# 國立交通大學

## 理學院網路學習在職專班

## 碩士論文

# 自動化文章敵意分級系統之初探研究

# A Pilot of Automatic Sorting System with Hostile Articles

研 究 生：林志鴻

指導教授：林珊如博士、劉旨峰博士

中 華 民 國 九 十 三 年 五 月

自動化文章敵意分級系統之初探研究

學生：林志鴻　　指導教授：林珊如博士

劉旨峰博士

國立交通大學網路學習（研究所）碩士班

## 摘　　要

隨著網路上文件的等比級數增加，如何精確地找出所需要文件成爲了重要的議題。在本文中，參酌自動化文件分類的相關研究，提出了利用向量模型對中文敵意文件的分類程序與方法。從學術網路 BBS 站的硬體討論版(tw.bbs.comp.hardware)抽樣 5000 篇文章，先以人工分類方式，將文章依敵意的程度分類後，再進行自動分類實驗，先輸入數篇文章，由系統分析出文章的關鍵詞，並計算權重，建立敵意文章中心向量，再依據輸入的文章會計算出與敵意文章相似度，最後將相似度高於門檻值的文章判定爲敵意文章，其他則爲非敵意文章，研究發現：

(1)利用同一主題文章作爲訓練文章，來計算敵意與非敵意文章與敵意中心向量的相似度時，其相似度具有明顯差異。

(2)訓練文章的主題不同時，所計算出的相似度亦有差距。

(3)利用門檻值實驗計算出的最佳門檻值 0.17 來進行分類時，對於非敵意文章有較佳的精確度，約爲 0.98，但對於敵意文章的分類精確度則較差，約爲 0.25。

(4)當門檻值降低至 0.136 左右時，可同時提高 HR 與 NHR 值至 0.7 左右。

A Pilot of Automatic Sorting System with Hostile Articles

student：Jyh-Horng Lin　　Advisors：Dr. Sunny S. J. Lin

Dr. Eric Zhi-Feng Liu

**Department（Institute）of E-learning**
**National Chiao Tung University**

## ABSTRACT

With the increasing of Website documents drastically, how to precisely find what are needed documents turns to be an important issue. In this article by referring to relevant study on the automatic document classification, it brings out to utilize the vector model to classify and process Chinese hostile documents.By sampling 5000 articles from the hardware discussion board in the academic BBS (tw.bbs.comp.hardware), we classify them by manual first, based on the degree of the hostile.Later on, proceeded automatic classification experiment. By ntering several articles in the beginning, the system can analysis key terms, and calculate the term weight ratio in order to establish the central vector of the hostile articles.Then, this system can calculate the similarity by comparing with the build-in hostile articles. Finally, if an article's similarity is higher than the threshold, then it will be classified into hostile articles. Other than that, it will be classified into articles without hostile. Some observations found through this study as following:

1. By using the same topic of articles for the purpose of training articles to calculate the similarity of the hostile central vector between hostile and unhostile articles, the similarity was obviously different.
2. When the topic of training articles was different, the similarity was different.
3. When using an optimum threshold value 0.17 to proceed classification, it came out a better accuracy for the articles without hostile with about 98%, but got a worse classification accuracy for the hostile articles, about 25%.
4.We can get better HR by decreasing the threshold.

# 誌　　謝

這篇論文的完成，首先要感謝的，是我的指導教授，林珊如教授與劉旨峰教授，兩位老師不辭辛勞的指導，讓這篇論文從無到有，從有到完整，也讓我真正的了解什麼是真正的求知，做學問，也感謝口試委員袁賢銘教授與莊祚敏教授對此論文的許多寶貴意見。

感謝大學同窗四年的好友煜庭，能在我遇到挫折的時候，幫我加油打氣，一起克服許多的困難。也感謝研究所的學弟妹，幸玲、佩芯、幸如、文婷、啓峰，在研究的過程中，給予我很多的協助與建議。

謝謝三姨媽還有四姨媽，常常調理許多美味可口的食物，讓我能補充體力後再出發，還要謝謝兩台可愛的伺服器，在這三年的時間能非常正常的運作，即使在我休息的時候，他們還是能盡忠職守，完成各項實驗工作。

最後要謝謝最辛苦的媽媽，大姊、二姊、碧緞，幫我處理許多的事情，讓我沒有後顧之憂，能專心的進行研究，在此向他們至上最崇高的謝意。　　^^

# 目錄

# 圖目錄

# 表目錄

# 一、前言

## 1.1 研究動機及目的：

由於網路的普及與方便性，許多的個人及團體紛紛將文件放置在網路上，提供需要者快速的取用途徑，當使用者要找尋某一類主題文件時，便可以利用搜尋引擎，快速找到所需要的資料。但由於網路上的文件數量極多，因此在尋找所要資訊時所遭遇到的問題，也從原來的不易取得文件，變成不易找到所需要文件(Belkin,1992)。爲了讓使用者可以很快速的找出想要的文件，我們通常會先將文件分類放置(Croft & Larkey, 1996; Liu & Yang, 1999;Tsay & Wang, 2000; Yang, 2001; Chao & Wai, 1998)，當使用者有查詢文件的需要時，便可以利用搜尋引擎，在搜尋引擎中輸入與文件相關的關鍵字或關鍵句，但是這樣的查詢方式主要是針對一些類別較明顯的文件，例如：要查詢關於電腦硬體的文件，可以直接輸入電腦硬體、主機板、或電腦週邊等關鍵字，即可找到相關文件。但如果想要查詢的文件是屬於類別較不清楚的，例如想要找出自傳，或者是表達謝意的文件，相關的討論就不是那麼多了。在公開的論壇或是留言版上，常常會看到一些惡意攻擊或是謾罵的文件，這類文件通常會偏離主題甚多，且容易引起網路上的論戰，而網管人員也常常要花費許多時間處理這些文件，因此若能設計一個能自動分類敵意文章的系統，就可幫助網管人員快速篩選出具有敵意的文章，減少網管人員處理這類文章的時間，提昇管理效率。本文的目的，即是希望能設計一個自動化敵意文章分級系統，由使用者先提供一些具有敵意的文章（以下簡稱訓練文件），系統會依照所提供的文章，找出想要搜尋文章類型的特徵，並利用此特徵計算文件庫中所有文件的特徵值，計算出每篇文件與敵意文章的相似度，並設定門檻，最後將敵意文章分類出來。

## 1.2 論文大綱：

本篇論文在第二章文獻探討部分，將介紹中文自動化文件分類與敵意的相關研究。在第三章研究架構方面，描述論文中敵意文章分類系統架構、設計方式及進行實驗的程序。在第四章中的結果分析及討論中，利用統計方法來分析實驗數據，並觀察論文中系統的分類效果，並給定結論。

# 二、文獻探討

## 2.1 自動化文件分類：

文件分類是資訊檢索(Information Retrieval，簡稱 IR)中的重要步驟(Jones, 1981; Jones & Rijsber-gen, 1976; Borlund & Ingwersen, 1997; Oddy, 1981)，主要目的則是透過各種模型將文件分類存放，以加快資料的搜尋，古典的分類模型有三種，布林模型(boolean model)、向量模型(vector space model)(Joachims, 2001; Jason & Rifkin, 2001)、機率模型(probabilistic model)，這些模型分類的方式是先給每篇文件一個特徵值，如布林模型則是將需分類文件給定 1、或 0 兩個值其中一個，1 表示此篇文件歸爲我們要的分類中，0 表示不是，因此需分類文件與該類別文件的相似度函數 sim(d，q)，函數值會等於 1 或 0，1 表示相似、0 表示不相似。對一般使用者而言，布林模型是分類模型中較容易被瞭解的，因爲它的概念淺顯易懂，但是布林模型缺乏部分相似的概念，因爲只能判別此文件“是”或者“不是”此一類別，而且很難將文件判別的方式轉換成布林表示式，也就是說，缺乏一具體方式或操作型定義來給定每篇文件一個或一組特徵值，因此要進行自動文件分類時，會產生相當大的困難。而在向量模型中，每篇文件皆以一個向量來表示，此向量的維度(dimension)等於關鍵詞的數目，而每一維度的值則爲關鍵字的權重(term weight)，需分類文件與該類別文件的相似程度則以相似度函數，如式(1)所示：

$$sim(\vec{d},\vec{q}) = \frac{\vec{d} \bullet \vec{q}}{\left|\vec{d}\right| \bullet \left|\vec{q}\right|} \tag{1}$$

$\vec{d}$ 代表此一類別文件的向量，$\vec{q}$ 代表需分類的文件，此相似度函數也就是在測量 $\vec{d}$ 與 $\vec{q}$ 在向量空間中的夾角，其值介於 0 與 1 之間，數字越大，相似程度越高，0 表示完全不相似，1 表示完全相似。此一模型具有兩個特點：(1)文件具有部分相似的可能性。(2)明確訂定了文件的特徵值計算方式，因此較適合做自動化文件分類使用，也成爲近代資訊檢索中常用之模型。而在機率模型中，文件的相似程度則以機率的方式來表示，假設 R 爲一群同類型的文件，而 $\overline{R}$ 爲 R 的補集，給定文件 $\vec{q}$ 與該類別文件 $\vec{d}$ 的相似度函數

$$sim(\vec{d},\vec{q}) = \frac{P(R\left|\vec{d}\right.)}{P(\overline{R}\left|\vec{d}\right.)} \tag{2}$$

，其中 $P(R|\vec{d})$ 表示文件 $\vec{d}$ 與文件 $\vec{q}$ 相關的機率，而 $P(\overline{R}|\vec{d})$ 表示文件 $\vec{d}$ 與文件 $\vec{q}$ 不相關的機率。

在以上三種模型中，布林模型(boolean model)爲最弱的模型(Ricardo & Berthier, 1999)，而機率模型的效率好壞會隨著樣本數的大小而變動，在 Salton 與 Buckley 的實驗中發現，向量模型的表現較機率模型爲佳，因此向量模型成爲近代資訊檢索中常用之方式，在本系統中亦將採用向量模型來進行敵意文章分類。

2.2 語詞權重：

在向量模型中，文件向量的每一維度即爲關鍵詞的權重(term weight)，關鍵詞的權重也代表關鍵詞在文件中的重要性程度，權重越高，則代表此關鍵詞越能代表此篇文件。例如某篇文件中，關鍵值”學習”的權重很高，則代表此篇文件有很高的機率是屬於教育類的文件。最早的權重計算方式，是看關鍵詞的出現與否，若出現，則將權重設爲 1，若沒有出現則設爲 0，如下所示：

若 w 代表文件 d 中關鍵字 k 的權重，則

$$w=\begin{cases}1 & \text{若k出現在d中}\\ 0 & \text{若k沒出現在d中}\end{cases} \tag{3}$$

換句話說，只要文件中出現過的關鍵詞，不論出現幾次，其權重皆相等，代表出現過的詞皆能代表此文件的類別，但實際狀況並非如此。例如：在一篇教育類的文件中若出現“學習電腦與學習數學一樣重要”，則依照此種權重計算方法，此篇文件屬於教育類、數學類或電腦類文件的可能性會一樣高，因此此種方式會造成文件分類的困難，而許多研究也顯示出權重可以是 1 與 0 之間的任何數，1 代表最高，0 代表最低的權重，如此較能符合一般狀況。

影響權重的要素有三個，第一個是關鍵詞出現的次數(term frequence)，出現次數越多，則表示此關鍵字越能代表此文件的類別。第二個是在所有文件中，出現此關鍵詞的篇數(collection frequence)，篇數越多，表示此關鍵詞越不能代表此文件的特性，例如一些常用的介詞或代名詞。在進行文件分類時，會將部份不同類的文件，分爲同一類。例如在電腦類的文件中會出現“學習”，而在數學類也會出現“學習”這個關鍵字，因此必須減少“學習”此關鍵字的權重，以免將電腦類的文件與數學類的文件視爲同一類。第三個要素則是文件向量的長度，由於每份文件的長短並不相同，但每份文件應視爲同樣重要，因此關鍵詞權重必須正規化(normalized)，也就是說關鍵詞的權重必須定義成

$\frac{w}{\sum_{i}^{n} w_i}$ 或 $\frac{w}{\sqrt{\sum_{i}^{n} w_i^2}}$ ， $w_i$ 爲文件中所有關鍵詞權重。

關鍵詞權重的計算方式即是上述三要素的乘積，將以上三個要素導入公式的計算方式有相當多種，而由於文件類別與特性的不同，因此計算此三要素的方式也有所不同，常見的計算方式有以下幾種方式：

表 1：影響關鍵詞權重的因素及常見計算方式

| 關鍵詞出現頻率(Term Frequency Component) | | |
|---|---|---|
| B | 1 | 若出現關鍵詞，則設爲 1，否則設爲 0。 |
| T | tf | 出現關鍵詞的次數。 |
| N | $0.5+0.5\frac{tf}{\max tf}$ | 將關鍵詞出現的次數正規化， $\max tf$ 表示在該文件中，出現頻率最高的關鍵詞次數，n 介於 1 與 0.5 之間。 |
| 關鍵詞在文件中出現頻率(Collection Frequency Component) | | |
| X | 1 | 不考慮在所有文件中出現的頻率。 |
| F | $\log \frac{N}{n}$ | 所有文件篇數除以出現該關鍵詞的次數後，再取對數。 |
| P | $\log \frac{N-n}{n}$ | 沒有出現該關鍵詞的次數篇數除以出現該關鍵詞的次數後，再取對數。 |
| 文件長度(Normalization Component) | | |
| X | 1 | 不考慮文件向量的長度。 |
| C | $\frac{1}{\sqrt{\sum_{vector} w_i^2}}$ | 避免文件的長度影響權重，因此將權重正規化。 |

資料來源：Salton, G., and Buckley, C., 1988a

上述的三個要素，若取的值爲 1，則表示不考慮此要素的影響。常見的組合方式則如表 2 所示：

表 2：常見的關鍵詞權重計算方式

| Weighting system | Document term weight | Query term weight |
|---|---|---|
| Best fully weighted system (TFC.NFX) | $\frac{tf \bullet \log\frac{N}{n}}{\sqrt{\sum_{vector}\left(tf, \bullet \log\frac{N}{n_1}\right)^2}}$ | $0.5+0.5\frac{tf}{\max tf} * \log\frac{N}{n}$ |
| Best weighted probabilistic weight (NXX.BXP) | $0.5+0.5\frac{tf}{\max tf}$ | $\log\frac{N-n}{n}$ |
| Classical weight (BFX.BFX) | $\log\frac{N}{n}$ | $\log\frac{N}{n}$ |
| Binary term independence (BXX.BPX) | 1 | $\log\frac{N-n}{n}$ |
| Standard if weight (TXC.TXX) | $\frac{tf}{\sqrt{\sum_{vector} w_i^2}}$ | tf |
| Coordination level (BXX.BXX) | 1 | 1 |

權重計算方式代碼說明：ABC.DEF，ABC 表示訓練文章的權重計算方式，DEF 表示實際文章的權重計算方式。

資料來源：Salton, G., and Buckley, C., 1988b

表2中的TFC.NFX的權重計算方式，是Salton and Buckley在進行文件分類實驗中，能產生最佳效能的計算方式，但是表2所使用的權重計算方式，主要是針對英文文件的分類，在Kwok and Luk (2002)針對中文自動化文件分類的實驗中發現，利用向量模型計算關鍵詞的權重方式時，關鍵詞權重的給定方式以

$$\log(tf+1)\bullet\log(\frac{N}{n}+1) \qquad (4)$$

的方式爲佳，且其效能較表2的計算方式有顯著的提升，因此本系統將以公式(4)計算關鍵詞權重。

2.3 文件相似度：

當我們利用向量模型計算出文件向量後、如果要找出想要的文件類型、就必須計算兩文件的相似度，常用的相似度計算方式如表 3 所示：

表 3：常見的文章相似度計算公式

| 序號 | 名稱 | 計算公式 |
|---|---|---|
| 1 | Simple matching (coordination level match) | $\sum_{j=1}^{m} d_j q_j$ |
| 2 | Dice’ s Coefficient | $\text{Dice(q,d)}=\frac{2\sum_{j=1}^{m} d_j q_j}{\sqrt{\sum_{j=1}^{m} q_j^2+\sum_{j=1}^{m} d_j^2}}$ |
| 3 | Jaccard’ s Coefficient | $\text{Jaccard(q,d)}=\frac{\sum_{j=1}^{m} d_j q_j}{\sqrt{\sum_{j=1}^{m} q_j^2+\sum_{j=1}^{m} d_j^2-\sum_{j=1}^{m} d_j q_j}}$ |
| 4 | Cosine Coefficient | $\cos ine(q,d)=\frac{\sum_{j=1}^{m} q_j \bullet d_j}{\sqrt{\sum_{j=1}^{m} (q_j)^2 \bullet \sum_{j=1}^{m} (d_j)^2}}$ |
| 5 | Overlap Coefficient | $overlap(q,d)=\frac{\sum_{j=1}^{m} q_j \bullet d_j}{\min(\sqrt{\sum_{j=1}^{m} q_j^2}, \sqrt{\sum_{j=1}^{m} d_j^2})}$ |

資料來源：Fred, 2002

2.3.1 Simple matching：

此相似度計算方式爲計算 q 與 d 重疊的部分，若 $q=(q_1,q_2.....q_m)$ ，$d=(d_1,d_2.....d_m)$ 則兩文章相似度爲 $\sum_{j=1}^{m} d_j q_j$ ，因此若有關鍵字 $k_i$ 未同時在 d 與 q 中出現，則 $d_i q_i=0$ ，共同出現的關鍵字越多，$\sum_{j=1}^{m} d_j q_j$ 的值越

大。

2.3.2 Dice’ s Coefficient：

此相似度計算方式爲計算 q 與 d 重疊部分占全部的比值，若 q 與 d 無重疊部分，則 Dice’ s Coefficient 爲 0，反之，若 q 與 d 是相同的文件，則值爲 1。

2.3.3 Jaccard’ s Coefficient：

此相似度計算方式爲計算 q 與 d 重疊部分占兩文件平均大小的比值。

2.3.4 Cosine Coefficient：

利用餘弦函數計算兩文件在向量空間中的餘弦值，若兩文件完全相同，則其值爲 1。

2.3.5 Overlap Coefficient：

計算 q 與 d 重疊部份占 q 與 d 中長度較短文件的比值。

2.4 敵意與論戰：

到目前爲止，心理學家還無法給敵意一個很清楚的定義，只知其爲一多向度的概念，其中可能包含了許多的概念，如 Buss, Fischer, and Simmoms (1968) 提出，敵意是對人與事物的負向評價。敵意產生的原因，常常是由於別人與自己的觀念不同，但卻又無法說服其他人時產生。Cook and Medley (1954) 認爲，個體一旦產生敵意後、會不喜歡與不信任他人，並認爲其他人是不道德、令人厭惡，必須接受處罰的，因此敵意常常跟生氣有關，也就是說，當一個人產生敵意時，常常也會有生氣的情緒表現，爲了保護自己或是證明自己是對的，個體就會產生攻擊的行爲，在網路上的具體呈現即是所謂的網路論戰，網路上的論戰是指被使用者利用公開〈如留言版〉或非公開〈如電子郵件〉的工具，利用攻擊性的文字與他人持續交談的過程。Reid (1995) 認爲，論戰中的文章通常有下列特徵：無理由的批評，包括侮辱、咒罵，以及敵意的陳述，因此當爭論的文章出現後，常常會引起一系列的攻擊性爭論。論戰是任何時候都會發生的，可能從文法、語詞或任何不重要的議題上開始，因此我們可以了解論戰的發生不一定是因爲議題，常常是因爲文字呈現方式的不同而發

生。而 Thompsen and Foulger (1996) 則提出論戰的五個過程：

1.分歧〈divergence〉：參與討論的人對同一個議題表達了至少二種以上不同的意見，而這些意見常呈現明顯的差異，甚至是相反的意義。

2.爭論〈disagreement〉：提出可支持自己意見的相關證據或是反對對方的相關證據，但並不會直接反對對方的意見。

3.緊張〈dension〉：直接反對對方的意見，並膨脹自己贊成的意見。

4.敵對〈antajonism〉：針對對方做人身攻擊，破壞對方的人格，以降低對方言論的可性度，此時雙方的焦點已漸漸脫離原本討論的主題。

5.尖銳敵對期〈profane antagonsm〉：雙方用大量誇大、且具攻擊性的言論來攻擊對方，此時雙方的焦點已完全脫離了原本討論的主題。

參與論戰的雙方或多方，常常是在討論一個主題時，由原本正常的討論狀況，逐漸演變至後來的無法接受他人意見，而產生敵意，進而利用文字攻擊他人，此處所謂的攻擊，包含了情緒性文字，如高台茜的網路言論情緒用詞資料庫(http://edu.ndhu.edu.tw/mkao/emotion)，及批評他人的負向字句，如：你全家死光吃屎、狗屎，而這類的發言，常常會使原本正常的討論文章，漸漸的偏離主題，變成具有敵意的文章。

2.5 語詞長度：

組成文件的基本單位爲字詞，因此若要分析文件的特性，就必須先將文件做斷詞處理，才能對文件做進一步的分析(Damashek, 1995)，而中文與英文在結構上有相當大的不同，英文的每個單字都是由26個字母組合成，且在英文句子中每個字都以空格(space)或是標點符號隔開，因此在擷取關鍵詞時，只要以空白或是標點符號來作爲斷詞依據即可，而中文則否，組成中文句子的最小單位爲字，而中文詞則由一個或多個的字組成，由於字與字之間並無明顯分隔，因此在處理中文文件時必須先做斷詞的處理，假設一篇有n個字的文件，由於詞的長度可從一個（如：水，書）到八、九個（如：後天免疫不全症候群），因此若要對此篇文件斷詞，且要找出所有的可能性，則需執行$2^n$次斷詞，在實際的應用上會產生困難，根據統計發現，在文件中二字詞出現的比率約佔全部詞的75%，在Kwok and Luk的研究中也發現，在自動化分類中文文件時，若採取向量模式，則斷詞方式

採二連字詞，其檢索效能較單字或多連字詞好，而楊允言、陳淑美、陳克健與謝清俊 (民88) 在中文文件自動分類的實驗中，也建議採用二連字詞，因此爲了兼顧準確度與速度，本文中所用的系統將以二連字詞爲主。曾元顯 (民91) 在進行文件主題自動分類成效因素探討實驗中發現，在文件中僅出現一次的詞，經常占一篇文章的60%~70%，刪掉之後雖然可以大幅減少文件的向量維度，但是保留的詞彙越多，效果越好，因此在本實驗中對於只出現一次的二連字詞將不進行刪除的動作，以提高分類效果。

2.6 測試文件集：

Salton 從 1961 年起，展開 SMART(System for Mechanical Analysis and Retrieval Text) 研究計畫，此計畫主要是利用 Cleverdon 在 1950 中期至 1960 中期完成的 Cranfield 研究的實驗文件爲基礎，建立大型電子文件資料庫，便於展開對自動化文件分類與全文資訊檢索理論的研究工作，而研究大型語料與檢索效能的 TREC(Text Retrieval Evaluate Conference)更是有史以來最大、參加者最多的資訊檢索實驗，具有以下特點：

(1)文件與辭彙數量龐大。

(2)資料庫內文件多爲全文。

(3)來自多個不同的學科領域。

(4)查詢句設計較長且具有結構性。

(5)對於查詢句與文件的相關性有較嚴格的標準，藉此增加相關判定的一致性。

(6)具有多種不同語言的語料。

TREC 成立的主要目的，是希望能讓研究者測試大規模語料實驗環境下，相關的檢索理論以及所設計之檢索系統的效能，甚至能更進一步地找出較適合的系統參數以及文件檢索方法。不過 TREC 雖然有建立中文的語料庫，但由於只有參與 TREC 實驗計劃的單位才能使用語料庫內的測試資料，且其使用中文方式，與國內使用中文的方式有相當大的差異，缺乏地域性。而國內在資訊檢索領域的研究起步較歐美國家晚，由中央研究院資訊科學研究所所組成的中文詞知識庫發展小組，從民國 75 年起，便開始結合計算機與語言學的中文詞知識庫計畫。目前的研究現況與應用發展以中文詞知識庫爲核心，主要發展中文語句分析、語音辨識、資訊檢索及語言學研究語料庫等。在此領域的基礎研究上，已有相當的成果 (楊允言, 民 82) 。中研院於 1984 年開始，開始推動史籍自動化的工作，並陸續將文件電子化，目前已經有總數近一億一千萬字的

文件上線 (謝清俊、林晰，民 86) 。系統並提供自由詞檢索、多詞同時檢索。因此，目前在中文全文資訊檢索研究上可以藉助於上述系統，但由於系統內的查詢主題與測試資料主要針對明確主題，且相關文件已經過整理，同質性高，不適合用於本系統中，因此本文將依據其他關於建立測試文件集的研究建議，及本研究主題的需要，建立一個針對敵意文件判別測試文件集，以提高本研究的信度及效度，並提供將來在研究資訊檢索領域，因個別需要，需自行建立測試文件集的研究者一些建議。江玉婷(民 89)認爲，理想的測試集，除了必須具備一定的規模外，在文件以及查詢的內容，型態，取得來源等方面要有相當的異質性，以下將就這幾個特點分析個別測試文件集的建立原則：

2.6.1 規模：

早期的測試集主要針對個別測試計劃，因此規模不大，與母群體的大小差異過大，樣本效度也因此偏低，下表所列，爲一些早期的測試文件集相關資料。

表 4：常用測試文件集

| 測試集 | 文件數 | 文體集大小 | 文件平均字數 | 查詢問題數 | 查詢問題平均字數 | 查詢問題平均相關文件數 | 主題 | 相關判斷次(相關) | 相關判斷次(不相關) | 語文 |
|---|---|---|---|---|---|---|---|---|---|---|
| Cranfield II | 1400 | 1.6 | 53.1 | 225 | 9.2 | 7.2 | 太空動力學 | 4 | 1 | 英文 |
| ADI | 82 | 0.04 | 27.1 | 35 | 14.6 | 9.5 | 文獻學 | N/A | N/A | 英文 |
| MEDLARS | 1033 | 1.1 | 51.6 | 30 | 10.1 | 23.2 | 醫學 | 2 | 2 | 英文 |
| TIME | 423 | 1.5 | 570 | 24 | 16.0 | 8.7 | 世界情勢 | N/A | N/A | 英文 |
| CACM | 3204 | 2.2 | 24.5 | 64 | 10.8 | 15.3 | ACM 通訊 | N/A | N/A | 英文 |
| CISI | 1460 | 2.2 | 46.5 | 112 | 28.3 | 49.8 | 資訊科 | N/A | N/A | 英文 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 學 | | | |

資料來源：江玉婷

表 4(續)：常用測試文件集

| 測試集 | 文件數 | 文體集大小 | 文件平均字數 | 查詢問題數 | 查詢問題平均字數 | 查詢問題平均相關文件數 | 主題 | 相關判斷次(相關) | 相關判斷次(不相關) | 語文 |
|---|---|---|---|---|---|---|---|---|---|---|
| NPL | 11429 | 3.1 | 20.0 | 100 | 7.2 | 22.4 | 電子，電腦，物理，地理 | N/A | N/A | 英文 |
| INSPEC | 12684 | N/A | 32.5 | 84 | 15.6 | 33.0 | 物理，電子控制 | 2 | 1 | 英文 |
| ISILT | 800 | N/A | N/A | 63 | N/A | 8.4 | 文獻學 | 1 | 1 | 英文 |
| UKCIS | 27361 | N/A | 182 | 193 | N/A | 57 | 生化 | 2 | 2 | 英文 |
| UKAEA | 12765 | N/A | N/A | 60 | N/A | N/A | 核子科學 | 2 | 1 | 英文 |
| LISA | 6004 | 3.4 | N/A | 35 | N/A | 10.8 | N/A | N/A | N/A | 英文 |
| Cystic Fibrosis | 1239 | N/A | 49.7 | 100 | 6.8 | 6.4-31.9 | 醫學 | 6 | 1 | 英文 |
| OHSUMED | 348566 | N/A | 250 | 101 | 10 | 17/19.4 | N/A | 2 | 1 | 英文 |
| TREC(TREC-1~6) | 175896 | 5GB | 481.6 | 350 | 105.8 | 185.3 | 多主題 | 1 | 1 | 英文 |
| AMARYLLIS | 336000 | 201 | N/A | 56 | N/A | N/A | 多主題 | N/A | N/A | 英文 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| NTCIR | 300000 | N/A | N/A | 100 | N/A | N/A | 多主題 | 2 | 1 | 英文 |
| IREX | N/A | N/A | N/A | N/A | N/A | N/A | 多主題 | 2 | 1 | 英文 |

