

# 中華民國第 54 屆中小學科學展覽會 作品說明書

---

高中組 生活與應用科學科

040808

可以不一樣—中文字推薦學習

學校名稱：國立花蓮高級中學

作者：  高一 黃少輔  高一 鄒秉均  高一 莊賀凱	指導老師：  溫健順
---	------------------

關鍵詞：中文學習、學習推薦、媒體文字

## 摘要

透過街頭民調找出生活中常被使用的媒體，計算出各媒體中各個文字被使用的次數並將之排序，重新安排國小國語課本生字學習順序，以在各媒體中出現次數高的中文字優先學習，並製作應用程式（App），讓學習者可以更有效率且隨時隨地學習中文。

## 壹、研究動機

外籍英文老師，雖然在台灣生活了好幾個月，但對於最基本的中文字卻非常不熟悉。而家中就讀小學的弟弟看報章雜誌還是有些吃力。一般的中文學習法因為缺乏效率，造成學習者許多不便與挫折，為此希望能找出最常使用的中文字並整理排序，提供中文學習者最佳的學習順序，讓學習者能快樂且有效率地學習中文

## 貳、研究目的

- 一、找出國小生字在日常生活中的使用頻率
- 二、重新安排中文生字學習順序，找出有效率的學習方式
- 三、撰寫應用程式，讓使用者更有效率且快樂的學習中文

## 參、研究設備及器材

- 一、32 位元筆記型電腦  
處理器：Intel(R) i5-2410 @ 2.3GHz  
記憶體：4.00GB  
作業系統：Windows 7
- 二、相關程式：
  - (一)編譯程式：Dev-C++、Java
  - (二)應用程式：Eclipse

## 肆、研究過程

### 一、緒論

透過街頭民調找出生活中常被使用的媒體，計算出各媒體中各個文字被使用的次數並將之排序，重新安排國小國語課本生字學習順序，在各媒體中出現次數高的中文字優先學習，並製作應用程式，讓學習者可以更有效率且隨時隨地學習中文。

### 二、文獻探討

#### (一)、如何學習中文

「聽、說、讀、寫」是學習語言的關鍵，「生活是學習語言的最佳場所」。本次

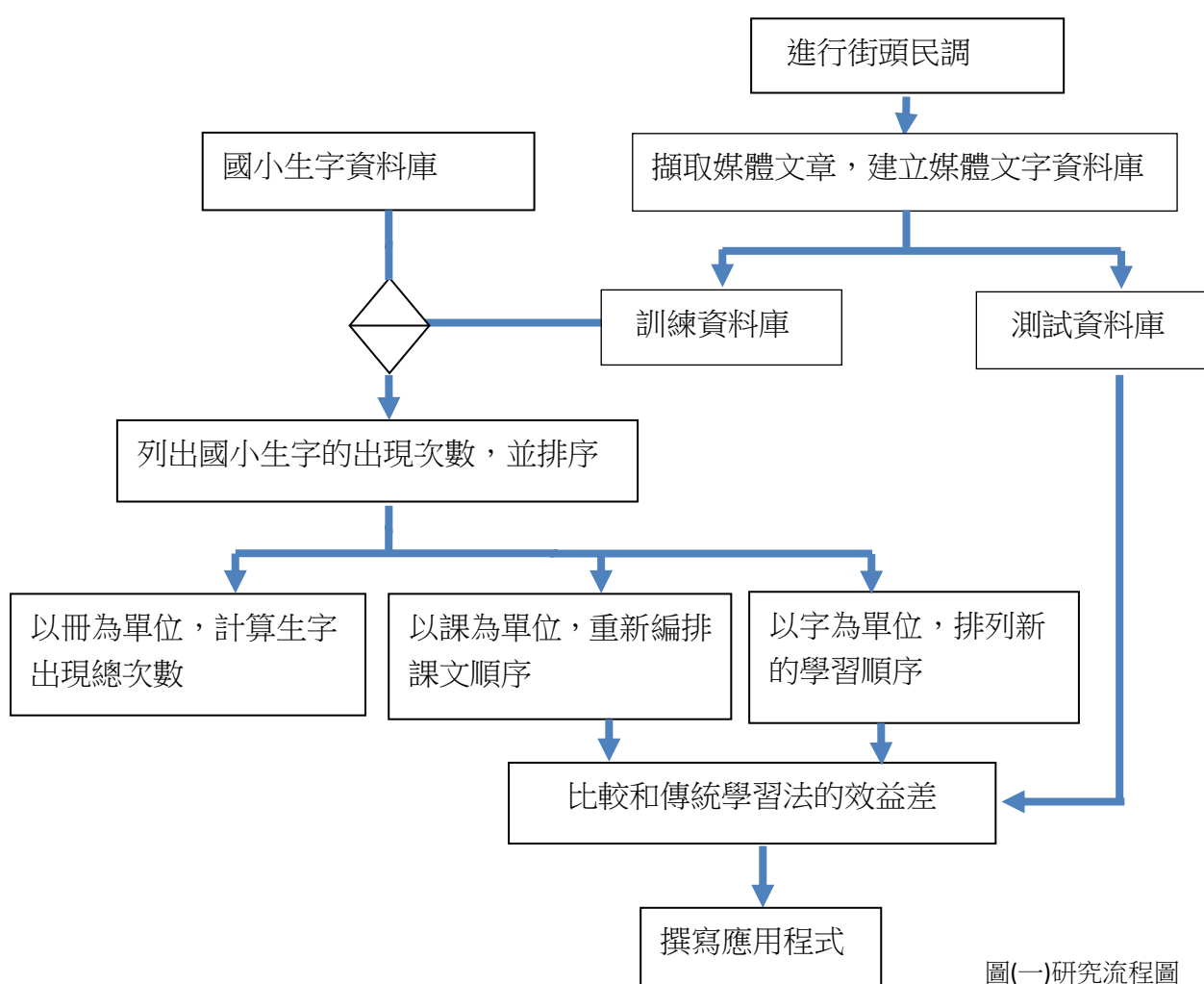
研究不管在選取資料來源或撰寫應用程式，都遵照「語音媒體(聽、說)」和「書面媒體(讀、寫)」並重，且貼近日常生活。

## (二)、中文的特色

每種語言都有其獨一無二的特色和不同的學習方法。就中文的本質而言，中文是一字多音多義的，字將會因為詞的用法不同而有不同的意思，如「步行」和「進步」的'步'，就有著不同的詞性和意思。本次研究所使用的學習法將會以“中文詞句”為主要的學習方式和研究方向。

## 三、研究過程

### (一)、研究流程



圖(一)研究流程圖

1. 實際民調：了解日常生活中媒體被使用情形，以建立媒體被使用的機率排序。
2. 建立資料庫：

本研究先建立二個資料庫(國小生字資料庫和媒體文字資料庫)，再將媒體文字資料庫分為訓練資料庫及測試資料庫，分別如下：

- (1) 國小生字資料庫  
以國小國語課本一到十二冊生字為底蘊，建立「國小生字資料庫」。
- (2) 媒體文字資料庫  
依民調結果擷取文字和影音媒體各五名之中的文字檔，建立「媒體文字資料庫」
  - a. 訓練資料庫  
依民調中媒體被使用率高低，從「媒體文字資料庫」中按比例分配，建立「訓練資料庫」，用來找出中文生字在媒體的出現次數
  - b. 測試資料庫  
將未被選取到「媒體文字資料庫」的資料歸類成為「測試資料庫」用來比較學習效益
3. 列出國小國語課本生字在訓練資料庫裡的出現次數排序
  - (1) 比對國小生字和訓練資料庫，計算各個生字在訓練資料庫中出現的總次數，按大小排列
  - (2) 分別列出以「冊」和「課」為單位的中文生字出現次數表
4. 分析國小生字出現次數排序
  - (1) 以冊為單位，觀察生字出現次數分布
  - (2) 以課為單位，計算每課各個生字在訓練資料庫中出現總次數，製作圖表加以觀察，並重新編排課文學習順序
  - (3) 化整為零，以單一中文生字為單位，製作圖表加以觀察
5. 利用測試資料庫，分別以課為單位、字為單位比較與傳統學習法的效益差
6. 撰寫中文推薦學習應用程式
  - (1) 以生字出現的頻率決定學習順序  
為了強調速學，向使用者推薦在測試資料庫中出現總次數較多的國小生字，希望能達到快速學習的目的
  - (2) 加入造詞  
讓使用者就中文一字多意的特點，使用以詞為單位學習方式學到單一字多種的用法
  - (3) 加入連續推薦  
以該推薦學習字的關聯性為中心，延伸學習與該字有關的其他詞。提高學習效益加入時間性推薦
  - (4) 加入時間性推薦  
在特定日期或時間下推薦在此時期或時間常用的字詞推薦給使用者，讓學習能貼近生活增加並增加在日常中應用可能性

