國立臺灣大學生物資源暨農學院農藝所生物統計學組

碩士論文

Division of Biometrics, Graduate Institute of Agronomy
College of Bioresources and Agriculture
National Taiwan University
Master Thesis

以文字探勘方法探討臺灣大學校務建言與回覆關聯性之研究

Applications of Text Mining to Studying the Association between Responses and Opinions from Opinion Web System of the National Taiwan University

廖子涵

Tzu-Han Liao

指導教授：劉仁沛 博士

Advisor：Jen-Pei Liu, Ph.D.

中華民國 105 年 6 月

June, 2016

doi:10.6342/NTU201603448

國立臺灣大學碩士學位論文
口試委員會審定書

以文字探勘方法探討臺灣大學校務建言與回覆關聯性之研究

Application of Text Mining to Studying the Association between Responses and Opinions from Opinion Web System of the National Taiwan University

本論文係 廖子涵 君（學號 R03621204）在國立臺灣大學農藝所生物統計學組完成之碩士學位論文，於民國 105 年 6 月 29 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

國立臺灣大學流行病學與預防醫學研究所教授

季瑋珠 博士 季瑋珠

國立臺灣大學語言學研究所副教授

謝舒凱 博士 謝舒凱

私立長庚大學臨床資訊與醫學統計研究中心助理教授

林志榮 博士 林志榮

國立臺灣大學農藝系(所)教授(指導教授)

劉仁沛 博士 劉仁沛

doi:10.6342/NTU201603448

誌謝

研究所的這兩年，首先感謝是我的指導教授-劉仁沛老師，不僅在我的課業上給予最專業的指導並灌輸大量新知識，也培養我自主學習的態度，更帶領我參加研討會及擔任統計學助教的機會，不時還會關心著我的近況，和老師在研究室中除了專業的統計話題外也能有閒話家常的時光，真的非常感謝老師。另外，非常感謝語言所的謝舒凱副教授，因著語言分析與資料科學這門通識課，而成為老師的學生，為了學習文字探勘這塊領域的知識及論文的研究，時常跑去老師的研究室和老師討論，謝謝老師總是在百忙之中抽空教導我，替我指點迷津，真的非常感謝老師。也非常感謝兩位口試委員季瑋珠教授及林志榮助理教授，感謝您們給予我的建議與指導，使我的論文更為完善。

謝謝這兩年來阿沛生統室的各位學長姐、同學及學弟妹們，不論在課業或是精神上都給予最大的幫忙與支持。感謝力維、冠霆、雅璇、思穎、欣宜學長姐，總是盡心盡力的從旁給予指導，在遇到問題時也提供許多建議，無微不至的照顧我這個學妹。感謝振豪同學，這兩年生活的陪伴，一同成長、互助及聊天，使得研究所生活中增添不少樂趣。感謝斯明、耘亘學弟妹，協助研究室的大小事宜，一起分擔共同事務。

最後，真的非常感謝我的父母親、妹妹、起豪及屬靈的家人心蘋、柔安等等，總是提醒我在忙碌之餘也要好好照顧身體健康，希望不要太辛苦累壞自己。對於自己投注在論文上的時間遠超過於陪伴家人的時間，真的深感愧疚，但未來回顧這些辛苦代價都會是值得的。我秉持著感恩、謙卑的心向身邊所有人學習，真的非常感謝每一個幫助過我的人，不論成果與否，至少在這過程中我都盡心盡力去學習，未來我也會保持虛心的態度，繼續努力!

doi:10.6342/NTU201603448

中文摘要

隨著網路資訊發達及行動通訊的重度使用發展趨勢，大眾們在網路上留下大量的數據，代表著大數據時代（Big Data）的來臨。根據國際數據資訊中心 IDC 公司統計，2020 年時全球總資料量將到達 40 Zettabyte（ZB），相當於約 43 兆 Gigabyte（GB），相較於 2010 年時超過五十倍的成長，且文字、圖片、視頻及音頻等非結構化資料的應用也會越來越頻繁。其中網路勢力崛起的鍵盤力量，讓文字儼然成為網路世界中大家溝通討論的媒介，重要性不可或缺。