資料來源：江玉婷

其中 TREC 是有史以來，文件數最多的測試集，在這麼大的測試文件集，要對所有測試文件進行相關判定是一件非常困難的事，但是由於有相當多的測試系統參與系統效能測試，因此在相關判定上利用 Pooling Method 法進行相關測試，將各效能系統送回結果的前 n 篇文件，剔除重複的文件後，回送給該查詢主題的原始建構者，再進行相關測試，此方法能有效的進行相關判定，並且能節省大量時間，由於本測試集只提供本文所建置的系統使用，因此在文件數的取樣數量上，將以非採用 Pooling Method 的測試集文件數量平均值爲依據，並逐一對每個查詢做相關比對，以彌補樣本效度的不足。

2.6.2 異質性：

早期測試集由於是先經由篩選，採用同質性相當高的文章，且文件長度差距不大，因此與真實檢索環境有相當大的不同，測試結果常受到質疑，故本測試集將直接在真實檢索環境取樣，以符合真實檢索環境的特性，並提高測試文件異質性。

2.6.3 相關判定：

在進行實際實驗時，必須先給定測試文件集與查詢問題的相關程度判別方式，作爲往後辨別文件分類方式結果準確率之依據，而相關判定有二元化與多元化等方式，二元化的方式是把文件區分爲相關與不相關兩類，如 TREC，其判別法則主要是觀察測試文件的某一部份是否與查詢問題有關，如果有關，則將其相關程度判定爲相關，否則即判定爲不相關，而多元判別的方式則是將文件相關程度區分爲幾個程度，例如分爲非常相關，相關，部分相關，不相關，相關程度區分等級越多，則區分難度越高，主要是因爲相關概念本身即是個相當主觀且模糊的概念，常會因爲判別情境及判別者的不同而產生相當大差異，且相關與不相關之間爲一個連續的，非離散地帶，無法非常清楚的劃分，再加上敵意概念本身即爲多向度概念，因此若採多元判定法，將影響相關判定的準確率，因此本測試集將以二元方式判斷文件與查詢主題相關程度，並且以多位領域專家進行相關判定，以提高相關判定的客觀度。

2.6.4　相關判定者：

Saracevic(1975)認爲，相關判定者通常是以資訊需求者擔任，以本文主題爲例，由於網站管理者或是討論板板主需要對討論區或留言版進行管理的工作，因此需要對文章進行敵意判別，而 Reid and Mizzaro(1998)認爲，判別人數可採一人單獨判斷或是多位需要相關資訊的人一起判斷，再利用加權或是其他方式來確定最後的相關程度。由於判別相關是相當主觀的工作，每個進行相關工作的人員常會因爲個人的個別認知差異，而對同一篇文章的相關判定產生很大的差異，據 TREC 的實驗結果顯示，不同的相關判斷者在判斷相關的一致性只有約 30%，而 Saracevic 也發現，(1)多位判斷者同時做相關判斷時，若判斷者的專長與需判斷文件所討論之主題相關性越高，則判斷結果的一致性越高。(2)若判斷者對討論主題較缺乏認識，則越容易將文件判定爲相關。(3)判斷爲不相關的一致性通常高於判斷爲相關的一致性。Voorhees(1998)根據 TREC 對相關判定是否會影響測試集的準確性實驗中發現，不同使用者對相關判定的差異並不會影響到被檢測之系統效益穩定性。因此在相關判定上，可採多位資訊需求者進行相關判定，對於判定爲相關的文章，可進行再次確認，以提高相關判定的客觀程度。

# 三、研究架構與方法

## 3.1 研究架構：

本研究經由問題分析與文獻探討後，確立了研究的流程與架構，初步先確定研究的目的與動機，再進行相關的文獻探討工作，同時進行系統的規劃與建置工作，並透過文獻不斷調整系統的規劃，選取樣本並進行實驗，最後依所得資訊進行分析，依實驗結果做出合理的結論，整體架構與流程如下圖所示：

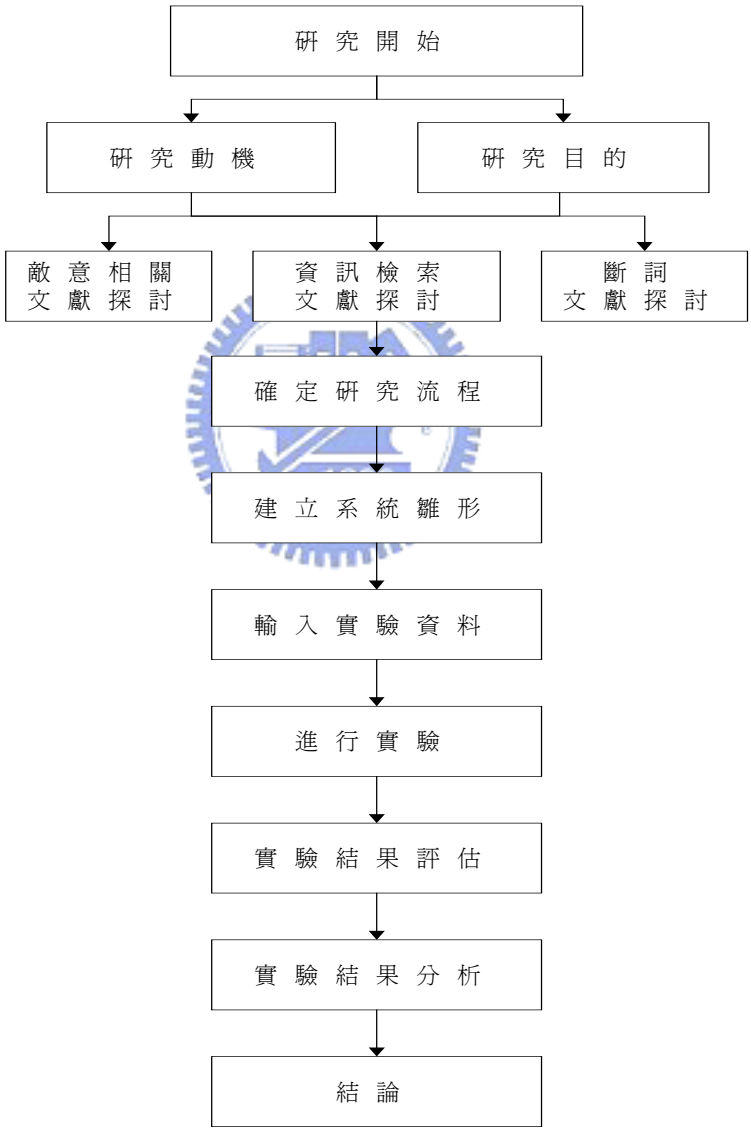


圖 1：研究架構圖

資料來源：本論文

3.2 研究工具：

本研究的主要工具爲一套建立在 web 上的敵意文章分類系統，系統環境如表 6 所示：

表 5：作業環境設置

| | |
|---|---|
| 作業系統 | RedHat Linux 7.2 (Enigma)Kernel 2.4.7-10 on an i686 |
| web 伺服器 | Apache 1.3.20-16 |
| 資料庫 | Postgresql 7.1.3-2 |
| 程式語言 | PHP-4.0.6-7、PHP-pgsql-4.0.6-7 |
| 系統網址 | http://163.25.180.120/cgi/nctu/i_r/index.html |

資料來源：本論文

爲了在實驗中觀察各步驟的影響以及控制變因，將敵意文章分類系統切割爲數個步驟實施，系統細項功能如下表所示：

敵意文章分類系統
- 步驟一:選擇訓練文章
  - 手動選擇
  - 隨機選擇
- 步驟二:進行文章斷詞
- 步驟三:計算關鍵詞權重
- 步驟四:計算敵意文章中心向量
- 步驟五:選擇實際文章
  - 手動選擇
  - 隨機選擇
- 步驟六:進行實際文章斷詞
- 步驟七:計算實際文章關鍵詞權重
- 步驟八:計算實際文章向量
- 步驟九:計算實際文章向量與中心向量相似
- 步驟十:結果統計

圖 2：敵意文章分類系統實行步驟

資料來源：本論文

步驟一與步驟二的主要目的是要建立能代表文章的關鍵詞詞庫，建立方式則是從訓練文件中找出文件特徵(即關鍵詞)，放入關鍵詞詞庫中，常用來建立詞庫的方法有

兩種，辭典式斷詞法與統計式斷詞法，辭典式斷詞法需事先將能判別文件類別的語詞放入詞庫中，如中央研究院詞庫小組(http://godel.iis.sinica.edu.tw/CKIP/)所建立的語料庫，但由於有些類別文件在定義上並不是那麼的清楚，以這次主題爲例，要找出具敵意文件的關鍵詞並不容易，況且由於地區、時間、討論主題的不同，即使他們都是具有敵意的文件，慣用語也會有不同的地方，而統計式斷詞法則是先蒐集一些同類型的文件(以下稱爲訓練文件)，經過斷詞後，計算出語詞的權重，並將可以代表此類文件的語詞選爲關鍵詞，並放入詞庫中，此方法可以解決辭典式斷詞法的缺點，因此在本系統中將採用統計式斷詞法。而取出語詞的長度，在中文文件中，由於較長的詞彙對文件分類並沒有明顯的效果〔22〕，因此取出語詞的長度將以長度爲 2 的二連字詞爲主。在電腦系統中，由於大小寫英文字母再加上常用符號，沒有超過 128 個，因此是以一個 byte 來儲存字母，但是常用中文字就將近 5000 個，所以必須用 2 個 byte 來儲存中文，但是這樣的方式，在斷詞時，會造成當大的困難，以要斷詞的長度爲兩個字爲例，如果一份文件中，只有英文字母和符號，在斷詞時的處理，只要從文件一開始，每次擷取 2 個 byte 的資料到文件結束即可將整份文件斷詞完成，但如果一份文件中含有中文及英文，則必須從文件一開始，先判別第一個 byte 的二進位碼是否大於 128，如果不是，再判別第二個 byte 是否大於 128，如果也不是，表示取出的資料不包含中文，故只要直接 2 個 byte 資料即可，但如果是的話，表示取出的資料，第一個字爲英文或符號，第二個字爲中文，因此總共需取出 3 個 byte 的資料，但是如果第一個 byte 大於 128，則需判別第三個 byte 的二進位碼是否大於 128，如果不是，表示取出的資料，第一個字爲中文，第二個字爲英文或符號，如果是的話，表示這 4 個 byte 總共包含 2 個中文字，因此需取出 4 個 byte 的資料，整個程序如圖 3 所示：

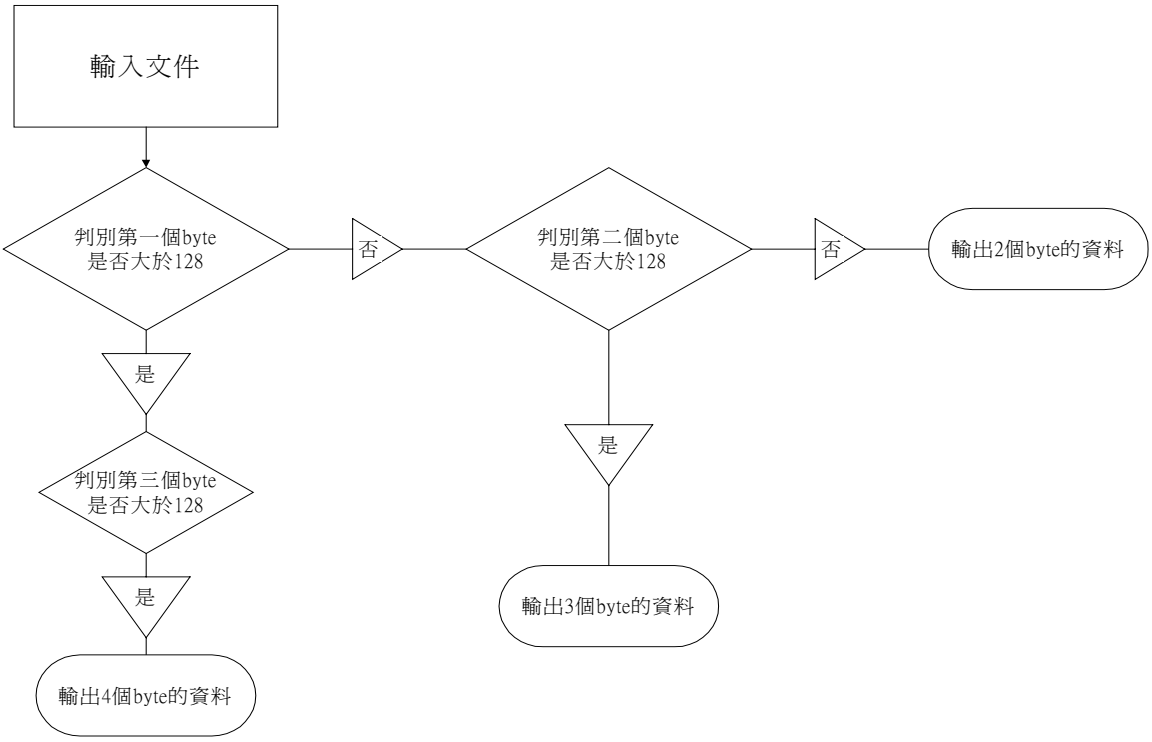


圖 3：中文關鍵詞斷詞流程(未導入 iconv 函數)

資料來源：本論文

我們可以發現，中文的斷詞比英文的斷詞要來的複雜許多，若要處理大量文件的斷詞，會耗費大量時間，由於本研究主要針對中文文件，因此本文中的系統將只擷取文章中的中文字，並利用 iconv 函數(http://www.iconv.com)，來加快斷詞的速度。擷取流程如圖 4：

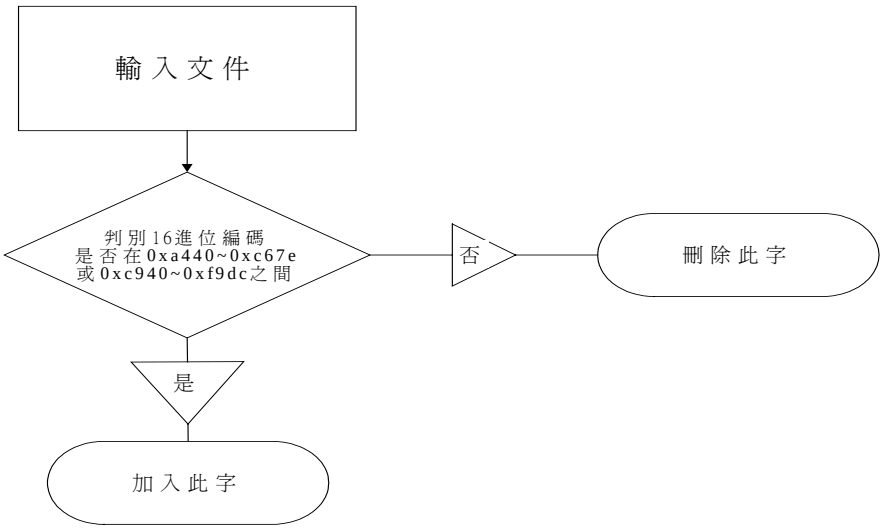


圖 4：中文關鍵詞斷詞流程(導入 iconv 函數)

資料來源：本論文

以下列文章爲例：

表 6：未經斷詞處理前的文章內容

| 作者 | 夏天 |
|---|---|
| 群組 | tw.bbs.comp.hardware |
| 標題 | Re: 微星主機板作弊被抓包了.. |
| 時間 | 2003-06-14 05:04:16 |

微星主機板作弊被抓包了..

※ 引述《(甲蟲)》之銘言：
> 你們爲何不買 ASUS P4C800 啊?
> .
> .
P4C800 除了賣那顆 875 還有什麼?

南橋不搭 ICH5R 要用 20378 作 S-ATA raid

那想用 IDE raid 的還要另買轉接頭?

NIC 也是，明明有 CSA 卻用 3Com

看了就不爽

除了 P4P800-D 看起來比較不錯之外(但是 VIA 的 raid...-_-?)

ASUS 其他 865/875 的板子都沒興趣
--
Origin: 精靈之城 ◆ From: vai.dorm4.ntnu.edu.tw

資料來源：tw.bbs.comp.hardware

系統會先將文章中的非中文文字，也就是標點符號、單位、英文字母、數字移除，移除後的文章內容如下表：

表 7：經斷詞處理後的文章內容

| | |
|---|---|
| 作者 | 熱@@ |
| 群組 | tw.bbs.comp.hardware |
| 標題 | Re: 微星主機板作弊被抓包了.. |
| 時間 | 2003-06-14 05:04:16 |
| 微星主機板作弊被抓包了引述風之銘言你們爲何不買啊除了賣那顆還有什麼南橋不搭要用作那想用的還要另買轉接頭也是，明明有卻用看了就不爽除了看起來比較不錯之外但是的其他的板子都沒興趣精靈之城 | |

資料來源：tw.bbs.comp.hardware

我們發現，在文章中除了使用者發表的本文外，通常還包含關於此主題的引言、及簽名，由於這兩部分無明顯的分界符號，或是較一致的特徵，因此並不容易用電腦進行精確的移除工作，所以在本系統中，對引言及簽名檔不進行移除的動作。在非中文文字移除後，系統會對此篇文章進行雙連字詞的擷取動作，並統計相關數據。雙連字詞擷取的結果如下表所示：

表 8：文章經斷詞後取出之關鍵詞列表

| |
|---|
| 微星、星主、主機、機板、板作、作弊、弊被、被抓、抓包、包了、了引、引述、述風、風之、之銘、銘言、言你、你們、們爲、爲何、何不、不買、買啊、啊除、除了、了賣、賣那、那顆、顆還、還有、有什、什麼、麼南、南橋、橋不、不搭、搭要、要用、用作、作那、那想、想用、用的、的還、還要、要另、另買、買轉、轉接、接頭、頭也、也是、是明、明明、明有、有卻、卻用、用看、看了、了就、就不、不爽、爽除、除了、了看、看起、起來、來比、比較、較不、不錯、錯之、之外、外但、但是、是的、的其、其他、他的、的板、板子、子都 都沒、沒興、興趣、趣精、精靈、靈之、之城 |

資料來源：本論文

步驟三、計算關鍵詞權重：在步驟二取出關鍵詞後，接下來要計算詞庫中每個關鍵

詞的權重，在本系統中，關鍵字的權值重是以 TF-IDF 方式取得，先假設關鍵字有 m 個，K 爲關鍵詞集合，

$$K=\{k_1,k_2,k_3.....k_m\} \tag{5}$$

而訓練文件有 n 份，將訓練文件集合設爲 D，則

$$D=\{\vec{d}_1,\vec{d}_2,\vec{d}_3.....\vec{d}_n \mid \vec{d}_i \in \vec{V}_m, \forall i=1,2..n\} \tag{6}$$

在訓練文件中出現關鍵字 $k_j$ 的篇數爲 $n_j$ ，設爲

$$N=\{n_1,n_2,n_3.....n_m\}\text{。} \tag{7}$$

$$\begin{aligned} w_{ij} &= tf_{ij}\bullet idf_{ij} \\ &= \frac{freq_{ij}}{\max_l freq_{il}}\bullet \log\frac{n}{n_j} \end{aligned} \tag{8}$$

在式子(8)中， $freq_{ij}$ 表示關鍵字 $k_j$ 在訓練文件 $d_i$ 中出現的次數， $freq_{ij}$ 爲正整數。

$\max_l freq_{il}$ 表示在訓練文件 $d_i$ 中出現最多次的關鍵字的次數， $\max_l freq_{il}$ 爲正整數。若關鍵詞 $k_j$ 未在某一文件 $d_i$ 中出現，則 $freq_{ij}=0$ ，因此 $w_{ij}=0$ 。

(2)、計算出每份訓練文件的向量：

在向量模型(vector model)中，每份文件皆以一個 m 維度的向量，

$$\vec{d}=(w_1,w_2,w_3.....w_m) \tag{9}$$

來表示〔16〕，其中 $w_i$ 表示關鍵字 $k_i$ 在文件 $d$ 中的權值重(term weight)，每一份訓練文件可以一個 m 維度的向量來表示，

$$\begin{aligned} \vec{d}_1 &= (w_{11},w_{12},w_{13}.....w_{1m}) \\ \vec{d}_2 &= (w_{21},w_{22},w_{23}.....w_{2m}) \\ &\bullet \\ &\bullet \\ &\bullet \\ \vec{d}_n &= (w_{n1},w_{n2},w_{n3}.....w_{nm}) \end{aligned} \tag{10}$$

其中 $w_{ij}$ 表示關鍵字 $k_j$ 在文件 $d_i$ 中的權值重(term weight)，

$$W=\{w_{ij}\,|\,i=1..n, j=1..m\} \tag{11}$$

步驟四、計算敵意文章中心向量：將所有訓練文件的向量平均後即得到敵意中心向量(hostility center vector)，以下簡稱 hcv。

設 $\vec{C}$ 爲 hcv，則

$$\vec{c}=(c_1,c_2,c_3.....c_m)$$
$$=(\frac{\sum_{i=1}^{n} w_{i1}}{n},\frac{\sum_{i=1}^{n} w_{i2}}{n},\frac{\sum_{i=1}^{n} w_{i3}}{n}.....\frac{\sum_{i=1}^{n} w_{im}}{n}) \tag{12}$$

步驟五、選擇實際文章：此步驟可以使用者需求，選擇自己所需要的實際文章，或指定文章類別與篇數兩個參數，由系統隨機選取文章：

步驟六到步驟八、計算實際文件向量：假設現在有一份文件 r，則此文件在文件向量空間中會以一個 m 維度的向量來表示

$$\vec{r}=(r_1,r_2,r_3.....r_m) \tag{12}$$

$r_i$ 表示關鍵字 $k_i$ 在文件 r 中的權重、權重的算法，則以 Salton and Buckley 所建議的計算方式

$$r_i = (0.5 + \frac{0.5\ freq_{ir}}{\max_l\ freq_{lr}})\bullet \log\frac{n}{n_i} \tag{13}$$
$$\forall\ i = 1..m$$

其中 $freq_{ir}$ 表示關鍵字 $k_j$ 在實際文件 r 中出現的次數，$\max_r\ freq_{ir}$ 表示在實際文件 r 中出現最多次的關鍵字的次數，而 n 與 $n_i$ 則與之前的定義相同。

步驟九、計算實際文章向量與敵意中心文章向量相似度：實際文章與敵意中心文章向量的相似度，則需利用餘弦函數來計算，文件 r 與敵意文章的相似度以 $sim(\vec{r},\vec{c})$ 來表示，函數 sim 定義如下：

$$sim(\vec{r},\vec{c}) = \frac{\vec{r}\bullet\vec{c}}{\left|\vec{r}\right|\bullet\left|\vec{c}\right|} = \frac{\sum_{i=1}^{m} r_i \bullet c_i}{\sqrt{\sum_{i=1}^{m} r_i^2} \bullet \sqrt{\sum_{i=1}^{m} c_i^2}} \quad (14)$$

sim 函數介於 0 與 1 之間，數值越大，表示實際文章與 hcv 的相似度越高。

3.3 研究步驟：

本實驗研究步驟分爲四個階段、第一階段爲實驗前之取樣及樣本分析：

3.3.1 研究樣本選取：

本實驗的訓練文件取樣來源爲實際文章中被分類爲論戰文章之文章，包含分歧(21 篇) 、爭論 (143 篇)、緊張(145 篇)、敵對(115 篇)、尖銳敵對(6 篇)合計共 430 篇文章。

表 9：實驗文件取樣來源

| | |
|---|---|
| 取樣來源 | 台灣 BBS 站的硬體版 tw.bbs.comp.hardware |
| 取樣方式 | 本實驗採連續抽樣，所有樣本的討論版文章之上傳時間爲 2003-06-06 04:59:05 至 2003-06-18 15:38:48。 |
| 文章長度 | 樣本文章長度介於 0 與 1371 之間。爲了避免文章長度影響文章自動化敵意值，因此本文所用分類方法將針對文章長度作正規化。 |
| 備註 | 本實驗只判斷文章是否具有敵意，因此若有不符合本版主題的文章，如廣告或發錯板面文意，皆不予刪除。 |

資料來源：本論文

3.3.2 實驗前之前置設定：

對訓練文章與實際文章進行敵意相關程度的設定，本實驗對文件的相關判定，採用次判斷者的判斷方式，本人爲判斷者 A、另外兩位爲次判斷者 B 與次判斷者 C，兩位次判斷者皆爲現職國中教師，且擔任資訊組長職務均超過 3 年，由於資訊組長需要擔任學校網站與討論板管理的職務，對判定敵意文件都具有

相當多的經驗，因此可提高相關認定的一致性與準確度。而由前面的說明，我們可以了解，具敵意的文章是由論戰中產生而來，但並不容易瞭解敵意是在論戰的哪一階段產生，因此如要直接判別一篇文章是否爲敵意文章難度較高，而論戰文章出現次數較高，較容易觀察，因此本文中對具敵意文章的認定，主要是先利用文章篇數出現較多的特性，找出論戰文章，再從論戰文章中，依照 Thompsen and Foulger 提出的論戰五種過程。將文章的敵意設爲 6~10，數字越高，代表越具有敵意。在相關程度的評定方式上，本實驗採取二階段評定，第一階段先由三位判斷者給予每篇文章 1 分或 0 分，1 分表示判斷者判定此篇文章具有敵意，0 分則否，三位判斷者給定的分數設爲 $S_A$、$S_B$、$S_C$，S 則爲三個分數的總和，$S = S_A + S_B + S_C$，由於在多位判斷者判定文章類別時，對於非主題的判定一致性較高，因此先將 S=0 的文件判定爲非敵意文章，再由判斷者 A 將非敵意文章分爲不完整文章、廣告文章、其他版面文章、發問或回覆文章，而 S>0 的文章則判定爲具有敵意文章，由判斷者 A 將敵意文章作進一步的確認，類別則以 Thompsen and Foulger 的分類爲基礎，分爲分歧、爭論、緊張、敵對、尖銳敵對五類，敵意判定流程如圖 5：

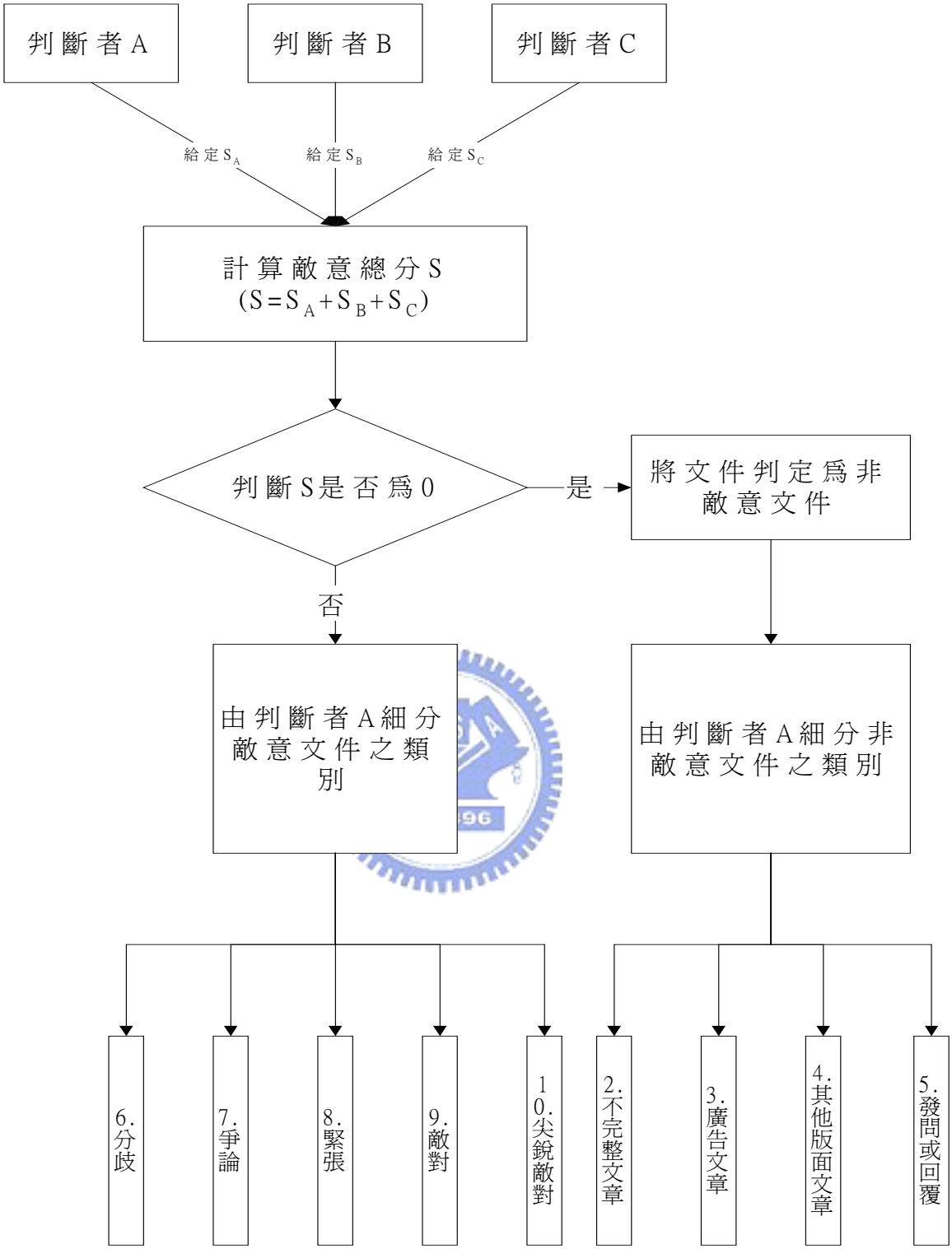


圖 5：文章敵意判別流程

資料來源：本論文

我們先將 5000 篇的文章以人工的方式給定敵意值，給定方式爲先將所有文件分類，先粗分有敵意文章、無敵意文章、及其他類文章，說明如下：

其他文章：

(1)不完整文章：此類文章的特性爲文章長度太短，無法呈現某個完整的概念，或是由於其他錯誤，導致文章不完整，例如撰寫文章時誤按送出，或是網路連線中斷導致不完整的文章。