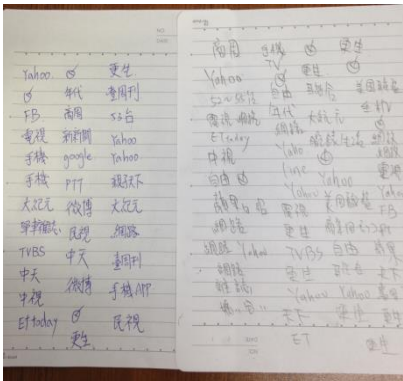
(5) 加入主題式推薦

依時事內容和學習者感興趣的方向，提供學習方向，滿足學習者的需求

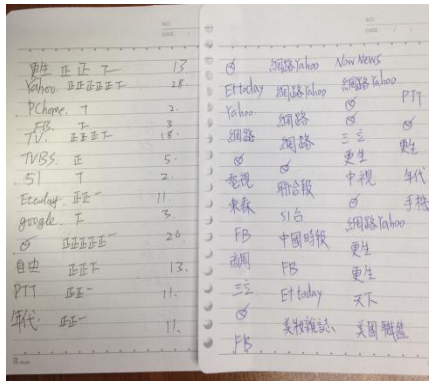
(二)、民調和媒體選用流程

1. 實際民調數據

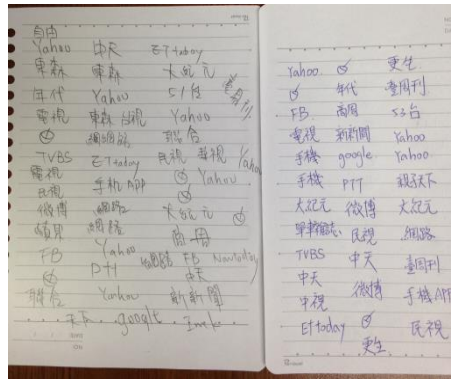
本次研究分別於 103/2/18、103/2/19 及 103/2/22 三日，以街頭問答方式進行，問題為「請問你（妳）最常使用哪種媒體看新聞或取得資訊？」，受訪 800 人，其中有效回答人數共 600 人。根據受訪者的答案，可分成五大類：網路、電視新聞、報章雜誌、電子報告欄、社群網站。整理出媒體使用情形與分類，如下圖(二、三、四)、表(一)



(圖二) 民調 103/2/18



(圖三) 民調 103/2/19

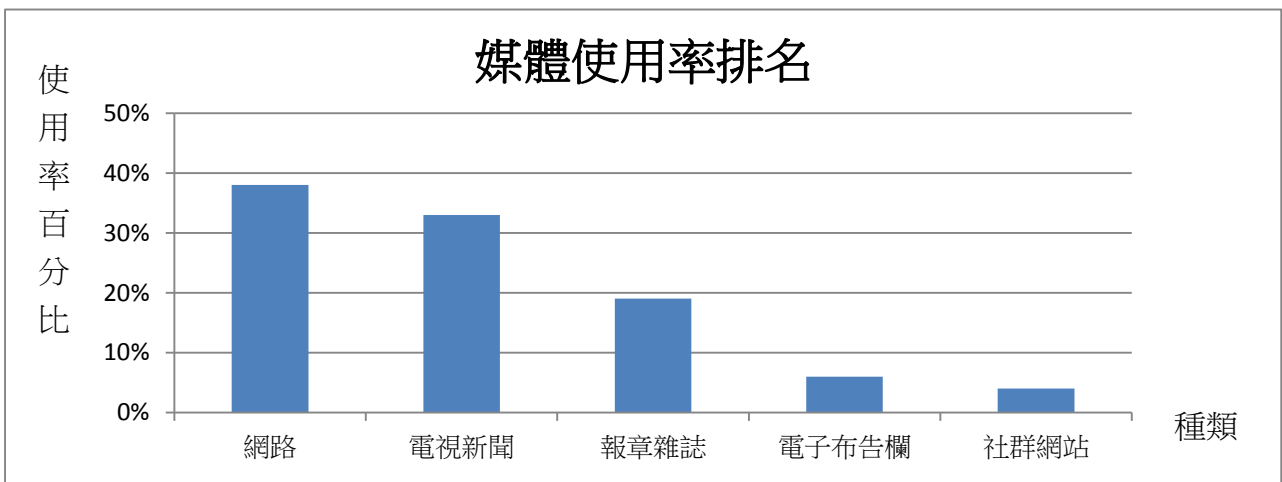


(圖四) 民調 103/2/22

2. 民調媒體使用率結果分析

	網路	電視新聞	報章雜誌	電子布告欄	社群網站	總數
人數	228	198	114	36	24	600
百分率	38%	33%	19%	6%	4%	100 %

表(一)民調統計圖



(圖五) 媒體使用率排名

### 3. 選擇媒體做為媒體文字資料庫樣本來源

由上圖(五)可看出網路媒體使用率最高，網路媒體又以「YAHOO 奇摩」所占比例最高，但其綜合文字媒體、影音媒體、超連結等，且內容大多來自各大書面或影音體。如下圖所示三則理由，可以發現網路媒體中文件與電視新聞及報章雜誌重疊性非常高

#### (1) 不使用「YAHOO 奇摩」的理由如下：

- a. 大部分為 YAHOO 奇摩新聞
- b. 其本身綜合文字、影音媒體和超連結，如下圖(六)



(圖六)網路媒體圖

#### c. 內容多來自各大新聞報章雜誌



資料來源

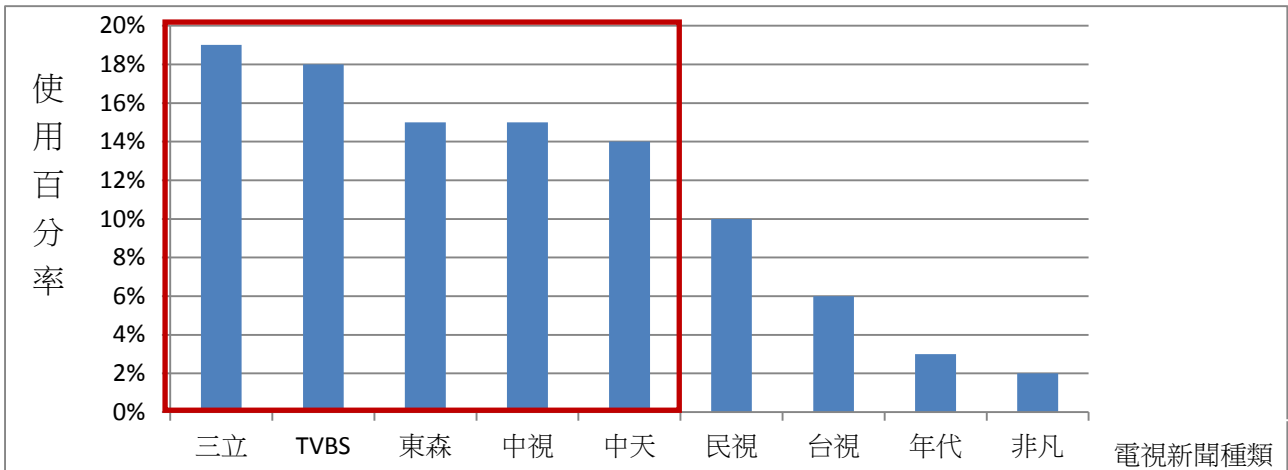
(圖七、八) 網路文章圖

#### (2) 不使用電子布告欄及社群網站的理由

電子布告欄及社群網站在受訪者的答案中，不為中文訊息的主要來源，因此本次研究以民調中第二及第三高的分類「電視新聞」和「報章雜誌」為建立媒體文字資料庫的樣本來源

#### (3) 選用媒體理由

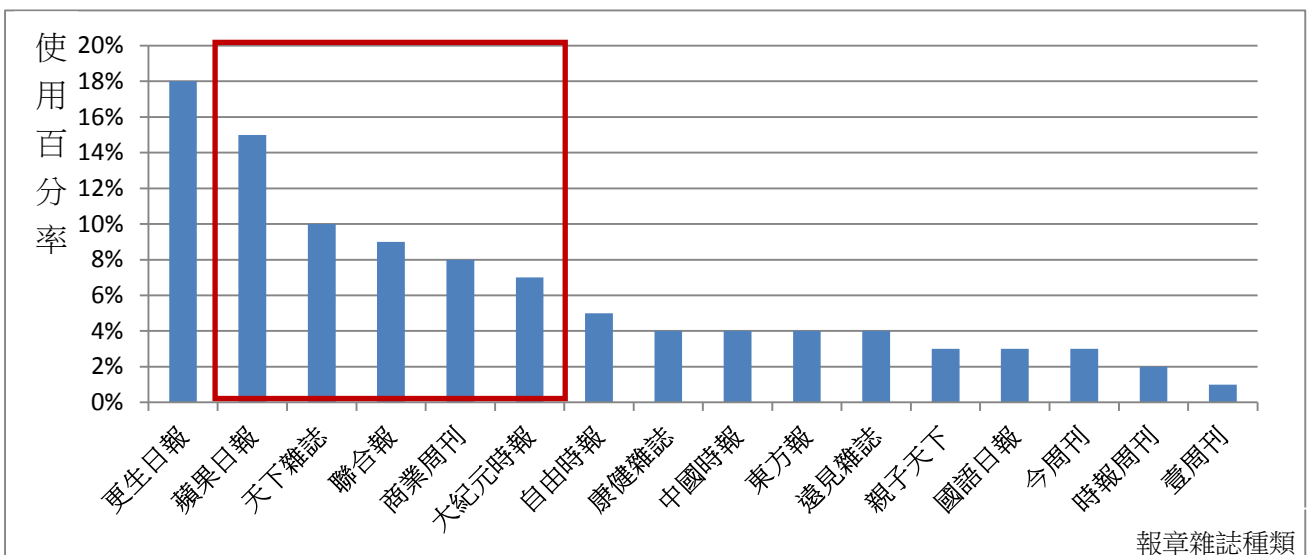
在電視新聞使用率排名中，發現中天新聞與民視新聞間出現較大差距，故採用三立新聞、TVBS 新聞、東森新聞、中視新聞、中天新聞等媒體做為來源，如下圖(九)



(圖九) 電視新聞使用率排名

#### (4) 選用媒體理由

在報章雜誌使用率排名中，發現大紀元時報和自由時報間出現較大差距，又因更生日報為地方報，不具有普遍性，故採用蘋果日報、天下雜誌、聯合報、商業週刊、大紀元時報做為媒體來源，如下圖(十)



