最後分析得到的字詞清單存於「特徵關鍵字詞」庫中，則必須按下「儲存」的按鈕。如果按下自動剖析，則會從第一題到最後一題自動完成字詞剖析及存檔到「特徵關鍵字詞」庫的動作。而綴字可以在綴字定義中設定，系統會在清除綴字的流程上判讀試題內容是否有綴字定義區的文字。如果有則該單字一律予以刪除。

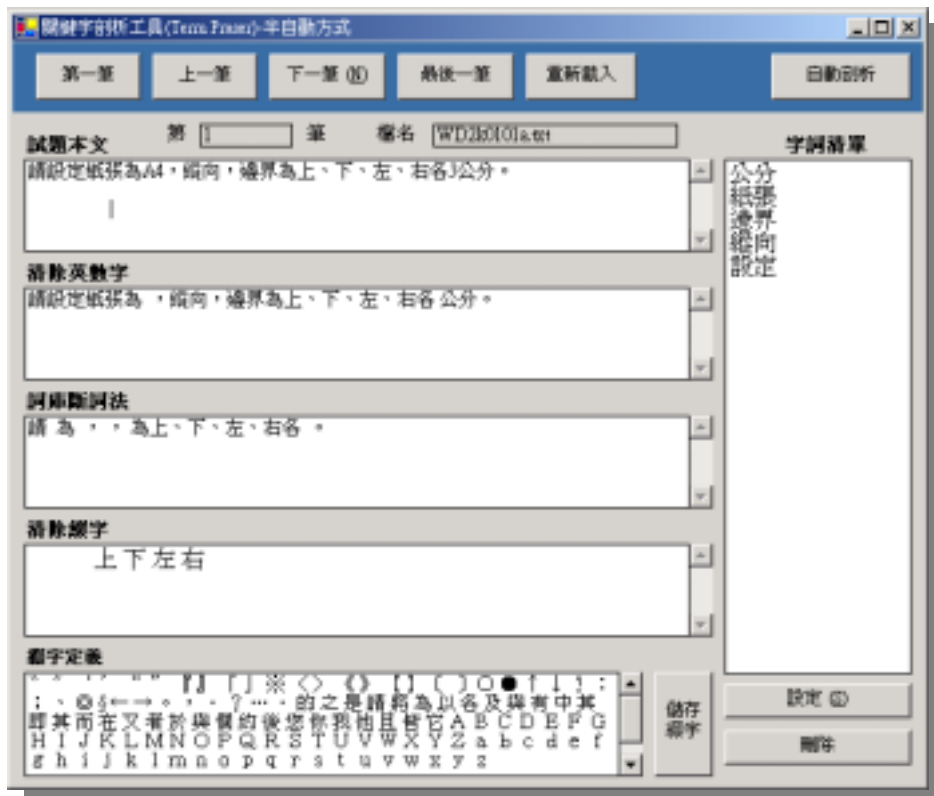


圖 4-12 關鍵字剖析工具 (Term Parser)



經字詞剖析處理之後，PreMOUS 的 91 道試題中，在不考慮跨能力項目的情況下，共可取得 752 個關鍵字詞；若將跨能力項目的關鍵字視為一筆時，則共有 415 筆。各能力項目所包含的關鍵字統計如下表所示：

表 4-7 各能力項目之關鍵字詞數量

能力項目碼	能力項目	數量
1	版面設定	16
2	字型設定	75
3	段落設定	52
4	框線及網底、頁面框線設定	33
5	分隔設定	26
6	欄位設定	14
7	加入項目符號	28
8	建立表格	9
9	表格內容設定	92
10	在文件中插入圖片	74
11	使用定位點	38
12	插入符號	12
13	加入文字方塊	43
14	插入日期	23
15	段落首字放大	11
16	設定文件頁首及頁尾	49
17	使用繪圖工具	23
18	取代文字功能	12
19	文字藝術師	19
20	使用拚字檢查	9
21	插入 Excel 物件	49
22	文字格式轉換成表格	8
23	直書/橫書	8
24	使用複製、貼上功能	29

各能力項目所擁有的關鍵字詞數，使用降冪排序，前四名分別為：

表 4-8 關鍵字詞數最多的前四個能力項目

能力項目碼	能力項目	數量
9	表格內容設定	92
2	字型設定	75
10	在文件中插入圖片	74
3	段落設定	52

在此假設關鍵字詞數量大的能力項目，會使其特徵意義變大，但是干擾也可能相對的變大。而這些干擾來自於比較平均分佈於各能力項目的關鍵字詞。由於上表的能力項目可能都擁有普遍性很高 (IDF 小) 的關鍵字詞，會導致該能力項目的特徵意義會被抵銷，在第六節採用 TF 字詞權重函數的實驗中，將會證明具有影響分類的干擾存在。

## 第四節 回歸測試處理

回歸測試處理用於評鑑權重向量矩陣，是否可以正確判讀樣本試題分類的能力。而藉由回歸率的高低，評斷分類品質的優劣。

依照計算字詞權重函數運算的積分排序，最高分者即是系統所判讀的分類，然而我們並無法保證系統可以百分之百的在最高分處顯示正確的分類，因此容許次高分及第三高分也可以算是正確分類。在此將最高分列為第一順位，次高分第二順位，第三高分則為第三順位。

以圖 4-14 為例，其採用 TF 作為字詞權重函數，得到的結果總題數有 91 題，在第一順位以內成功分類的有 67 筆，回歸率佔 73.6%，在第二順位以內可以成功分類的筆數有 86 筆，回歸率佔 94.5%，在第三順位以內成功分類的有 81 筆，佔 95.60%

而畫面的左下方，為每一個單題經計算之後，其各能力項目的權值排名。右下方為回歸測試分析圖，縱軸能力項目分佈比例，取 TF 最大值為 100%，其餘能力項目的 TF 與  $TF_{max}$  求得其比例：