(2)廣告文章：文章內容爲宣傳某樣產品或有交易行爲之文章。

(3)其他版面文章：本次實驗所使用文章爲 tw.bbs.comp.hardware 上的文章，主要內容爲討論電腦硬體的文章，若文章主題不符，如討論文學或政治，則歸類爲其他版面文章。

敵意文章：由前面的討論，我們發現敵意文章是在論戰過程中產生的，當某些因素引起使用者敵意後，使用者會將其敵意藉由文章表現出來，但是並無法得知敵意是在論戰的哪一個時期產生，而論戰文章由於篇數較多，較容易被觀察，因此先計算對於某一主題的文章篇數，若文章篇數明顯高於討論其他主題的文章篇數，則表示此一主題文章屬於論戰文章，並將其初步定爲具有敵意的文章。

非敵意文章：此類文章包含一般討論文章，即發問與討論。

分類方式如表 10 所示：

表 10：所有文章分類表

5000篇*tw.bbs.comp.hardware*的文章

- 02.不完整文章
- 完整文章
  - 不符合版面特性文章
    - 03.廣告文章
    - 04.其他版面文章
  - 符合版面特性文章
    - 非論戰文章
      - 05.發問或回覆文章
    - 論戰文章
      - 06.分歧(雙方表達不同意見)
      - 07.爭論(提出證據,支持自己)
      - 08.緊張(提出證據,攻擊對方)
      - 09.敵對(攻擊對方論點,指明道姓做人身攻擊)
      - 10.尖銳敵對期(使用大量攻擊性文字攻擊對方)

資料來源：本論文

經由此分類方式分類後，各文章的篇數如表 11：

表 11：各類文章篇數及所佔比例

| 類別代碼 | 文章類別 | 篇數 | 所佔比例 |
|---|---|---|---|
| 2 | 不完整文章 | 45 | 0.90 |
| 3 | 廣告文章 | 78 | 1.56 |
| 4 | 其他版面文章 | 5 | 0.10 |
| 5 | 發問或回覆文章 | 4442 | 88.84 |
| 6 | 分歧 | 21 | 0.42 |
| 7 | 爭論 | 143 | 2.86 |
| 8 | 緊張 | 145 | 2.90 |
| 9 | 敵對 | 115 | 2.30 |
| 10 | 尖銳敵對 | 6 | 0.12 |

資料來源：本論文

而文章類別代碼爲 6、7、8、9、10 的文章屬於論戰文章，共有 12 個討論主題，其中關於各主題的文章類別分布如下：

表 12：論戰文章中各主題所佔篇數及比例

| 序號 | 主題代碼 | 討論主題 | 類別代碼 | 篇數 | 比例 |
|---|---|---|---|---|---|
| 1 | A | [問題]HITACHI 的硬碟 | 7 | 2 | 8.60% |
| | | | 8 | 5 | |
| | | | 9 | 25 | |
| | | | 10 | 5 | |
| 2 | B | Lite-on 真好 | 8 | 1 | 0.23% |

資料來源：本論文

表 12(續)：論戰文章中各主題所佔篇數及比例

| 序號 | 主題代碼 | 討論主題 | 類別代碼 | 篇數 | 比例 |
|---|---|---|---|---|---|
| 3 | C | 台中市哪裡賣電腦的便宜又好的?? | 8 | 1 | 0.70% |
| | | | 9 | 2 | |
| 4 | D | 台灣....好貴的"寬頻" >"< | 6 | 3 | 48.14% |
| | | | 7 | 83 | |
| | | | 8 | 73 | |
| | | | 9 | 48 | |
| 5 | E | 全世界只有台灣收什麼鬼電路費 | 8 | 2 | 8.60% |
| | | | 9 | 18 | |
| | | | 10 | 17 | |
| 6 | F | 全國產電腦，可能嗎? | 6 | 6 | 10.93% |
| | | | 7 | 30 | |
| | | | 8 | 9 | |
| | | | 9 | 2 | |
| 7 | G | 如何讓 3DMark2001 的分數破萬?? | 8 | 1 | 1.40% |
| | | | 9 | 4 | |
| | | | 10 | 1 | |
| 8 | H | 青雲的產品網頁沒有中文 | 6 | 3 | 8.60% |
| | | | 7 | 20 | |
| | | | 8 | 15 | |
| 9 | I | 現在的顯示卡 | 8 | 1 | 0.23% |
| 10 | J | 微星主機板作弊被抓包了.. | 6 | 12 | 10.93% |
| | | | 7 | 5 | |
| | | | 8 | 20 | |
| | | | 9 | 10 | |
| 11 | K | 請問 so-net 的連線品質好嗎 | 7 | 5 | 1.16% |
| 12 | L | 請問一下 k6-2 的 cpu | 9 | 1 | 0.23% |
| | | | 總計 | 430 | 100% |

資料來源：本論文

3.3.3 門檻值設定：

每篇實際文章經過系統計算後，會產生與敵意文章的相似度，此相似度經過正規化之後會介於 0 與 1 之間，數值越大表示與敵意文章越相似，反之則否，由於相似度是介於 0 與 1 之間的數值，因此必須設定敵意文章的門檻值，才能依門檻值判別是否爲敵意文章，而門檻值越高，判定的精確度越高，但有越多的文章會被判定爲非敵意文章，反之，若門檻值越低，則能找出更多敵意文章，但也會因此而降低精確度，因此必須先進行門檻值實驗，找出較佳的敵意門檻值，同時提高判定敵意文章與非敵意文章的精確度。爲了解以不同主題爲訓練文章時的相似度分布，我們從 12 個主題中，選出 6 個討論篇數較多的主題，也就是文章篇數高於 8%的文章，來觀察敵意相似度的分布，此六個主題爲 A( [問題]HITACHI 的硬碟)、D (台灣....好貴的"寬頻" >"<)、E(全世界只有台灣收什麼鬼電路費)、F(全國產電腦，可能嗎?)、H(青雲的產品網頁沒有中文)、J(微星主機板作弊被抓包了..)，我們分別以此六主題文章爲訓練文章 ，每次在同一主題中隨機選出三篇文章爲訓練文章，再分別從敵意與非敵意文章中，隨機選取 10、20、30 篇文章，進行敵意相似度的計算，詳細流程如下圖所示：

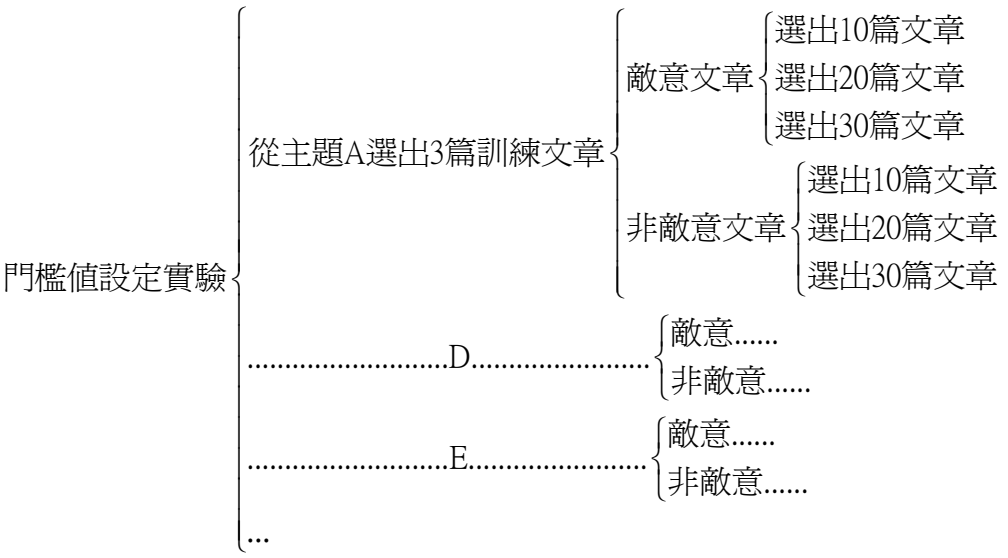


圖 6：門檻值實驗流程

資料來源：本論文

實驗後的各主題平均值如下圖所示：

表 13：各組平均值

| | 主題 A | 主題 D | 主題 E | 主題 F | 主題 H | 主題 J |
|---|---|---|---|---|---|---|
| 敵意 10 | 0.210303 | 0.160904 | 0.241688 | 0.108914 | 0.273305 | 0.147637 |
| 非敵意 10 | 0.179063 | 0.099368 | 0.115494 | 0.090258 | 0.123775 | 0.106237 |
| 敵意 20 | 0.223465 | 0.161734 | 0.290873 | 0.132015 | 0.208256 | 0.139688 |
| 非敵意 20 | 0.147783 | 0.085326 | 0.118047 | 0.098903 | 0.128751 | 0.099666 |
| 敵意 30 | 0.194430 | 0.164296 | 0.244877 | 0.116721 | 0.216210 | 0.138887 |
| 非敵意 30 | 0.139652 | 0.092839 | 0.120687 | 0.096279 | 0.131310 | 0.107615 |

資料來源：本論文

( 1)同一主題時，敵意與非敵意文章相似度差異檢定：

我們可以發現，敵意文章的門檻值分布在 0.290873 與 0.108914 之間，而非敵意文章的門檻值分布在 0.179063 與 0.085326 之間，且在所有實驗中，敵意文章的平均相似度都比非敵意文章的平均相似度要高，爲了了解在同一主題時，計算出的敵意文章與非敵意文章的相似度，是否具有明顯差異，以及因此先假設 $H_0$ 爲” 訓練文章爲同一主題時，所計算出的敵意文章與非敵意文章的相似度，沒有明顯差異”， $H_1$ 爲” 訓練文章爲同一主題時，所計算出的敵意文章與非敵意文章的相似度，具有明顯差異”，並利用獨立樣本 t 檢定來進行檢定，結果如表 14，利用同一主題文章作爲訓練文章時，計算出的敵意文章敵意值與非敵意文章敵意值的差異，50%的***p<0.001、50%的*p<0.05，達到統計上的顯著水準，可以拒絕虛無假設，也就是當利用同一主題文章作爲訓練文件，來計算敵意文章與非敵意文章的相似度時，所計算出的相似度具有明顯差距。這也表示利用同一主題文章所計算出的 hcv，能對敵意與非敵意文章產生明顯的分辨效果。

表 14：不同主題時，敵意文章與非敵意文章敵意值的差異顯著水準

| | 主題 A | 主題 D | 主題 E | 主題 F | 主題 H | 主題 J |
|---|---|---|---|---|---|---|
| 10 篇 | 0.230 | 0.011 | 0.001 | 0.107 | 0.012 | 0.004 |
| 20 篇 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.004 |
| 30 篇 | 0.002 | 0.000 | 0.000 | 0.048 | 0.000 | 0.004 |

資料來源：本論文

( 2)不同主題時，計算出的敵意文章相似度差異檢定：

以不同的主題作爲訓練文章，來計算敵意文章的相似度，是否也會產生差異？我們的虛無假設爲不同的主題作爲訓練文章，來計算敵意文章的相似度沒有差異。利用單因子變異數分析(one-way ANOVA)來進行假設檢定，結果如表 15 所示：

表 15：針對敵意文章以主題別作爲因子之單因子變異數分析摘要表

| Levene 統計 | 分子自由度 | 分母自由度 | 顯著性 |
|---|---|---|---|
| 3.886 | 5 | 174 | .002 |

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 組間 | .35 | 5 | .07 | 27.7*** |
| 組內 | .44 | 174 | .002527 | |
| 整體 | .79 | 179 | | |

資料來源：本論文

***p< .001

顯著性***p<.001，達非常顯著，但同質性檢定亦達顯著，無法將此六組資料視爲相等，因此我們利用多重比較的 LSD 法，來觀察各組資料間之相對差異，發現每個主題對於其他主題的差異度皆不同，有些差異甚大，有些組別幾乎沒有差異，但最少都與一個主題有明顯差異，檢定達顯著水準的總數爲 23，平均主題數爲 4.75 個。

表 16：針對敵意文章以主題別作爲因子之單因子變異數分析多重比較 LSD 法

| | 主題 A | 主題 D | 主題 E | 主題 F | 主題 H | 主題 J |
|---|---|---|---|---|---|---|
| 主題 A | | .021* | .000*** | .000*** | .095 | .000*** |
| 主題 D | .021* | | .000*** | .000*** | .000*** | .052 |
| 主題 E | .000*** | .000*** | | .000*** | .029* | .000*** |
| 主題 F | .000*** | .000*** | .000*** | | .000*** | .089 |
| 主題 H | .095 | .000*** | .029* | .000*** | | .000*** |
| 主題 J | .000*** | .052 | .000*** | .089 | .000*** | |

資料來源：本論文

六個主題所判別出的敵意平均值折線圖如圖 7 所示：

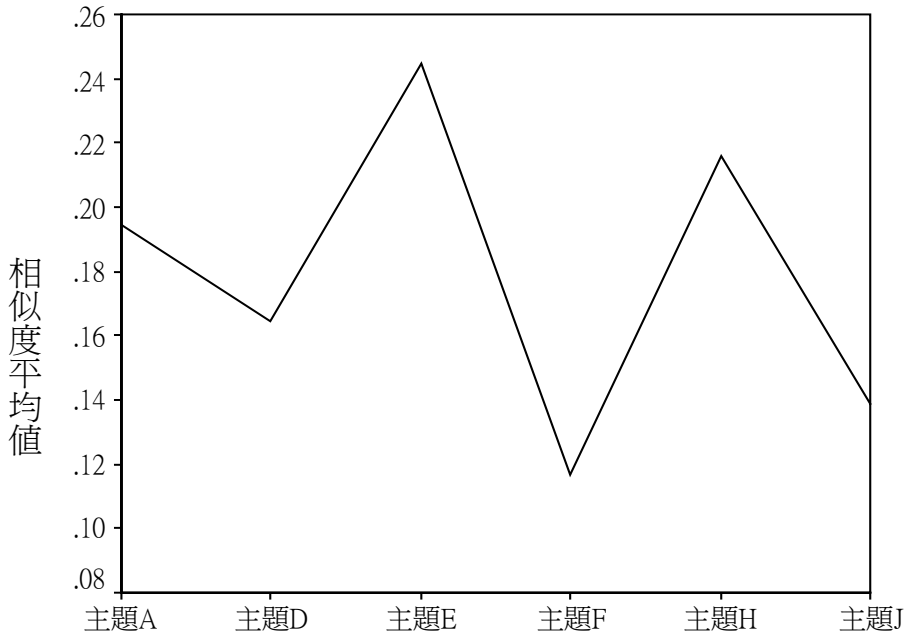


圖 7：主題與非敵意文章敵意平均值關係折線圖

資料來源：本論文

(3)不同主題時，計算出的非敵意文章相似度差異檢定：

以不同的主題作爲訓練文章，來計算非敵意文章的敵意值時，是否也會產生差異？我們虛無假設是利用不同的主題作爲訓練文章，所計算敵意文章的相似度沒有差異，利用單因子變異數分析(one-way ANOVA)來進行

假設檢定，結果如表17所示：

表 17：針對非敵意文章以主題別作爲因子之單因子變異數分析摘要

| Levene 統計 | 分子自由度 | 分母自由度 | 顯著性 |
|---|---|---|---|
| 12.730 | 5 | 174 | .000 |

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 組間 | .054 | 5 | .011 | 4.85*** |
| 組內 | .389 | 174 | .002 | |
| 整體 | .44 | 179 | | |

資料來源：本論文

***p< .001

顯著性***p<.001，達非常顯著，但同質性檢定亦達顯著，無法將此六組資料視爲相等，因此我們利用多重比較的LSD法，來觀察各組資料間之相對差異，發現每個主題對於其他主題的差異度皆不同，有些差異甚大，有些組別幾乎沒有差異，檢定達顯著水準的總數爲23，平均主題數爲4.75個，但最少都與一個主題有明顯差異，且具有差異的主題數較敵意文章少。

表 18：針對非敵意文章以主題別作爲因子之單因子變異數分析多重比較 LSD 法

| | 主題 A | 主題 D | 主題 E | 主題 F | 主題 H | 主題 J |
|---|---|---|---|---|---|---|
| 主題 A | | .000*** | .122 | .000*** | .495 | .009** |
| 主題 D | .000*** | | .024* | .788 | .002** | .227 |
| 主題 E | .122 | .024* | | .047* | .385 | .285 |
| 主題 F | .000*** | .778 | .047* | | .005** | .354 |
| 主題 H | .495 | .002** | .385 | .005** | | .054 |
| 主題 J | .009** | .227 | .285 | .354 | .054 | |

資料來源：本論文

六個主題所判別出的相似度平均值折線圖如圖8所示：

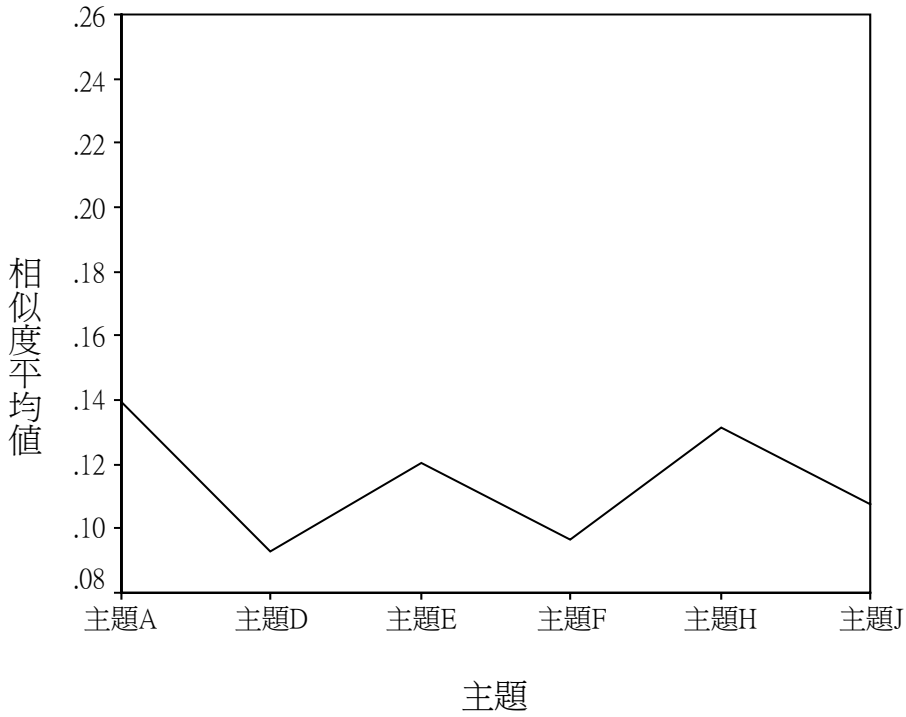


圖 8：主題與敵意文章敵意平均值關係折線圖

資料來源：本論文

由以上討論，我們可以知道，利用此六個主題作爲訓練文件，能夠使敵意文章與非敵意文章的敵意值，產生明顯差異，而不同主題所計算出之敵意值，亦具有相當差距，顯示每一主題文章在評估敵意時會產生不同影響，所以我們在選擇門檻值的時候，不採取加權方式，也就是以所有能夠判斷正確的敵意與非敵意文章總數，作爲門檻值的選擇標準，以同時提高敵意文章與非敵意文章的辨識正確率。爲了解相似度的分佈情形，我們以實驗篇數爲30篇時的實驗數據，來說明相似度的分布情形。