因此，在現今社會中除了利用量化的資料進行分析外，質化的資料含有更大量的資訊，其分析的結果也更具備價值性。故本研究延續「國立台灣大學校務會議及校務建言資料之分析研究」量化性的研究結果，進一步對台大校務建言的內容進行質化的資料分析。

透過文字探勘（Text Mining）進行中文斷詞、潛在語意分析（Latent Semantic Analysis）及情緒分析（Sentiment Analysis），從眾多繁雜、尚未處理的文字中，找出被隱藏在字裡行間的重要資訊。來探討學生使用校務建言系統來表達意見，究竟為了什麼樣的溝通目的及需求；校方面對這樣的問題及建言，究竟如何回應學生，其是否有真正回答並處理問題，還是僅僅只是敷衍學生罷了。再者，兩方在溝通的過程中是否確實地落實真實的雙向「理性」溝通。

在校務建言的橋樑中，若能透過文字探勘的分析，從中探討雙方在溝通上的問題並給予建議，使學生更能以理性的思辨與態度，提出建言及問題；學校更能以積極的誠意與態度，處理及回應建言。讓行政單位和學生透過校務建言這個網路意見交流的平台，彼此間有良性的互動溝通，這將是使學校的運作有更好的發展。

關鍵字:資料採礦、文字探勘、詞頻統計、詞雲、中文斷詞、潛在語意分析、情緒分析及相關分析。

doi:10.6342/NTU201603448

English Abstract

Advanced network information and growing mobile communications have resulted in an increase of publicly available data on the internet. This indicates the arrival of Big Data. According to statistics from the International Data Corporation （IDC）, the overall volume of data worldwide will reach 40 Zettabytes（ZB）or, 43 trillion Gigabytes（GB）by 2020. This will generate a 50-fold growth from 2010. The frequent use of unstructured information such as text, images, video and audio will also become greater. Specifically, the rise of the power of keyboard has made text-based communication an essential channel for the public to discuss and exchange information online.

Aside from the commonly used quantitative analysis, qualitative data incorporates extensive information to provide additional value of analyzes. This study follows Yao's work: "Statistical Analysis of the Data from University Assembly Meetings and the Opinion Web System of the National Taiwan University" and utilizes its quantitative analysis results to further conduct qualitative analysis on the National Taiwan University's opinion web system.

This research aims to search for hidden information in complicated unprocessed text through text mining which involves text segmentation, latent semantic analysis and sentiment analysis. By using these approaches, it examines the issues that students used the system to express their opinions; and whether the responses from the university effectively and adequately responded and resolved these issues. The research also examines whether both sides have actually communicated in a rational manner.

To better the communication between students and the university on the opinion web system; this research used the technic of text mining to uncover the problems that occurred in the process of information exchange. It gives further recommendations for

doi:10.6342/NTU201603448

students to raise questions through rational and critical thinking and for the university to respond with a positive and genuine attitude. This can enhance the operations of the university and lead to a better development in the future.

Keyword: Data mining, Text mining, Word frequency, Word cloud, Text segmentation, Latent semantic analysis and sentiment analysis

doi:10.6342/NTU201603448

目錄

doi:10.6342/NTU201603448

doi:10.6342/NTU201603448

doi:10.6342/NTU201603448

表目錄

doi:10.6342/NTU201603448

圖目錄

doi:10.6342/NTU201603448

第一章 緒論

第一節　研究背景與動機

隨著網路資訊發達及行動通訊的重度使用發展趨勢，大眾們在網路上留下大量的數據，代表著大數據（Big Data）時代的來臨。根據國際數據資訊中心 IDC 公司統計[1]，在 2020 年時全球總資料量將到達 40 Zettabyte（ZB），相當於約 43 兆 Gigabyte（GB），相較於 2010 年時超過五十倍的成長，且文字、圖片、視頻及音頻等非結構化資料的應用也會越來越頻繁。其中網路勢力崛起的鍵盤力量，讓文字儼然成為網路世界中大家溝通討論的媒介，重要性不可或缺。

因此，在現今社會中除了利用量化的資料進行分析外，質化的資料含有更大量的資訊，透過文字探勘（Text Mining）與統計分析技術的融合，從眾多繁雜、尚未處理的文字中，找出被隱藏在字裡行間的重要資訊，這使得文字探勘成為近幾年的主流，其分析的結果更具備價值性。