$$Rate = TF/TF_{max} \quad (4-2)$$

峰值最高者，該能力項目的特徵越明顯，其餘的依照高低，判別是否有隱含的能力項目。

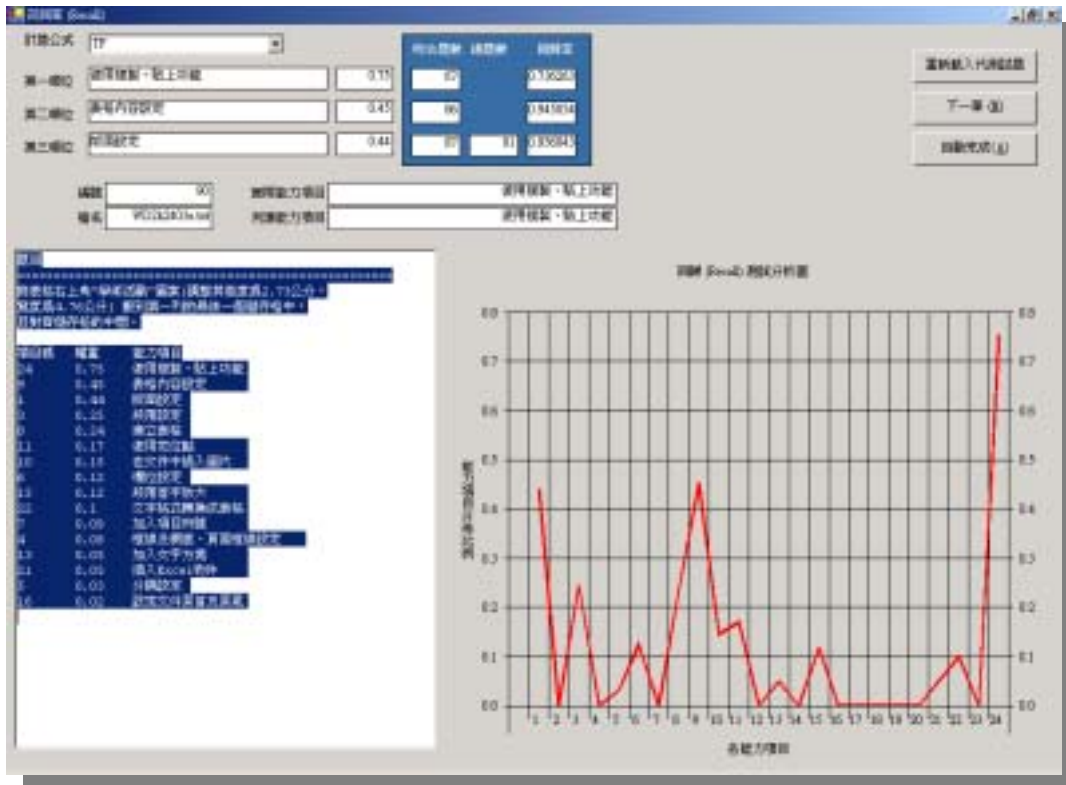


圖 4-13 回歸測試處理工具

## 第五節 回歸測試結果

回歸測試的主要目的是採用關鍵字剖析工具所產生的關鍵字對應能力項目之權重向量表，並透過各種試題分類的公式（如 TF、TFxIDF、WIDF....等字詞權重函數），將樣本試題重新作能力項目分類測試，以計算各能力項目正確分類的筆數。

由於回歸測試的結果，不見得百分之百精準，因為本實驗定義計算而得之數據由大到小前三名，即可視為正確能力項目分類。為了能夠更清楚的理解各種公式的優缺，保留第一順位到第三順位的正確分類筆數。實驗結果如下：

實驗採用 TF、TFIDF、WIDF、TF/WIDF、TFIDF2 及 WIDF 等六個字詞權重函數，並比較其優缺，實驗結果如下：

### 一、採用 TF 字詞權重函數

依照實驗結果，91 道試題可以在第一順位正確分類者，有 67 筆，佔了 73.6%，可以在第二順位正確分類者共 86 筆，佔 94.5%，可以在第三順位正確分類者，有 87 筆，佔 95.6%。

表 4-9 採用 TF 字詞權重函數的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	4	11	11	11	36.36%	100.00%	100.00%
段落設定	3	7	10	10	11	63.64%	90.91%	90.91%
框線及網底、頁面框線設	4	6	6	6	6	100.00%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	5	11	12	15	33.33%	73.33%	80.00%
在文件中插入圖片	10	4	5	5	5	80.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	5	6	6	6	83.33%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拼字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	2	3	3	3	66.67%	100.00%	100.00%
		67	86	87	91	73.63%	94.51%	95.60%

由上述的實驗數據，可以得到圖 4-15 的結果。由圖表中可以了解，幾乎所有的能力項目分類都可以正確分類，然而在 2、3、9、10 及 24，分類的狀況反而不良好，尤其以第 9 項只有 33.33% 可正確回歸最為嚴重（而其關鍵字詞的數量卻是最多的），也就是說，第九項在第一順位有 66.66% 的試題無法正確分類。

而整體而言，在第一順位回歸率可達到 73.6% 的水準，第二級第三順位分別可以達到 94.5% 及 95.60% 的水準。

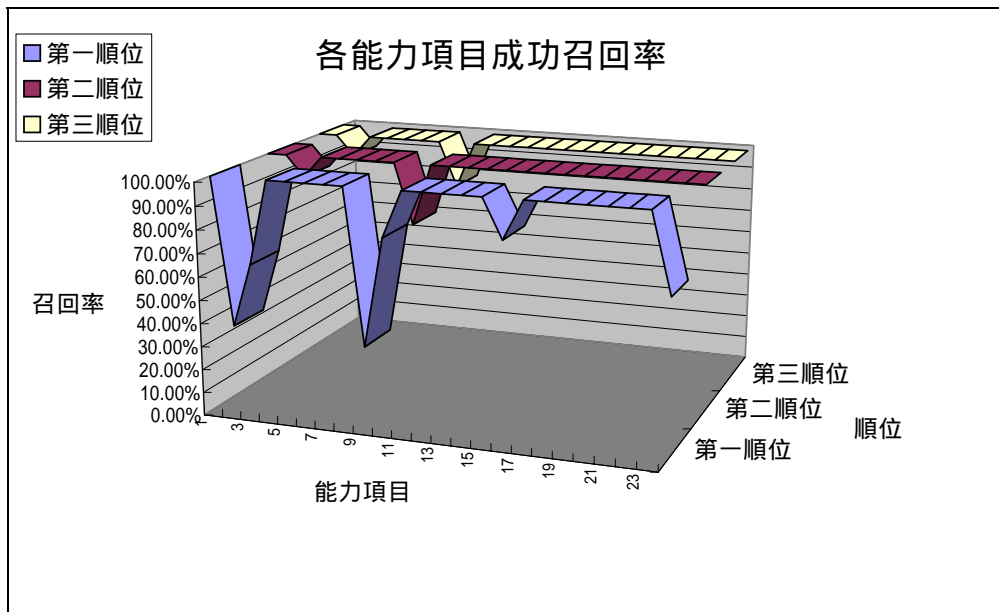


圖 4-14 採用 TF 字詞權重函數成功分類之分佈示意圖

為了更清楚的了解原因，可以從圖 4-16 清楚的看出，試題分類效果較差者，其共同的特徵是樣本筆數都比較多，根據此一現象，本實驗作一個假設：「在 TF 測試中，樣本的數量越多則回歸測試的效率會越衰減」。