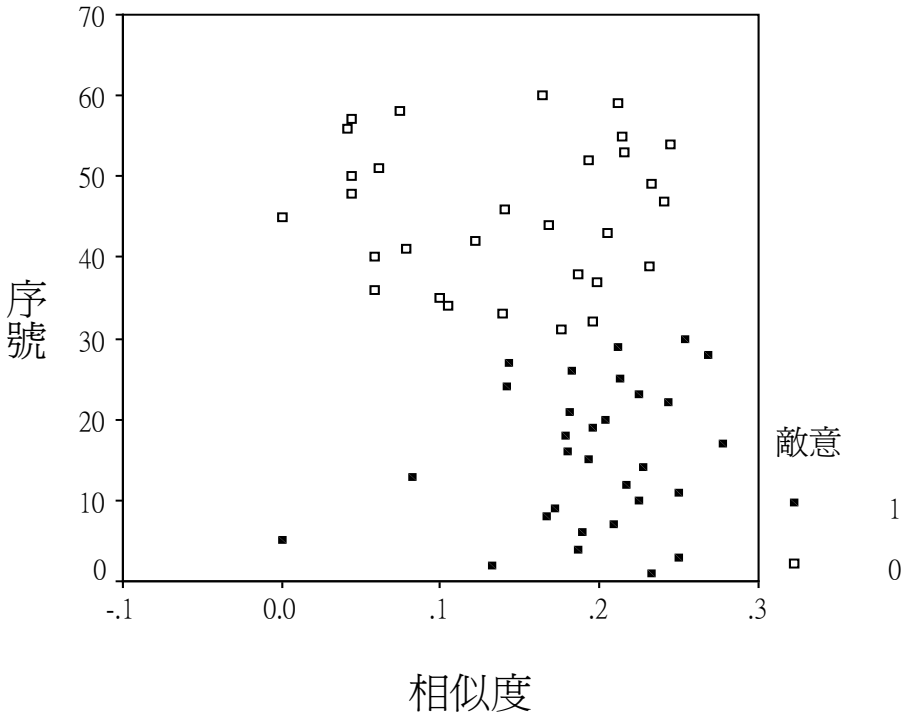


圖 9：主題 A 相似度分布圖

資料來源：本論文

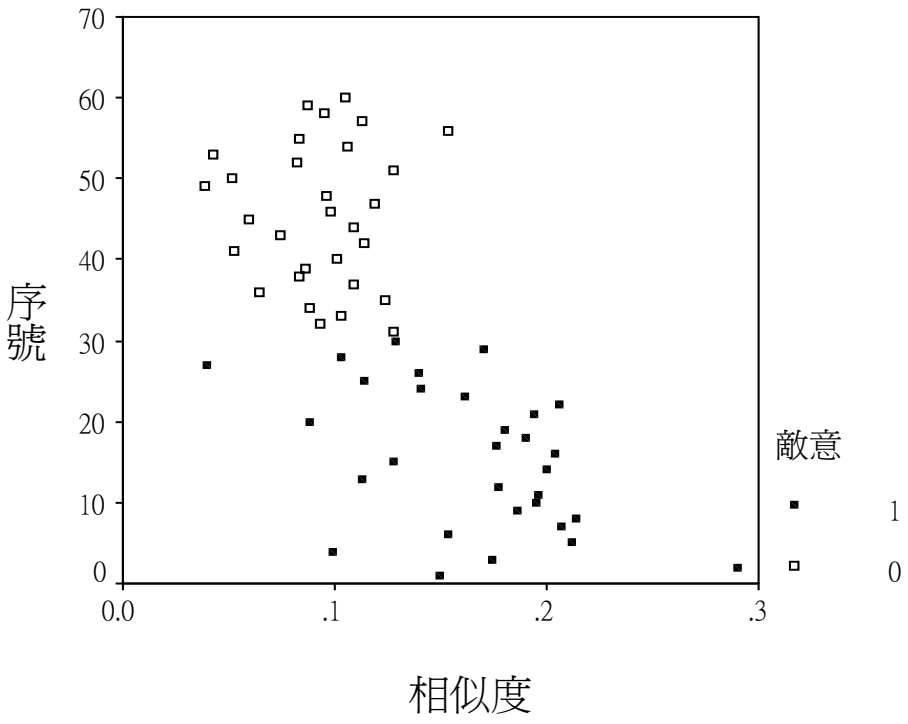


圖 10：主題 D 相似度分布圖

資料來源：本論文

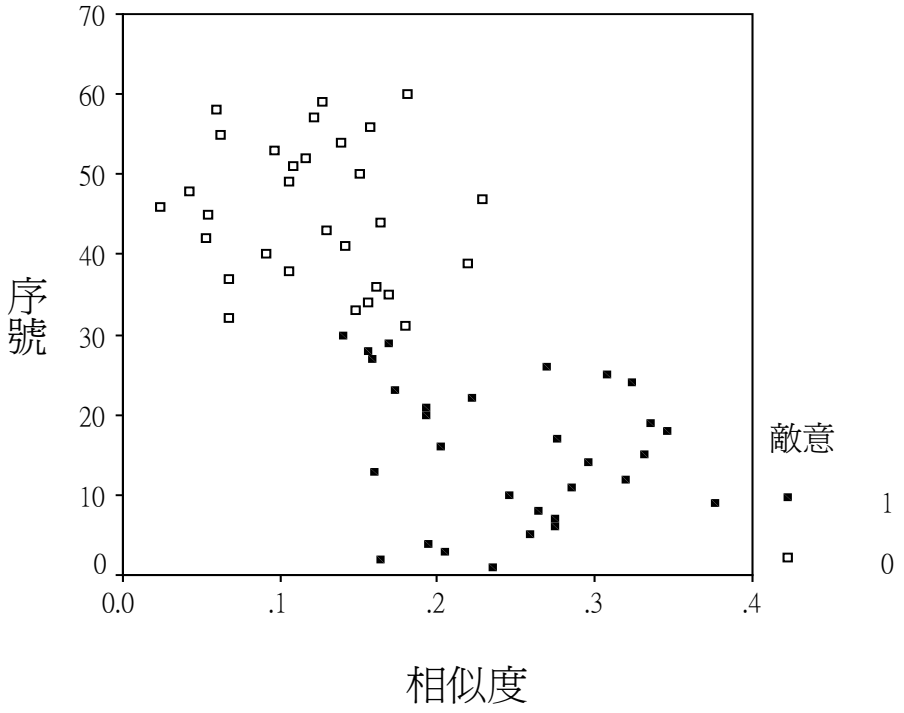


圖 11：主題 E 相似度分布圖

資料來源：本論文

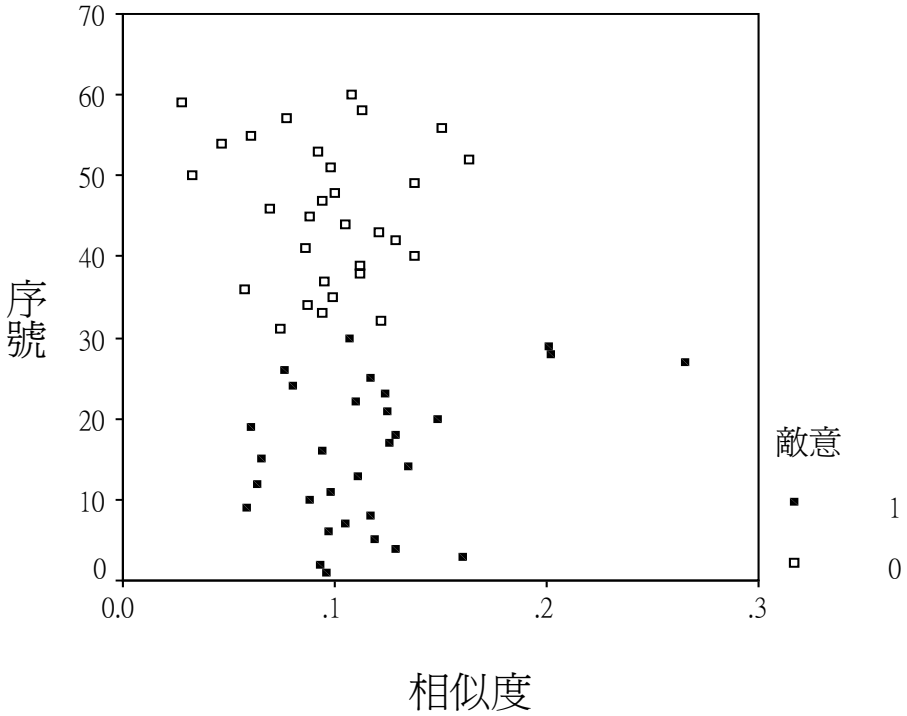


圖 12：主題 F 相似度分布圖

資料來源：本論文

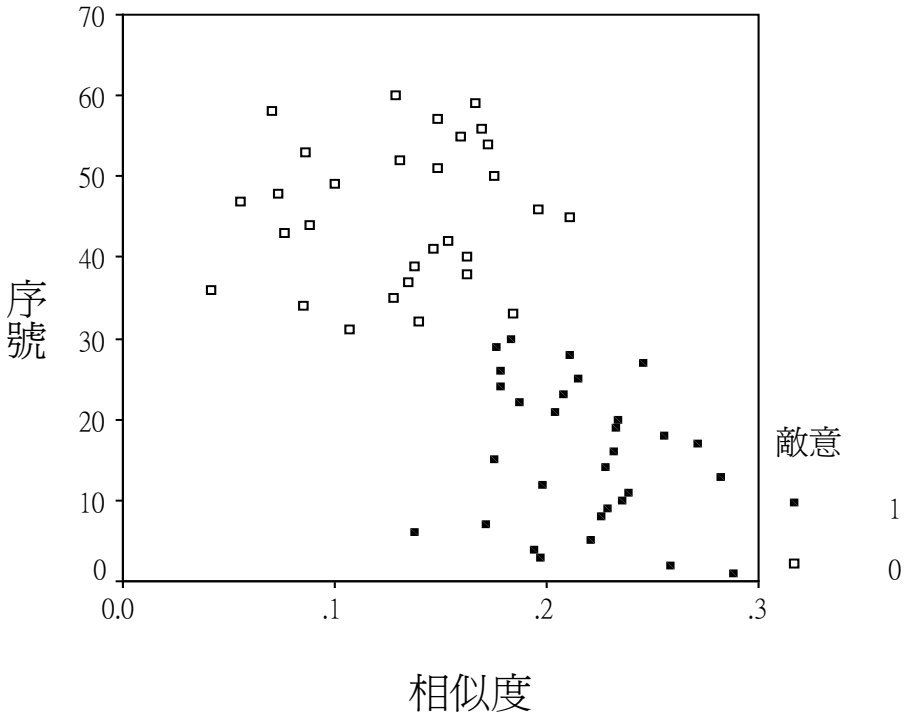


圖 13：主題 H 相似度分布圖

資料來源：本論文

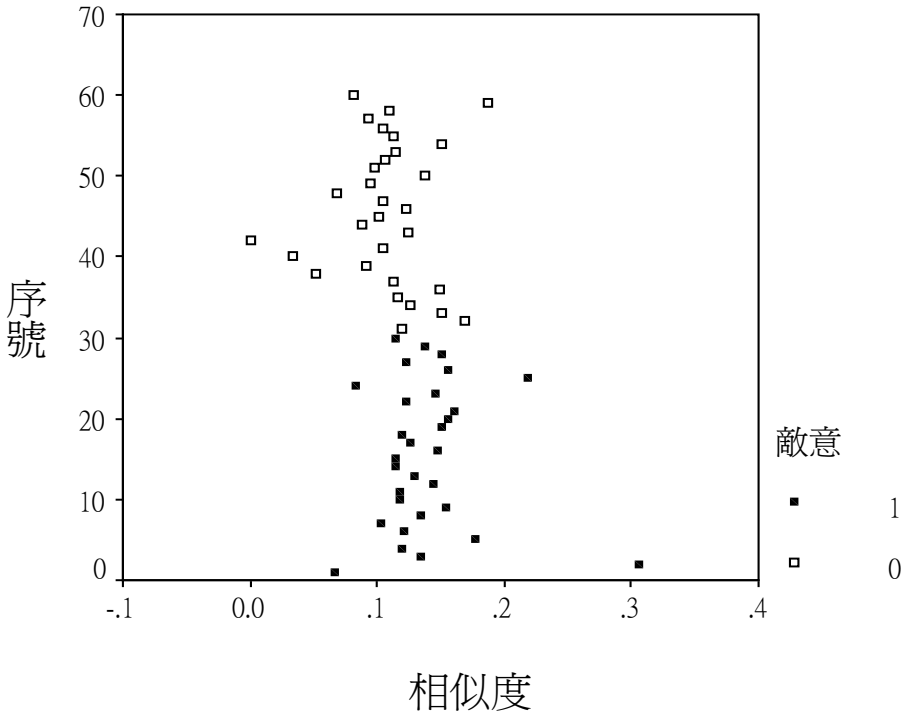


圖 14：主題 J 相似度分布圖

資料來源：本論文

我們可以發現，主題不同會影響相似度的分布離散情形，例如以主題 A 爲訓練文章時，敵意與非敵意文章的相似度分布的相當廣泛，而若以主題 J 作爲訓練文章，則相似度的分布會相當集中。而敵意文章與非敵意文章的重疊程度也隨主題的不同而有所改變，如以主題 H 爲訓練文章時，敵意文章與非敵意文章的重疊程度較低，而以主題 F 爲訓練文章時，敵意文章與非敵意文章的重疊程度就顯得較高。

經由上述說明，我們可以了解敵意文章與非敵意文章的相似度大部分皆分布在 0~0.3 之間，爲了設定判別敵意文章與非敵意文章的門檻值，我們將把門檻值從 0.04 開始，每次以 0.01 爲單位遞增到 0.3，來觀察能分辨出敵意文章與非敵意文章的程度，評量標準則是以在 6 個文章類別中，能正確判斷敵意文章與非敵意文章的篇數和，如下表所示：

表 17：各門檻值能正確判別之文章總數

| 序號 | 門檻 | A-1 | A-0 | D-1 | D-0 | E-1 | E-0 | F-1 | F-0 | H-1 | H-0 | J-1 | J-0 | tot-1 | tot-0 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.04 | 29 | 1 | 29 | 1 | 30 | 1 | 30 | 2 | 30 | 0 | 30 | 2 | 178 | 7 | 185 |
| 2 | 0.05 | 29 | 5 | 29 | 2 | 30 | 2 | 30 | 3 | 30 | 1 | 30 | 2 | 178 | 15 | 193 |
| 3 | 0.06 | 29 | 7 | 29 | 5 | 30 | 5 | 28 | 4 | 30 | 2 | 30 | 3 | 176 | 26 | 202 |
| 4 | 0.07 | 29 | 8 | 29 | 6 | 30 | 8 | 26 | 6 | 30 | 2 | 29 | 4 | 173 | 34 | 207 |
| 5 | 0.08 | 29 | 10 | 29 | 7 | 30 | 8 | 25 | 8 | 30 | 5 | 29 | 4 | 172 | 42 | 214 |
| 6 | 0.09 | 28 | 10 | 28 | 13 | 30 | 8 | 23 | 11 | 30 | 8 | 28 | 6 | 167 | 56 | 223 |
| 7 | 0.1 | 28 | 11 | 27 | 17 | 30 | 10 | 18 | 18 | 30 | 8 | 28 | 10 | 161 | 74 | 235 |
| 8 | 0.11 | 28 | 12 | 26 | 23 | 30 | 13 | 15 | 20 | 30 | 10 | 27 | 16 | 156 | 94 | 250 |
| 9 | 0.12 | 28 | 12 | 24 | 26 | 30 | 14 | 11 | 23 | 30 | 10 | 21 | 21 | 144 | 106 | 250 |
| 10 | 0.13 | 28 | 13 | 22 | 29 | 30 | 17 | 6 | 26 | 30 | 12 | 15 | 24 | 131 | 121 | 252 |
| 11 | 0.14 | 27 | 14 | 21 | 29 | 30 | 18 | 5 | 28 | 29 | 16 | 12 | 25 | 124 | 130 | 254 |
| 12 | 0.15 | 25 | 15 | 19 | 29 | 29 | 20 | 4 | 28 | 29 | 19 | 9 | 26 | 115 | 137 | 252 |
| 13 | 0.16 | 25 | 15 | 18 | 30 | 26 | 23 | 4 | 29 | 29 | 21 | 4 | 28 | 106 | 146 | 252 |

資料來源：本論文

表 17(續)：各門檻值能正確判別之文章總數

| 序號 | 門檻 | A-1 | A-0 | D-1 | D-0 | E-1 | E-0 | F-1 | F-0 | H-1 | H-0 | J-1 | J-0 | tot-1 | tot-0 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 0.17 | 24 | 17 | 17 | 30 | 24 | 26 | 3 | 30 | 29 | 25 | 3 | 29 | 100 | 157 | 257＊ |
| 15 | 0.18 | 21 | 18 | 13 | 30 | 23 | 27 | 3 | 30 | 24 | 27 | 2 | 29 | 86 | 161 | 247 |
| 16 | 0.19 | 17 | 19 | 11 | 30 | 23 | 28 | 3 | 30 | 22 | 28 | 2 | 30 | 78 | 165 | 243 |
| 17 | 0.2 | 15 | 22 | 6 | 30 | 20 | 28 | 3 | 30 | 19 | 29 | 2 | 30 | 65 | 169 | 234 |
| 18 | 0.21 | 13 | 23 | 3 | 30 | 18 | 28 | 1 | 30 | 17 | 29 | 2 | 30 | 54 | 170 | 224 |
| 19 | 0.22 | 10 | 26 | 1 | 30 | 18 | 29 | 1 | 30 | 15 | 30 | 1 | 30 | 46 | 175 | 221 |
| 20 | 0.23 | 7 | 26 | 1 | 30 | 17 | 30 | 1 | 30 | 11 | 30 | 1 | 30 | 38 | 176 | 214 |
| 21 | 0.24 | 6 | 28 | 1 | 30 | 16 | 30 | 1 | 30 | 6 | 30 | 1 | 30 | 31 | 178 | 209 |
| 22 | 0.25 | 3 | 30 | 1 | 30 | 15 | 30 | 1 | 30 | 5 | 30 | 1 | 30 | 26 | 180 | 206 |
| 23 | 0.26 | 2 | 30 | 1 | 30 | 14 | 30 | 1 | 30 | 3 | 30 | 1 | 30 | 22 | 180 | 202 |
| 24 | 0.27 | 1 | 30 | 1 | 30 | 12 | 30 | 0 | 30 | 3 | 30 | 1 | 30 | 18 | 180 | 198 |
| 25 | 0.28 | 0 | 30 | 1 | 30 | 9 | 30 | 0 | 30 | 2 | 30 | 1 | 30 | 13 | 180 | 193 |
| 26 | 0.29 | 0 | 30 | 0 | 30 | 8 | 30 | 0 | 30 | 0 | 30 | 1 | 30 | 9 | 180 | 189 |
| 27 | 0.3 | 0 | 30 | 0 | 30 | 7 | 30 | 0 | 30 | 0 | 30 | 1 | 30 | 8 | 180 | 188 |

資料來源：本論文

＊爲正確判別知文章篇數總和的最大值。

我們可以發現，當門檻值由 0.04 遞增到 0.14 時，正確判別的文章總數也隨之提高，而門檻值由 0.17~0.3 時，正確判別的文章總數會隨著降低，而在門檻值等於 0.17 時，能正確判別的文章總數和爲 257，爲所有能判別正確的文章總數和的最高值，如下圖所示，因此在接下來所進行的主要實驗中，將以 0.17 作爲敵意文章與非敵意文章的門檻值，相似度大於或等於 0.17 的文章，判定爲敵意文章，而低於 0.17 的文章，則判定爲非敵意文章。

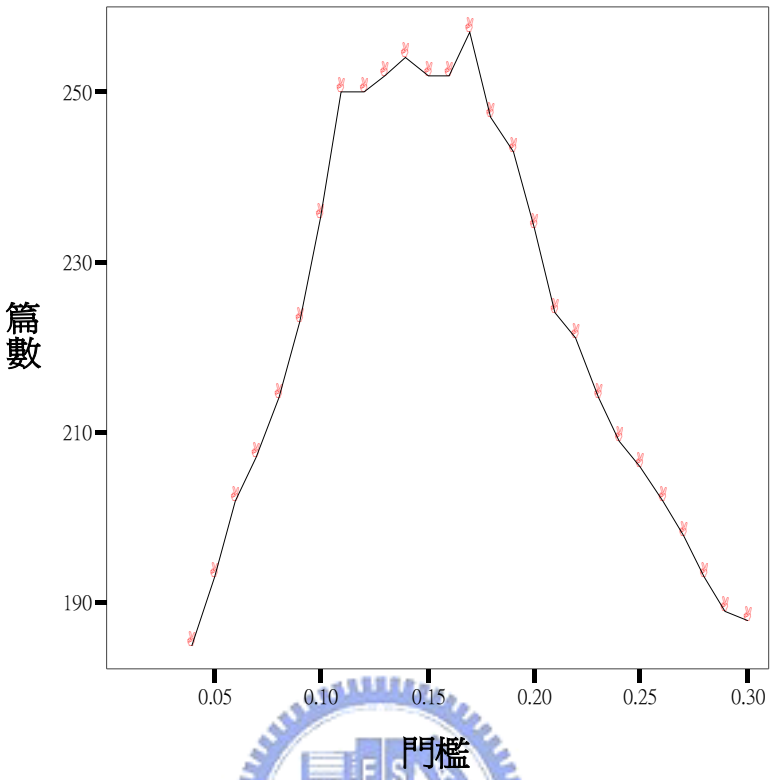


圖 15：門檻值與正確判別文章之關係圖

資料來源：本論文

第二階段(實驗)：隨機在訓練文件中選出 3 篇不同主題文章後，利用 5 份文章來計算 hcv，接下來在 5000 篇實際文章中，分別對敵意文章與非敵意文章進行實驗，每次隨機抽樣 10、20 篇文章，由系統計算出與 hcv 的相似度，將相似度大於或等於 0.17 的文章敵意值設爲 1，其他則設爲 0，依此程序進行實驗，總計需進行 20 次子實驗，選取 $(10+20)\times 10$ 篇，也就是 300 篇文章，每次實驗後所得結果如表 18 範例所示：

表 18：實驗結果範例

| 文章序號 | 文章編號 | 文章類別 | 相似度 | 電腦敵意 | 人工敵意 |
|---|---|---|---|---|---|
| 1 | 61824 | 4 | 0.234848 | 1 | 0 |
| 2 | 62117 | 3 | 0.163132 | 0 | 0 |
| 3 | 62392 | 5 | 0.204557 | 1 | 0 |
| 4 | 62521 | 7 | 0.194616 | 1 | 1 |
| 5 | 62577 | 2 | 0.259172 | 1 | 0 |
| 6 | 63019 | 3 | 0.074889 | 0 | 0 |
| 7 | 63033 | 8 | 0.074351 | 0 | 1 |
| 8 | 63055 | 2 | 0.064196 | 0 | 0 |
| 9 | 63176 | 3 | 0.376834 | 1 | 0 |

資料來源：本論文

第三階段(實驗後結果分析)：爲了解此方式所計算出之文件相似度與人類思考方式之差異，計畫中將採取實驗法，以系統判讀爲實驗組，人工判讀爲對照組，將電腦所判別出來的相似度與人工判讀的相似度做對照。精確率(prescion)及召回率(recall)是常用來評估文件分類系統效能的兩個數值，精確率(prescion)的意思是，所有擷取出的文件中，與搜尋 q 有關的文件比率。而召回率(recall)的意思是，所有與搜尋 q 有關的文件，能被擷取出的比率。例如現在共有 100 篇文件，其中與搜尋 q 有關的文件共有 40 篇，假設某次的擷取共取出 50 篇，而此 50 篇文件中有 30 篇與搜尋 q 有關，則

$$\begin{cases} \text{精確率}(precison) = \dfrac{30}{50} = 0.6 \\ \text{召回率(recall)} = \dfrac{30}{40} = 0.75 \end{cases} \tag{16}$$

本實驗由於實驗設計的因素，文章分類只有敵意與非敵意兩種，需對評估指標作修正，以符合本實驗所需，評估的指標將使用敵意正確率(HR)與非敵意正確率(NHR)，HR 與 NHR 主要在測量敵意文章與非敵意文章的正確率，數值越高，表示精確度越高。

$$\begin{cases} \text{HR}=\dfrac{\text{人工與系統皆認爲具有敵意的文章篇數}}{\text{抽出文章總篇數}} \\ NHR=\dfrac{\text{人工與系統皆認爲不具有敵意的文章篇數}}{\text{抽出文章總篇數}} \end{cases} \tag{17}$$

以上述實驗方式爲例，假設在某次實驗中取出 n 篇文章，經由實驗後，人工與系統對認定文章是否有敵意的組合，可能有下列四種：

（1）人工敵意＝1，系統敵意＝1，人工與系統皆認定此篇文章具有敵意。

（2）人工敵意＝1，系統敵意＝0，人工認爲具有敵意，系統不認爲具有敵意。

（3）人工敵意＝0，系統敵意＝0，人工認爲不具敵意，但電腦認爲具有敵意。

（4）人工敵意＝0，系統敵意＝0，人工與系統皆認爲不具有敵意。

則 $\begin{cases} HR=\dfrac{a}{n} \\ NHR=\dfrac{d}{n} \end{cases}$ (18)

表 19：人工與系統對敵意認定的可能組合

| 組合代碼 | 人工 | 系統 | 篇數 |
|---|---|---|---|
| 1 | 1 | 1 | a |
| 2 | 1 | 0 | b |
| 3 | 0 | 1 | c |
| 4 | 0 | 0 | d |
| | | | n |

資料來源：本論文

# 四、結果與討論

## 4.1 實驗結果：

非敵意文章的相似度分布在 0 到 0.1695 之間，平均爲 0.109569，各敘述統計量如表 21 所示：

表 20：非敵意文章，每次取樣 10 篇，進行 10 次實驗之敘述統計量

| 最小值 | 最大值 | 平均數 | 標準差 |
|---|---|---|---|
| 0 | .17 | .11 | .036 |

資料來源：本論文

由於門檻值爲0.17，因此所有文章皆被判別爲非敵意文章，NHR值皆爲1，與人工認定方式完全符合。各次實驗的NHR值如表22所示：

表 21：非敵意文章，每次取樣 10 篇，進行 10 次實驗之實驗結果

| id | a | b | c | d | NHR |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 10 | 1 |
| 2 | 0 | 0 | 0 | 10 | 1 |
| 3 | 0 | 0 | 0 | 10 | 1 |
| 4 | 0 | 0 | 0 | 10 | 1 |
| 5 | 0 | 0 | 0 | 10 | 1 |
| 6 | 0 | 0 | 0 | 10 | 1 |
| 7 | 0 | 0 | 0 | 10 | 1 |
| 8 | 0 | 0 | 0 | 10 | 1 |
| 9 | 0 | 0 | 0 | 10 | 1 |
| 10 | 0 | 0 | 0 | 10 | 1 |

資料來源：本論文

敵意文章的相似度分布在 0.091 到 0.2135 之間，平均爲 0.152683，各敘述統計量如表 23 所示：

表 22：敵意文章，每次取樣 10 篇，進行 10 次實驗之敘述統計量

| 最小值 | 最大值 | 平均數 | 標準差 |
|---|---|---|---|
| .091 | .214 | .153 | .024 |

資料來源：本論文

只有少部分文章皆被判別爲敵意文章，NR 介於 0.1 與 0.4 之間，平均值爲 0.25。各次實驗的 HR 值如表 23 所示：

表 23：敵意文章，每次取樣 10 篇，進行 10 次實驗之實驗結果

| id | a | b | c | d | HR |
|---|---|---|---|---|---|
| 1 | 3 | 0 | 7 | 0 | 0.3 |
| 2 | 1 | 0 | 9 | 0 | 0.1 |
| 3 | 2 | 0 | 8 | 0 | 0.2 |
| 4 | 3 | 0 | 7 | 0 | 0.3 |
| 5 | 3 | 0 | 7 | 0 | 0.3 |
| 6 | 4 | 0 | 6 | 0 | 0.4 |
| 7 | 3 | 0 | 7 | 0 | 0.3 |
| 8 | 1 | 0 | 9 | 0 | 0.1 |
| 9 | 1 | 0 | 9 | 0 | 0.1 |
| 10 | 4 | 0 | 6 | 0 | 0.4 |

資料來源：本論文

非敵意文章取樣 20 篇時，敵意文章的相似度分布在 0 到 0.213 之間，平均值爲 0.108197，各敘述統計量如表 24 所示：

表 24：非敵意文章，每次取樣 20 篇，進行 10 次實驗之敘述統計量

| 最小值 | 最大值 | 平均數 | 標準差 |
|---|---|---|---|
| 0 | .213 | .108 | .037 |

資料來源：本論文

大部分文章皆被判別爲非敵意文章，NR 介 0.1 與 0.9 之間，平均值爲 0.975。各次實驗的 NHR 值如表 26 所示：

表 25：非敵意文章，每次取樣 20 篇，進行 10 次實驗之實驗結果

| id | a | b | c | d | NHR |
|---|---|---|---|---|---|
| 51 | 0 | 1 | 0 | 19 | 0.95 |
| 53 | 0 | 0 | 0 | 10 | 1 |
| 58 | 0 | 0 | 0 | 20 | 1 |
| 59 | 0 | 1 | 0 | 19 | 0.95 |
| 60 | 0 | 0 | 0 | 20 | 1 |
| 61 | 0 | 0 | 0 | 20 | 1 |
| 62 | 0 | 2 | 0 | 18 | 0.9 |
| 63 | 0 | 1 | 0 | 19 | 0.95 |
| 64 | 0 | 0 | 0 | 20 | 1 |
| 65 | 0 | 0 | 0 | 20 | 1 |

資料來源：本論文

敵意文章取樣 20 篇時，敵意文章的相似度分布在 0.0469 到 0.3194 之間，平均值爲 0.155687，各敘述統計量如表 27 所示：

表 26：敵意文章，每次取樣 20 篇，進行 10 次實驗之敘述統計量

| 最小值 | 最大值 | 平均數 | 標準差 |
|---|---|---|---|
| .047 | .312 | .156 | .035 |

資料來源：本論文

大部分文章皆被判別爲非敵意文章，NR 介 0.1 與 0.45 之間，平均值爲 0.29。各次實驗的 HR 值如表 28 所示：

表 27：敵意文章，每次取樣 20 篇，進行 10 次實驗之實驗結果

| id | a | b | c | d | HR |
|---|---|---|---|---|---|
| 1 | 9 | 0 | 11 | 0 | 0.45 |
| 2 | 9 | 0 | 11 | 0 | 0.45 |
| 3 | 6 | 0 | 14 | 0 | 0.3 |
| 4 | 5 | 0 | 15 | 0 | 0.25 |
| 5 | 2 | 0 | 18 | 0 | 0.1 |
| 6 | 6 | 0 | 14 | 0 | 0.3 |
| 7 | 3 | 0 | 17 | 0 | 0.15 |
| 8 | 6 | 0 | 14 | 0 | 0.3 |
| 9 | 6 | 0 | 14 | 0 | 0.3 |
| 10 | 6 | 0 | 14 | 0 | 0.3 |

資料來源：本論文

我們發現，在分類非敵意文章時，每次取樣 10 篇，進行 10 次實驗之 NHR 的值皆爲 1，因此 hcv 對於非敵意文章具有良好分類效果，而在分類敵意文章時，平均的準確率只有 0.25，可能的原因是敵意文章門檻值過低，而系統在判定敵意文章時，是依據敵意文章門檻值來作爲判斷的標準，若門檻值設定過高，會減少人工認定具敵意文章被系統判定爲非敵意文章的篇數，因此準確率就會隨著降低。因此我們以門檻值爲自變項，HR 與 NHR 的平均值爲依變項，來觀察準確率的變化情形。

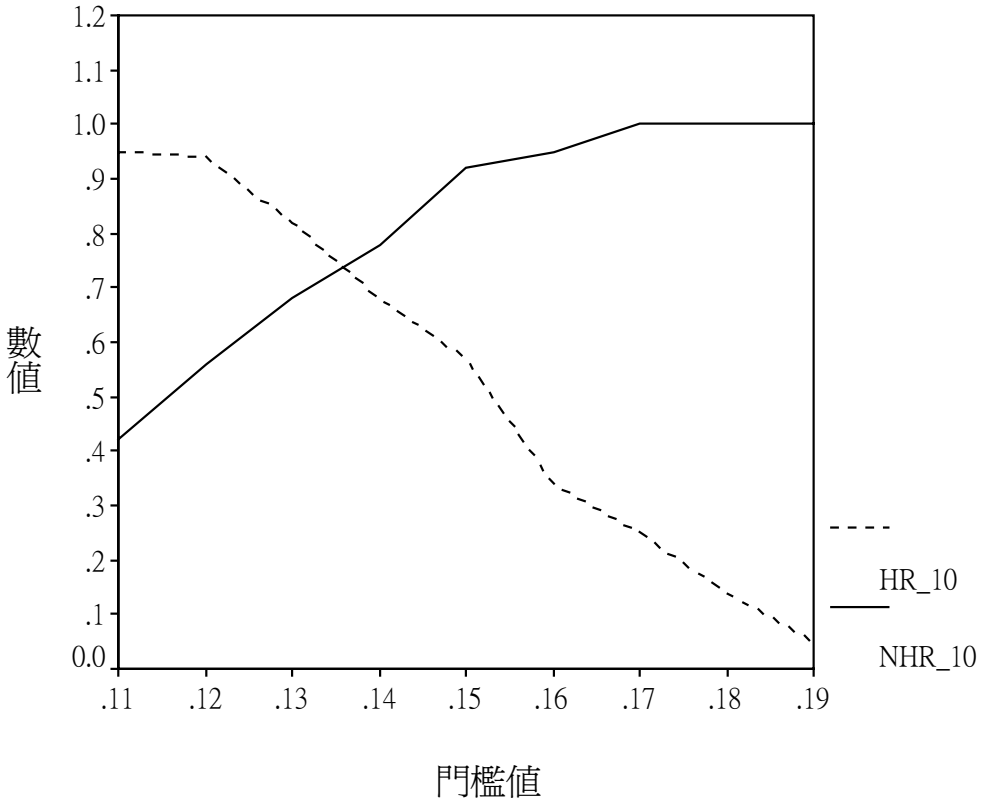


圖 16：取樣10篇時，門檻值與HR、NHR之關係折線圖

資料來源：本論文

當門檻值設定爲0.17時，對敵意文章的判別準確率，平均值爲0.25，對非敵意文章的判別準確率，平均值爲0.97，若降低門檻值，HR與NHR的平均值會同時提昇，當門檻值降低爲0.136時，HR會等於與NHR，約爲0.7，對於提昇HR與NHR，具有相當明顯的效果。而當取樣篇數爲20篇時，亦呈現類似情形。

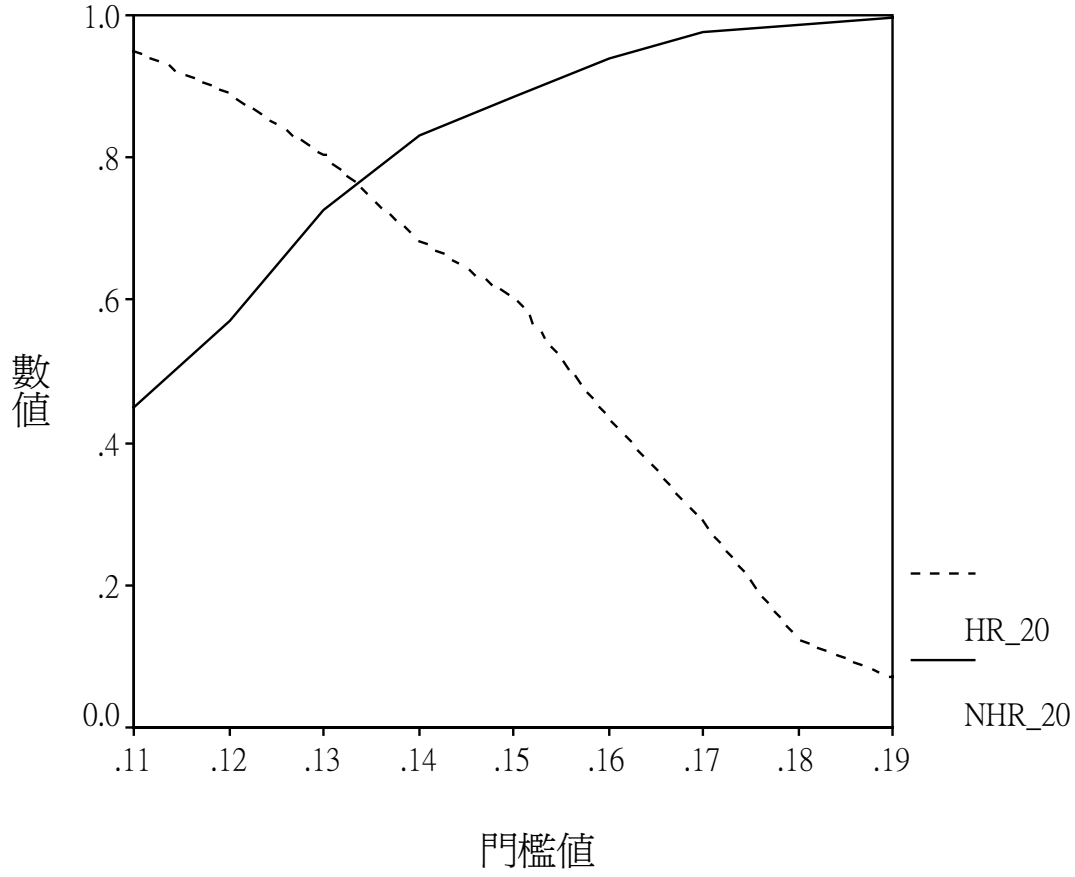


圖 17：取樣 20 篇時，門檻值與 HR、NHR 之關係折線圖

資料來源：本論文

4.2 結論：

本文進行了敵意文章分級系統的初探研究，藉由向量模型，將每篇文章對應到一個向量，向量的維度即爲 hcv 的維度，利用餘弦函數計算向量與 hcv 的餘弦值，此值即爲對應文章與 hcv 之相似度，而相似度高於門檻值的文章稱爲敵意文章。並在 tw.bbs.comp.hardware 中，以連續取樣方式，取樣 5000 篇文章進行實驗，門檻值的設定方式則依門檻值實驗取得，由人工評定具敵意的文章中，選出討論篇數高於 8%的六個主題文章，並分別利用六個主題文章來取得 hcv，分別針對敵意與非敵意文章進行相似度計算，利用獨立樣本 t 檢定與單因子變異數分析來檢測相似度的變化情形，研究發現：