本研究延續「國立台灣大學校務會議及校務建言資料之分析研究」量化性的研究結果[2]，進一步對台大校務建言的內容進行質化的資料分析。過去，學生向校方表達對於校園事務的意見或看法，多半是採取與校方面對面溝通、投書校園意見箱，或是由學生會代表代為傳達其需求。如今，因網路的盛行，校方建立公開討論平台－校務建言系統，提供大學內部一個行政、教學、研究和溝通的橋樑讓學生們可以更便利地透過網路來表達校務意見，使校方了解其需求。

而各行政單位管理者所要面對與處理的，是一個大量資訊快速流通的溝通媒介，回覆的速度與內容，關係著校園組織溝通的進行，也就是校方是否能充分利用

doi:10.6342/NTU201603448

電腦中介傳播的特性，來進行校園資訊的流通傳遞、或是回應同學關於校園事務的討論，使其成為一個雙向溝通、具備即時性、互動性的溝通管道。

在校務建言的橋樑中，若能透過文字探勘的分析，從中探討雙方在溝通上的問題並給予建議，使學生更能以理性的思辨與態度，提出建言及問題；學校更能以積極的誠意與態度，處理及回應建言。讓行政單位和學生透過校務建言這個網路意見交流的平台，彼此間有良性的互動溝通，這將是使學校的運作有更好的發展。

doi:10.6342/NTU201603448

第二節　研究目的

本研究利用西元2005至2012年台灣大學校務建言系統的文字資料，進行文字探勘（Text Mining）分析，透過中文斷詞、潛在語意分析（Latent Semantic Analysis）及情緒分析（Sentiment Analysis）。從中探討學生使用這樣的管道表達意見，究竟為了什麼樣的溝通目的及需求；校方面對這樣的問題及建言，究竟如何回應學生，其是否有真正回答並處理問題，還是僅僅只是敷衍學生罷了。再者，兩方在溝通的過程中是否確實地落實真實的雙向「理性」溝通。因此，本研究的研究目的可以整理成如下：

一、瞭解學生表達校務建言的動機、需求。

二、瞭解各行政單位於校務建言中常被建議的內容類型。

三、探討校方及學生的用字習慣及文章長度。

四、探討學生所表達的意見，校方是否有真實回答學生所需。

五、探討校方與學生兩者在溝通的過程中，兩方的情緒是否達到理性的溝通。

doi:10.6342/NTU201603448

第三節　研究架構

本篇論文共分為五個章節：第一章為緒論，說明本研究的背景與動機、目的及整體架構。第二章為文獻探討，針對研究議題及使用方法進行文獻探討，包含大數據、資料探勘、文字探勘及台灣大學校務建言之量化研究的相關介紹。第三章為研究方法，依照本研究之目的擬定研究流程，並說明其資料處理及分析方法。第四章為實證分析，說明本研究資料分析之結果。第五章為結論與建議，說明本研究得到的結論與建議後續研究者的方向。本研究架構圖，如圖 1-3-1 所示。

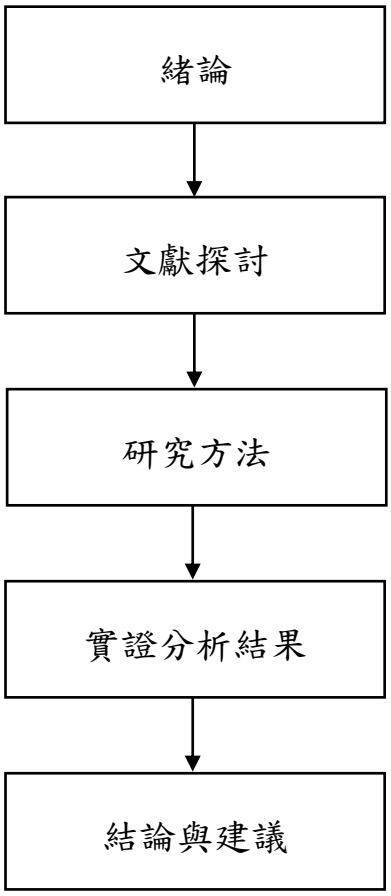


圖 1-3-1　研究架構圖

doi:10.6342/NTU201603448

第二章 文獻探討

第一節 大數據（Big Data）

隨著科技的發達，在數位化的世界裡，人們無時無刻都在產生著多元且大量的數據資料。根據國際數據資訊中心 IDC 公司估計[1]，在 2020 年時全球總資料量將到達 40 Zettabyte（ZB），相當於約 43 兆 Gigabyte（GB），相較於 2010 年時超過五十倍的成長，如圖 2-1-2 所示。若把這些資料全都裝在容量 128GB 的 iPad Air 平板電腦堆疊起來，高度可達 6.6 個地球至月亮的距離。