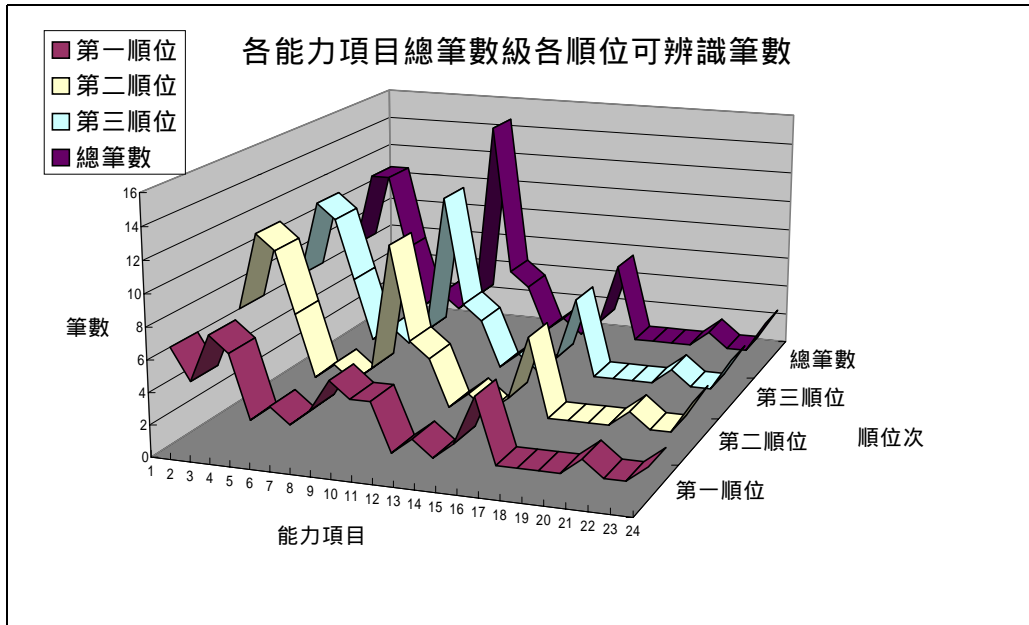


圖 4-15 各能力項目總筆數及各順位可辨識筆數

## 二、採用 TFIDF 字詞權重函數

使用  $TF \times IDF$ ，已經開始將判別關鍵字詞分佈普遍性的  $IDF$  考慮進來。前面提到，當使用的樣本試題多的時候，干擾也會相對的變大，而造成回歸率降低，因此使用  $TF$  與  $IDF$  相乘主要是為了衰減這些干擾。實驗結果如下：



表 4-10 採用 TFxIDF 字詞權重函數的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	9	10	11	11	81.82%	90.91%	100.00%
段落設定	3	11	11	11	11	100.00%	100.00%	100.00%
框線及網底、頁面框線設	4	6	6	6	6	100.00%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	8	12	13	15	53.33%	80.00%	86.67%
在文件中插入圖片	10	5	5	5	5	100.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	6	6	6	6	100.00%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拚字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	3	3	3	3	100.00%	100.00%	100.00%
總計		82	87	89	91	90.11%	95.60%	97.80%

而由上表可以看到，2、3、9、10 的回歸率均已經明顯提高，第 2 項由 36.36% 提高到 81.82%，第 3 項由 63.94% 提高到 100%，第 9 項由 33.33% 到 53.33%，第 16 項由 83.33% 提高到 100% 的水準。而整體而言，第一順位的回歸率可以達到 90.11%，第二順位分別都是 95.6%，第三順位可以達到 97.8%。三個順位的平均回歸率而言，總共提昇了 6.58%。由此可證明：「考慮到普遍性之後，可明顯改善回歸率」。

### 三、採用 TF/WIDF 字詞權重函數

採用 IDF 的方式，在文獻探討及研究設計均有提到，可以表達普遍性質，但卻無法表達更細微的普遍性，如各能力項目均有出現的關

關鍵字詞，當中出現較多的，應該比較具有特徵意義，使用 WIDF 可以將較細微的普遍性考慮進來。

表 4-11 採用 TF/ WIDF 公式的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	10	11	11	11	90.91%	100.00%	100.00%
段落設定	3	10	11	11	11	90.91%	100.00%	100.00%
框線及網底、頁面框線設	4	5	6	6	6	83.33%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	11	14	14	15	73.33%	93.33%	93.33%
在文件中插入圖片	10	5	5	5	5	100.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	5	6	6	6	83.33%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拚字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	3	3	3	3	100.00%	100.00%	100.00%
總計		83	90	90	91	91.21%	98.90%	98.90%

比較起 TFIDF，第 2 項由 81.82% 提升到 90.91%，第 3 項由 100% 降為到 90.91%，第四項由 100% 降為 83.33%，第 9 項由 53.33% 提升到 73.33%，第 16 項則由 100% 降到 83.33% 的水準。第一順位有 91.21% 的回歸率，高於 TFxIDF，然而第二及第三順位，均可以達到 98.9% 的回歸率。三個順位的平均回歸率比起 TFIDF 提昇了 1.21%。

顯然有些微的提昇，可以證明，考慮到字詞出現頻率多寡的 WIDF 確實可以得到比較好的回歸率。

四、TFIDF<sup>2</sup> 字詞權重函數

採用 IDF 的方式，在文獻探討及研究設計均有提到，可以表達普遍性質，如果將 IDF 作乘方，則可以更加突顯其普遍性

表 4-12 採用 TFIDF<sup>2</sup> 字詞權重函數的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	10	10	11	11	90.91%	90.91%	100.00%
段落設定	3	11	11	11	11	100.00%	100.00%	100.00%
框線及網底、頁面框線設	4	5	6	6	6	83.33%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	11	14	14	15	73.33%	93.33%	93.33%
在文件中插入圖片	10	5	5	5	5	100.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	6	6	6	6	100.00%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拼字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	3	3	3	3	100.00%	100.00%	100.00%
總計		85	89	90	91	93.41%	97.80%	98.90%

比較起 TFIDF，第 2 項由 81.82% 提高到 90.91%，第 3 項由 100% 降到到 83.33%，第 9 項由 53.33% 到 73.33%。整體的回歸率而言第一順位有 93.41，第二順位有 97.8%，第三順位有 98.9%。三個順位的平均回歸率，比起 TFIDF 提昇的 2.2%。可見採用加強特徵的 TFIDF<sup>2</sup> 可以提昇回歸率。