(圖十) 報章雜誌使用率排名

#### (二)、資料庫建立流程

##### 1. 建立國小生字資料庫

以國小國語課本一到十二冊生字為藍本，建立「國小生字資料庫」

##### 2. 建立媒體文字資料庫

###### (1) 媒體文字的來源

選擇電視新聞媒體五名和報章雜誌媒體五名（共十名）

分別為 電視新聞：東森新聞、TVBS 新聞、三立新聞、中天新聞、中視新聞

報章雜誌：蘋果日報、天下雜誌、商業週刊、聯合報、大紀元時報

(2) 選擇的方式：

a. 時間日期依據

因為一年為一較完整時間周期，故以 2013 年 4 月至 2014 年 3 月為期（共採記一年）。而一年內不同日期選取的文章性質完全不同，為使其公平，使用日期亂數程式，隨機挑選該媒體要採用的文章的日期

b. 項目分類

使用項目分類亂數程式，隨機挑選要選擇媒體文章之分類項目

c. 公平性

因為各個媒體都有其不同數目的項目分類，為了達到選取文章種類的公平性，故以十項媒體各個分類項目的最小公倍數（120），做為每個媒體文章擷取的份數。以商業周刊為例：如下表(2)、圖(11)所示，商業周刊共有八項分類，故每項分類各取 15 份文章



項目分類

圖(十一) 商業周刊項目分類圖

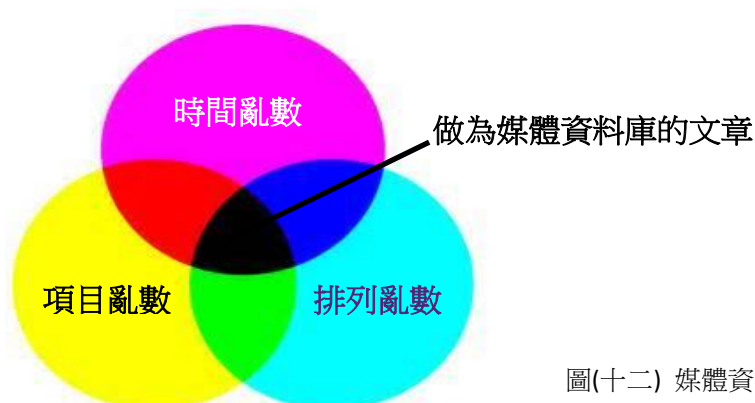
媒體	東森	TVBS	三立	中天	中視	蘋果	天下	商周	聯合	大紀元	備註
媒體分類項目總數	10	12	8	8	10	8	10	8	20	8	最小公倍數 120
各媒體選取份數	120	120	120	120	120	120	120	120	120	120	為媒體文字資料庫共 1200 份
平均分類選取篇數	12	10	15	15	12	15	12	15	6	15	120/各媒體分類項目數

表(二) 媒體平均分類選取文章數目



d. 文章排列：

選取各媒體 120 份文章後，利用排列亂數程式，將其重新隨機排列，如下圖(十二)



3. 建立訓練資料庫和測試資料庫

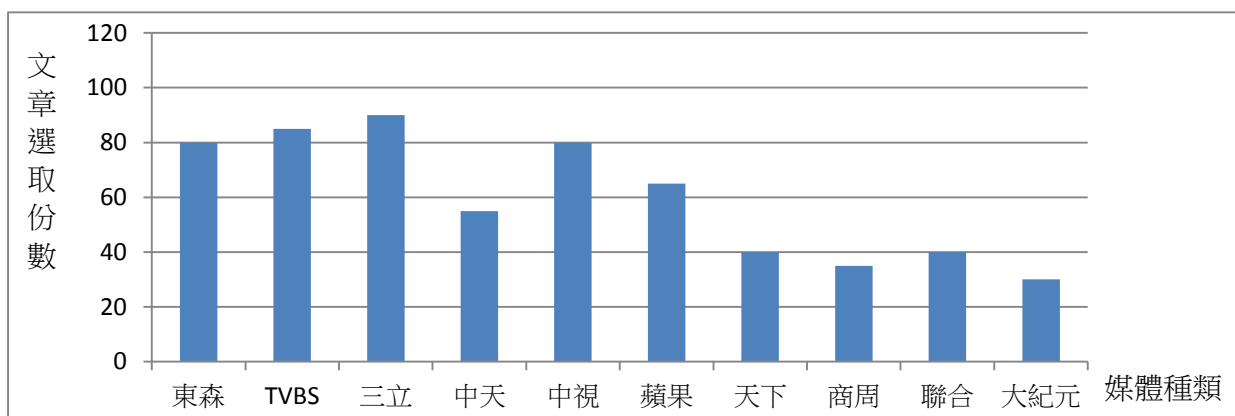
(1) 文章總份數

將媒體文字資料庫（1200 份文章）分為一半，建立訓練資料庫（600 份）和測試資料庫（600 份）

(2) 選取依據

a. 民調比例：

依據各項媒體被使用的比例乘以所採取文章總份數（600 份），做為媒體文章在訓練資料庫中的文章份數，剩下的文章則做為測試資料庫，如下圖(十三)



圖(十三) 媒體樣本選取份數圖

b. 選擇方式：

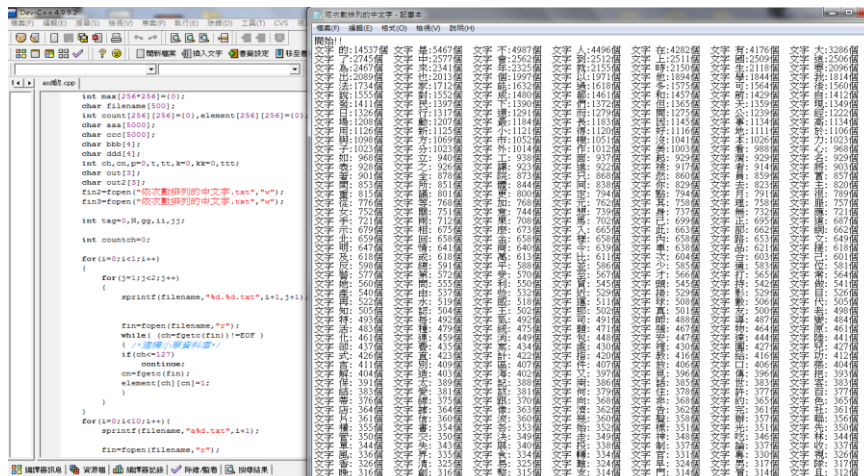
在其媒體所占媒體資料庫中的總份數（120 份），選取所佔比例份數，從第一份取到要採取的份數（文章排序已打亂，故達到隨機選取的目的），做為該媒體在訓練資料庫中的文章。

媒體項目	東森	TVBS	三立	中天	中視	蘋果	天下	商周	聯合	大紀元	備註
選取份數	120	120	120	120	120	120	120	120	120	120	媒體資料庫 1200 篇
民調百分比	12.8%	14.1%	14.7%	9.6%	12.8%	10.8%	7.2%	6.1%	6.5%	5.4%	共 100%
選取文章數	80	85	90	55	80	65	40	35	40	30	訓練資料庫 600 篇
未選取文章數	40	35	30	65	40	55	80	85	80	90	測試資料庫 600 篇

表(三) 媒體文字資料庫內容分類表

### (三) 國小國語課本生字在訓練資料庫中出現總次數排序

1. 比對國小生字資料庫及訓練資料庫，計算各個國小生字在訓練資料庫中出現總次數，結果如下圖(十四)



圖(十四) 國小生字出現總次數統計程式圖

2. 以課為單位，計算每課生字出現總次數

#### (1) 以課為單位，生字出現總次數表

橫軸:冊數

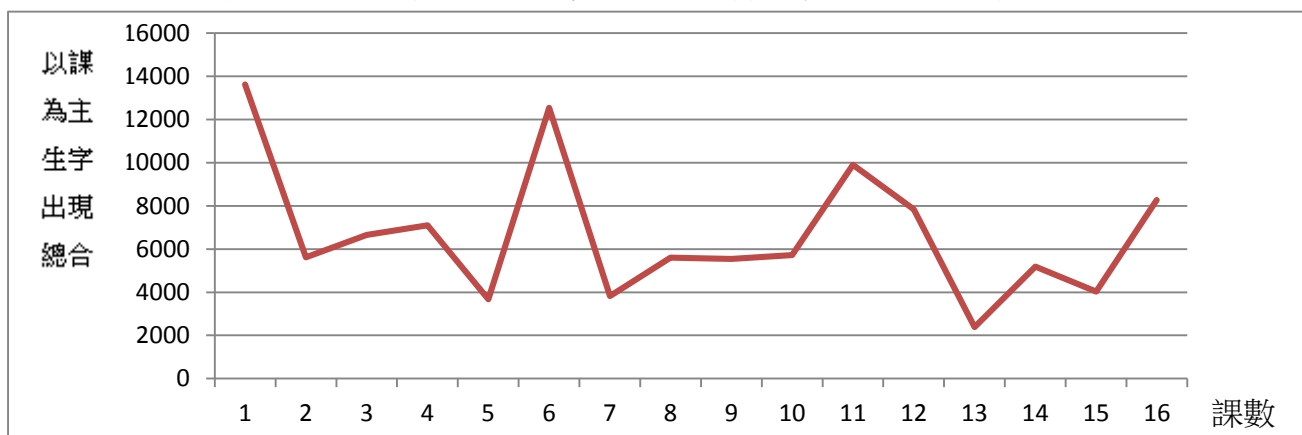
縱軸:課數

	1	2	3	4	5	6	7	8	9	10	11	12
1	9584	13631	7558	4779	2678	490	1110	85	1145	100	224	128
2	13766	5613	3951	3803	4512	1735	636	638	460	239	454	273
3	5985	6648	5830	2977	3393	999	829	1042	362	765	454	152
4	26782	7093	9258	1881	1585	1930	998	623	297	802	841	157
5	11721	3680	2216	3011	3935	1510	656	932	629	223	514	43
6	4519	12546	5606	1566	2188	2045	467	525	828	940	100	26
7	8730	3816	4974	2266	2240	1792	338	917	157	198	80	332
8	10108	5601	4027	3062	2678	1977	2026	1707	1676	501	161	213
9	0	5548	2757	2104	3374	1418	1163	223	452	597	183	122
10	0	5727	5949	3022	2066	1546	865	2272	477	71	261	331
11	0	9898	6793	1135	1809	600	1116	1174	645	198	423	81
12	0	7845	1957	1733	977	2078	302	1124	434	1134	29	125
13	0	2381	3807	2656	2443	834	485	265	344	313	194	0
14	0	5193	1462	4260	3551	793	825	446	493	377	245	0
15	0	4029	3714	5159	0	0	0	0	0	0	0	0
16	0	8275	3083	2589	0	0	0	0	0	0	0	0
總和	91195	106233	72945	46003	37429	19747	11816	11759	8399	6458	4163	1983