商業智慧分析平台 DOMO 公司的最新網路數據資料調查中[3]，更指出在每分鐘之內，Uber 可獲得 694 個訂單、蘋果用戶會下載 51 萬個應用程式、YouTube 會有 300 個小時的影片被上傳、Facebook 增加 400 萬個讚等，如圖 4-1-2 所示。這些都代表著大數據時代的來臨。

大數據在近年來被廣泛的熱烈討論，其又稱為「巨量資料」、「海量資料」。然而大數據至今沒有明確的定義，但大多數說法是「超過典型資料庫工具的硬體與軟體環境所能獲取、存儲、管理和分析能力者」，換句話說，「所謂的大數據，就是用現有的一般技術難以管理的大量資料。」。「用現有的一般技術難以管理」，指的是目前企業資料庫主流的關聯資料庫已無法管理結構複雜的資料；或是因為量的增加，導致查詢資料的反應時間超過容許範圍等等的龐大資料[4]。

一般而言，海量資料分析包含以下四大特性，簡稱 4V：

一. 巨量性（Volume）-存放數據量超過 PB

人類數據儲存量呈爆炸性成長，其以 PB～ZB 為儲存單位，如表 2-1-1 所示。

二. 即時性（Velocity）-數據擷取時間不到一秒

即時變動的流動資料（In-motion Data），表示這些數據產出快、變化也快，譬如在數據串流的環境下，數據不斷快速流入，而且還不斷更新變動，數據能夠被擷取而且被進一步應用的時間，甚至連一秒都不到，其反應時間僅短短幾秒至百萬分之一秒。

doi:10.6342/NTU201603448

三. 多樣性（Variety）-數據庫管理人員只處理了 20%的結構化數據

一直以來數據庫管理人員把大多數時間花在處理僅 20%格式整齊的結構化數據資料。然而，現今的資料種類繁雜，除了以前結構化資料外，其餘 80%以上的數據是存在於社交網路、物聯網（Internet of Things）屬於非結構、純文字、多媒體資料等，如圖 2-1-3 所示[5]。

四. 不確定性（Veracity）or 價值（Value）-全球有 80%數據不可靠

過去，企業是最主要的數據來源，企業通常會仔細查核內部的數據，故數據可靠度較高。但自 2010 年以來，在網路通訊、社群網站和感測器技術蓬勃發展下，破碎的、不完整的、不可靠的數據越來越多，甚至有分析師預估，到 2015 年時，在全球搜集的所有資訊中，將有超過八成屬於不確定可靠與否的資訊，如圖 2-1-4 所示[6]。數據可靠性若不高，採集到的數據價值也會受到影響，例如，消費廠商想從社交網站中找出消費者對其產品的喜好，但社群網站上充滿了使用假身分、發表假言論，以及任意轉貼網路謠言或過時資訊的使用者，若缺乏篩選和判斷的機制，就難以挖掘出真正有價值的資訊。真偽存疑、不確定的資料，因資料不完整、不一致、時間差、意義不明、蓄意欺騙而導致。