## 五、採用 IDF 字詞權重函數

接著我們再假設，在關鍵字均已經純化到能力項目範圍內，沒有其他非相關的關鍵字，則 TF 反而會降低其特徵性質。因此只考慮到普遍性反而可以提昇其回歸率。

表 4-13 採用 IDF 公式的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	11	11	11	11	100.00%	100.00%	100.00%
段落設定	3	11	11	11	11	100.00%	100.00%	100.00%
框線及網底、頁面框線設	4	6	6	6	6	100.00%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	15	15	15	15	100.00%	100.00%	100.00%
在文件中插入圖片	10	5	5	5	5	100.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	6	6	6	6	100.00%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拚字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	3	3	3	3	100.00%	100.00%	100.00%
總計		91	91	91	91	100.00%	100.00%	100.00%

而由上表的結果可以看出，三個順位均可以回到 100% 的回歸率。

## 六、採用 WIDF 字詞權重函數

與 IDF 同樣的理由來看，使用 WIDF 應該也可以有效的提昇回歸

率。

表 4-14 採用 WIDF 公式的實驗結果

能力項目	能力項目碼	第一順位	第二順位	第三順位	總筆數	第一順位	第二順位	第三順位
版面設定	1	6	6	6	6	100.00%	100.00%	100.00%
字型設定	2	11	11	11	11	100.00%	100.00%	100.00%
段落設定	3	11	11	11	11	100.00%	100.00%	100.00%
框線及網底、頁面框線設定	4	6	6	6	6	100.00%	100.00%	100.00%
分隔設定	5	2	2	2	2	100.00%	100.00%	100.00%
欄位設定	6	3	3	3	3	100.00%	100.00%	100.00%
加入項目符號	7	2	2	2	2	100.00%	100.00%	100.00%
建立表格	8	3	3	3	3	100.00%	100.00%	100.00%
表格內容設定	9	15	15	15	15	100.00%	100.00%	100.00%
在文件中插入圖片	10	5	5	5	5	100.00%	100.00%	100.00%
使用定位點	11	4	4	4	4	100.00%	100.00%	100.00%
插入符號	12	1	1	1	1	100.00%	100.00%	100.00%
加入文字方塊	13	2	2	2	2	100.00%	100.00%	100.00%
插入日期	14	1	1	1	1	100.00%	100.00%	100.00%
段落首字放大	15	2	2	2	2	100.00%	100.00%	100.00%
設定文件頁首及頁尾	16	6	6	6	6	100.00%	100.00%	100.00%
使用繪圖工具	17	1	1	1	1	100.00%	100.00%	100.00%
取代文字功能	18	1	1	1	1	100.00%	100.00%	100.00%
文字藝術師	19	1	1	1	1	100.00%	100.00%	100.00%
使用拚字檢查	20	1	1	1	1	100.00%	100.00%	100.00%
插入Excel物件	21	2	2	2	2	100.00%	100.00%	100.00%
文字格式轉換成表格	22	1	1	1	1	100.00%	100.00%	100.00%
直書/橫書	23	1	1	1	1	100.00%	100.00%	100.00%
使用複製、貼上功能	24	3	3	3	3	100.00%	100.00%	100.00%
	總計	91	91	91	91	100.00%	100.00%	100.00%

由上表的結果可以看出，三個順位均可以回到 100% 的回歸率。

由本回歸率測試的結果可以得知，使用 PreMOUS 為測驗樣本的作回歸測試，採用 IDF 及 WIDF 字詞權重函數可以得到 100% 的回歸率，而其他的字詞權重函數除了 TF 之外，也都可以達到 90% 以上的回歸率水準。

## 第六節 關鍵字詞數對回歸率的影響

在上一節的 TF 實驗中，可以發現只要是樣本試題多的能力項目，往往會導致回歸率的降低，特徵關鍵字詞數量是否會影響回歸率的優缺呢？

在這個實驗中，將樣本試題數量分別使用 24、40、51、58 四組。其採用樣本試題的原則是在各能力項目中，平均取出試題。因為部份能力項目試題數量不夠，因此試題樣本數量，並不是依照等差級數增加。而各組的關鍵字詞數，經字詞剖析處理後，分別有 318、473、536、588 筆。

將 91 道試題分別用上述四組所產生權重向量矩陣運算，經實驗結果可以得到下表的結果：

表 4-15 不同樣本試題數的回歸率測試結果

樣本數 / 關鍵字數	24 / 318				40 / 473				51 / 536				58 / 588			
	順位1	順位2	順位3	平均	順位1	順位2	順位3	平均	順位1	順位2	順位3	平均	順位1	順位2	順位3	平均
TF	61.8	78	79.1	72.97	62.6	81.3	82.4	75.43	67	84.6	85.7	79.1	68.1	89	90	82.37
IDF	63.7	72.5	73.6	69.93	80.2	85.7	86.8	84.23	83.5	91.2	92.3	89	85.7	92.3	93.4	90.47
TFxIDF	64.8	79.1	80.2	74.7	76.9	84.6	85.7	82.4	80.2	85.7	86.8	84.23	82.4	90.1	91.2	87.9
WIDF	67	76.9	78	73.97	72.5	82.4	83.5	79.47	74.7	87.9	89	83.87	76.9	90.1	91.2	86.07
TF/WIDF	65.9	79.1	80.2	75.07	75.8	84.6	85.7	82.03	76.9	85.7	86.8	83.13	76.9	87.9	89	84.6
TFxIDF2	63.7	78	79.1	73.6	75.8	83.5	84.6	81.3	79.1	85.7	86.8	83.87	83.5	89	90.1	87.53