(表五) 以課為單位的生字出現次數表

(2) 製作以課為單位之生字出現總次數折線圖 (以第二冊為例):

發現在傳統學習法中，每課的生字在訓練資料庫中出現的總合多寡並無規律性，如下圖(十五)，出現次數總和最高的課數為第一課，依次為第六課、第十一課……是否以生字出現總次數較高的課文優先學習，會有最好的學習效果呢？

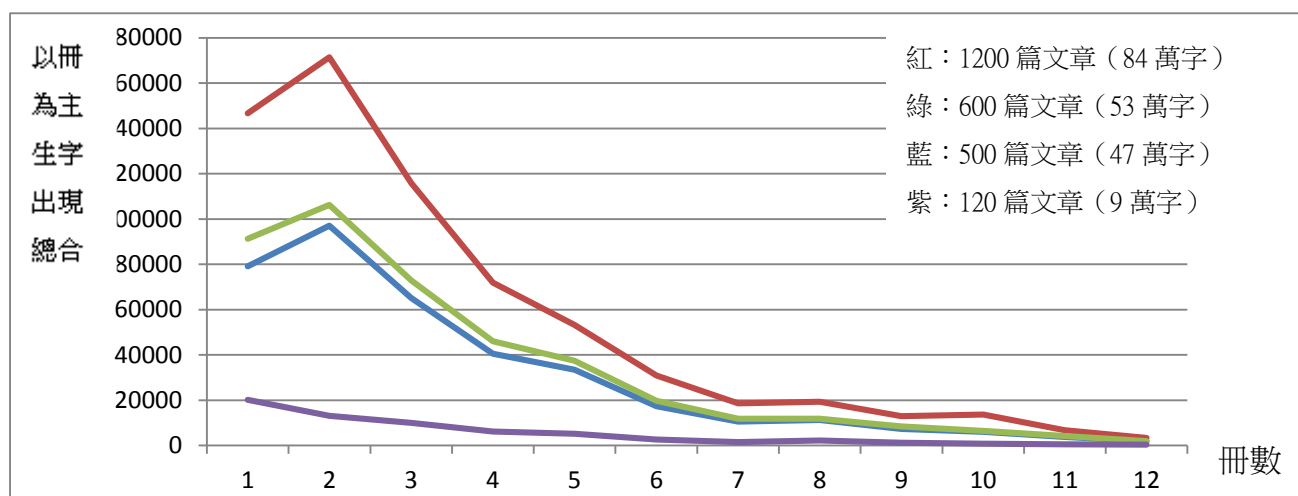


圖(十五) 以課為單位製作之課文生字出現總次數統計折線圖

3. 以冊為單位，計算每冊生字出現總次數，並繪出折線圖加以觀察

(1)分別取 4 種不同份數的訓練資料、每份多次平均。整理生字出現總次數。以冊為單位，

製作國小生字出現總折線圖，發現生字出現次數的趨勢走向與訓練資料總份數、字數呈正相關。



(圖十六) 以冊為單位之文字出現頻率折線圖

## 伍、結論

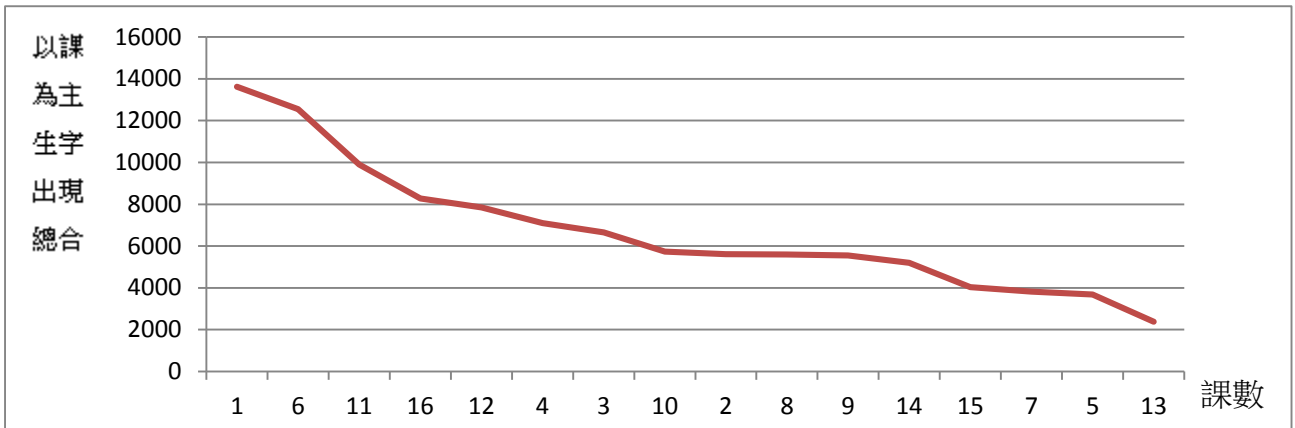
- 一、透過街頭民調，找出媒體使用率排名，選出本次研究所需媒體。
- 二、以國小國語課本一到十二冊生字為藍本，建立出「國小生字資料庫」。透過街頭民調，找出媒體使用率排名，選出本次研究所需媒體。
- 三、以上述媒體文章為來源，在民調比例和公平性的條件下，建立出「訓練資料庫」及「測試資料庫」。
- 四、比對國小生字資料庫及訓練資料庫，列出各國小生字在訓練資料庫中出現的總次數。
- 五、分別以冊為單位、以課為單位及以字為單位，製作圖表觀察各個生字出現次數的分布。
- 六、分別以課為單位、以字為單位重新排列國小國語課本生字出現次序，出現總次數高的生字優先學習。
- 七、測試出以課，字為單位的新學習順序與傳統學習法的效益差(討論部分)。
- 八、撰寫應用程式，讓使用者快樂且更有效率得學習中文。

## 陸、討論及應用

(一) 以課、字為單位，改變學習中文生字順序

1. 以課為單位，重新編排課文學習順序(以第二冊為例)

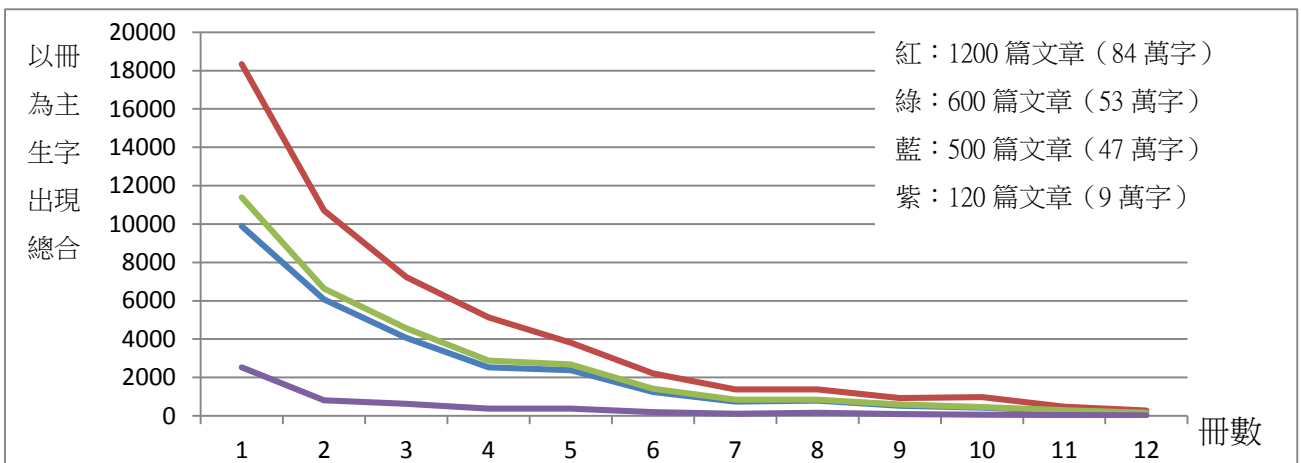
依生字出現總次數由高到低排列，如下圖(十七)如果從生字出現總次數最高的課數依序學習，學習者將可在比傳統學習法更短的時間內學會出現總次數較多的國小生字。



圖(十七) 以課為主重排課文生字出現總次數統計折線圖

2. 以冊為單位，觀察生字出現次數分布

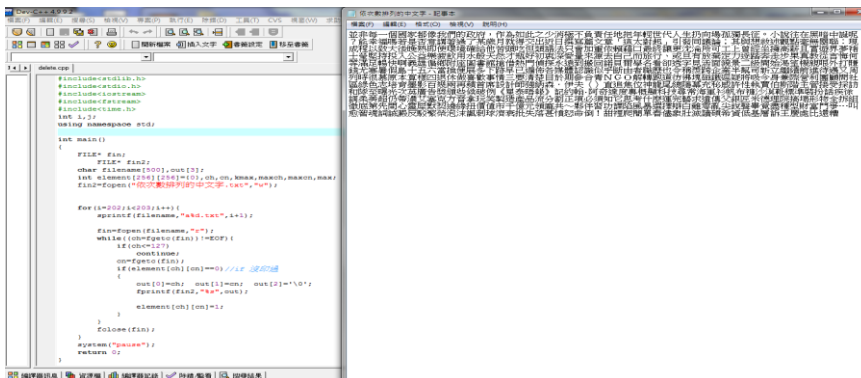
每冊課數不同導致缺乏公平性，如圖(十六)，第一冊因為只有八課，而第二冊有十六課，可能導致第一冊生字出現總次數低於第二冊的趨勢。故將每冊生字出現總次數除以每冊總課數，以達到公平性。並以冊為單位，製作國小生字出現總次數折線圖。