(1)同一主題時，計算出的敵意文章相似度平均值，皆比非敵意文章相似度平均值要高，且敵意與非敵意文章相似度具有顯著差異，50%的***p<0.001、50%的*p<0.05，表示利用同一主題文章所建立的 hcv，能明顯的區分出敵意與非敵意文章。

(2)不同主題時，利用多重比較的 LSD 法，來觀察各組資料間之相對差異，發現每個主題對於其他主題的差異度皆不同，有些差異甚大，有些組別幾乎沒有差異，但最少都與一個主題有明顯差異。

(3)不同主題時，計算出的非敵意文章相似度差異性，與(2)的結果相同，但具有差異的主題數較敵意文章少。

(4)六個主題所計算出的相似度離散差異甚大，有些相似度較集中，有些較分散。因此在設定門檻值，以六個主題所能正確分辨出的敵意與非敵意文章篇數和爲標準，得到最佳的門檻值爲 0.17，因此以 0.17 作爲正式實驗的門檻值。進行敵意與非敵意文章的分類實驗，並以 HR 與 NHR 作爲分類效能的指標，發現利用門檻值門檻值 0.17 來進行敵意文章判別並利用不同主題文章所產生的 hcv 時，對於非敵意文章有較佳的精確度，約爲 0.98，但對於敵意文章的分類精確度則較差，約爲 0.25，但當門檻值調整爲 0.136 時，非敵意文章與敵意文章的精確度皆爲 0.73，具有良好分類成效。

(5)實際分類時，由於只有兩種分類，且對於非敵意文章的分類準確度接近 1，因此可利用反向分類的方式，先將非敵意文章分類出來，再以人工方式從剩下的文章中找出敵意文章。

4.3 研究限制：

(1)研究設計的限制：本系統設計主要是針對中文文章進行分類，如權重計算方式或是斷詞方式亦針對中文的特性設計，無法確定對於其他語文文章，亦具有相同效果。

(2)研究對象限制：本研究因時間、人力的限制，只採用 tw.bbs.comp.hardware 版面上的文章，作爲研究對象，文章的取樣篇數亦只有 5000 篇，樣本並不多，再由樣本分析來看，取樣方式爲按照時間的連續抽樣，因此對實驗結果亦會產生影響，在推論研究結果時，必須考慮這些因素。

(3)研究應用的限制：本實驗對象主要是 tw.bbs.comp.hardware 版面上的文章，文章具有硬體討論版的特性，無法保證亦適用在其他版面。

4.4 未來研究方向：

(1)降低 HCV 的維度：在計算 hcv 時，隨機抽樣出 10 篇文章進行實驗，所需時間約 2 小時 20 分左右，計算一篇文章相似度的平均時間約 14 分左右，未來研究可朝向如何在不影響分類成效的前提下，降

低 hcv 的維度，以提昇分類效率。

(2)敵意文章組成方式：本研究並未對會形成敵意文章的二連字詞進行研究，僅在附錄中列出出現次數高於 10 次的二連字詞，未來可研究敵意文章中某些特別語詞是否出現頻率較高，以深入了解敵意文章組成。

(3)自動取得最佳門檻值：本研究是以半自動的方式取得門檻值，需要先進行門檻值實驗，求出第一次最佳門檻值後，在真實環境中進行實驗，觀察分類結果，並進一步調整較佳門檻值，未來可研究將門檻值設定程序自動化的方法，降低人為誤差，以提昇分類成效。

(4)敵意文章的定義：本研究為敵意文章分級系統初探，在進行文章分級時，僅透過敵意的外顯行為，也就是論戰，來將敵意與非敵意文章進行初步分類，未來可研究針對敵意文章進行直接的定義或是給定操作型定義，以便針對敵意文章，做進一步的分級。

(5)不同主題所造成 hcv 的差異：本研究僅利用統計方法，證明不同主題所形成的 hcv 具有差異，但並未針對造成差異的原因，進行質化研究，未來可研究造成不同主題 hcv 產生差異的原因，以便找出較能代表敵意文章的 hcv。

(6)心理敵意與敵意文章之關聯：敵意是人類心理的狀態，而敵意文章則是敵意的外顯行為，未來可研究兩者之間的關聯性，也就是個體在具有敵意的狀態下，是如何產生敵意文章，以及敵意文章是如何使個體產生敵意，來釐清敵意與非敵意文章之間的關聯性。

(7)減少敵意文章之機制：大量的敵意文章，常會造成網管人員的困擾，也會提高個體產生衝突的可能性，因此一般的管理模式，僅能被動的找出敵意文章後，將其刪除，但其負面影響通常已產生，未來可研究敵意文章自動消除的機制，在敵意文章出現之初即由系統進行消除的動作，以減少負面影響。

# 參考文獻


01. 江玉婷 (民 89)。*中文資訊檢索測試集之設計與製作*。資訊傳播與圖書館學，6 卷，3 期，61-80。

02. 黃雲龍 (民 86)。*中文全文文件群集索引理論研究--向量空間模型(Vector-Space Model)的建構*。國立台灣大學商學研究所博士論文。

03. 曾元顯 (民 91)。*文件主題自動分類成效因素探討*。中國圖書館學會會報，68 期，62-83。

04. 楊允言 (民82)。*文件自動分類及其相似性排序*。國立清華大學資訊科學研究所碩士論文。

05. 楊允言、陳淑美、陳克健、謝清俊 (民88)。*中文文件自動分類之探討*。大漢學報第13期，241-256。

06. 謝清俊、林晰 (民86)。「*中央研究院古籍全文資料庫的發展概要*」，台北：中央研究院資訊科學研究所文獻處理實驗室技術報告，2-3。

07. Belkin. (1992). ***Information Filtering and Information Retrieval***. Communications of the ACM. 35(12).

08. Buss, A., J. Fischer, & A. Simmons. (1968). ***Aggression and hostility in psychiatric patients***. Journal of Consulting and Clinical Psychology 32: 21.

09. Cook, W., & Medley, D. (1954). ***Proposed hostility and pharisaic-virtue for the MMPI***. Journal of Applied Psychology, 38, 414-418.

10. Ellen M. Voorhees. (1998). ***Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness***. In Proceedings of the 21st ACM-SIGIR International Conference on Research and Development in Information Retrieval, Melbourne, 315-323.

11. Foo, S., & Li, H. (2001). ***Chinese Word Segmentation Accuracy and Its Effects on Information Retrieval***. TEXT Technology.

12. Fred Annexstein. (2002). ***Indexing and Representation:The Vector Space Model Retrieved***, December 25, 2003, from the World Wide Web: http://www.ececs.uc.edu/~annexste/Courses/cs690/Indexing%20and%20Representation.ppt

13. Jane Reid and Stefano Mizzaro. (1998). ***On the Consensus between Relevance Judges in a Multi-media Context***. In Proceeding of the 6th Mira Workshop, Dublin, October 20-30, http://www.dcs.gla.ac.uk/mira/workshops/dublin/procs/mr.pdf.

14. Jason D. M.Rennie, & Ryan Rifkin. (2001). ***Improving Multiclass Text Classification with the Support Vector Machine***, Massachusetts Institute of Technology. AI MemoAIM-2001-026. http://www.ai.mit.edu/~jrennie/papers/aimemo2001.ps.gz

15. Jhy-Jong Tsay, & Jing-Doo Wang. (2000). ***Improving Automatic Chinese Text Categorization by Error Correction***, Proceedings of the 5th International Workshop on Information Retrieval with Asian Languages, pp. 1-8.

16. Karan Sparck Jones, & C. J. van Rijsber-gen. (1976). ***Information Retrieval Test Collections***. Journal of Documentation 32 : 63-73.

17. Karan Sparck Jones. (1981). ***The Cranfield Tests***. In Information Retrieval Experiment. ed. Karan Sparck Jones London; Boston: Butteerworths.

18. Leah S. Larkey, & W. Bruce Croft. (1996). ***Combining Classifiers in Text Categorization***. Proceedings of the 19th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. pp. 289-297.

19. Marc Damashek. (1995). ***Gauging Similarity with N-grams: Language-Independent Categoriza-tion of Text***. Science 267， pp.843-848.

20. Pia Borlund, & Peter Ingwersen. (1997). ***The Development of a Method for the Evaluation of Interactive Information Retrieval Systems***. Journal of Documentation 53. no. 3: 226.

21. Reid, E. (1995). ***Virtual worlds：culture and imagination***. From Jones, S. G.（Ed.）. Cybersociey：Computer-Mediated Communication and Community, California：Sage Publications, Inc.

22. Ricardo, B.Y., & Berthier, R. N. (1999). ***Modern Information Retrieval***. Don Mills. New York: ACM PRESS.

23. Robert N. Oddy. (1981). ***Laboratory Tests: Automatic Systems***. In Information Retrieval Experiment. ed. Karan Sparck Jones. London: Boston, Butterworths. 161.

24. Salton, G., & Buckley, C. (1988). ***Term-weighting approaches in automatic retrieval***. Information Processing and Management. pp.323-328.

25. Salton, G., & Buckley, C. (1988). ***On the use of spreading activation methods in automatic information retrieval***. In Proceedings of the 11th International Conference on Research and Development in Information Retrieval, pp. 147-160.

26. Tefko Saracevic. (1975). ***Relevance: A Review of and a Framwork for the Thinking on the Notion in Information Science***. Jurnal of the American Society for Information Science 26. 341-342.

27. Thompsen, P. A. & Foulger, D. A. (1996). ***Effects of pictographs and quoting on flaming in Electronic mail***. Computers in Human Behavior, 12（2）pp.225-243.

28. Thorsten Joachims. (2001). ***A Statistical Learning Model of Text Classification for Support Vector Machines.*** Proceedings of the 23rd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp. 128-136.

29. Yiming Yang, and Xin Liu. (2001). ***A Study on Thresholding Strategies for Text Categorization***, Proceedings of the 23rd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval (2001)， pp. 137-145.

30. Yang, Y., & Liu, X. (1999). ***A Re-Examination of Text Categorization Methods***, Proceedings of the 22nd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp. 42-49.

31. Wai Lam, and Chao Yang Ho. (1998). ***Using a Generalized Instance Set for Automatic Text Categorization***, Proceedings of the 21st Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp. 81-89.

32. Robert W.P. Luk & K.L. Kwok (2002). ***A Comparison of Chinese Document Indexing Strategies and Retrieval Models.*** ACM Transactions on Asian Language Information Processing, Vol. 1, No. 3, pp. 225-268.

附錄一 出現次數大於十次之敵意詞 (1/11)