採用圖表表示，結果如圖 4-16 所示：

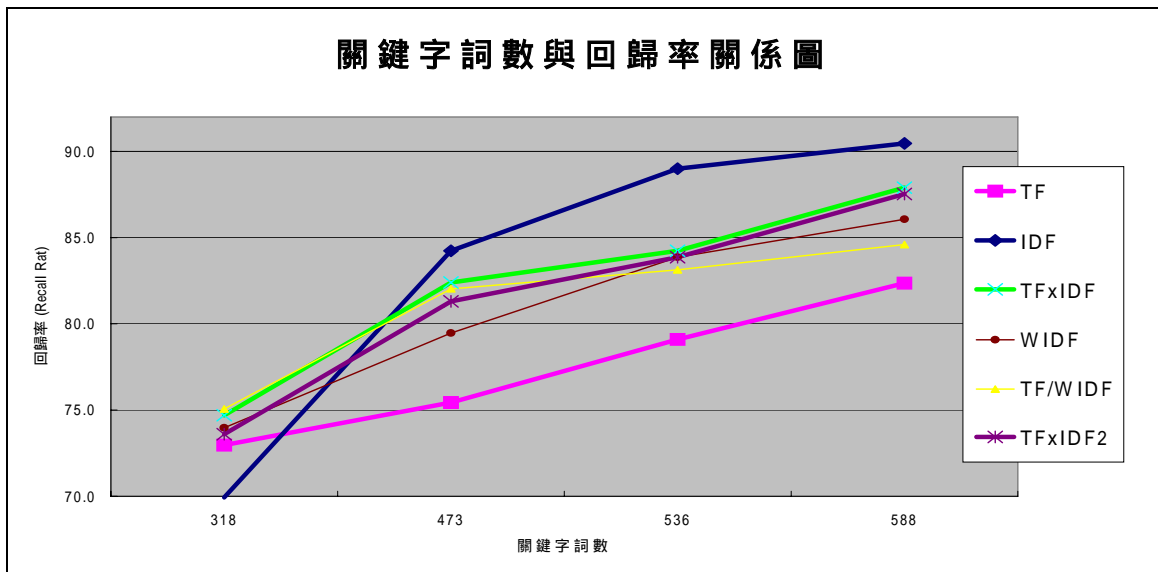


圖 4-16 關鍵字數與回歸率關係及趨勢圖

由圖 4-17 的結果，可以看出所有的函數的趨勢都是上升的狀態，因此可以得知關鍵字數的多寡與回歸率有正比的關係。

而比較這六種函數，IDF 與關鍵字詞數量多寡的反應會比較敏銳，由圖 4-17 可以得知，IDF 在 318 筆關鍵字詞的位置，回歸率最低，然而當增加關鍵字數量之後，IDF 立即超越所有的字詞權重函數，達到最佳的回歸率。TF 對於關鍵字詞數量的增加反應最不敏銳。除了 318 筆的位置曾經超越 IDF，其餘的時候均落後其他函數。

在 PreMOUS Word 2000 的 91 道試題樣本中，使用 58 道試題作為樣本，IDF 依然可以達到 85.7% 的回歸率，遠高於其他函數。因此在 PreMOUS Word 2000 的樣本領域範圍內，IDF 的分類效果最好。

## 第七節 新增試題分類及分析實驗設計

新增試題分析工具主要作用是將新增的試題放入 QCS，並藉由分類系統得知相關的能力項目分類。本實驗以 MOCC Word 專業級的術科題庫，作為模擬新增試題的實驗樣本，採用的理由有三：

1. 因該試題已經透過專家作能力項目分類，具有專家效度。
2. 該題庫有 32 項能力項目，比起 PreMOUS Word 的 24 項能力項目多出了 8 個能力項目，以模擬或考驗題庫能力項目不完整的狀況。
3. 一題 MOCC Word 專業級的試題等於多組 PreMOUS Word 的綜合試題，具有高複雜度，可以充分的考驗 QCS 的分類能力。



表 4-16 經專家分析的 MOCC Word 2000 專業級能力項目表

項目碼	2k															80															合計
	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	4c	5a	5b	5c	1a	1b	1c	2a	2b	2c	3a	3b	3c	4a	4b	4c	5a	5b	5c	
1	1	1				1										1	1	1	1	1	1	1		1	1	1	1	1	1	16	
2	1		1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21	
3	1	1	1				1		1	1			1			1	1	1	1	1	1	1		1	1	1	1	1	1	17	
4						1			1				1	1								1	1	1	1	1	1	1	1	9	
5									1														1		1	1	1	1	1	3	
6		1					1									1	1								1	1	1	1	1	5	
7	1															1	1					1							4		
8	1		1													1		1			1								5		
9	1	1	1	1	1			1				1				1		1	1	1	1	1		1					11		
10		1	1	1			1	1		1	1					1	1		1	1	1	1		1	1	1	1	1	13		
11				1												1	1	1											3		
12	1											1														1			3		
13					1								1															1	3		
14								1																1				1	3		
15						1	1							1								1	1						5		
16								1																1	1	1	1	1	4		
17				1			1	1		1	1	1	1	1		1		1	1	1	1	1							9		
18										1										1	1									2	
19							1							1					1	1									3		
20						1																		1					2		
21														1													1	1	3		
22								1																	1				2		
23		1								1																		1	3		
24												1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3		
	3	5	5	3	4	3	4	6	6	5	6	4	4	5	6	6	5	8	4	4	5	5	3	6	5	5	9	4	6	8	152
25		1					1										1					1							4		
26		1	1				1	1		1	1		1				1		1	1	1	1			1	1	1	1	12		
27										1									1										2		
28																									1				1		
29								1																	1				2		
30						1																							1		
31												1																	1		
32												1	1																2		
	3	5	7	3	5	4	4	7	8	6	7	6	5	6	8	6	5	10	4	4	7	5	3	7	5	6	11	4	7	9	177

在表 4-16 的內容中，整個 MOCC Word 專業級共有 177 項次的能力項目，1~24 屬於 PreMOUS Word 與 MOCC Word 專業級共有的能力項目。在所有的 MOCC Word 專業級當中，這一個部份共有 152 項次的能力項目，佔整個題庫的 86%。而 25~32 為 MOCC Word 加入的能力項目，一共多出 25 項次的能力項目，此一部份佔整個題庫的 14%。

本實驗將採用回歸率測試中表現最好的 IDF 函數。比對系統與專家分類間的差異，以考驗分類效果的優劣。但過程必須考慮整個 MOCC Word 專業級題庫中有 14% 的內容，沒有出現在以 PreMOUS Word 作為樣本的權重向量矩陣。此 14% 的內容是用於模擬題庫能力項目不完整的狀況。

## 一、MOCC Word 專業級試題的內容

MOCC Word 專業級是一種綜合型、高複雜度的題組形式實作題，每一道試題均有數量不等的操作步驟，以第一題的試題內容為例，其內容共有四個步驟。

1. 請設定紙張為 A4，  
邊界為上、下各 3cm，左、右各 3cm，  
與頁緣距離：頁首 1cm，頁尾 1cm。
2. 請將所有的段落，設定左右對齊，  
首行縮排 0.85cm，與前段距離 9pt，  
固定行高 18pt。
3. 請將段落「職業的本質，.....富強康樂。」，  
文字字型改為標楷體，斜體加粗，大小為 14，  
並加底線，文字色彩改為淺藍色。
4. 請將第一段標題"「生活」與「生存」"文字，  
設定字體為細明體，粗體，大小為 18，  
置中對齊，文字字元網底樣式為 15%。