目前對於第 4V 是不確定性（Veracity），各界持不同看法，有些人認為第 4V 為價值（Value），表示若我們懂得妥善應用大數據，將可以從大數據資料中獲得極大的商業價值。

doi:10.6342/NTU201603448

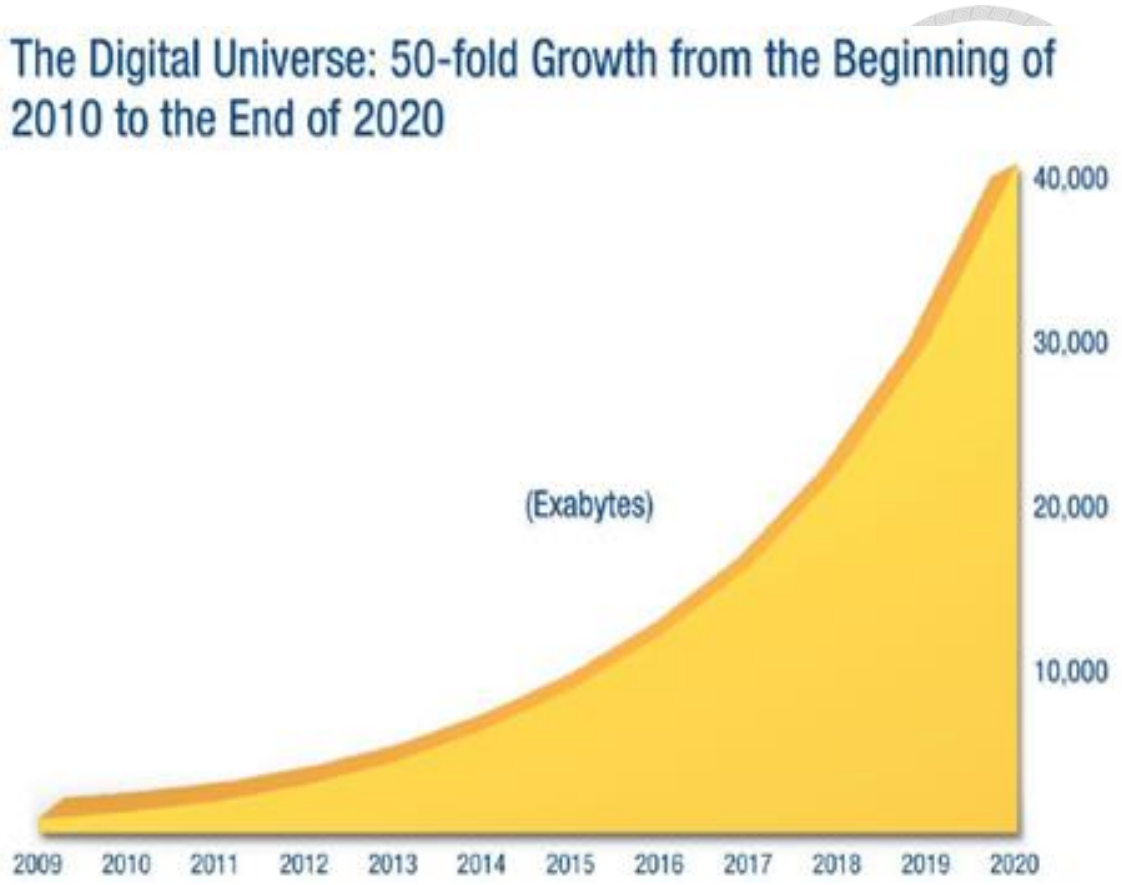


圖 2-1-1　IDC 全球資料量成長圖

doi:10.6342/NTU201603448

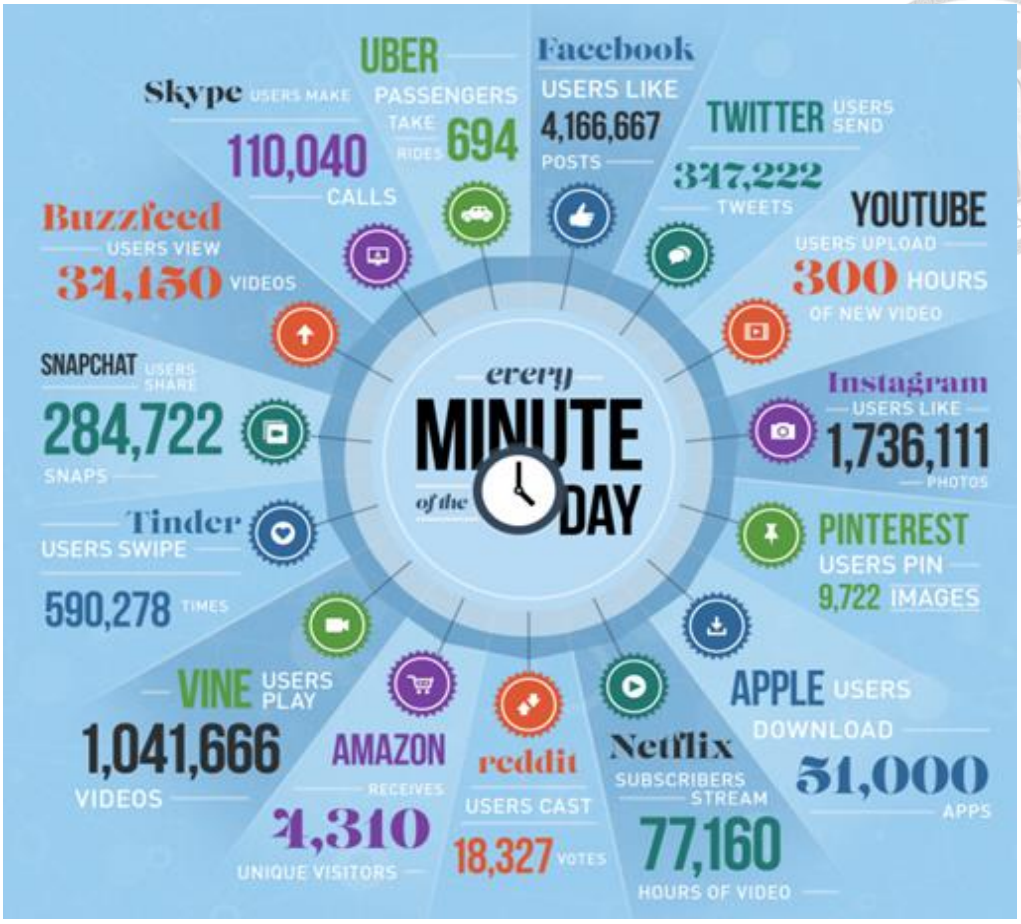


圖 2-1-2　DOMO 網路一分鐘資訊圖

doi:10.6342/NTU201603448

表 2-1-1　資料的儲存單位

儲存單位（英）		說明
Byte		檔案儲存容量的最小單位元
Kilobyte	(KB)	1024 Bytes
Megabyte	(MB)	1024 KB
Gigabyte	(GB)	1024 MB
Terabyte	(TB)	1024 GB
Petabyte	(PB)	1024 TB
Exabyte	(EB)	1024 PB
Zettabyte	(ZB)	1024 ZB