(圖十八) 以冊為單位之文字出現頻率折線圖

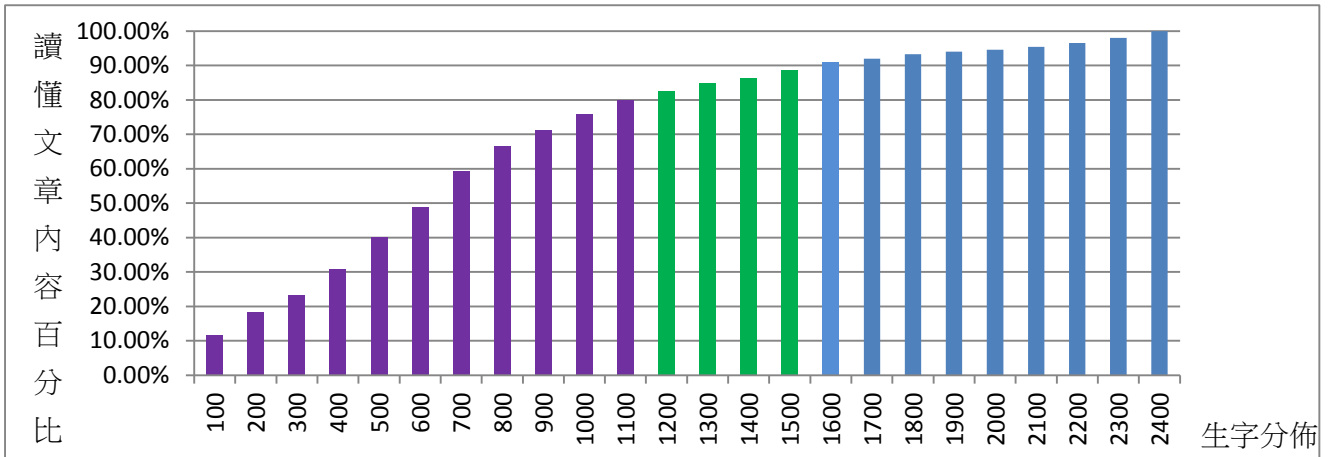
(二) 找出新學習方式，並比較與原先方式的效益差

1. 使用亂數方式從測試資料庫中選取文章
2. 利用程式將文章中重複的字刪除，避免重複的字影響數據。



圖(十七) 刪除文章重複文字程式圖

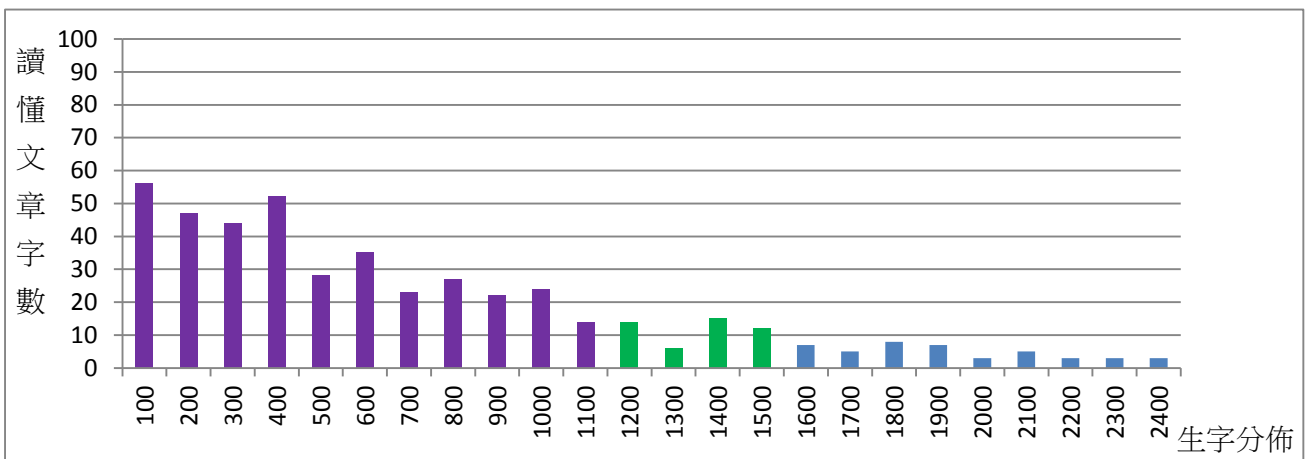




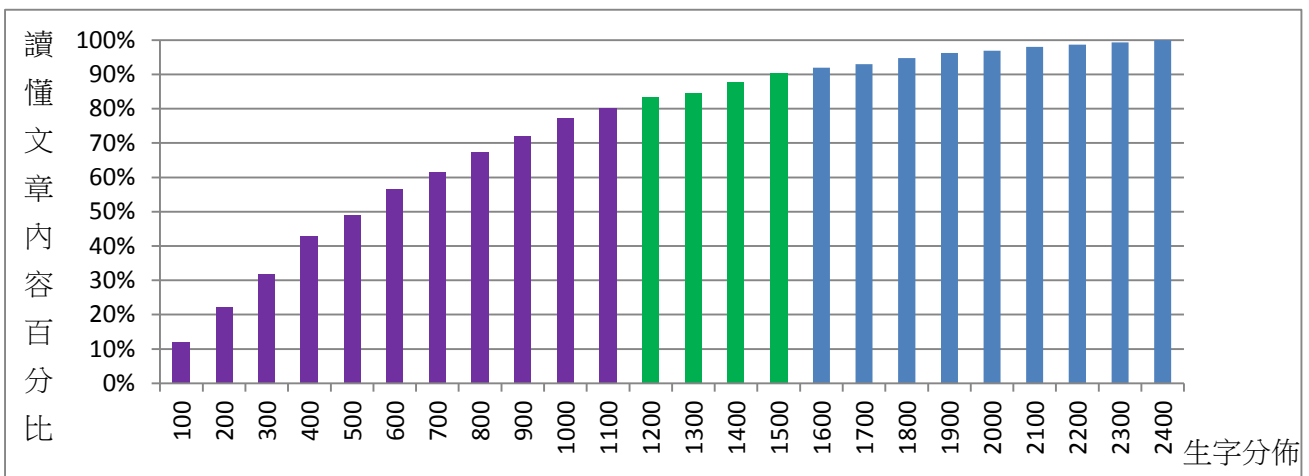
圖(二十) 文章文字在傳統學習法中的分布累加長條圖

(3) 繪出文章文字在以課為單位(新)的學習法的分布長條圖、文字百分比累加長條圖

如圖(二十一、二十二)。由圖(二十一)發現在以課為主(新)的學習法中，雖然讀懂文章的百分比和學習生字的字數較具規律性，但在觀看總體效益可以發現和傳統學習法無太大差異，如圖(二十二)，讀懂文章的80%均須學完1100個國小生字，讀懂文章的90%均須學完1500個國小生字。為何差異不大，我們推測和課文不跨冊關。