判讀 MOCC Word 專業級的試題必須注意的事項：

1. 每個步驟均有一到多個能力項目
2. 每一個步驟可能是一題到多題 PreMOUS Word 的試題，如果系統採用不分步驟整體判讀的話，則會造成步驟與步驟之間的能力項目重疊或互相干擾，判讀的效率會降低。

## 二、受測試題分析方式

依據上述的注意事項，MOCC 專業級的試題，必須採用每一步驟獨立判讀的方式，而字詞權重函數則採用表現最佳的 IDF。受測試題的

判讀方法及步驟：

- (一) 讀取試題的單一步驟。
- (二) 採用權重值降冪排序後前三名的能力項目。
- (三) 第二名或第三名能力項目的權重，如果大於與第一名能力項目權重的 95% 容許比例，則也視同該操作步驟的能力項目。

例 4-1：假設步驟  $n$  判獨到  $a$ 、 $b$ 、 $c$  的能力項目

權重分別為  $W_a$ 、 $W_b$ 、 $W_c$

若  $W_a > W_b > W_c$

則能力項目  $a$  為操作步驟  $n$  的能力項目

如果  $W_b \geq 0.95W_a$

則能力項目  $b$  亦為操作步驟  $n$  的能力項目

如果  $W_c < 0.95W_a$

則能力項目  $c$  不是操作步驟  $n$  的能力項目

由上述的結果可以得知， $W_a$  為權重值最大，因此必然為操作步驟  $n$  的能力項目，而  $W_b$  的權值在  $W_a$  為權重值得 95% 以上，在此也定義為操作步驟  $n$  的能力項目。

- (四) 如果不是最後的作答步驟,則讀取試題的下一步驟,回到(二)  
繼續執行

### 三、受測試題品質的檢測

受測試題的品質檢測必須先找出試題  $i$  的錯誤率  $E_i$  及正確率  $R_i$ 。

- (一) 錯誤率  $E_i$ : 試題  $i$  的判讀錯誤的次數  $C_{err}$  與專家分析總能力  
項目數  $C_{exp}$  之比。

$$E_i = \frac{C_{err}}{C_{exp}} \quad (4-2)$$

錯誤率在此反映的並非 QCS 的分類品質,也包含測試試題  
本身內容存在的問題,因此可以配合正確率作試題品質的  
檢測。

- (二) 正確率  $R_i$ ：試題  $i$  的判讀正確的次數  $C_{right}$  與專家分析總能力項目數  $C_{exp}$  之比。

$$R_i = \frac{C_{right}}{C_{exp}} \quad (4-3)$$

試題分類系統無法分類的因素有下列三種

1. 可能是函數的精密度不足 (缺判、誤判)
2. 能力項目定義不清或不周全 (誤判)
3. 試題本文的品質問題 (誤判)

為了讓系統可以檢驗出無法分類的原因，定義檢驗分類品質指標有

內因品質指標及外因品質指標：

- (一) QCS 缺判的比例，定義為內因品質指標  $Q_{in}$

則試題  $i$  的內因品質指標為

$$Q_{in} = 1 - R_i \quad (4-4)$$

而品質指標數值所代表的意義如下

- |                    |  |
|--------------------|--|
| 若 $Q_{in} = 1$     | QCS 本身無法做任何分類                          |
| 若 $1 > Q_{in} > 0$ | QCS 本身可針對部份內容正確分類<br>$Q_{in}$ 越小分類品質越高 |
| 若 $Q_{in} = 0$     | QCS 本身可以完整分類(最佳狀況)                     |

(二) QCS 誤判的比例，定義為外因品質指標  $Q_{out}$

則試題  $i$  的外因品質指標為

$$Q_{out} = E_i + R_i - 1 \quad (4-5)$$

而品質指標數值所代表的意義如下

若  $1 - Q_{out} < 0$  誤判導致 QCS 分類錯誤

$Q_{out}$  越大誤判越嚴重

若  $Q_{out} = 0$  沒有誤判導致 QCS 分類錯誤

(最佳狀況)

### 第八節 新增試題分類及分析實驗結果

QCS 使用 IDF 的函數測試後，實驗結果如表 4-17 所示，1 表示該題所具有的能力項目，空白表示該題不具有的能力項目。

表 4-17 使用 MOCC Word 2000 專業級作新增試題樣本的測驗結果

試題 能力項目	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1		1				1									1	1		1	1		1		1	1	1	1		1	1
2	1					1		1		1	1	1				1		1	1			1		1	1					
3	1	1	1				1			1			1			1			1			1			1	1	1	1	1	1
4												1		1		1		1			1									
5											1							1						1						
6			1					1										1										1		
7		1															1						1					1		
8				1													1							1						
9		1		1	1	1			1					1	1		1			1	1	1	1			1	1	1	1	
10			1		1			1	1		1	1				1	1			1		1		1			1	1		
11					1												1		1				1			1		1		
12		1																											1	
13						1								1										1				1		1
14					1																							1		
15							1	1								1									1	1				
16										1																		1	1	1
17															1							1								
18																							1							
19														1	1						1									
20							1														1					1				
21						1										1												1		1
22									1																			1		
23					1						1								1											
24													1	1			1													

接下來考驗分類效果的好壞，在此項實驗嘗試使用錯誤率及正確率分別構成一個 E 矩陣及 R 矩陣，由矩陣中的每一行去取得每一道試題的正確率  $E_i$  與錯誤率  $R_i$ ，最後藉由內因品質指標  $Q_{in}$  及外因品質指標  $Q_{out}$  來診斷試題品質不良的因素，進而去找出錯誤處加以修改。

一、 正確率

上述的結果中，如果只是單純沒有判讀到能力項目的分類，則只會出現一次的判讀錯誤；而如果是判讀到另一個能力項目，會出現兩次判讀錯誤，而造成更高的錯誤率。

與專家分析的能力項目表互相比較，假設測試結果為  $D_1$  的矩陣，專家分析之能力項目表為  $D_2$  矩陣。如果要得到正確判斷的結果，可以將  $D_1$  及  $D_2$  做 AND 運算，運算結果得 R 矩陣如下表所示：