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 台灣 | 672 | 41 | 也不 | 156 | 81 | 主機 | 100 | 121 | 提供 | 75 | 161 | 們的 | 59 |
| 2 | 電信 | 663 | 42 | 也是 | 154 | 82 | 國外 | 98 | 122 | 麼多 | 73 | 162 | 當然 | 59 |
| 3 | 銘言 | 639 | 43 | 是不 | 154 | 83 | 很多 | 97 | 123 | 政府 | 73 | 163 | 了吧 | 59 |
| 4 | 引述 | 639 | 44 | 不知 | 154 | 84 | 機板 | 97 | 124 | 這種 | 71 | 164 | 上傳 | 58 |
| 5 | 之銘 | 632 | 45 | 他們 | 152 | 85 | 便宜 | 97 | 125 | 就不 | 71 | 165 | 費撥 | 58 |
| 6 | 中華 | 598 | 46 | 電腦 | 149 | 86 | 撥接 | 96 | 126 | 風之 | 71 | 166 | 第一 | 58 |
| 7 | 華電 | 519 | 47 | 作弊 | 149 | 87 | 還有 | 94 | 127 | 怎麼 | 71 | 167 | 的事 | 58 |
| 8 | 寬頻 | 409 | 48 | 應該 | 148 | 88 | 來就 | 94 | 128 | 國家 | 70 | 168 | 請問 | 58 |
| 9 | 路費 | 347 | 49 | 民營 | 146 | 89 | 免費 | 93 | 129 | 的網 | 70 | 169 | 帳號 | 58 |
| 10 | 可以 | 338 | 50 | 真的 | 145 | 90 | 比較 | 92 | 130 | 學風 | 69 | 170 | 號密 | 58 |
| 11 | 的寬 | 301 | 51 | 這樣 | 143 | 91 | 灣的 | 92 | 131 | 元智 | 69 | 171 | 時代 | 58 |
| 12 | 不是 | 295 | 52 | 不過 | 143 | 92 | 頻寬 | 91 | 132 | 智大 | 69 | 172 | 資訊 | 58 |
| 13 | 貴的 | 291 | 53 | 都是 | 138 | 93 | 這麼 | 90 | 133 | 之塔 | 69 | 173 | 不良 | 58 |
| 14 | 線路 | 290 | 54 | 一個 | 137 | 94 | 不要 | 90 | 134 | 在的 | 69 | 174 | 密碼 | 58 |
| 15 | 好貴 | 281 | 55 | 現在 | 136 | 95 | 技術 | 90 | 135 | 樣的 | 68 | 175 | 每個 | 57 |
| 16 | 灣好 | 278 | 56 | 不會 | 133 | 96 | 的話 | 89 | 136 | 是在 | 68 | 176 | 清楚 | 57 |
| 17 | 知道 | 264 | 57 | 只是 | 132 | 97 | 一家 | 89 | 137 | 是中 | 68 | 177 | 的硬 | 56 |
| 18 | 公司 | 262 | 58 | 提到 | 131 | 98 | 廠商 | 88 | 138 | 市場 | 67 | 178 | 申請 | 56 |
| 19 | 網路 | 257 | 59 | 那麼 | 131 | 99 | 使用 | 88 | 139 | 競爭 | 67 | 179 | 大的 | 56 |
| 20 | 電路 | 247 | 60 | 日本 | 130 | 100 | 的費 | 87 | 140 | 文章 | 67 | 180 | 也沒 | 56 |
| 21 | 言引 | 246 | 61 | 自己 | 129 | 101 | 出來 | 86 | 141 | 有的 | 67 | 181 | 的不 | 56 |
| 22 | 電話 | 235 | 62 | 國產 | 127 | 102 | 的線 | 85 | 142 | 全國 | 66 | 182 | 良牛 | 56 |
| 23 | 就是 | 230 | 63 | 的大 | 124 | 103 | 覺得 | 83 | 143 | 說的 | 65 | 183 | 一堆 | 55 |
| 24 | 如果 | 224 | 64 | 所以 | 123 | 104 | 東西 | 83 | 144 | 我想 | 65 | 184 | 然後 | 55 |
| 25 | 述之 | 222 | 65 | 可能 | 122 | 105 | 是一 | 82 | 145 | 你的 | 65 | 185 | 被抓 | 55 |
| 26 | 什麼 | 214 | 66 | 大學 | 120 | 106 | 一下 | 82 | 146 | 他家 | 64 | 186 | 之前 | 55 |
| 27 | 因爲 | 206 | 67 | 價格 | 120 | 107 | 成本 | 82 | 147 | 設備 | 64 | 187 | 可是 | 55 |
| 28 | 業者 | 204 | 68 | 降價 | 120 | 108 | 馬達 | 81 | 148 | 的錢 | 64 | 188 | 這是 | 54 |
| 29 | 其他 | 200 | 69 | 中提 | 118 | 109 | 個月 | 81 | 149 | 信的 | 64 | 189 | 大大 | 54 |
| 30 | 沒有 | 195 | 70 | 的電 | 117 | 110 | 用的 | 81 | 150 | 香港 | 63 | 190 | 雖然 | 54 |
| 31 | 頻引 | 194 | 71 | 這個 | 113 | 111 | 作中 | 81 | 151 | 只要 | 62 | 191 | 的時 | 54 |
| 32 | 的是 | 188 | 72 | 不用 | 110 | 112 | 大作 | 81 | 152 | 來的 | 61 | 192 | 美國 | 53 |
| 33 | 華碩 | 185 | 73 | 不能 | 110 | 113 | 一定 | 80 | 153 | 多少 | 61 | 193 | 太多 | 53 |
| 34 | 微星 | 175 | 74 | 一樣 | 109 | 114 | 他的 | 80 | 154 | 來源 | 61 | 194 | 是用 | 53 |
| 35 | 還是 | 174 | 75 | 我們 | 107 | 115 | 討論 | 80 | 155 | 市話 | 61 | 195 | 的嗎 | 53 |
| 36 | 費用 | 166 | 76 | 硬碟 | 104 | 116 | 費的 | 78 | 156 | 那個 | 60 | 196 | 這些 | 53 |
| 37 | 大家 | 163 | 77 | 地方 | 104 | 117 | 根本 | 77 | 157 | 是要 | 60 | 197 | 國際 | 53 |
| 38 | 但是 | 161 | 78 | 的人 | 104 | 118 | 起來 | 76 | 158 | 好像 | 60 | 198 | 已經 | 53 |
| 39 | 固網 | 161 | 79 | 只有 | 101 | 119 | 而且 | 75 | 159 | 那些 | 59 | 199 | 難道 | 53 |
| 40 | 問題 | 156 | 80 | 本來 | 101 | 120 | 開放 | 75 | 160 | 顯示 | 59 | 200 | 話線 | 53 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 201 | 的地 | 52 | 241 | 弊被 | 47 | 281 | 國庫 | 42 | 321 | 的廠 | 37 | 361 | 的東 | 34 |
| 202 | 爲了 | 52 | 242 | 腦可 | 47 | 282 | 東森 | 42 | 322 | 一起 | 37 | 362 | 外國 | 34 |
| 203 | 電機 | 52 | 243 | 到引 | 47 | 283 | 家電 | 41 | 323 | 都沒 | 37 | 363 | 總局 | 34 |
| 204 | 不可 | 52 | 244 | 不同 | 47 | 284 | 多的 | 41 | 324 | 公共 | 37 | 364 | 哪裡 | 34 |
| 205 | 好了 | 52 | 245 | 心站 | 46 | 285 | 情況 | 41 | 325 | 速度 | 36 | 365 | 所有 | 34 |
| 206 | 感謝 | 51 | 246 | 原電 | 46 | 286 | 消費 | 41 | 326 | 能力 | 36 | 366 | 話費 | 34 |
| 207 | 家的 | 51 | 247 | 測試 | 46 | 287 | 哥大 | 41 | 327 | 十年 | 36 | 367 | 進步 | 34 |
| 208 | 本的 | 51 | 248 | 機心 | 46 | 288 | 上面 | 40 | 328 | 硬引 | 36 | 368 | 信總 | 34 |
| 209 | 手機 | 51 | 249 | 或是 | 46 | 289 | 機不 | 40 | 329 | 的就 | 36 | 369 | 分享 | 33 |
| 210 | 都有 | 51 | 250 | 的文 | 45 | 290 | 一些 | 40 | 330 | 有他 | 36 | 370 | 就有 | 33 |
| 211 | 抓包 | 51 | 251 | 爲什 | 45 | 291 | 服務 | 40 | 331 | 還不 | 36 | 371 | 太貴 | 33 |
| 212 | 有一 | 50 | 252 | 不多 | 45 | 292 | 用戶 | 40 | 332 | 的只 | 36 | 372 | 時候 | 33 |
| 213 | 言在 | 50 | 253 | 會有 | 45 | 293 | 人家 | 40 | 333 | 是有 | 36 | 373 | 動態 | 33 |
| 214 | 包了 | 50 | 254 | 是台 | 45 | 294 | 是現 | 39 | 334 | 的價 | 36 | 374 | 我也 | 33 |
| 215 | 超頻 | 50 | 255 | 鐵路 | 44 | 295 | 我的 | 39 | 335 | 了我 | 36 | 375 | 要買 | 33 |
| 216 | 大哥 | 50 | 256 | 不到 | 44 | 296 | 物價 | 39 | 336 | 想降 | 36 | 376 | 是微 | 33 |
| 217 | 需要 | 50 | 257 | 有人 | 44 | 297 | 的我 | 39 | 337 | 是那 | 36 | 377 | 有台 | 33 |
| 218 | 與繁 | 50 | 258 | 原來 | 44 | 298 | 是說 | 39 | 338 | 用錢 | 36 | 378 | 該不 | 33 |
| 219 | 頻在 | 50 | 259 | 的一 | 44 | 299 | 沒錯 | 39 | 339 | 有線 | 36 | 379 | 國內 | 33 |
| 220 | 繁星 | 50 | 260 | 錢的 | 44 | 300 | 話帳 | 39 | 340 | 如何 | 36 | 380 | 是國 | 33 |
| 221 | 盈月 | 50 | 261 | 今天 | 44 | 301 | 接電 | 39 | 341 | 下載 | 35 | 381 | 場不 | 32 |
| 222 | 月與 | 50 | 262 | 反正 | 43 | 302 | 上的 | 39 | 342 | 是我 | 35 | 382 | 同學 | 32 |
| 223 | 能嗎 | 50 | 263 | 買了 | 43 | 303 | 多了 | 39 | 343 | 部分 | 35 | 383 | 事實 | 32 |
| 224 | 了一 | 49 | 264 | 的情 | 43 | 304 | 章中 | 39 | 344 | 謝謝 | 35 | 384 | 無法 | 32 |
| 225 | 都不 | 49 | 265 | 哈哈 | 43 | 305 | 除了 | 38 | 345 | 本上 | 35 | 385 | 是作 | 32 |
| 226 | 中原 | 49 | 266 | 信業 | 43 | 306 | 這不 | 38 | 346 | 是這 | 35 | 386 | 吧元 | 32 |
| 227 | 而已 | 49 | 267 | 一次 | 43 | 307 | 投資 | 38 | 347 | 不然 | 35 | 387 | 以爲 | 32 |
| 228 | 家都 | 49 | 268 | 營業 | 43 | 308 | 是你 | 38 | 348 | 甚至 | 35 | 388 | 是爲 | 32 |
| 229 | 路的 | 49 | 269 | 基本 | 43 | 309 | 本不 | 38 | 349 | 所謂 | 35 | 389 | 開始 | 32 |
| 230 | 一點 | 48 | 270 | 謝大 | 42 | 310 | 際電 | 38 | 350 | 外的 | 35 | 390 | 說不 | 32 |
| 231 | 賺錢 | 48 | 271 | 大陸 | 42 | 311 | 補助 | 37 | 351 | 晶片 | 35 | 391 | 星的 | 32 |
| 232 | 我是 | 48 | 272 | 看看 | 42 | 312 | 真是 | 37 | 352 | 嗎引 | 35 | 392 | 有這 | 32 |
| 233 | 機房 | 48 | 273 | 題的 | 42 | 313 | 話那 | 37 | 353 | 維護 | 35 | 393 | 正常 | 32 |
| 234 | 就可 | 48 | 274 | 以前 | 42 | 314 | 是只 | 37 | 354 | 總不 | 34 | 394 | 還要 | 32 |
| 235 | 產電 | 48 | 275 | 差不 | 42 | 315 | 年前 | 37 | 355 | 時間 | 34 | 395 | 的頻 | 32 |
| 236 | 效能 | 47 | 276 | 的還 | 42 | 316 | 看到 | 37 | 356 | 一直 | 34 | 396 | 交大 | 32 |
| 237 | 是因 | 47 | 277 | 道要 | 42 | 317 | 了引 | 37 | 357 | 世界 | 34 | 397 | 當初 | 32 |
| 238 | 板作 | 47 | 278 | 同一 | 42 | 318 | 費者 | 37 | 358 | 信不 | 34 | 398 | 到底 | 32 |
| 239 | 發現 | 47 | 279 | 來說 | 42 | 319 | 面的 | 37 | 359 | 暴利 | 34 | 399 | 隻好 | 32 |
| 240 | 星主 | 47 | 280 | 交通 | 42 | 320 | 明明 | 37 | 360 | 就算 | 34 | 400 | 港的 | 31 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 401 | 了不 | 31 | 441 | 不對 | 30 | 481 | 又在 | 29 | 521 | 我不 | 27 | 561 | 去了 | 26 |
| 402 | 火車 | 31 | 442 | 的有 | 30 | 482 | 用這 | 29 | 522 | 下都 | 27 | 562 | 這兩 | 25 |
| 403 | 所得 | 31 | 443 | 不懂 | 30 | 483 | 才知 | 28 | 523 | 核心 | 27 | 563 | 只能 | 25 |
| 404 | 不就 | 31 | 444 | 的設 | 30 | 484 | 一條 | 28 | 524 | 記得 | 27 | 564 | 自由 | 25 |
| 405 | 天下 | 31 | 445 | 利潤 | 30 | 485 | 差那 | 28 | 525 | 保證 | 27 | 565 | 不一 | 25 |
| 406 | 做的 | 31 | 446 | 民所 | 30 | 486 | 他業 | 28 | 526 | 價錢 | 27 | 566 | 家獨 | 25 |
| 407 | 哪一 | 31 | 447 | 而不 | 30 | 487 | 獨大 | 28 | 527 | 光是 | 27 | 567 | 維修 | 25 |
| 408 | 好嗎 | 31 | 448 | 下來 | 30 | 488 | 我同 | 28 | 528 | 費是 | 27 | 568 | 起步 | 25 |
| 409 | 月租 | 31 | 449 | 高頻 | 30 | 489 | 麼好 | 28 | 529 | 搞不 | 27 | 569 | 不好 | 25 |
| 410 | 另外 | 31 | 450 | 能全 | 30 | 490 | 也可 | 28 | 530 | 一張 | 27 | 570 | 要錢 | 25 |
| 411 | 至於 | 31 | 451 | 強手 | 30 | 491 | 關係 | 28 | 531 | 在台 | 27 | 571 | 最後 | 25 |
| 412 | 的問 | 31 | 452 | 隻超 | 30 | 492 | 有很 | 28 | 532 | 有那 | 27 | 572 | 你不 | 25 |
| 413 | 財團 | 31 | 453 | 路是 | 30 | 493 | 是跟 | 28 | 533 | 才有 | 27 | 573 | 的公 | 25 |
| 414 | 大部 | 31 | 454 | 牧場 | 29 | 494 | 意思 | 28 | 534 | 等於 | 27 | 574 | 說是 | 25 |
| 415 | 賺的 | 31 | 455 | 長痘 | 29 | 495 | 結果 | 28 | 535 | 台固 | 27 | 575 | 他說 | 25 |
| 416 | 用中 | 31 | 456 | 痘痘 | 29 | 496 | 得到 | 28 | 536 | 現有 | 27 | 576 | 下的 | 25 |
| 417 | 營固 | 31 | 457 | 他是 | 29 | 497 | 以做 | 28 | 537 | 態超 | 27 | 577 | 您的 | 25 |
| 418 | 錢也 | 31 | 458 | 期乳 | 29 | 498 | 有什 | 28 | 538 | 痘之 | 27 | 578 | 又不 | 25 |
| 419 | 說總 | 31 | 459 | 酸恐 | 29 | 499 | 牛免 | 28 | 539 | 比我 | 27 | 579 | 龍之 | 25 |
| 420 | 如此 | 31 | 460 | 乳酸 | 29 | 500 | 要是 | 28 | 540 | 雙向 | 26 | 580 | 到我 | 25 |
| 421 | 威盛 | 30 | 461 | 抱怨 | 29 | 501 | 會不 | 28 | 541 | 不得 | 26 | 581 | 現象 | 25 |
| 422 | 在長 | 30 | 462 | 理由 | 29 | 502 | 對對 | 28 | 542 | 連線 | 26 | 582 | 爲何 | 25 |
| 423 | 韓國 | 30 | 463 | 希望 | 29 | 503 | 要求 | 28 | 543 | 至少 | 26 | 583 | 全民 | 25 |
| 424 | 產品 | 30 | 464 | 場上 | 29 | 504 | 員工 | 28 | 544 | 國民 | 26 | 584 | 才是 | 25 |
| 425 | 是指 | 30 | 465 | 代表 | 29 | 505 | 到的 | 28 | 545 | 幾年 | 26 | 585 | 是公 | 25 |
| 426 | 不太 | 30 | 466 | 的也 | 29 | 506 | 大電 | 28 | 546 | 得比 | 26 | 586 | 也都 | 25 |
| 427 | 論的 | 30 | 467 | 己做 | 29 | 507 | 通通 | 28 | 547 | 維持 | 26 | 587 | 辦法 | 25 |
| 428 | 的那 | 30 | 468 | 裡面 | 29 | 508 | 以上 | 28 | 548 | 共財 | 26 | 588 | 行動 | 25 |
| 429 | 全買 | 30 | 469 | 就好 | 29 | 509 | 以不 | 28 | 549 | 的分 | 26 | 589 | 不清 | 25 |
| 430 | 三隻 | 30 | 470 | 是的 | 29 | 510 | 費那 | 28 | 550 | 空間 | 26 | 590 | 產的 | 24 |
| 431 | 好說 | 30 | 471 | 堆人 | 29 | 511 | 牛牧 | 28 | 551 | 其實 | 26 | 591 | 中山 | 24 |
| 432 | 超強 | 30 | 472 | 恐龍 | 29 | 512 | 人的 | 28 | 552 | 的國 | 26 | 592 | 可不 | 24 |
| 433 | 這三 | 30 | 473 | 是很 | 29 | 513 | 並不 | 28 | 553 | 多人 | 26 | 593 | 收電 | 24 |
| 434 | 買哪 | 30 | 474 | 上網 | 29 | 514 | 是電 | 28 | 554 | 全球 | 26 | 594 | 己的 | 24 |
| 435 | 哪隻 | 30 | 475 | 股票 | 29 | 515 | 的速 | 27 | 555 | 也要 | 26 | 595 | 之後 | 24 |
| 436 | 來都 | 30 | 476 | 是全 | 29 | 516 | 線的 | 27 | 556 | 的市 | 26 | 596 | 才會 | 24 |
| 437 | 前的 | 30 | 477 | 幾百 | 29 | 517 | 灣固 | 27 | 557 | 網公 | 26 | 597 | 主要 | 24 |
| 438 | 示卡 | 30 | 478 | 我家 | 29 | 518 | 租費 | 27 | 558 | 是其 | 26 | 598 | 網的 | 24 |
| 439 | 過期 | 30 | 479 | 的主 | 29 | 519 | 經營 | 27 | 559 | 科技 | 26 | 599 | 永遠 | 24 |
| 440 | 不想 | 30 | 480 | 部份 | 29 | 520 | 述又 | 27 | 560 | 那種 | 26 | 600 | 是外 | 24 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 601 | 別家 | 24 | 641 | 給民 | 23 | 681 | 國的 | 22 | 721 | 了那 | 21 | 761 | 自行 | 20 |
| 602 | 頻的 | 24 | 642 | 下去 | 23 | 682 | 合理 | 22 | 722 | 你可 | 21 | 762 | 了嗎 | 20 |
| 603 | 是否 | 24 | 643 | 的部 | 23 | 683 | 是假 | 22 | 723 | 其它 | 21 | 763 | 台幣 | 20 |
| 604 | 品的 | 24 | 644 | 家不 | 23 | 684 | 的成 | 22 | 724 | 假的 | 21 | 764 | 無恥 | 20 |
| 605 | 通部 | 24 | 645 | 要不 | 23 | 685 | 定要 | 22 | 725 | 出錢 | 21 | 765 | 是是 | 20 |
| 606 | 狀況 | 24 | 646 | 買的 | 23 | 686 | 發展 | 22 | 726 | 好的 | 21 | 766 | 不代 | 20 |
| 607 | 是如 | 24 | 647 | 站中 | 23 | 687 | 限制 | 22 | 727 | 不需 | 21 | 767 | 那一 | 20 |
| 608 | 表示 | 24 | 648 | 影響 | 23 | 688 | 把他 | 22 | 728 | 那大 | 21 | 768 | 嗎我 | 20 |
| 609 | 是真 | 24 | 649 | 是被 | 23 | 689 | 贈品 | 21 | 729 | 近來 | 20 | 769 | 的通 | 20 |
| 610 | 中站 | 24 | 650 | 找不 | 23 | 690 | 有不 | 21 | 730 | 個網 | 20 | 770 | 要多 | 20 |
| 611 | 路本 | 24 | 651 | 速的 | 22 | 691 | 營的 | 21 | 731 | 個電 | 20 | 771 | 要收 | 20 |
| 612 | 信公 | 24 | 652 | 是比 | 22 | 692 | 果是 | 21 | 732 | 大約 | 20 | 772 | 速率 | 20 |
| 613 | 不降 | 24 | 653 | 具有 | 22 | 693 | 裝機 | 21 | 733 | 到台 | 20 | 773 | 才對 | 20 |
| 614 | 大臣 | 23 | 654 | 的說 | 22 | 694 | 本狗 | 21 | 734 | 收費 | 20 | 774 | 會讓 | 20 |
| 615 | 真正 | 23 | 655 | 本是 | 22 | 695 | 會這 | 21 | 735 | 推推 | 20 | 775 | 一年 | 20 |
| 616 | 海軍 | 23 | 656 | 系列 | 22 | 696 | 發信 | 21 | 736 | 的台 | 20 | 776 | 多錢 | 20 |
| 617 | 後來 | 23 | 657 | 楓橋 | 22 | 697 | 完全 | 21 | 737 | 情形 | 20 | 777 | 是繳 | 20 |
| 618 | 軍大 | 23 | 658 | 言我 | 22 | 698 | 以自 | 21 | 738 | 隨便 | 20 | 778 | 信賺 | 20 |
| 619 | 同樣 | 23 | 659 | 橋驛 | 22 | 699 | 電子 | 21 | 739 | 普及 | 20 | 779 | 言這 | 20 |
| 620 | 生存 | 23 | 660 | 付錢 | 22 | 700 | 種東 | 21 | 740 | 讀取 | 20 | 780 | 說明 | 20 |
| 621 | 的這 | 23 | 661 | 驛站 | 22 | 701 | 過了 | 21 | 741 | 平均 | 20 | 781 | 罷了 | 20 |
| 622 | 比台 | 23 | 662 | 麼不 | 22 | 702 | 製造 | 21 | 742 | 股東 | 20 | 782 | 信局 | 20 |
| 623 | 文不 | 23 | 663 | 突然 | 22 | 703 | 就要 | 21 | 743 | 碟機 | 20 | 783 | 要看 | 20 |
| 624 | 福利 | 23 | 664 | 的但 | 22 | 704 | 做得 | 21 | 744 | 的其 | 20 | 784 | 奇摩 | 20 |
| 625 | 加的 | 23 | 665 | 烏龜 | 22 | 705 | 幾個 | 21 | 745 | 是第 | 20 | 785 | 被電 | 20 |
| 626 | 都被 | 23 | 666 | 不加 | 22 | 706 | 對的 | 21 | 746 | 心技 | 20 | 786 | 又是 | 20 |
| 627 | 跟你 | 23 | 667 | 該也 | 22 | 707 | 看不 | 21 | 747 | 了之 | 20 | 787 | 第二 | 20 |
| 628 | 是怎 | 23 | 668 | 變成 | 22 | 708 | 用就 | 21 | 748 | 哪家 | 20 | 788 | 收的 | 20 |
| 629 | 在中 | 23 | 669 | 在那 | 22 | 709 | 的意 | 21 | 749 | 述靠 | 20 | 789 | 的馬 | 20 |
| 630 | 者的 | 23 | 670 | 的動 | 22 | 710 | 靠場 | 21 | 750 | 早就 | 20 | 790 | 減資 | 19 |
| 631 | 程式 | 23 | 671 | 個地 | 22 | 711 | 上場 | 21 | 751 | 一台 | 20 | 791 | 同的 | 19 |
| 632 | 我覺 | 23 | 672 | 述過 | 22 | 712 | 專利 | 21 | 752 | 的核 | 20 | 792 | 家業 | 19 |
| 633 | 都會 | 23 | 673 | 在一 | 22 | 713 | 述我 | 21 | 753 | 摩大 | 20 | 793 | 多數 | 19 |
| 634 | 別人 | 23 | 674 | 大多 | 22 | 714 | 小烏 | 21 | 754 | 摩域 | 20 | 794 | 看中 | 19 |
| 635 | 網業 | 23 | 675 | 負責 | 22 | 715 | 是華 | 21 | 755 | 大摩 | 20 | 795 | 路價 | 19 |
| 636 | 謂的 | 23 | 676 | 每次 | 22 | 716 | 有些 | 21 | 756 | 場下 | 20 | 796 | 灣可 | 19 |
| 637 | 我用 | 23 | 677 | 花錢 | 22 | 717 | 亂講 | 21 | 757 | 目前 | 20 | 797 | 的沒 | 19 |
| 638 | 成績 | 23 | 678 | 的利 | 22 | 718 | 我看 | 21 | 758 | 市售 | 20 | 798 | 的上 | 19 |
| 639 | 訊站 | 23 | 679 | 壟斷 | 22 | 719 | 是每 | 21 | 759 | 送測 | 20 | 799 | 元的 | 19 |
| 640 | 外頻 | 23 | 680 | 得好 | 22 | 720 | 好不 | 21 | 760 | 在哪 | 20 | 800 | 的火 | 19 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 801 | 包含 | 19 | 841 | 地區 | 18 | 881 | 的確 | 18 | 921 | 聲損 | 17 | 961 | 費不 | 17 |
| 802 | 建設 | 19 | 842 | 的業 | 18 | 882 | 了但 | 18 | 922 | 包括 | 17 | 962 | 不少 | 17 |
| 803 | 參數 | 19 | 843 | 鍛鍊 | 18 | 883 | 了台 | 18 | 923 | 領先 | 17 | 963 | 得這 | 17 |
| 804 | 貴但 | 19 | 844 | 達到 | 18 | 884 | 收線 | 18 | 924 | 算是 | 17 | 964 | 在電 | 17 |
| 805 | 了只 | 19 | 845 | 的月 | 18 | 885 | 不只 | 18 | 925 | 售的 | 17 | 965 | 實際 | 17 |
| 806 | 所長 | 19 | 846 | 大概 | 18 | 886 | 鍊身 | 18 | 926 | 五萬 | 17 | 966 | 多年 | 17 |
| 807 | 的都 | 19 | 847 | 的看 | 18 | 887 | 身體 | 18 | 927 | 除非 | 17 | 967 | 要就 | 17 |
| 808 | 下一 | 19 | 848 | 灣也 | 18 | 888 | 業務 | 18 | 928 | 相當 | 17 | 968 | 費就 | 17 |
| 809 | 看的 | 19 | 849 | 信站 | 18 | 889 | 用他 | 18 | 929 | 給你 | 17 | 969 | 星歡 | 17 |
| 810 | 就知 | 19 | 850 | 都自 | 18 | 890 | 片的 | 18 | 930 | 跟我 | 17 | 970 | 臨參 | 17 |
| 811 | 有電 | 19 | 851 | 你在 | 18 | 891 | 接帳 | 18 | 931 | 而是 | 17 | 971 | 星提 | 17 |
| 812 | 更多 | 19 | 852 | 出的 | 18 | 892 | 看過 | 18 | 932 | 客戶 | 17 | 972 | 供您 | 17 |
| 813 | 有說 | 19 | 853 | 有點 | 18 | 893 | 電死 | 18 | 933 | 雷聲 | 17 | 973 | 您免 | 17 |
| 814 | 是沒 | 19 | 854 | 邊是 | 18 | 894 | 降到 | 18 | 934 | 到他 | 17 | 974 | 的關 | 17 |
| 815 | 容易 | 19 | 855 | 然跳 | 18 | 895 | 那是 | 18 | 935 | 心的 | 17 | 975 | 全省 | 17 |
| 816 | 說了 | 19 | 856 | 選擇 | 18 | 896 | 你是 | 18 | 936 | 宜這 | 17 | 976 | 觀盈 | 17 |
| 817 | 抗議 | 19 | 857 | 明大 | 18 | 897 | 規格 | 18 | 937 | 都說 | 17 | 977 | 通用 | 17 |
| 818 | 考慮 | 19 | 858 | 多久 | 18 | 898 | 撰寫 | 17 | 938 | 少錢 | 17 | 978 | 話的 | 17 |
| 819 | 卻不 | 19 | 859 | 剩下 | 18 | 899 | 損及 | 17 | 939 | 你有 | 17 | 979 | 路不 | 17 |
| 820 | 源交 | 19 | 860 | 機費 | 18 | 900 | 較貴 | 17 | 940 | 和國 | 17 | 980 | 省通 | 17 |
| 821 | 電控 | 19 | 861 | 人民 | 18 | 901 | 月的 | 17 | 941 | 詳細 | 17 | 981 | 碼電 | 17 |
| 822 | 台鐵 | 19 | 862 | 必須 | 18 | 902 | 郵件 | 17 | 942 | 的所 | 17 | 982 | 蒞臨 | 17 |
| 823 | 網站 | 19 | 863 | 的華 | 18 | 903 | 算的 | 17 | 943 | 取頭 | 17 | 983 | 內電 | 17 |
| 824 | 也有 | 19 | 864 | 華的 | 18 | 904 | 就比 | 17 | 944 | 三倍 | 17 | 984 | 作的 | 17 |
| 825 | 歡迎 | 19 | 865 | 有哪 | 18 | 905 | 市內 | 17 | 945 | 麼會 | 17 | 985 | 室內 | 17 |
| 826 | 過去 | 19 | 866 | 增加 | 18 | 906 | 全沒 | 17 | 946 | 貴就 | 17 | 986 | 不收 | 17 |
| 827 | 者不 | 19 | 867 | 然不 | 18 | 907 | 一段 | 17 | 947 | 在這 | 17 | 987 | 話全 | 17 |
| 828 | 本就 | 19 | 868 | 絕對 | 18 | 908 | 馬上 | 17 | 948 | 言不 | 17 | 988 | 參觀 | 17 |
| 829 | 要用 | 19 | 869 | 老實 | 18 | 909 | 講的 | 17 | 949 | 是想 | 17 | 989 | 迎蒞 | 17 |
| 830 | 是什 | 19 | 870 | 實在 | 18 | 910 | 還比 | 17 | 950 | 不值 | 17 | 990 | 的如 | 17 |
| 831 | 高速 | 19 | 871 | 成華 | 18 | 911 | 設定 | 17 | 951 | 各地 | 17 | 991 | 來看 | 17 |
| 832 | 到一 | 19 | 872 | 跳成 | 18 | 912 | 版子 | 17 | 952 | 一文 | 17 | 992 | 到你 | 17 |
| 833 | 願意 | 19 | 873 | 保護 | 18 | 913 | 沒人 | 17 | 953 | 成的 | 17 | 993 | 及電 | 17 |
| 834 | 是看 | 19 | 874 | 在其 | 18 | 914 | 來跑 | 17 | 954 | 了啦 | 17 | 994 | 吧不 | 17 |
| 835 | 硬要 | 19 | 875 | 說過 | 18 | 915 | 都在 | 17 | 955 | 一般 | 17 | 995 | 星是 | 17 |
| 836 | 的討 | 19 | 876 | 獨佔 | 18 | 916 | 製的 | 17 | 956 | 要把 | 17 | 996 | 有沒 | 17 |
| 837 | 大計 | 19 | 877 | 像是 | 18 | 917 | 研究 | 17 | 957 | 有誰 | 17 | 997 | 言因 | 16 |
| 838 | 計中 | 19 | 878 | 了中 | 18 | 918 | 是還 | 17 | 958 | 條線 | 17 | 998 | 看了 | 16 |
| 839 | 對阿 | 19 | 879 | 收取 | 18 | 919 | 怎樣 | 17 | 959 | 你家 | 17 | 999 | 吧光 | 16 |
| 840 | 輔英 | 18 | 880 | 了在 | 18 | 920 | 果中 | 17 | 960 | 小的 | 17 | 1000 | 他不 | 16 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1001 | 就完 | 16 | 1041 | 正的 | 16 | 1081 | 的專 | 16 | 1121 | 前中 | 15 | 1161 | 看你 | 15 |
| 1002 | 貴了 | 16 | 1042 | 點是 | 16 | 1082 | 電之 | 16 | 1122 | 路維 | 15 | 1162 | 令人 | 15 |
| 1003 | 資本 | 16 | 1043 | 重點 | 16 | 1083 | 力的 | 16 | 1123 | 方面 | 15 | 1163 | 機會 | 15 |
| 1004 | 那裡 | 16 | 1044 | 碩可 | 16 | 1084 | 一開 | 16 | 1124 | 然是 | 15 | 1164 | 果你 | 15 |
| 1005 | 算算 | 16 | 1045 | 房的 | 16 | 1085 | 我在 | 16 | 1125 | 讓中 | 15 | 1165 | 你用 | 15 |
| 1006 | 人一 | 16 | 1046 | 嗎不 | 16 | 1086 | 讓他 | 16 | 1126 | 不爽 | 15 | 1166 | 不該 | 15 |
| 1007 | 要調 | 16 | 1047 | 是民 | 16 | 1087 | 的要 | 16 | 1127 | 奇怪 | 15 | 1167 | 算不 | 15 |
| 1008 | 要有 | 16 | 1048 | 的的 | 16 | 1088 | 不起 | 16 | 1128 | 感覺 | 15 | 1168 | 的吧 | 15 |
| 1009 | 然知 | 16 | 1049 | 直接 | 16 | 1089 | 的機 | 16 | 1129 | 爲微 | 15 | 1169 | 格低 | 14 |
| 1010 | 是顯 | 16 | 1050 | 系統 | 16 | 1090 | 品牌 | 16 | 1130 | 就會 | 15 | 1170 | 來比 | 14 |
| 1011 | 大樓 | 16 | 1051 | 思是 | 16 | 1091 | 述鍛 | 16 | 1131 | 死了 | 15 | 1171 | 政策 | 14 |
| 1012 | 水準 | 16 | 1052 | 技嘉 | 16 | 1092 | 體之 | 16 | 1132 | 在本 | 15 | 1172 | 簡單 | 14 |
| 1013 | 因此 | 16 | 1053 | 想可 | 16 | 1093 | 跑了 | 15 | 1133 | 比不 | 15 | 1173 | 爲近 | 14 |
| 1014 | 不了 | 16 | 1054 | 公路 | 16 | 1094 | 人只 | 15 | 1134 | 灣人 | 15 | 1174 | 也只 | 14 |
| 1015 | 的資 | 16 | 1055 | 在收 | 16 | 1095 | 還便 | 15 | 1135 | 通過 | 15 | 1175 | 些香 | 14 |
| 1016 | 一種 | 16 | 1056 | 格的 | 16 | 1096 | 面前 | 15 | 1136 | 了你 | 15 | 1176 | 大資 | 14 |
| 1017 | 家競 | 16 | 1057 | 要的 | 16 | 1097 | 的連 | 15 | 1137 | 是兩 | 15 | 1177 | 論區 | 14 |
| 1018 | 爲大 | 16 | 1058 | 記憶 | 16 | 1098 | 的能 | 15 | 1138 | 是可 | 15 | 1178 | 狗面 | 14 |
| 1019 | 它們 | 16 | 1059 | 長途 | 16 | 1099 | 還蠻 | 15 | 1139 | 費中 | 15 | 1179 | 區才 | 14 |
| 1020 | 不成 | 16 | 1060 | 跟國 | 16 | 1100 | 是大 | 15 | 1140 | 也想 | 15 | 1180 | 佈告 | 14 |
| 1021 | 信自 | 16 | 1061 | 去問 | 16 | 1101 | 降電 | 15 | 1141 | 網頁 | 15 | 1181 | 道原 | 14 |
| 1022 | 以作 | 16 | 1062 | 要說 | 16 | 1102 | 路這 | 15 | 1142 | 搞清 | 15 | 1182 | 告欄 | 14 |
| 1023 | 些固 | 16 | 1063 | 學系 | 16 | 1103 | 斷線 | 15 | 1143 | 好啦 | 15 | 1183 | 來網 | 14 |
| 1024 | 左右 | 16 | 1064 | 提高 | 16 | 1104 | 的作 | 15 | 1144 | 殺人 | 15 | 1184 | 告訴 | 14 |
| 1025 | 動電 | 16 | 1065 | 嗎如 | 16 | 1105 | 外公 | 15 | 1145 | 百億 | 15 | 1185 | 多雖 | 14 |
| 1026 | 了幾 | 16 | 1066 | 有錢 | 16 | 1106 | 能夠 | 15 | 1146 | 就跟 | 15 | 1186 | 宜的 | 14 |
| 1027 | 家應 | 16 | 1067 | 的唷 | 16 | 1107 | 問的 | 15 | 1147 | 等民 | 15 | 1187 | 道台 | 14 |
| 1028 | 對手 | 16 | 1068 | 沒得 | 16 | 1108 | 家就 | 15 | 1148 | 用盈 | 15 | 1188 | 強者 | 14 |
| 1029 | 供的 | 16 | 1069 | 那你 | 16 | 1109 | 偶推 | 15 | 1149 | 人說 | 15 | 1189 | 在才 | 14 |
| 1030 | 也知 | 16 | 1070 | 調整 | 16 | 1110 | 是他 | 15 | 1150 | 好康 | 15 | 1190 | 是樂 | 14 |
| 1031 | 灣寬 | 16 | 1071 | 有華 | 16 | 1111 | 用力 | 15 | 1151 | 角度 | 15 | 1191 | 格差 | 14 |
| 1032 | 一切 | 16 | 1072 | 營公 | 16 | 1112 | 爲這 | 15 | 1152 | 之器 | 15 | 1192 | 以下 | 14 |
| 1033 | 麼貴 | 16 | 1073 | 通大 | 16 | 1113 | 謝感 | 15 | 1153 | 每一 | 15 | 1193 | 道差 | 14 |
| 1034 | 我只 | 16 | 1074 | 定會 | 16 | 1114 | 得是 | 15 | 1154 | 板子 | 15 | 1194 | 過他 | 14 |
| 1035 | 師大 | 16 | 1075 | 的好 | 16 | 1115 | 大分 | 15 | 1155 | 還好 | 15 | 1195 | 多香 | 14 |
| 1036 | 新聞 | 16 | 1076 | 有其 | 16 | 1116 | 得出 | 15 | 1156 | 你們 | 15 | 1196 | 的現 | 14 |
| 1037 | 品質 | 16 | 1077 | 好吧 | 16 | 1117 | 司的 | 15 | 1157 | 家有 | 15 | 1197 | 用雙 | 14 |
| 1038 | 一半 | 16 | 1078 | 台北 | 16 | 1118 | 偷加 | 15 | 1158 | 信也 | 15 | 1198 | 都作 | 14 |
| 1039 | 是了 | 16 | 1079 | 算還 | 16 | 1119 | 能怪 | 15 | 1159 | 想到 | 15 | 1199 | 向同 | 14 |
| 1040 | 腦的 | 16 | 1080 | 但也 | 16 | 1120 | 了是 | 15 | 1160 | 人不 | 15 | 1200 | 分和 | 14 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1201 | 同速 | 14 | 1241 | 是從 | 14 | 1281 | 分的 | 14 | 1321 | 述海 | 13 | 1361 | 論是 | 13 |
| 1202 | 喔我 | 14 | 1242 | 認爲 | 14 | 1282 | 個國 | 14 | 1322 | 只會 | 13 | 1362 | 這也 | 13 |
| 1203 | 信根 | 14 | 1243 | 我這 | 14 | 1283 | 他也 | 14 | 1323 | 用費 | 13 | 1363 | 說真 | 13 |
| 1204 | 的出 | 14 | 1244 | 我對 | 14 | 1284 | 路月 | 14 | 1324 | 錢給 | 13 | 1364 | 裝的 | 13 |
| 1205 | 是暴 | 14 | 1245 | 第四 | 14 | 1285 | 當時 | 14 | 1325 | 上下 | 13 | 1365 | 嗎還 | 13 |
| 1206 | 重要 | 14 | 1246 | 者我 | 14 | 1286 | 這裡 | 14 | 1326 | 會是 | 13 | 1366 | 貴是 | 13 |
| 1207 | 利吧 | 14 | 1247 | 們自 | 14 | 1287 | 是對 | 14 | 1327 | 錢建 | 13 | 1367 | 是價 | 13 |
| 1208 | 塊的 | 14 | 1248 | 運作 | 14 | 1288 | 商品 | 14 | 1328 | 的可 | 13 | 1368 | 安全 | 13 |
| 1209 | 月算 | 14 | 1249 | 等到 | 14 | 1289 | 有可 | 14 | 1329 | 到現 | 13 | 1369 | 算國 | 13 |
| 1210 | 上只 | 14 | 1250 | 缺點 | 14 | 1290 | 些人 | 14 | 1330 | 家公 | 13 | 1370 | 有問 | 13 |
| 1211 | 這中 | 14 | 1251 | 寫到 | 14 | 1291 | 我之 | 14 | 1331 | 冤有 | 13 | 1371 | 錯了 | 13 |
| 1212 | 道你 | 14 | 1252 | 多地 | 14 | 1292 | 不去 | 14 | 1332 | 要比 | 13 | 1372 | 說我 | 13 |
| 1213 | 比阿 | 14 | 1253 | 經濟 | 14 | 1293 | 收一 | 14 | 1333 | 有頭 | 13 | 1373 | 美麗 | 13 |
| 1214 | 那邊 | 14 | 1254 | 次都 | 14 | 1294 | 子佈 | 14 | 1334 | 接下 | 13 | 1374 | 山大 | 13 |
| 1215 | 的言 | 14 | 1255 | 之下 | 14 | 1295 | 太可 | 14 | 1335 | 頭債 | 13 | 1375 | 話是 | 13 |
| 1216 | 你去 | 14 | 1256 | 發表 | 14 | 1296 | 連國 | 14 | 1336 | 實質 | 13 | 1376 | 別說 | 13 |
| 1217 | 整參 | 14 | 1257 | 有用 | 14 | 1297 | 看起 | 14 | 1337 | 債有 | 13 | 1377 | 英文 | 13 |
| 1218 | 有多 | 14 | 1258 | 不管 | 14 | 1298 | 參考 | 14 | 1338 | 界第 | 13 | 1378 | 頻真 | 13 |
| 1219 | 做而 | 14 | 1259 | 爲的 | 14 | 1299 | 請不 | 14 | 1339 | 有主 | 13 | 1379 | 才能 | 13 |
| 1220 | 你說 | 14 | 1260 | 比起 | 14 | 1300 | 到中 | 14 | 1340 | 道在 | 13 | 1380 | 沒聽 | 13 |
| 1221 | 有同 | 14 | 1261 | 不肯 | 14 | 1301 | 屬於 | 14 | 1341 | 建的 | 13 | 1381 | 路也 | 13 |
| 1222 | 你怎 | 14 | 1262 | 用線 | 14 | 1302 | 的每 | 14 | 1342 | 麼大 | 13 | 1382 | 就已 | 13 |
| 1223 | 上我 | 14 | 1263 | 或者 | 14 | 1303 | 在市 | 14 | 1343 | 任何 | 13 | 1383 | 中文 | 13 |
| 1224 | 是主 | 14 | 1264 | 費才 | 14 | 1304 | 是網 | 14 | 1344 | 述雷 | 13 | 1384 | 該是 | 13 |
| 1225 | 子女 | 14 | 1265 | 也就 | 14 | 1305 | 我認 | 13 | 1345 | 四台 | 13 | 1385 | 主義 | 13 |
| 1226 | 是連 | 14 | 1266 | 外加 | 14 | 1306 | 的很 | 13 | 1346 | 掌握 | 13 | 1386 | 民族 | 13 |
| 1227 | 供給 | 14 | 1267 | 費一 | 14 | 1307 | 只需 | 13 | 1347 | 段時 | 13 | 1387 | 族主 | 13 |
| 1228 | 來是 | 14 | 1268 | 民間 | 14 | 1308 | 是英 | 13 | 1348 | 的你 | 13 | 1388 | 標題 | 13 |
| 1229 | 檔案 | 14 | 1269 | 少的 | 14 | 1309 | 者還 | 13 | 1349 | 還真 | 13 | 1389 | 就沒 | 13 |
| 1230 | 你這 | 14 | 1270 | 果真 | 14 | 1310 | 已通 | 13 | 1350 | 支持 | 13 | 1390 | 更高 | 13 |
| 1231 | 一倍 | 14 | 1271 | 到了 | 14 | 1311 | 寬服 | 13 | 1351 | 受到 | 13 | 1391 | 過路 | 13 |
| 1232 | 然有 | 14 | 1272 | 家還 | 14 | 1312 | 過認 | 13 | 1352 | 腦之 | 13 | 1392 | 過電 | 13 |
| 1233 | 也會 | 14 | 1273 | 法想 | 14 | 1313 | 只看 | 13 | 1353 | 帳目 | 13 | 1393 | 到這 | 13 |
| 1234 | 加得 | 14 | 1274 | 會把 | 14 | 1314 | 批踢 | 13 | 1354 | 要一 | 13 | 1394 | 聽過 | 13 |
| 1235 | 你也 | 14 | 1275 | 是靠 | 14 | 1315 | 格也 | 13 | 1355 | 的經 | 13 | 1395 | 那不 | 13 |
| 1236 | 批評 | 14 | 1276 | 光碟 | 14 | 1316 | 學美 | 13 | 1356 | 天使 | 13 | 1396 | 就看 | 13 |
| 1237 | 信所 | 14 | 1277 | 看一 | 14 | 1317 | 功能 | 13 | 1357 | 目列 | 13 | 1397 | 說法 | 13 |
| 1238 | 測的 | 14 | 1278 | 大附 | 14 | 1318 | 站批 | 13 | 1358 | 是好 | 13 | 1398 | 樣子 | 13 |
| 1239 | 鄉下 | 14 | 1279 | 附中 | 14 | 1319 | 型電 | 13 | 1359 | 先進 | 13 | 1399 | 還沒 | 13 |
| 1240 | 強的 | 14 | 1280 | 定的 | 14 | 1320 | 正他 | 13 | 1360 | 居然 | 13 | 1400 | 看板 | 13 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1401 | 想要 | 13 | 1441 | 發奇 | 12 | 1481 | 的固 | 12 | 1521 | 是最 | 12 | 1561 | 長一 | 12 |
| 1402 | 麼都 | 13 | 1442 | 到那 | 12 | 1482 | 先推 | 12 | 1522 | 麻煩 | 12 | 1562 | 材料 | 12 |
| 1403 | 有國 | 13 | 1443 | 奇想 | 12 | 1483 | 言發 | 12 | 1523 | 寫於 | 12 | 1563 | 教育 | 12 |
| 1404 | 比喻 | 13 | 1444 | 一項 | 12 | 1484 | 國人 | 12 | 1524 | 千分 | 12 | 1564 | 果然 | 12 |
| 1405 | 來那 | 13 | 1445 | 能組 | 12 | 1485 | 用韓 | 12 | 1525 | 於郵 | 12 | 1565 | 倍的 | 12 |
| 1406 | 去之 | 13 | 1446 | 專線 | 12 | 1486 | 本人 | 12 | 1526 | 你沒 | 12 | 1566 | 爭議 | 12 |
| 1407 | 方就 | 13 | 1447 | 組一 | 12 | 1487 | 幫幫 | 12 | 1527 | 爛了 | 12 | 1567 | 問一 | 12 |
| 1408 | 國有 | 13 | 1448 | 的技 | 12 | 1488 | 在不 | 12 | 1528 | 啊不 | 12 | 1568 | 這就 | 12 |
| 1409 | 得有 | 13 | 1449 | 台全 | 12 | 1489 | 比好 | 12 | 1529 | 路使 | 12 | 1569 | 費如 | 12 |
| 1410 | 後面 | 13 | 1450 | 是會 | 12 | 1490 | 發明 | 12 | 1530 | 幅度 | 12 | 1570 | 是哪 | 12 |
| 1411 | 你就 | 13 | 1451 | 產包 | 12 | 1491 | 不上 | 12 | 1531 | 下下 | 12 | 1571 | 你多 | 12 |
| 1412 | 是誰 | 13 | 1452 | 者也 | 12 | 1492 | 磁頭 | 12 | 1532 | 的狀 | 12 | 1572 | 局不 | 12 |
| 1413 | 星這 | 13 | 1453 | 括晶 | 12 | 1493 | 飆到 | 12 | 1533 | 難以 | 12 | 1573 | 原因 | 12 |
| 1414 | 好一 | 13 | 1454 | 吧我 | 12 | 1494 | 然要 | 12 | 1534 | 計算 | 12 | 1574 | 打死 | 12 |
| 1415 | 單位 | 13 | 1455 | 腦首 | 12 | 1495 | 韓貨 | 12 | 1535 | 者會 | 12 | 1575 | 連到 | 12 |
| 1416 | 後的 | 13 | 1456 | 好我 | 12 | 1496 | 下就 | 12 | 1536 | 倒是 | 12 | 1576 | 費也 | 12 |
| 1417 | 事情 | 13 | 1457 | 首先 | 12 | 1497 | 在外 | 12 | 1537 | 各位 | 12 | 1577 | 的呀 | 12 |
| 1418 | 的了 | 13 | 1458 | 用到 | 12 | 1498 | 欄系 | 12 | 1538 | 到不 | 12 | 1578 | 有關 | 12 |
| 1419 | 我有 | 13 | 1459 | 先硬 | 12 | 1499 | 路卡 | 12 | 1539 | 能用 | 12 | 1579 | 一公 | 12 |
| 1420 | 降的 | 13 | 1460 | 快一 | 12 | 1500 | 的貴 | 12 | 1540 | 從來 | 12 | 1580 | 最好 | 12 |
| 1421 | 業坊 | 13 | 1461 | 碟好 | 12 | 1501 | 有威 | 12 | 1541 | 另一 | 12 | 1581 | 是屬 | 12 |
| 1422 | 踢實 | 13 | 1462 | 不出 | 12 | 1502 | 了接 | 12 | 1542 | 小小 | 12 | 1582 | 最大 | 12 |
| 1423 | 實業 | 13 | 1463 | 像就 | 12 | 1503 | 矽統 | 12 | 1543 | 前我 | 12 | 1583 | 繳庫 | 12 |
| 1424 | 死你 | 13 | 1464 | 還會 | 12 | 1504 | 了有 | 12 | 1544 | 網咖 | 12 | 1584 | 訊科 | 12 |
| 1425 | 就夠 | 13 | 1465 | 就找 | 12 | 1505 | 定義 | 12 | 1545 | 的商 | 12 | 1585 | 光一 | 12 |
| 1426 | 多收 | 13 | 1466 | 爲他 | 12 | 1506 | 的效 | 12 | 1546 | 莫名 | 12 | 1586 | 個人 | 12 |
| 1427 | 跟日 | 13 | 1467 | 台製 | 12 | 1507 | 下我 | 12 | 1547 | 想像 | 12 | 1587 | 反而 | 12 |
| 1428 | 麗之 | 13 | 1468 | 碩送 | 12 | 1508 | 名目 | 12 | 1548 | 錢是 | 12 | 1588 | 熱死 | 12 |
| 1429 | 的版 | 13 | 1469 | 路頻 | 12 | 1509 | 後一 | 12 | 1549 | 吧台 | 12 | 1589 | 很久 | 12 |
| 1430 | 島已 | 13 | 1470 | 意外 | 12 | 1510 | 跟他 | 12 | 1550 | 懷疑 | 12 | 1590 | 跟電 | 12 |
| 1431 | 之島 | 13 | 1471 | 上可 | 12 | 1511 | 三度 | 12 | 1551 | 題是 | 12 | 1591 | 的而 | 12 |
| 1432 | 愛上 | 13 | 1472 | 問你 | 12 | 1512 | 了國 | 12 | 1552 | 了沒 | 12 | 1592 | 相關 | 12 |
| 1433 | 認證 | 13 | 1473 | 麼低 | 12 | 1513 | 度空 | 12 | 1553 | 由化 | 12 | 1593 | 說你 | 12 |
| 1434 | 前也 | 13 | 1474 | 話說 | 12 | 1514 | 比的 | 12 | 1554 | 照你 | 12 | 1594 | 回來 | 12 |
| 1435 | 的到 | 13 | 1475 | 不具 | 12 | 1515 | 臣之 | 12 | 1555 | 會受 | 12 | 1595 | 況下 | 12 |
| 1436 | 踢踢 | 13 | 1476 | 是以 | 12 | 1516 | 了還 | 12 | 1556 | 好成 | 12 | 1596 | 說一 | 12 |
| 1437 | 明顯 | 13 | 1477 | 很明 | 12 | 1517 | 能提 | 12 | 1557 | 造成 | 12 | 1597 | 以我 | 12 |
| 1438 | 本香 | 13 | 1478 | 關鍵 | 12 | 1518 | 了這 | 12 | 1558 | 價降 | 12 | 1598 | 篇文 | 12 |
| 1439 | 突發 | 12 | 1479 | 了他 | 12 | 1519 | 言你 | 12 | 1559 | 過我 | 12 | 1599 | 嘿嘿 | 12 |
| 1440 | 幾十 | 12 | 1480 | 不完 | 12 | 1520 | 的比 | 12 | 1560 | 在測 | 12 | 1600 | 上買 | 12 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1601 | 者是 | 12 | 1641 | 會被 | 11 | 1681 | 達就 | 11 | 1721 | 不開 | 11 | 1761 | 例子 | 11 |
| 1602 | 有權 | 12 | 1642 | 所能 | 11 | 1682 | 爭市 | 11 | 1722 | 每年 | 11 | 1762 | 家會 | 11 |
| 1603 | 用那 | 12 | 1643 | 的測 | 11 | 1683 | 夠電 | 11 | 1723 | 前面 | 11 | 1763 | 們有 | 11 |
| 1604 | 民眾 | 12 | 1644 | 寬根 | 11 | 1684 | 不容 | 11 | 1724 | 非常 | 11 | 1764 | 述阿 | 11 |
| 1605 | 要怪 | 12 | 1645 | 永恆 | 11 | 1685 | 蠢裝 | 11 | 1725 | 言對 | 11 | 1765 | 以一 | 11 |
| 1606 | 過這 | 12 | 1646 | 只如 | 11 | 1686 | 被上 | 11 | 1726 | 的道 | 11 | 1766 | 者要 | 11 |
| 1607 | 沒收 | 12 | 1647 | 明的 | 11 | 1687 | 一分 | 11 | 1727 | 快的 | 11 | 1767 | 訊息 | 11 |
| 1608 | 某些 | 12 | 1648 | 此他 | 11 | 1688 | 此中 | 11 | 1728 | 司公 | 11 | 1768 | 家固 | 11 |
| 1609 | 上大 | 12 | 1649 | 確定 | 11 | 1689 | 況且 | 11 | 1729 | 誰不 | 11 | 1769 | 開的 | 11 |
| 1610 | 格是 | 12 | 1650 | 說以 | 11 | 1690 | 低於 | 11 | 1730 | 公家 | 11 | 1770 | 再說 | 11 |
| 1611 | 次愛 | 12 | 1651 | 子的 | 11 | 1691 | 術是 | 11 | 1731 | 是收 | 11 | 1771 | 繳國 | 11 |
| 1612 | 了如 | 12 | 1652 | 備只 | 11 | 1692 | 他國 | 11 | 1732 | 頻一 | 11 | 1772 | 果用 | 11 |
| 1613 | 內的 | 12 | 1653 | 想說 | 11 | 1693 | 不算 | 11 | 1733 | 在是 | 11 | 1773 | 回國 | 11 |
| 1614 | 且價 | 11 | 1654 | 以現 | 11 | 1694 | 線費 | 11 | 1734 | 道路 | 11 | 1774 | 錢大 | 11 |
| 1615 | 雲的 | 11 | 1655 | 說反 | 11 | 1695 | 青雲 | 11 | 1735 | 碼來 | 11 | 1775 | 對外 | 11 |
| 1616 | 長進 | 11 | 1656 | 則所 | 11 | 1696 | 都要 | 11 | 1736 | 鐵的 | 11 | 1776 | 及率 | 11 |
| 1617 | 要算 | 11 | 1657 | 頭技 | 11 | 1697 | 麼問 | 11 | 1737 | 定是 | 11 | 1777 | 持有 | 11 |
| 1618 | 所所 | 11 | 1658 | 數則 | 11 | 1698 | 中的 | 11 | 1738 | 路線 | 11 | 1778 | 的物 | 11 |
| 1619 | 給嘔 | 11 | 1659 | 到第 | 11 | 1699 | 我說 | 11 | 1739 | 企業 | 11 | 1779 | 速博 | 11 |
| 1620 | 怪中 | 11 | 1660 | 度馬 | 11 | 1700 | 術還 | 11 | 1740 | 萬多 | 11 | 1780 | 低的 | 11 |
| 1621 | 麼呆 | 11 | 1661 | 個硬 | 11 | 1701 | 畢竟 | 11 | 1741 | 要超 | 11 | 1781 | 租用 | 11 |
| 1622 | 信具 | 11 | 1662 | 以達 | 11 | 1702 | 創見 | 11 | 1742 | 證金 | 11 | 1782 | 山研 | 11 |
| 1623 | 前耍 | 11 | 1663 | 既然 | 11 | 1703 | 麼蠢 | 11 | 1743 | 是像 | 11 | 1783 | 讓人 | 11 |
| 1624 | 具我 | 11 | 1664 | 到爲 | 11 | 1704 | 碩技 | 11 | 1744 | 二十 | 11 | 1784 | 究所 | 11 |
| 1625 | 康訊 | 11 | 1665 | 後大 | 11 | 1705 | 呆給 | 11 | 1745 | 美日 | 11 | 1785 | 說寬 | 11 |
| 1626 | 識中 | 11 | 1666 | 不做 | 11 | 1706 | 多是 | 11 | 1746 | 不論 | 11 | 1786 | 比日 | 11 |
| 1627 | 一天 | 11 | 1667 | 你如 | 11 | 1707 | 嘔旁 | 11 | 1747 | 沒什 | 11 | 1787 | 的投 | 11 |
| 1628 | 信裡 | 11 | 1668 | 而價 | 11 | 1708 | 腦有 | 11 | 1748 | 董事 | 11 | 1788 | 低價 | 11 |
| 1629 | 酪洞 | 11 | 1669 | 得只 | 11 | 1709 | 分數 | 11 | 1749 | 了美 | 11 | 1789 | 就一 | 11 |
| 1630 | 認識 | 11 | 1670 | 格又 | 11 | 1710 | 來吧 | 11 | 1750 | 事長 | 11 | 1790 | 讓其 | 11 |
| 1631 | 話不 | 11 | 1671 | 閣下 | 11 | 1711 | 對啊 | 11 | 1751 | 道但 | 11 | 1791 | 再多 | 11 |
| 1632 | 發佈 | 11 | 1672 | 又那 | 11 | 1712 | 卡還 | 11 | 1752 | 信是 | 11 | 1792 | 商業 | 11 |
| 1633 | 過是 | 11 | 1673 | 買下 | 11 | 1713 | 事業 | 11 | 1753 | 學資 | 11 | 1793 | 後就 | 11 |
| 1634 | 佈門 | 11 | 1674 | 爲其 | 11 | 1714 | 的定 | 11 | 1754 | 爲民 | 11 | 1794 | 且網 | 11 |
| 1635 | 弊的 | 11 | 1675 | 了也 | 11 | 1715 | 速公 | 11 | 1755 | 科學 | 11 | 1795 | 要線 | 11 |
| 1636 | 門的 | 11 | 1676 | 假使 | 11 | 1716 | 在海 | 11 | 1756 | 並沒 | 11 | 1796 | 額外 | 11 |
| 1637 | 信一 | 11 | 1677 | 怪微 | 11 | 1717 | 板之 | 11 | 1757 | 美其 | 11 | 1797 | 息提 | 11 |
| 1638 | 人表 | 11 | 1678 | 使中 | 11 | 1718 | 臣的 | 11 | 1758 | 民國 | 11 | 1798 | 老闆 | 11 |
| 1639 | 你認 | 11 | 1679 | 來自 | 11 | 1719 | 有在 | 11 | 1759 | 天地 | 11 | 1799 | 其名 | 11 |
| 1640 | 示中 | 11 | 1680 | 衝擊 | 11 | 1720 | 雜誌 | 11 | 1760 | 吧這 | 11 | 1800 | 路上 | 11 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1801 | 沒看 | 11 | 1841 | 乳酪 | 11 | 1881 | 到嚴 | 10 | 1921 | 能低 | 10 | 1961 | 麗台 | 10 |
| 1802 | 正站 | 11 | 1842 | 香香 | 11 | 1882 | 做不 | 10 | 1922 | 麼做 | 10 | 1962 | 亂的 | 10 |
| 1803 | 旁邊 | 11 | 1843 | 線了 | 11 | 1883 | 嚴重 | 10 | 1923 | 爭廠 | 10 | 1963 | 做台 | 10 |
| 1804 | 忘了 | 11 | 1844 | 裝什 | 11 | 1884 | 了它 | 10 | 1924 | 得多 | 10 | 1964 | 被污 | 10 |
| 1805 | 老鼠 | 11 | 1845 | 站好 | 11 | 1885 | 重衝 | 10 | 1925 | 到市 | 10 | 1965 | 造出 | 10 |
| 1806 | 百萬 | 11 | 1846 | 的規 | 11 | 1886 | 以合 | 10 | 1926 | 頻網 | 10 | 1966 | 頻也 | 10 |
| 1807 | 解釋 | 11 | 1847 | 恥政 | 11 | 1887 | 司會 | 10 | 1927 | 於其 | 10 | 1967 | 用過 | 10 |
| 1808 | 佔的 | 11 | 1848 | 法令 | 11 | 1888 | 會倒 | 10 | 1928 | 錯的 | 10 | 1968 | 取暴 | 10 |
| 1809 | 啦之 | 11 | 1849 | 陸的 | 11 | 1889 | 信被 | 10 | 1929 | 竟然 | 10 | 1969 | 推偶 | 10 |
| 1810 | 洞電 | 11 | 1850 | 言那 | 11 | 1890 | 本沒 | 10 | 1930 | 言請 | 10 | 1970 | 誤會 | 10 |
| 1811 | 是下 | 11 | 1851 | 藉口 | 11 | 1891 | 面指 | 10 | 1931 | 機器 | 10 | 1971 | 讚啦 | 10 |
| 1812 | 沒辦 | 11 | 1852 | 是給 | 11 | 1892 | 己去 | 10 | 1932 | 嚮往 | 10 | 1972 | 站來 | 10 |
| 1813 | 期待 | 11 | 1853 | 個馬 | 11 | 1893 | 指示 | 10 | 1933 | 司才 | 10 | 1973 | 推用 | 10 |
| 1814 | 類的 | 11 | 1854 | 立正 | 11 | 1894 | 用是 | 10 | 1934 | 是接 | 10 | 1974 | 二條 | 10 |
| 1815 | 的香 | 11 | 1855 | 邊立 | 11 | 1895 | 示在 | 10 | 1935 | 就像 | 10 | 1975 | 力推 | 10 |
| 1816 | 香乳 | 11 | 1856 | 求不 | 11 | 1896 | 有辦 | 10 | 1936 | 對中 | 10 | 1976 | 則貴 | 10 |
| 1817 | 對他 | 11 | 1857 | 耍什 | 11 | 1897 | 者具 | 10 | 1937 | 言論 | 10 | 1977 | 下了 | 10 |
| 1818 | 言的 | 11 | 1858 | 度不 | 11 | 1898 | 是硬 | 10 | 1938 | 代的 | 10 | 1978 | 成大 | 10 |
| 1819 | 收了 | 11 | 1859 | 鋪設 | 10 | 1899 | 樣能 | 10 | 1939 | 太爛 | 10 | 1979 | 山之 | 10 |
| 1820 | 喜歡 | 11 | 1860 | 科大 | 10 | 1900 | 述嚮 | 10 | 1940 | 往箱 | 10 | 1980 | 管的 | 10 |
| 1821 | 壓不 | 11 | 1861 | 好賺 | 10 | 1901 | 力以 | 10 | 1941 | 的太 | 10 | 1981 | 顆粒 | 10 |
| 1822 | 下吧 | 11 | 1862 | 以那 | 10 | 1902 | 發言 | 10 | 1942 | 房端 | 10 | 1982 | 的小 | 10 |
| 1823 | 道不 | 11 | 1863 | 信降 | 10 | 1903 | 得事 | 10 | 1943 | 個就 | 10 | 1983 | 接時 | 10 |
| 1824 | 數據 | 11 | 1864 | 呀我 | 10 | 1904 | 各大 | 10 | 1944 | 得捨 | 10 | 1984 | 找到 | 10 |
| 1825 | 爲太 | 11 | 1865 | 倒不 | 10 | 1905 | 事出 | 10 | 1945 | 過中 | 10 | 1985 | 沒想 | 10 |
| 1826 | 權利 | 11 | 1866 | 下水 | 10 | 1906 | 箱根 | 10 | 1946 | 後又 | 10 | 1986 | 光光 | 10 |
| 1827 | 我沒 | 11 | 1867 | 眾的 | 10 | 1907 | 出更 | 10 | 1947 | 跑不 | 10 | 1987 | 快了 | 10 |
| 1828 | 在說 | 11 | 1868 | 用也 | 10 | 1908 | 事尚 | 10 | 1948 | 是明 | 10 | 1988 | 我倒 | 10 |
| 1829 | 有是 | 11 | 1869 | 率先 | 10 | 1909 | 寬到 | 10 | 1949 | 集團 | 10 | 1989 | 你還 | 10 |
| 1830 | 供免 | 11 | 1870 | 奇妙 | 10 | 1910 | 組件 | 10 | 1950 | 護做 | 10 | 1990 | 用者 | 10 |
| 1831 | 層面 | 11 | 1871 | 先開 | 10 | 1911 | 場中 | 10 | 1951 | 技大 | 10 | 1991 | 壞了 | 10 |
| 1832 | 誰就 | 11 | 1872 | 陣子 | 10 | 1912 | 全天 | 10 | 1952 | 取勝 | 10 | 1992 | 費只 | 10 |
| 1833 | 難不 | 11 | 1873 | 放高 | 10 | 1913 | 且就 | 10 | 1953 | 述倒 | 10 | 1993 | 新興 | 10 |
| 1834 | 跟中 | 11 | 1874 | 固定 | 10 | 1914 | 我錯 | 10 | 1954 | 弊都 | 10 | 1994 | 多還 | 10 |
| 1835 | 多不 | 11 | 1875 | 務很 | 10 | 1915 | 中而 | 10 | 1955 | 握在 | 10 | 1995 | 給公 | 10 |
| 1836 | 也說 | 11 | 1876 | 信去 | 10 | 1916 | 朋友 | 10 | 1956 | 有比 | 10 | 1996 | 但有 | 10 |
| 1837 | 的中 | 11 | 1877 | 信率 | 10 | 1917 | 有價 | 10 | 1957 | 恩雅 | 10 | 1997 | 司可 | 10 |
| 1838 | 了網 | 11 | 1878 | 界各 | 10 | 1918 | 無聊 | 10 | 1958 | 比是 | 10 | 1998 | 的民 | 10 |
| 1839 | 了就 | 11 | 1879 | 顯的 | 10 | 1919 | 就現 | 10 | 1959 | 雅麗 | 10 | 1999 | 旅遊 | 10 |
| 1840 | 鼠的 | 11 | 1880 | 全保 | 10 | 1920 | 一本 | 10 | 1960 | 樣他 | 10 | 2000 | 鋪好 | 10 |