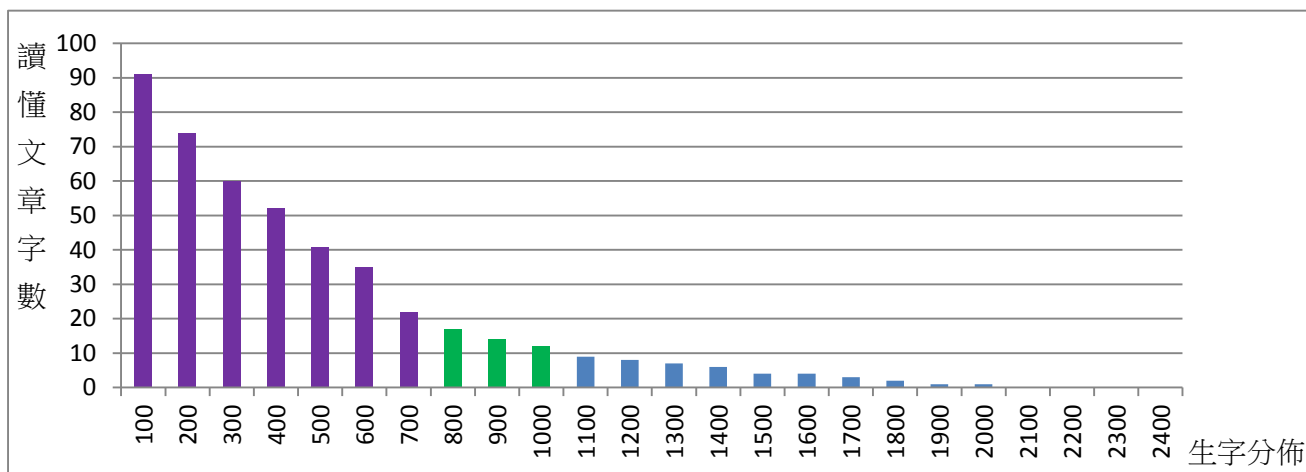
圖(二十一) 文章文字在新學習法中的分布長條圖



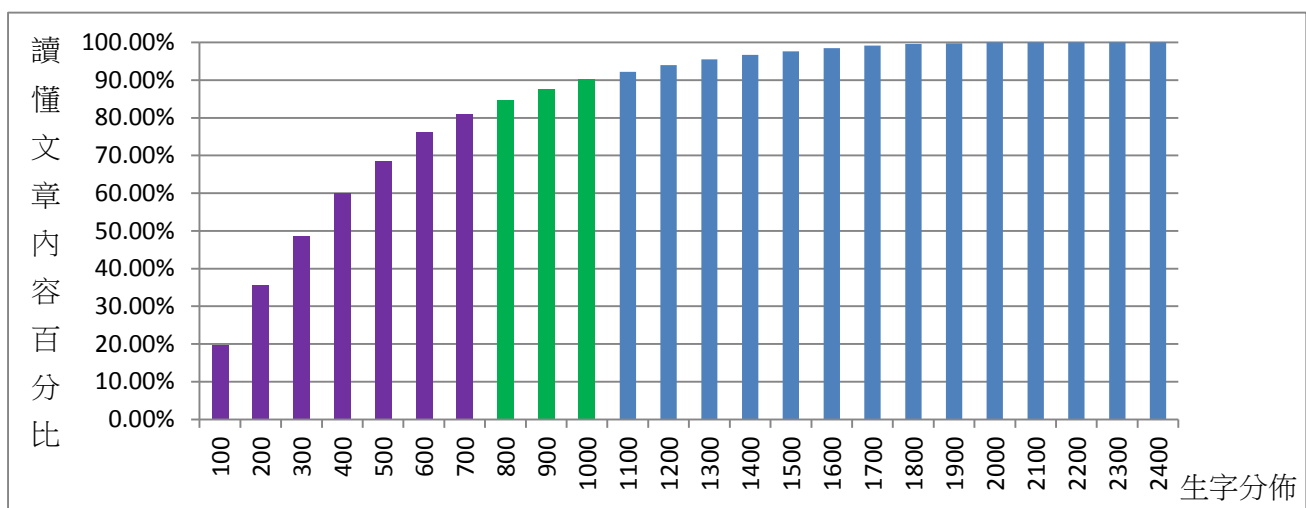
圖(二十二) 文章文字在新學習法中的分布累加長條圖

#### (4) 製作文章文字在以字為單位的學習法中的分布長條圖、文章文字百分比累加長條圖

發現以字為單位的學習方法，相較於以課為單位的學習方法，和傳統學習法有更明顯的差距，如圖(二十三)，以字為單位的學習方法能讓學習者在學完 700 個生字後，讀懂文章的 80%，在學完 1000 個生字後，就可讀懂文章的 90%，與圖(二十)的傳統學習方法有明顯的差距。



圖(二十三) 文章文字在以字為單位的學習法中的分布長條圖

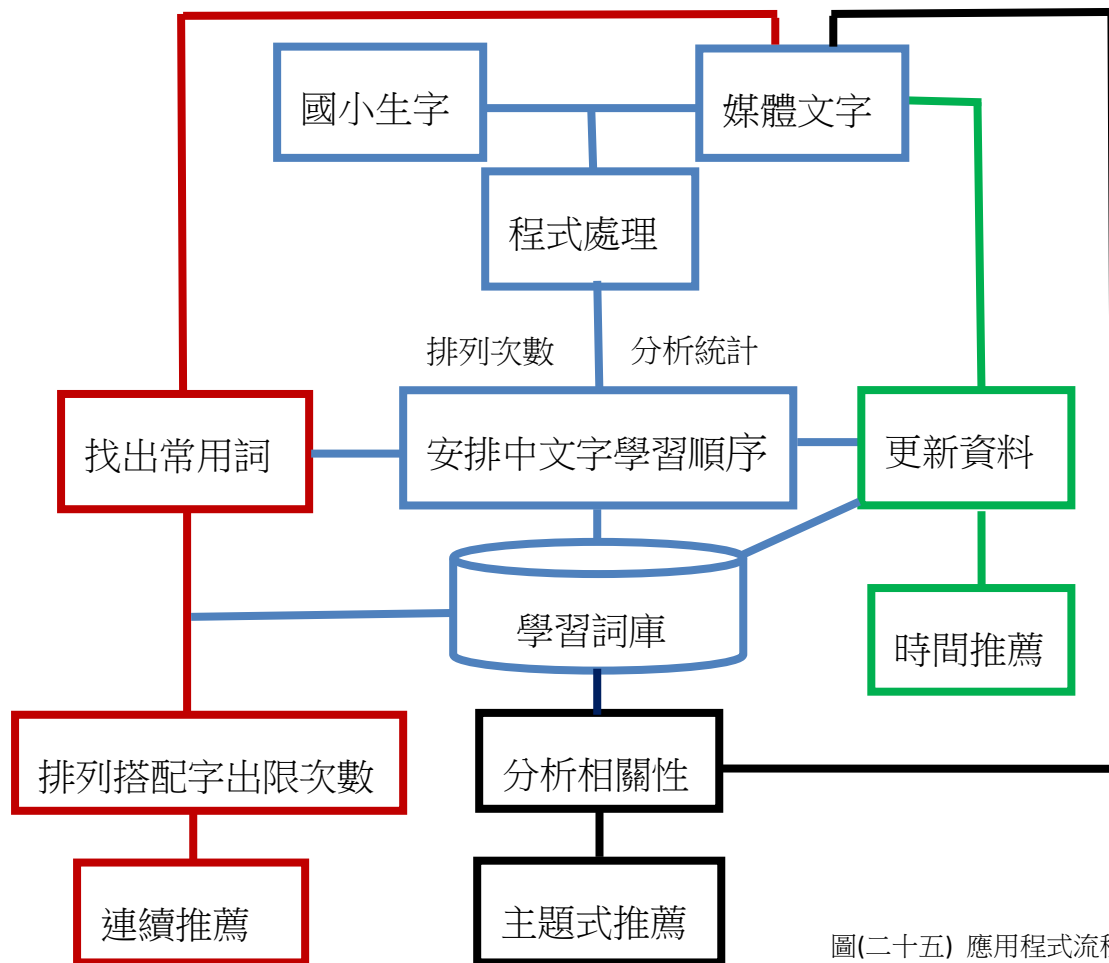


圖(二十四) 文章文字在以字為單位的學習法中的分布累加長條圖

### (三) 製作中文推薦學習應用程式

#### 1. 流程圖





圖(二十五) 應用程式流程圖

## 2. 安排中文字學習順序

期望快速學習，發現在課文中必定會有一些字很常被使用，也有一些字幾乎不用。透過程式將「國小生字」和「媒體文字」比對，安排生字學習順序(使用頻率越高的字越先學習)，縮短學習時間。

## 3. 學習詞庫

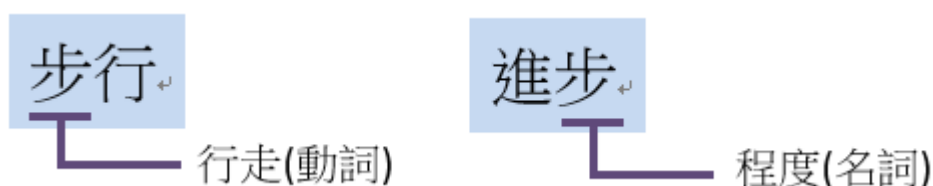
就中文一字多音多意的特點，發現日常中使用的中文字會因與其搭配成詞，不同的字而有不同的意思，並以詞為日常使用中文的單位。所以本次研究將以詞為學習中文的單位，希望抓住學習的關鍵達到有效果且實用的學習方法。

### (1) 詞為日常使用中文的單位



圖(二十六) 詞為日常使用中文的單位圖

(2) 中文會因為用法不同有不同意思



#### 4. 常用詞

圖(二十七) 一字多意示意圖

##### (1) 辨識困難

將國小生字和媒體文章比對，截取此生字的前一個字或後一個字，作為此生字所搭成詞的文字。但此情形下所獲得的字的組合不一定是完整的詞，造成辨識方面的難。

##### (2) 教育部國文字典

為了解決上述的困難，本次研究先從教育部國文字典取得造詞，再透過程式回推至媒體資料庫中，找出最常被使用的詞，推薦給使用者，讓使用者能夠學到「使用頻率最高的字」和「所搭配使用率最高的詞」。

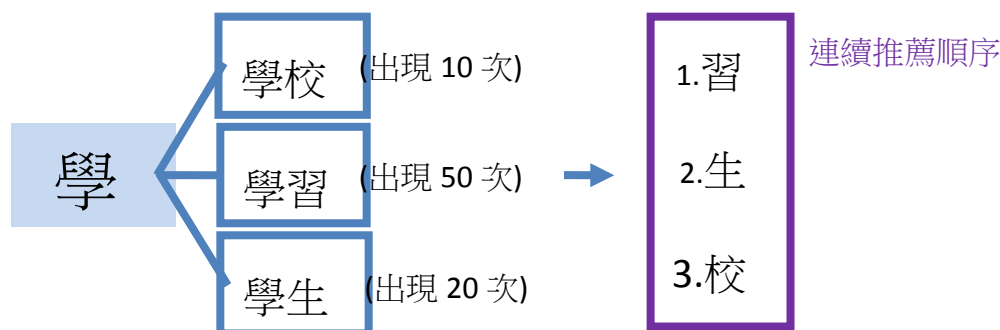
英特爾 3D 列印**機器人** 能走能講  
2014 年 5 月 29 日  
(路透加州蘭喬帕洛斯維第斯 28 日電) 美國科技公司英特爾 (Intel) 今天推出會走路、會講話的**機器人**。特別的是，這個**機器人**零件是由 3D 列印而成。消費者今年就可以買到，但前提是他們願意花 1600 美元左右買工具零件包，自己組裝**機器人**。  
這個**機器人**名叫做吉米 (Jimmy)，英特爾執行長科再奇 (Brian Krzanich) 在它陪伴下走上加州蘭喬帕洛斯維第斯 (Rancho Palos Verdes) 的程式碼大會 (Code Conference) 舞台。  
英特爾說吉米是研究型**機器人**，將不收費提供消費者比較低階的 3D 列印**機器人**方案，合作夥伴則出售非 3D 列印的零件組，例如引擎與 1 個英特爾艾迪生電腦處理器。  
使用者可以透過電腦程式讓吉米唱歌、翻譯語言、發送推特推文，甚至替主人拿罐冰涼的啤酒。