表 4-18 找出正確分類的矩陣 R

試題 能力項目	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	正確次數	
1	1		1				1									1	1		1	1		1			1	1	1		1	1	13	
2	1						1		1	1	1	1				1		1	1			1		1	1						12	
3	1	1	1				1			1			1			1			1			1			1	1	1	1	1		14	
4													1			1		1													3	
5										1															1						2	
6			1					1										1									1				4	
7		1														1						1									3	
8				1													1							1							3	
9		1		1	1	1			1				1			1			1	1		1			1					11		
10			1		1			1	1		1	1				1	1			1		1		1		1				11		
11					1												1		1												3	
12		1																										1			2	
13						1							1																	1	3	
14																											1				1	
15							1	1							1										1	1					5	
16									1																			1	1	1		4
17																						1										1
18																							1									1
19															1							1										2
20							1																			1						2
21																1													1	1		3
22									1																		1					2
23											1																					1
24													1																			1
Cright	3	4	4	2	3	2	4	4	3	3	4	2	3	3	3	5	5	4	4	2	4	4	5	2	4	5	4	6	2	4	4	107
Cexp	3	5	5	3	4	3	4	6	6	5	6	4	4	5	6	6	5	8	4	4	5	5	3	6	5	5	9	4	6	8	152	
Ri	1	0.8	0.8	0.7	0.8	0.7	1	0.7	0.5	0.6	0.7	0.5	0.8	0.6	0.5	0.8	1	0.5	1	0.5	0.8	1	0.7	0.7	1	0.8	0.7	0.5	0.7	0.5	0.72	

表 4-18 的行代表試題，列代表能力項目，整個 R 矩陣中，含有 1 的元素便是判斷的部份，每一題判讀正確的次數紀錄於  $C_{right}$ ，而經專



家分類每一題共有能力項目紀錄  $C_{exp}$  ,  $R_i$  即為試題  $i$  的正確率。

表 4-19 正確率的區間分佈比例

正確率	筆數	比例
0.0 ~ 0.2	0	0%
0.2 ~ 0.4	0	0%
0.4 ~ 0.6	9	30%
0.6 ~ 0.8	14	47%
0.8 ~ 1.0	7	23%

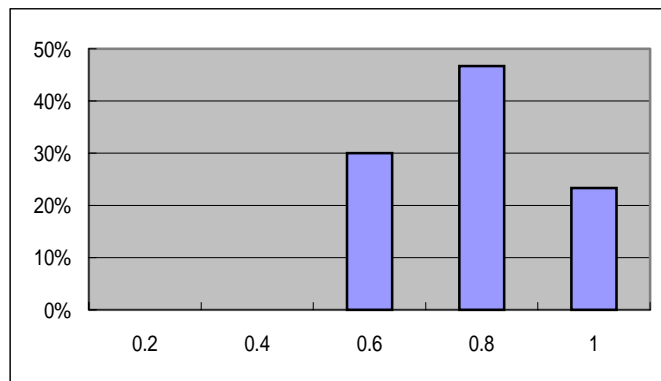


圖 4-17 正確率區間分佈圖

由正確率的區間來看，大部份的試題，都偏向於高正確率，其平均正確率則高達 72%，具有不錯的分類效果。

## 二、 錯誤率

與專家分析的能力項目表互相比較，假設測試結果為  $D_1$  的矩陣，專家分析之能力項目表為  $D_2$  矩陣。如果要得到正確判斷的結果，可以將  $D_1$  及  $D_2$  做 AND 運算，運算結果的  $E$  矩陣如下表所示：

表 4-20 運算後所得的 E 矩陣

試題 能力項目	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	錯誤數		
1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	
2	0	0	0	0	1	0	0	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	9	
3	0	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	8	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	
6	0	0	0	1	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	1	10	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	2	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	3	
9	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
10	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	4	
11	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	4	
14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	4	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	3	
16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	2	
18	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	5	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	2	
21	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cerr	0	1	1	1	2	4	0	2	3	2	2	3	1	4	5	1	1	6	0	2	2	1	3	3	0	2	7	5	3	5	72		
Cexp	3	5	5	3	4	3	4	6	6	5	6	4	4	5	6	6	5	8	4	4	5	5	3	6	5	5	9	4	6	8	152		
Ei	0	02	02	03	05	13	0	03	05	04	03	08	03	08	08	02	02	08	0	05	04	02	1	05	0	04	08	13	05	06	0.47		

表 4-20 的行代表試題，列代表能力項目，整個 E 矩陣中，含有 1 的元素便是錯誤的部份，每一題判讀錯誤的次數紀錄於  $C_{err}$ ，而經專家分類每一題共有能力項目紀錄  $C_{exp}$ ， $E_i$  即為試題  $i$  的錯誤率。

由錯誤率可以判斷，第 6、23、28 號的試題，數值都大於 1，表示判斷錯誤超過該題原有的能力項目分類；而錯誤率為 0%，則表示完全判斷正確。表中顯示總錯誤率有 47%，然而這一個錯誤率所代表的意義不應該只是分類效率的問題而已，如果參考表 4-16 可以發現，MOCC Word 2000 專業級共有 32 個能力項目，比起 PreMOUS Word 2000 多了 8 個。因此若精確的考慮多出來的能力項目，應該可以降低錯誤率。

表 4-21 實驗所得之錯誤率區間比較表

錯誤率	題數	比例
0.0 ~ 0.2	9	30%
0.2 ~ 0.4	7	23%
0.4 ~ 0.6	5	17%
0.6 ~ 0.8	5	17%
0.8 ~ 1.0	2	2%
1.0 ~	2	2%

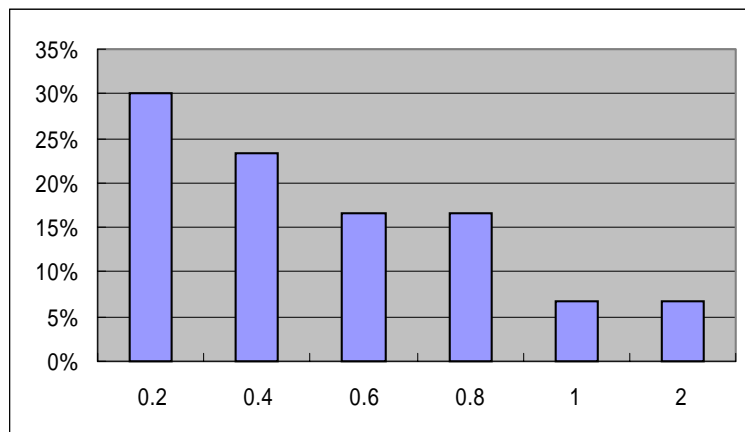


圖 4-18 錯誤率區間分佈圖

由錯誤率區間比較圖，可以發現大部份的試題都偏向於少數，因此可了解本分類系統的分類效果偏向於中低錯誤率。

### 三、 試題分類品質

本 QCS 可高達 72% 的正確率，確實可以對複雜題組型的綜合能力項目術科試題，作可靠的能力項目分類。而本實驗中所使用的實驗組樣本 MOCC Word 2000 有 14% 的內容與目前分類系統使用的分類權重向量矩陣不同，目的是要模擬能力項目分類不周全的狀況。