| s.n | term | num | s.n | term | num | s.n | term | num | s.n | term | num |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2001 | 各項 | 10 | 2041 | 存到 | 10 | 2081 | 信上 | 10 | 2121 | 就把 | 10 |
| 2002 | 真搞 | 10 | 2042 | 國營 | 10 | 2082 | 的顯 | 10 | 2122 | 們不 | 10 |
| 2003 | 項費 | 10 | 2043 | 森寬 | 10 | 2083 | 大可 | 10 | 2123 | 靈之 | 10 |
| 2004 | 問他 | 10 | 2044 | 樂殺 | 10 | 2084 | 有大 | 10 | 2124 | 的保 | 10 |
| 2005 | 灣大 | 10 | 2045 | 分別 | 10 | 2085 | 要等 | 10 | 2125 | 多麼 | 10 |
| 2006 | 沒開 | 10 | 2046 | 灣沒 | 10 | 2086 | 回答 | 10 | 2126 | 是使 | 10 |
| 2007 | 錢都 | 10 | 2047 | 有自 | 10 | 2087 | 等其 | 10 | 2127 | 些錢 | 10 |
| 2008 | 話了 | 10 | 2048 | 的啊 | 10 | 2088 | 君子 | 10 | 2128 | 啊我 | 10 |
| 2009 | 的政 | 10 | 2049 | 讓你 | 10 | 2089 | 喪禮 | 10 | 2129 | 牽線 | 10 |
| 2010 | 要發 | 10 | 2050 | 收過 | 10 | 2090 | 的方 | 10 | 2130 | 你對 | 10 |
| 2011 | 會降 | 10 | 2051 | 比人 | 10 | 2091 | 方法 | 10 | 2131 | 年的 | 10 |
| 2012 | 好處 | 10 | 2052 | 府出 | 10 | 2092 | 後再 | 10 | 2132 | 低廉 | 10 |
| 2013 | 學費 | 10 | 2053 | 步比 | 10 | 2093 | 都用 | 10 | 2133 | 弊我 | 10 |
| 2014 | 大台 | 10 | 2054 | 年輕 | 10 | 2094 | 過不 | 10 | 2134 | 只剩 | 10 |
| 2015 | 資源 | 10 | 2055 | 那樣 | 10 | 2095 | 信電 | 10 | 2135 | 言沒 | 10 |
| 2016 | 憶體 | 10 | 2056 | 公尺 | 10 | 2096 | 定這 | 10 | 2136 | 財但 | 10 |
| 2017 | 台塑 | 10 | 2057 | 的服 | 10 | 2097 | 謀取 | 10 | 2137 | 今日 | 10 |
| 2018 | 麼我 | 10 | 2058 | 將軍 | 10 | 2098 | 述求 | 10 | 2138 | 在管 | 10 |
| 2019 | 工作 | 10 | 2059 | 要花 | 10 | 2099 | 是多 | 10 | 2139 | 電總 | 10 |
| 2020 | 樣都 | 10 | 2060 | 誰都 | 10 | 2100 | 我就 | 10 | 2140 | 想太 | 10 |
| 2021 | 個問 | 10 | 2061 | 別的 | 10 | 2101 | 一件 | 10 | 2141 | 嗎電 | 10 |
| 2022 | 的板 | 10 | 2062 | 不表 | 10 | 2102 | 自大 | 10 | 2142 | 他電 | 10 |
| 2023 | 月五 | 10 | 2063 | 兩家 | 10 | 2103 | 步的 | 10 | 2143 | 成你 | 10 |
| 2024 | 零組 | 10 | 2064 | 兵者 | 10 | 2104 | 錢這 | 10 | | | |
| 2025 | 三萬 | 10 | 2065 | 賺太 | 10 | 2105 | 本都 | 10 | | | |
| 2026 | 正確 | 10 | 2066 | 不祥 | 10 | 2106 | 想在 | 10 | | | |
| 2027 | 萬二 | 10 | 2067 | 網寬 | 10 | 2107 | 捨不 | 10 | | | |
| 2028 | 斷的 | 10 | 2068 | 祥之 | 10 | 2108 | 來不 | 10 | | | |
| 2029 | 機關 | 10 | 2069 | 視訊 | 10 | 2109 | 在用 | 10 | | | |
| 2030 | 份有 | 10 | 2070 | 以把 | 10 | 2110 | 推出 | 10 | | | |
| 2031 | 唯一 | 10 | 2071 | 議的 | 10 | 2111 | 修費 | 10 | | | |
| 2032 | 處之 | 10 | 2072 | 灣做 | 10 | 2112 | 聯強 | 10 | | | |
| 2033 | 題應 | 10 | 2073 | 都低 | 10 | 2113 | 無限 | 10 | | | |
| 2034 | 禮處 | 10 | 2074 | 年繳 | 10 | 2114 | 爲主 | 10 | | | |
| 2035 | 死光 | 10 | 2075 | 始就 | 10 | 2115 | 採購 | 10 | | | |
| 2036 | 軍居 | 10 | 2076 | 以解 | 10 | 2116 | 碟的 | 10 | | | |
| 2037 | 不讓 | 10 | 2077 | 路就 | 10 | 2117 | 的爲 | 10 | | | |
| 2038 | 以喪 | 10 | 2078 | 繳回 | 10 | 2118 | 實說 | 10 | | | |
| 2039 | 華降 | 10 | 2079 | 讓大 | 10 | 2119 | 來我 | 10 | | | |
| 2040 | 備是 | 10 | 2080 | 看來 | 10 | 2120 | 根之 | 10 | | | |

附錄二 系統執行畫面 (1/5)