圖(二十八) 詞尋找示意圖

#### 5. 連續推薦(紅)

##### (1) 找出常用詞

了解中文字最常被使用的方式，獲得與此中文字「同時出現頻率」較高的字詞，利用一連串推薦相關且常用之生字或字詞的方式，讓使用者能以一點加以延伸，增加記憶與應用靈活度。

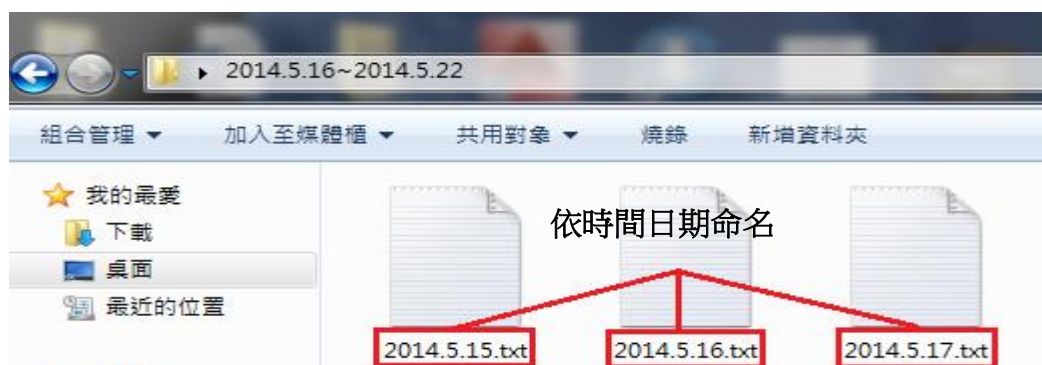


圖(二十九) 連續推薦順序圖

## 6. 時間性推薦(綠)

### (1)加入時間因子

將媒體資料庫的文章以蒐集時間做為文檔名稱，分類文章出現的時段，即可以一年、一月或一周作為媒體資料庫內容，鎖定學習時間範圍內的字詞。如下圖所示。



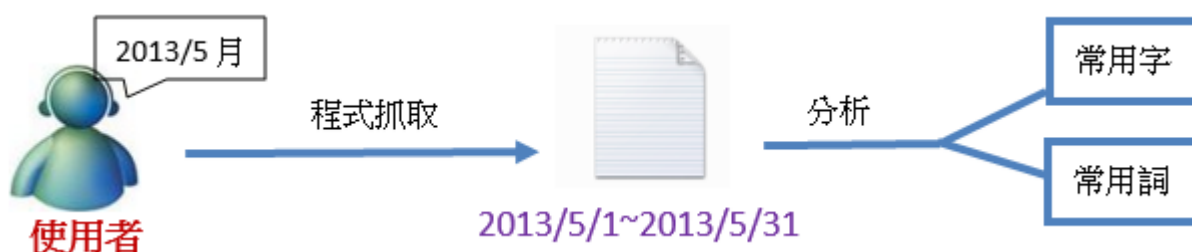
圖(三十) 詞尋找示意圖

### (2)更新資料

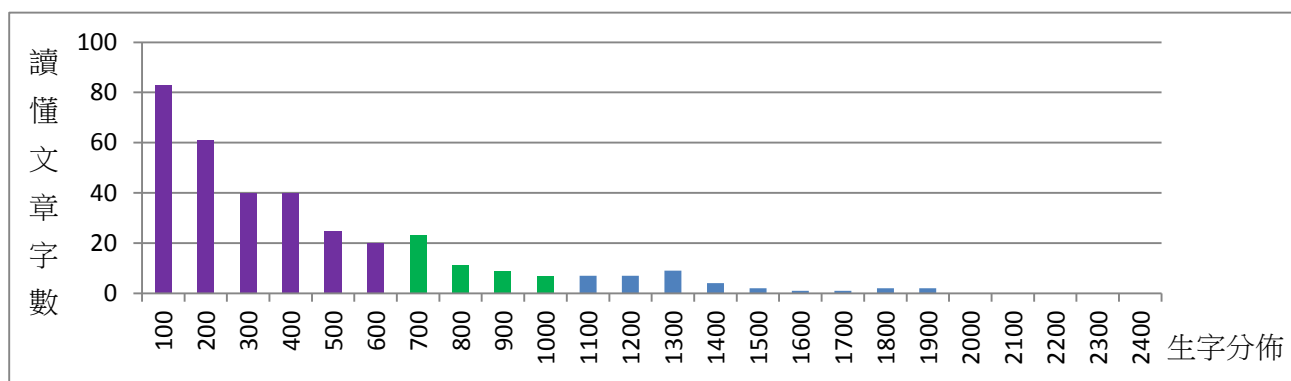
建立媒體文章資料夾，並設定程式截取媒體文章的位置。透過定期存入文檔的方式更新文章資料

### (3)學習者的喜好

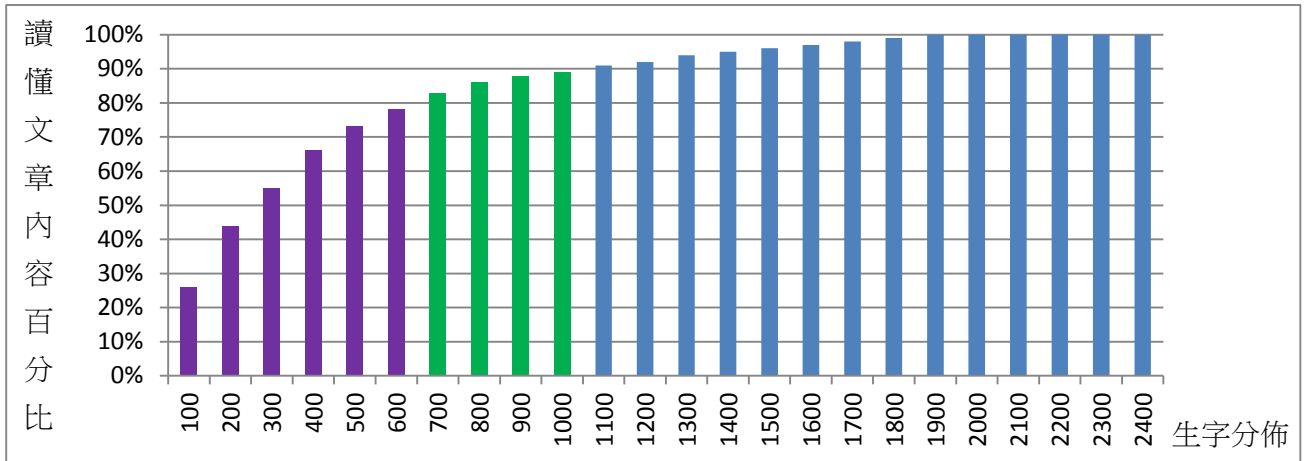
不同的學習者會有不同所需要學習的時間點，為了讓使用者可以對要學習的時間點做指定，我們改為每日都取媒體文章。在學習者指定的時間內，找出使用頻率較多的字詞，並優先推薦學習。以 2013 年為例，每日蒐集媒體文章能讓使用者任意選擇時間範圍，由圖(三十二、三十三)也可再次證明以字為單位的學習成效。