使用品質指標檢測得

$$Q_{in} = 1 - 0.72 = 0.28$$

$$Q_{out} = 0.47 + 0.72 - 1 = 0.19$$

可以得知 28% 來自於缺判，在於權重向量矩陣或字詞權重函數的辨識能力的問題。而 19% 的可能來自於誤判，包括可能來自於能力項目定義不清(如本實驗組樣本的能力項目數多餘系統本身可辨識的能力項目數)或不周全或試題本文的品質問題(如試題的字詞錯誤或太多無關於試題本身主要內涵的內容...等)。

例 4-2：當系統判讀之後，可以得到下表的結果，其中紀錄各題的  $E_i$ 、

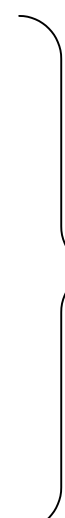

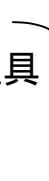

$R_i$ 、 $Q_{in}$  及  $Q_{out}$ 。

表 4-22 藉由錯誤率與正確率之和找出誤判的試題

題號	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$E_i$	0.00	0.20	0.20	0.33	0.50	1.33	0.00	0.33	0.50	0.40	0.33	0.20	1.00	0.50	0.00	0.40	0.78	1.25	0.50	0.63	0.00	0.67	0.67	1.00	0.80	0.67	0.50	0.67	0.50	
$R_i$	1.00	0.80	0.80	0.67	0.75	0.67	1.00	0.67	0.50	0.60	0.67	0.00	0.67	0.67	1.00	0.80	0.67	0.50	0.67	0.50	0.00	0.33	0.33	0.00	0.20	0.33	0.50	0.33	0.50	
$Q_{in}$	0.00	0.20	0.20	0.33	0.25	0.33	0.00	0.33	0.50	0.40	0.00	0.00	0.33	0.33	0.00	0.20	0.33	0.50	0.33	0.50	0.00	0.33	0.33	0.00	0.20	0.33	0.50	0.33	0.50	
$Q_{out}$	0.00	0.00	0.00	0.00	0.25	1.00	0.00	0.00	0.00	0.00	0.00	0.20	0.67	0.17	0.00	0.20	0.44	0.75	0.17	0.13	0.00	0.67	0.17	0.00	0.20	0.44	0.75	0.17	0.13	

結果可以看出，5、6、12、14、15、17、18、21、22、23、24、26、27、28、29、30 題有誤判，其中第 6 題  $Q_{out}=1$  及第 28 題  $Q_{out}=0.75$  最嚴重。

以最嚴重的第 6 題為例，題目中共有 7 個步驟，內容如下：

<ol style="list-style-type: none"> <li>1. 請使用合併列印功能，並建立主文件為 "套印信件" 類型，資料來源請讀取 "安裝本系統的所在資料夾下的 Data 資料夾中 "客戶資料檔.doc" 檔案"。</li> <li>2. 在文件「您好」的前面，利用「插入合併欄位」 功能插入「客戶名稱」這個欄位，在內文第一行 「參加」「一隊」中插入「團名」這個欄位， 在第二行「十五日」「親臨」之間插入 「說明會時間」這個欄位，在第三行 「由」「為您」之間插入「解說人員」這個欄位。</li> </ol>		<p>合併列印</p>
<ol style="list-style-type: none"> <li>3. 在「填妥下面表格:」下方的表格中， 分別在第一列第一欄填入「身份證字號」， 第二列第一欄填入「習慣暱稱」，第三列第一欄填入 「聯絡電話」，第四列第一欄填入「出團日期」， 文字對齊方式設定為「分散對齊」。</li> <li>4. 請設定表格第一欄欄寬 3cm，第二欄欄寬 5cm， 第三欄欄寬 4cm。</li> <li>5. 請將「身份證字號」其後的兩個儲存格先合併 再分割成 10 格，每格寬度為 0.6cm。</li> </ol>		<p>表格內容 設定</p>
<ol style="list-style-type: none"> <li>6. 請將第二列第三欄到第四列第三欄合併， 再利用繪圖工具列中的「垂直文字方塊」工具 填入"浮貼照片"文字，字型為新細明體， 大小為 12，文字方塊大小為 高度 1.27cm，寬度 1.97cm，色彩及框線為 無填滿，無線條。</li> </ol>		<p>文字方塊 +表格內容設定 +字型設定</p>
<ol style="list-style-type: none"> <li>7. 請將整個表格自動設定為"現代的"格式。</li> </ol>		<p>表格內容設定</p>

將試題內容與表 4-23 互相比照，可以分析結果如下：

1. 該題第 1、2 項為『合併列印』功能，包含 PreMOUS 的 24 個  
能力項目並沒有『合併列印』的項目，系統會判讀錯誤的能力

項目，因此這都是誤判。

2. 第 3、4、5、6 項沒有誤判。
3. 第 7 項很明顯是表格內容設定，誤判成字型設定、插入日期兩個能力項目。

表 4-23 各項次權重排行前三名的能力項目

作答步驟	權重	能力項目碼	能力項目
1	88.00	21	插入 Excel 物件
	76.00	10	在文件中插入圖片
	40.00	9	表格內容設定
2	168.00	9	表格內容設定
	66.00	11	使用定位點
	48.00	14	插入日期
3	139.56	9	表格內容設定
	109.16	13	加入文字方塊
	76.51	16	設定文件頁首及頁尾
4	98.65	9	表格內容設定
	12.00	10	欄位設定
	9.85	6	在文件中插入圖片
5	72.00	9	表格內容設定
	24.00	16	設定文件頁首及頁尾
	12.00	24	使用複製、貼上功能
	12.00	5	分隔設定
6	173.94	13	加入文字方塊
	137.94	9	表格內容設定
	41.94	17	使用繪圖工具
7	12.85	2	字型設定
	12.85	14	插入日期
	11.00	16	設定文件頁首及頁尾

藉由  $Q_{in}$  與  $Q_{out}$  可以了解到試題分類的品質，並了解試題的品質不

良是來自於缺判及誤判，缺判大部份可依靠系統設計者改良 QCS 來達成；而誤判方面，雖然有可能是函數本身的問題，但試題本文內容不良絕對會導致判斷錯誤。而這方面可由系統立即回饋  $Q_{out}$  的數據大於 0 的試題，供命題人員進行參考並進行試題的修正。