圖 1 子功能列表

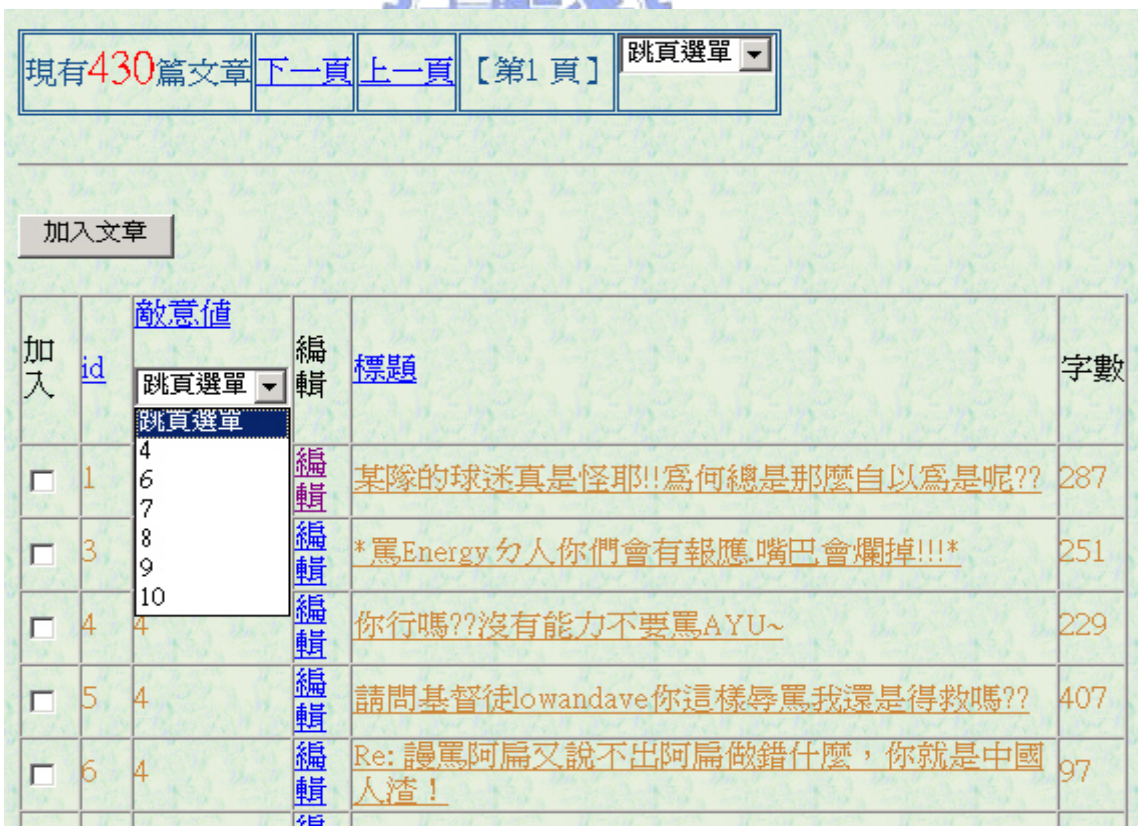


圖 2 手動選擇訓練文章

附錄二 系統執行畫面 (2/5)

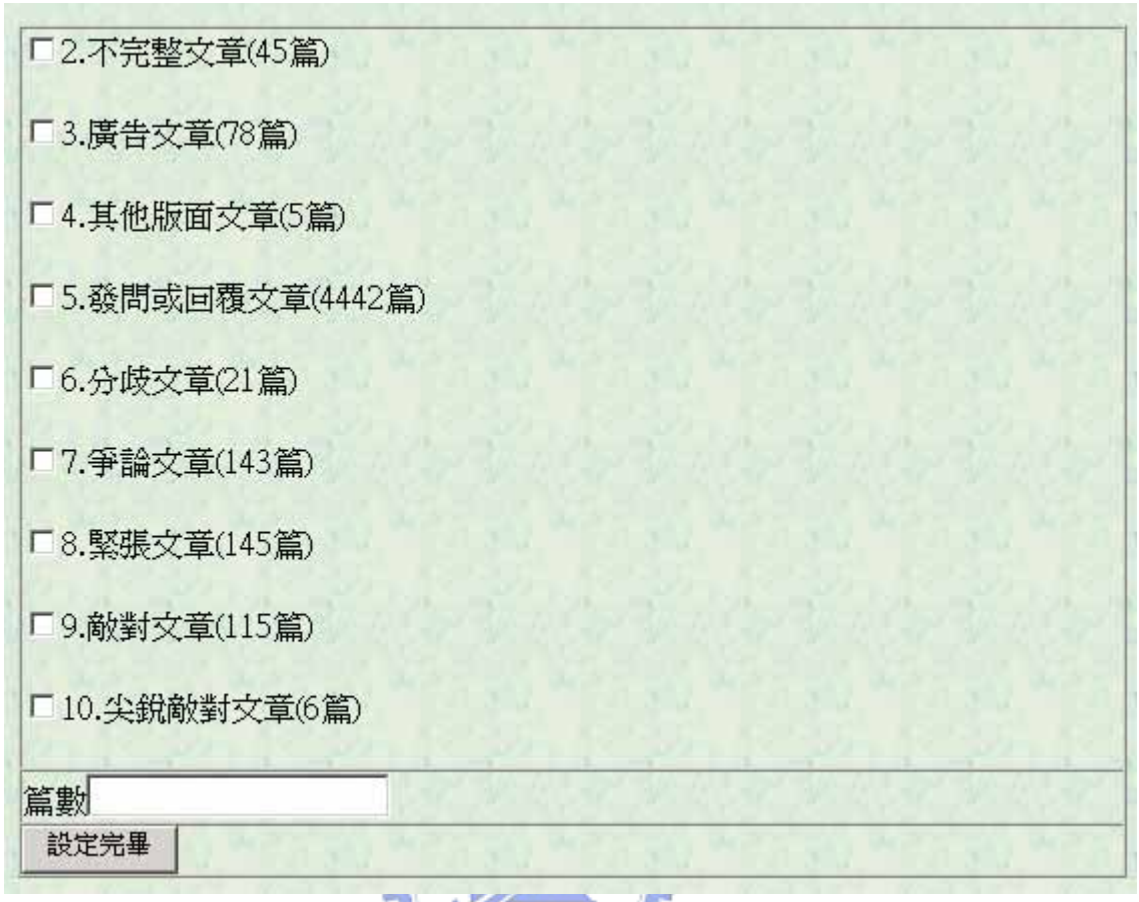


圖 3 隨機選擇訓練文章


圖 4 訓練文章斷詞

附錄二 系統執行畫面 (3/5)

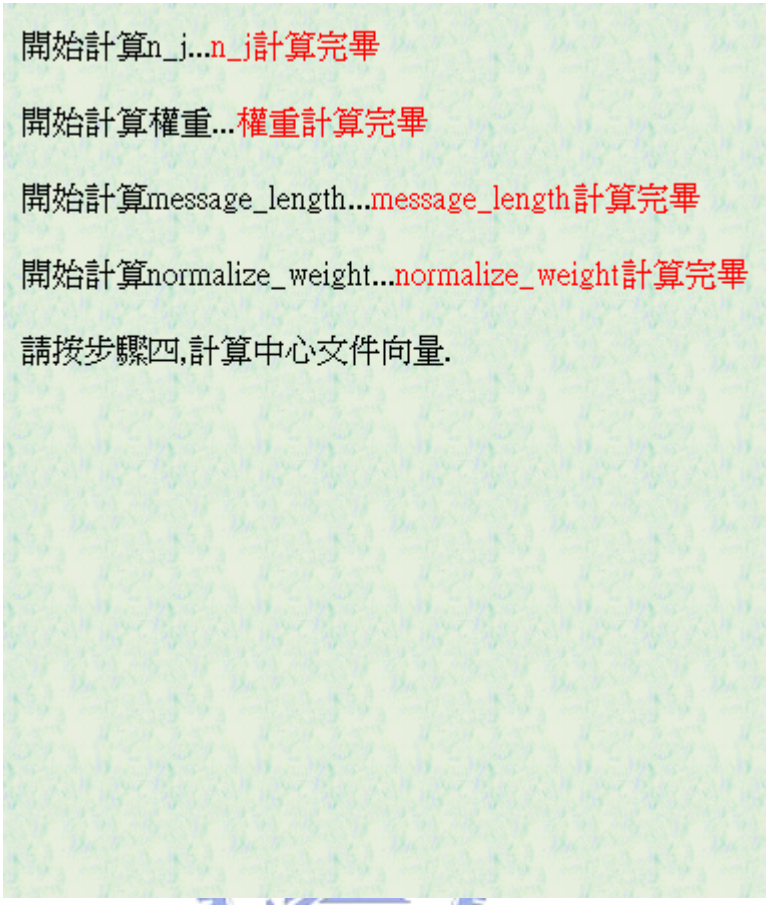


圖 5 計算訓練文章權重


圖 6 計算敵意文章中心向量(hcv)

附錄二 系統執行畫面 (4/5)

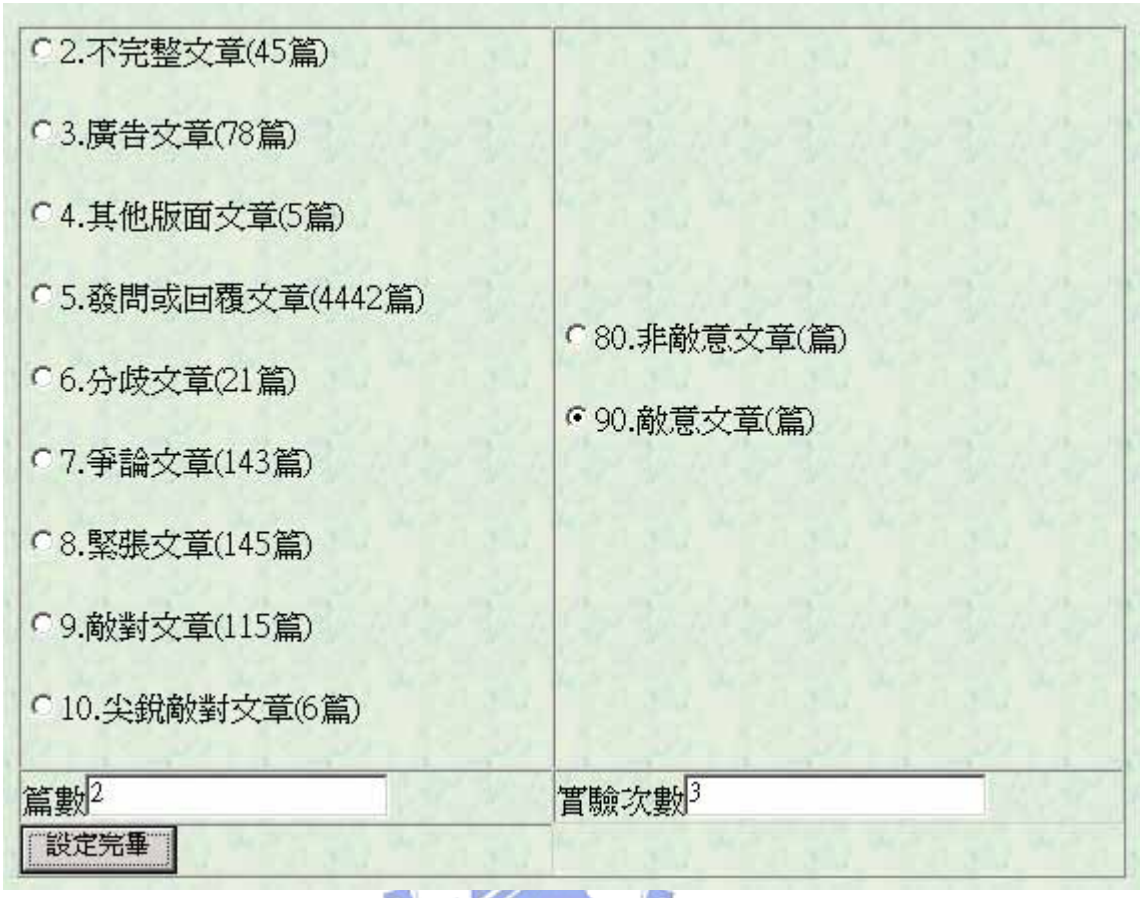


圖 7 隨機選擇實際文章

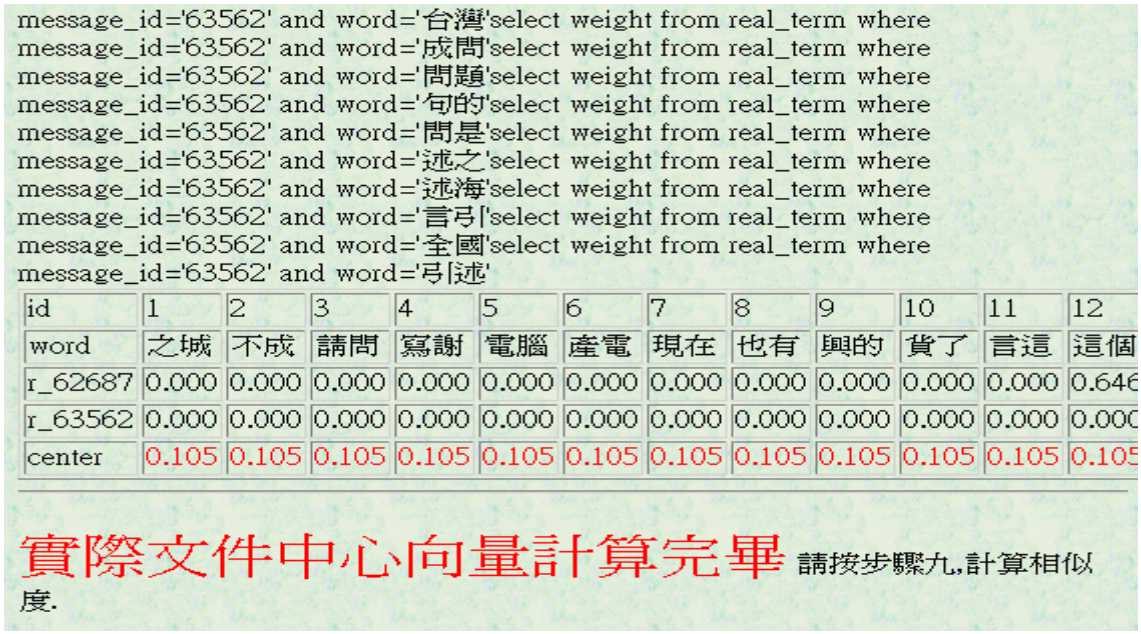


圖 8 計算實際文章向量

| id | msg_id | sn | 標題 | 字數 | 類別 | 相似度 | 電腦敵意 | 人工敵意 |
|---|---|---|---|---|---|---|---|---|
| 1 | 62687 | yes_263 | Re: 台灣....好貴的"寬頻">"< | 269 | 7 | 0.177575 | 1 | 1 |
| 2 | 63562 | yes_292 | Re: 台灣....好貴的"寬頻">"< | 66 | 7 | 0.339272 | 1 | 1 |

結束=13:21:29

圖 9 計算實際文章與 hcv 的相似度

附錄三 討論篇數大於十篇之主題 (1/2)

| 序號 | 群組 | 篇數 |
|---|---|---|
| 1 | Re: 台灣....好貴的"寬頻" >"< | 207 |
| 2 | Re: 記憶體多，有何好處？ | 59 |
| 3 | Re: 微星主機板作弊被抓包了.. | 53 |
| 4 | Re: 全國產電腦，可能嗎? | 48 |
| 5 | Re: 新買的 CPU 完蛋了.... | 40 |
| 6 | Re: [問題]HITACHI 的硬 | 40 |
| 7 | Re: 台中市哪裡賣電腦的便宜又好的?? | 38 |
| 8 | Re: 請大家幫我組台電腦,謝謝 | 29 |
| 9 | Re: 請問 so-net 的連線品質好嗎 | 27 |
| 10 | Re: 幫人組電腦 請告訴我這次組的合不合適 | 27 |
| 11 | Re: 請推薦 17 吋 LCD with DVI | 26 |
| 12 | Re: 請問 GF3 TI200 DDR64MB 這張顯示卡好嗎?? | 26 |
| 13 | Re: 請問 CPU 溫度 | 25 |
| 14 | Re: intel and AMD | 24 |
| 15 | Re: 硬碟也有分老大嗎 | 23 |
| 16 | Re: 關於處理大的影像檔之硬體升級疑問 | 22 |
| 17 | Re: 請推薦 amd 雙通道主機板 | 21 |
| 18 | Re: 請問哪一品牌的 Power 安靜又穩定的 | 21 |
| 19 | Re: 推個燒錄機吧... | 20 |
| 20 | Re: P3-800 雙 cpu、p4 誰快!! | 20 |
| 21 | Re: SiS748 | 20 |
| 22 | Re: [問題] 請問組裝電腦建議 XP 或 winMe/ | 20 |
| 23 | Re: 一般的顯示器可用多少年? | 19 |
| 24 | Re: 請問 TI200 和 9000PRO 那一塊比較好 | 18 |
| 25 | Re: CPU 若壞了開機後還會有文字嗎? | 18 |
| 26 | Re: [問題]各位組電腦的高手給我個建議吧~ | 18 |
| 27 | Re: 有人在用 centrino 的嗎 | 17 |
| 28 | Re: 給我一台 52X CD-ROM | 16 |
| 29 | Re: 高手看一下.. | 16 |
| 30 | Re: DDR 漲好多天了！到底何時會跌？ | 16 |
| 31 | Re: 自己裝主機會很難嗎... | 16 |
| 32 | Re: 請問 Ghost 在 Dos 底下的速度,有人比我高的嗎? | 15 |
| 33 | Re: 請問信件一直積在 Outlook 裡會不會造成問題呢？ | 15 |
| 34 | Re: 求助：將 moden 撥接設成無聲的指令 | 15 |
| 35 | Re: [測試] Liteon 燒錄機真的很耐操~ | 15 |
| 36 | Re: [問題]And xp2000+ 的溫度 | 15 |
| 37 | Re: 請問這麼組合好嗎? | 14 |
| 38 | Re: 推薦一下隨身碟吧～ | 14 |
| 39 | Re: 音效卡 用 2 年後掛了...算夠本了嘛? | 14 |
| 40 | Re: AMD 和 Intel 的浮點運算會差異很大嗎? | 14 |

附錄三 討論篇數大於十篇之主題 (2/2)

| 序號 | 群組 | 篇數 |
|---|---|---|
| 41 | Re: AMD Opteron 64 出了嗎?? | 14 |
| 42 | Re: 如何讓 3DMark2001 的分數破萬?? | 14 |
| 43 | Re: 請推薦一張適合我的顯示卡.... | 13 |
| 44 | Re: 請問 1.44 磁片.哪家的最好啊.!? | 13 |
| 45 | Re: 請問現在 SDRAM 和 DDR 哪個貴? | 13 |
| 46 | Re: 請問 35000 這樣配好咩? | 13 |
| 47 | Re: 請問 ATA100 跟 ATA133 效能會差很多嗎?? | 13 |
| 48 | Re: 請問怎麼更改顯示卡的頻率????? | 13 |
| 49 | Re: 請問 MX440SE 是不是比 MX440 效能差阿? | 13 |
| 50 | Re: 現在的顯示卡 | 13 |
| 51 | Re: 矽統子公司圖誠科技取得 Trident 繪圖晶片設計團?… | 13 |
| 52 | Re: [問題]請問一下 MX480E,9000 PRO,FX5200 那塊好 | 13 |
| 53 | Re: 請問 866 的 cpu 配上 ti4200 的顯示卡會太浪費嗎？ | 12 |
| 54 | Re: 請問以購買下哪一款燒錄機較好？ | 12 |
| 55 | Re: 請問我 xp1800 直接超到 166*11.5 會不會太冒險??? | 12 |
| 56 | Re: 19 吋 CRT 還是 17 吋 LCD? | 12 |
| 57 | Re: 用過 BenQ FP767 的人請進(很急..HELP~) | 12 |
| 58 | Re: 今天的地震有沒有人壞硬碟啊?! | 12 |
| 59 | Re: 三千元以內的顯示卡 | 12 |
| 60 | Re: [問題] 想組台電腦 請問大家的意見 | 12 |
| 61 | Re: 問個 LCD.. | 12 |
| 62 | Re:請推薦一台拋棄式的印表機！謝 | 11 |
| 63 | Re: SCSI 的春天在那裡? | 11 |
| 64 | Re: [問題]電腦輻射 | 11 |
| 65 | Re: 請推薦噴墨印表機... | 10 |
| 66 | Re: Asus P4P Deluxe 出貨了嗎?? | 10 |
| 67 | Re: 雙 cpu 的問題 | 10 |
| 68 | Re: BIOS 錯誤訊息...求救 | 10 |
| 69 | Re: 二千元到三千元 cpu+mb ?? | 10 |
| 70 | Re: Ti-4200 跑 3DMark2003 的分數應該多少? | 10 |

## 附錄四 台灣學術網路 BBS 站管理使用公約 (1/2)

86.04.22 第一次修定

BBS(Bulletin Board System)具有訊息交換、線上交談、問題解答、經驗交流等多項功能，舉凡校園資訊、圖書館服務、學術活動、交通資訊都盡在其中，爲學校學生之最愛，在台灣學術網路上甚爲流行，因此爲使網路資訊品質不流於浮濫，擬定以下規範做爲 BBS 站管理者及使用者遵守之依據。各學校應爲其 BBS 站負起督導責任，而各站管理者需能配合督導其站內使用品質。

一、 管理方面

(一) 各學校應盡告知本公約之義務，並應爲其 BBS 站等各類網路服務負起督導責任。

(二) 必須記錄遠端主機 ( remote host)及遠端使用者( remote username)以便追蹤問題來源。

(三) 版面名稱必須定義清楚俾利使用者選擇適合的討論區。

(四) 討論區之設立與刪除由各站自行決定辦法。

(五) 版主(Board Manager)之產生、任期、罷免或辭職等辦法由各站自行決定。

(六) 各站之管理人與相關版主須爲其版內之文章發佈做適切地選擇，促使使用者確實針對討論區主題參予討論，必要時得刪除不適切的文章並於適當時機說明理由。

(七) 除有完善管理能力之單位建議不要使用 BBSnet 的功能。

(八) 各單位依據本公約，自訂管理辦法，並提報學校或機關之權責單位核備後公佈之。

二、 使用方面

(一) 使用者不得使用他人帳號，並且只有註冊者才能張貼文章，使用者應爲自己所張貼的每一篇文章負責，並遵守下列五點要求：

禁止利用 BBS 做爲傳送或發表具威脅性、猥褻性、攻擊性的資料及文章。

禁止利用 BBS 做爲傳送未經各站之管理單位核准之商業性資料。

禁止利用 BBS 做爲傳送耗用大量傳送頻寬及儲存空間之資料。

禁止利用 BBS 做爲干擾或破壞網路上其他使用者或節點之硬軟體系統，例如散佈電腦病毒、嘗試侵入未經授權之電腦系統、或其他類似之情形者，皆在禁止範圍內。

避免在公眾討論區討論私人事務，發佈文章時，請尊重他人的權益及隱私。

(二) 註冊時，使用者必須註冊完全，必須告之〝真實姓名〞、〝地址〞與〝電子郵件地址 (e-mail address)〞，註冊不全或違規使用者，系統管理者(SYSOP)有權清除其帳號。

三、其他

附錄四 台灣學術網路 BBS 站管理使用公約 (1/2)

(一) 各站的使用者所公開發表之著作，如涉嫌侵害他人之權利時，自負民事與刑事責任，必要時各站可主動依法處理。

(二) 本公約之修訂需經台灣學術網路(TANet)管理委員會通過後施行。

## 附錄五 台灣學術網路使用規範 (1/1)

台灣學術網路之目的，係爲支援台灣地區學校及研究機構間之教學研究活動，以相互分享資源並相互提供合作機會。本使用規範主要敘述 TANet 資料傳輸使用之可接受性範圍，若資料傳輸跨越其它網路時，TANet 之使用者仍有義務遵守其它網路之使用規範。所有 TANet 使用者皆必須遵守及履行下列事項：

一、所有使用必須符合 TANet 之目的。

二、禁止使用 TANet 做爲傳送具威脅性的、猥褻性的、不友善性的資料。爲愛惜使用網路頻寬，未得 TANet 骨幹網路相關節點的合作允許，禁止大量傳送及登載與原設立目的不符的資訊。

三、商業性的合法資訊或軟體，若原創者或智慧財產權擁有者願意免費或優惠方式供 TANet 使用者使用，但必須由該節點之學校與資訊提供單位訂定相關合作事宜，方得放置於 TANet 之節點上，必要時得提 TANet 管理委員會協調處理。

四、禁止使用 TANet 做爲干擾或破壞網路上其它使用者或節點之硬軟體系統，此種干擾與破壞如散佈電腦病毒、嘗試侵入未經授權之電腦系統、或其它類似之情形者皆在禁止範圍內。

五、網路上所可存取到之任何資源，皆屬其擁有之個人或單位所有，除非已正式開放或已獲授權使用，否則 TANet 使用者禁止使用此等資源。

六、若使用目的與 TANet 相符，則直接支援該使用之相關資訊，亦在可接受範圍內，如校務行政資訊等。

附錄六 教育部校園網路使用規範 (1/3)

教育部 90 電創 184016 號　文中華民國 90 年 12 月 26 日核定

一、規範目的

爲充分發揮校園網路（以下簡稱網路）功能、普及尊重法治觀念，並提供網路使用者可資遵循之準據，以促進教育及學習，特訂定本規範。

二、網路規範與委員會

各校應參考本規範訂定網路使用規範，並視實際需要設置委員會或指定專人辦理下列事項：

（一）協助學校處理網路相關法律問題。

（二）採取適當之措施以維護網路安全。

（三）宣導網路使用之相關規範，並引導網路使用者正確使用資訊資源、重視網路相關法令及禮節。

（四）其他與網路有關之事項。

三、尊重智慧財產權

網路使用者應尊重智慧財產權。

學校應宣導網路使用者避免下列可能涉及侵害智慧財產權之行爲：

（一）使用未經授權之電腦程式。

（二）違法下載、拷貝受著作權法保護之著作。

（三）未經著作權人之同意，將受保護之著作上傳於公開之網站上。

（四）BBS 或其他線上討論區上之文章，經作者明示禁止轉載，而仍然任意轉載。

（五）架設網站供公眾違法下載受保護之著作。

（六）其他可能涉及侵害智慧財產權之行爲。

四、 禁止濫用網路系統

使用者不得爲下列行爲：

（一）散布電腦病毒或其他干擾或破壞系統機能之程式。

（二）擅自截取網路傳輸訊息。

（三）以破解、盜用或冒用他人帳號及密碼等方式，未經授權使用網路資源，或無故洩

漏他人之帳號及密碼。

（四）無故將帳號借予他人使用。

（五）隱藏帳號或使用虛假帳號。但經明確授權得匿名使用者不在此限。

（六）窺視他人之電子郵件或檔案。

（七）以任何方式濫用網路資源，包括以電子郵件大量傳送廣告信、連鎖信或無用之信息，或以灌爆信箱、掠奪資源等方式，影響系統之正常運作。

（八）以電子郵件、線上談話、電子佈告欄（BBS）或類似功能之方法散布詐欺、誹謗、侮辱、猥褻、騷擾、非法軟體交易或其他違法之訊息。

（九）利用學校之網路資源從事非教學研究等相關之活動或違法行爲。

五、 網路之管理

學校爲執行本規範之內容，其有關網路之管理事項如下：

（一）協助網路使用者建立自律機制。

（二）對網路流量應爲適當之區隔與管控。

（三）對於違反本規範或影響網路正常運作者，得暫停該使用者使用之權利。

（四）BBS 及其他網站應設置專人負責管理、維護。違反網站使用規則者，負責人得刪除其文章或暫停其使用。情節重大、違反校規或法令者，並應轉請學校處置。

（五）其他有關校園網路管理之事項。

使用者若發現系統安全有任何缺陷，應儘速報告網路管理單位。

六、 網路隱私權之保護

學校應尊重網路隱私權，不得任意窺視使用者之個人資料或有其他侵犯隱私權之行爲。但有下列情形之一者，不在此限：

（一）爲維護或檢查系統安全。

（二）依合理之根據，懷疑有違反校規之情事時，爲取得證據或調查不當行爲。

（三）爲配合司法機關之調查。

（四）其他依法令之行爲。

七、違反之效果

網路使用者違反本規範者，將受到下列之處分：

（一）停止使用網路資源。

（二）接受校規之處分。

網路管理者違反本規範者，應加重其處分。

依前兩項規定之處分者，其另有違法行爲時，行爲人尚應依民法、刑法、著作權法或其他相關法令負法律責任。

八、處理原則及程序

各校訂定之校園網路使用規範應明定於校規。

前項校規和網路管理單位對違反本規範之行爲人，或爲防範違反本規範，對行爲人或非特定對象所採取之各項管制措施，應符合必要原則、比例原則及法律保留原則。

各校對違反本規範之行爲人所爲之處分，應依正當法律程序，提供申訴和救濟機制。

學校處理相關網路申訴或救濟程序時，應徵詢校內網路委員會或指定專人之意見。