圖(三十一) 時間性推薦示意圖



圖(三十二) 時間性推薦示意圖



圖(三十三) 時間性推薦示意圖

#### (4)時間推薦優點

讓使用者能在短時間內學到最熱門的中文生字，在生活中立即應用並獲得更深刻的學習記憶，如：端午節將至，向使用者推薦字詞「粽子」、「龍舟」。

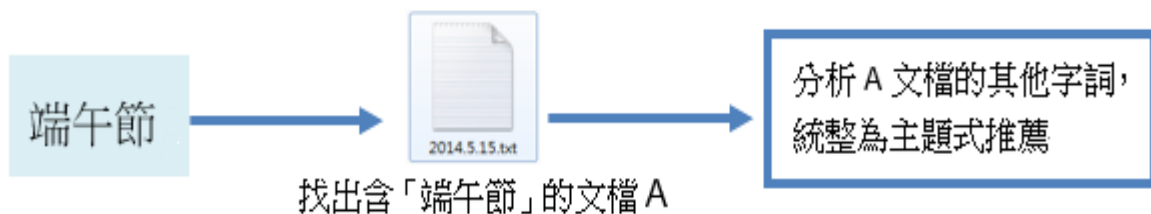


圖(三十四) 時間性推薦示意圖

### 7. 主題式推薦(黑)

#### (1)分析相關性

將詞對照回媒體資料庫，找出所含此造詞的文章。透過分析此文章中其他文字的出現順序，找出詞和詞之間的相關性，統整與此造詞相關的其他字詞。



圖(三十五) 詞尋找示意圖

#### (2)主題式推薦優點

使應用程式能因字和詞的使用場合歸類出相關項目，提供特定的學習內容。

#### (四) 應用價值

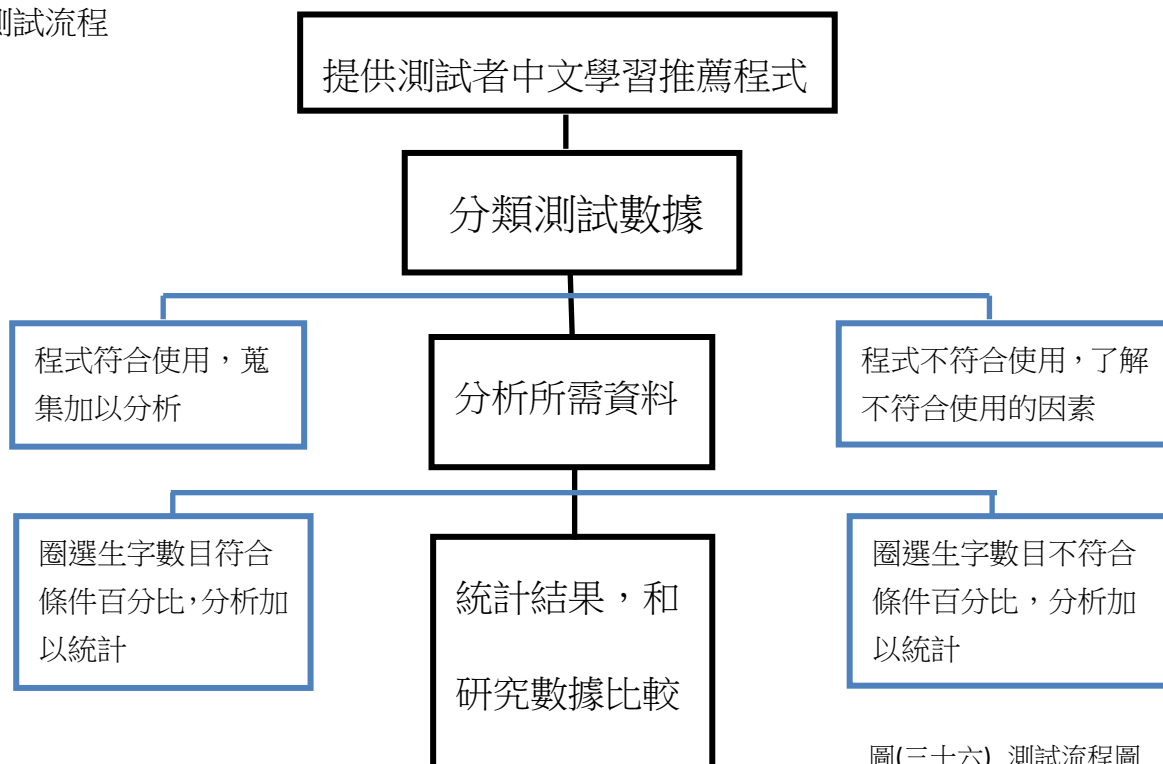
1. 提供中文學習者更有效率的學習方式  
透過「日常生活中越常使用的字越先學習」，達到速學的目的。
2. 幫助學習者有效果的學習字詞  
由「詞」的改念學習中文，引入主題式推薦、連續推薦、時間推薦。加強學習的記憶、字和字的使用連結。

#### (五) 未來展望

1. 資料庫方面
  - (1) 增加「媒體文字資料庫」的來源豐富度和蒐集文章的時間範圍，使研究更具普遍性及準確性
  - (2) 提高選取文章的公平性，使研究結果更加準確
2. 測試方面
  - (1) 增加測試次數，使結果更具可信度
  - (2) 提高選取文章的公平性，使測試結果更加準確
  - (3) 實際測試數據

根據研究圖表統計，使用以字為主的學習方式，能在學完 700 字時讀懂文章平均 80% 的字、學完 1000 字時讀懂文章平均 90% 的字。看似具有較好的效率但在日常生活中不一定符合。本研究將統計實際測試者的數據，將研究數據貼近日常生活情況。

#### 1. 測試流程



圖(三十六) 測試流程圖

## 2.測試流程詳述

### (1)測試目的

從測試者回饋的資料中，分析本研究統計出的學會 700 字是否能讀懂文章的 80%

### (2)提供測試者中文學習推薦程式

提供測試者使用中文學習推薦程式，使測試數據盡量符合本研究所需，並取得程式使用數據。

### (3)分類測試數據

#### a. 如何測試

從媒體資料庫中，隨機挑選文章內的句子讓測試者閱讀。請測試者圈選出句中不會的生字，並在句尾以打圈和打差的方式表示是否能讀懂整句句意。

#### b. 分類測試數據

依測試者所圈選部會的生字，將出現以下兩種情形，「所圈字為推薦學習前 700 字以內」(A 情形、B 情形)和「所圈字為推薦學習前 700 字以外」(C 情形)。本測試目的在於驗證特定情況的結果，對此 A、B 情形採取不同的處理方式。

##### (a) 情形 A

圈選的生字為中文推薦程式的前 700 字，與已學會推薦程式前 700 個字的情形不符，計算測試者所圈不符合的生字比例，蒐集資料加以使用。

##### (b) 情形 B

未圈選出不是中文推薦程式的前 700 的生字，與已學會推薦程式前 700 個字的情形不符，計算測試者所圈不符合的生字比例，蒐集資料加以使用。

##### (c) 情形 C

圈選的生字不為中文推薦程式前 700 字，符合已學會推薦程式前 700 個字的形，蒐集資料加以使用。

### (4)分析符合資料

分析與所需情形相符之資料中「句中生字讀懂的多寡」、「是否理解句意」。統計出生字學習量和文句解讀能力的關係，以達到證明本研究數據的目的。

#### a.句中所圈生字多寡

本測試中，測試者將圈出不懂的生字。以未被測試者所圈的生字多寡為主，又本測驗為討論讀懂 80%的文句和實際理解句意的關係。將資料分為兩種情形。「大於整句生字總數的 80%」(D 情形)和「小於等於整句生字總數的 80%」(E 情形)。

##### (a) 情形 D

未被測試者所圈的生字大於等於整句生字總數的 80%，意即讀懂文句的百分比大於等於 80%。符合本研究所需情形，個別分析，並計算實際讀懂句意的次數多寡。



## 捌、附錄

### 一、中文字計數程式

```
#include<stdlib.h>
#include<stdio.h>
#include<iostream>
#include<fstream>
using namespace std;
int main(){
    FILE* fin;
    char filename[500];
    int count[256][256]={0},element[256][256]={0};
    int ch,cn,l,j,countch=0;
    char out[3];
    for(i=0;i<xxxx;i++){
        for(j=0;j<xxxx;j++){
            sprintf(filename,"%d.%d.txt",i+1,j+1);

            fin=fopen(filename,"r");
            while( (ch=fgetc(fin))!=EOF ){ /*建構小學國語課本生字資料庫*/
                if(ch<=127)
                    continue;
                cn=fgetc(fin);
                element[ch][cn]=1;
            }
        }
    }
    for(i=0;i<3;i++){
        sprintf(filename,"a%d.txt",i+1);

        fin=fopen(filename,"r");
        while((ch=fgetc(fin))!=EOF){
            if(ch<=127)
                continue;
            cn=fgetc(fin);
            if(element[ch][cn]==1){
                count[ch][cn]+=1;
            }
        }
    }
}
```



```

        countch+=1;
    }
    //printf("%d %d\n",ch,cn);
}
fclose(fin);
}
printf("開始!!\n");
for(i=0;i<256;i++){
    for(j=0;j<256;j++){
        if(count[i][j]!=0){
            out[0]=i;
            out[1]=j;
            out[2]='\0';

            printf("文字:%s  %d 個\t",out,count[i][j]);
        }
    }
}
printf("\n 總數:%d 個\n",countch);
system("pause");
return 0;
}

```

## 二、去除文章重複字程式

```

#include<stdlib.h>
#include<stdio.h>
#include<iostream>
#include<fstream>
#include<time.h>
int i,j;
using namespace std;
int main()
{
FILE* fin;
FILE* fin2;
char filename[500],out[3];

```

```

int element[256][256]={0},ch,cn,kmax,maxch,maxcn,max;
fin2=fopen("依次數排列的中文字.txt","w");

for(i=99;i<100;i++){
    sprintf(filename,"a%d.txt",i+1);

    fin=fopen(filename,"r");
    while((ch=fgetc(fin))!=EOF){
        if(ch<=127)
            continue;
        cn=fgetc(fin);
        if(element[ch][cn]==0)//if 沒印過
        {
            out[0]=ch;   out[1]=cn;   out[2]='\0';
            fprintf(fin2,"%s",out);

            element[ch][cn]=1;
        }
    }
    fclose(fin);
}
system("pause");
return 0;
}

```

### 三、計算文章生字出現點程式

```

#include<stdlib.h>
#include<stdio.h>
#include<iostream>
#include<fstream>
#include<time.h>
int i,j;
using namespace std;
int main()
{
    FILE* fin;
    FILE* fin2;

```

```

char filename[500],out[3];
int element[256][256][2]={0},ch,cn,kmax,maxch,maxcn,max;
fin2=fopen("依次數排列的中文字.txt","w");

sprintf(filename,"1.2.txt");
fin=fopen(filename,"r");
i=1;

while( (ch=fgetc(fin))!=EOF )
{ /*建構小學資料庫*/
    if(ch<=127) continue;
    cn=fgetc(fin);

    element[ch][cn][0]=i;
    i++;
}
max=i;
kmax=i;

for(i=203;i<204;i++){
    sprintf(filename,"a%d.txt",i+1);

    fin=fopen(filename,"r");
    while((ch=fgetc(fin))!=EOF){
        if(ch<=127)
            continue;
        cn=fgetc(fin);
        if(element[ch][cn][0]!=0)//if 存在
            element[ch][cn][1]=1;//標記
    }
    fclose(fin);
}
while(max>=1)
{
    for(i=0;i<256;i++)
    for(j=0;j<256;j++)
        if(element[i][j][1]==1 && element[i][j][0]==max)
        {

```

```
maxch=i;
maxcn=j;
out[0]=maxch;out[1]=maxcn;out[2]='\0';
fprintf(fin2,"第%d 難字： %s\n",element[i][j][0],out); //要看前面的話改成檔案輸出
    }
max--;
}
system("pause");
return 0;
}
```

## 【評語】 040808

本作品可以推薦中文字頻率高出現率給初學者，如能配合適當的教學工具及方法，可以有效提升中文字的學習速率及效果。