doi:10.6342/NTU201603448

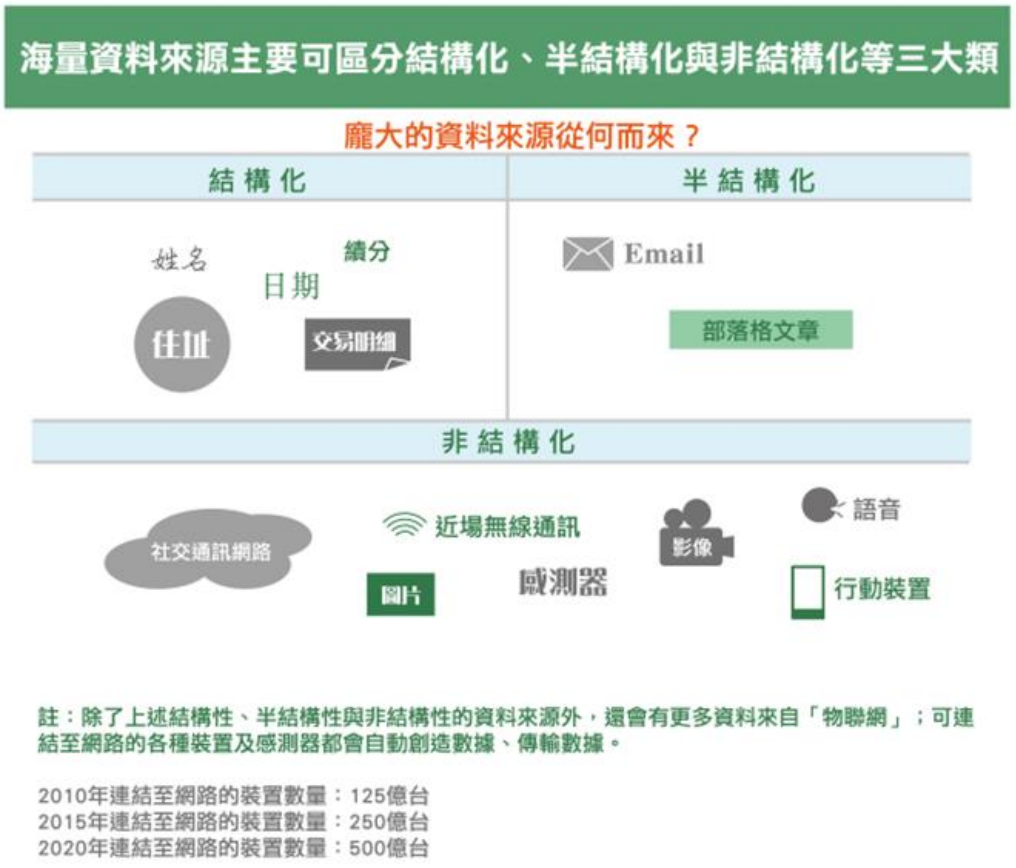


圖 2-1-3 結構化、半結構化、非結構化資料

doi:10.6342/NTU201603448

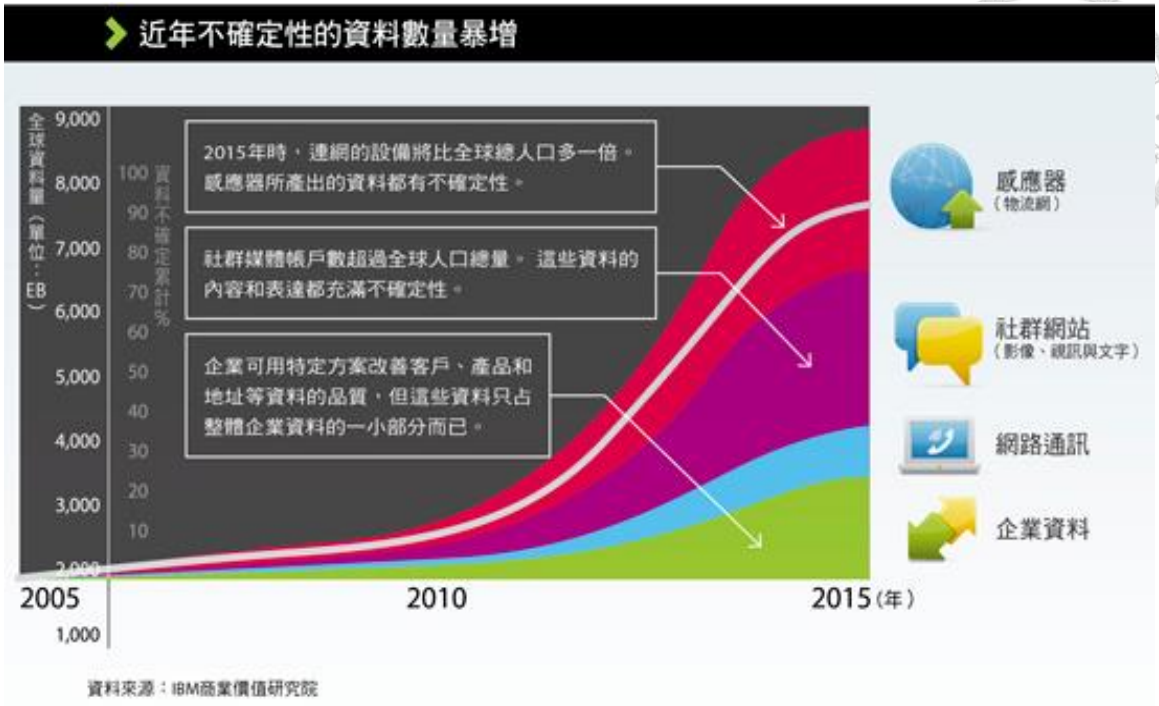


圖 2-1-4 不確定性的資料量爆增

doi:10.6342/NTU201603448

第二節 資料探勘（Data Mining）

一. 資料探勘的定義

資料探勘是指在龐大的資料庫當中，利用各種技術與統計方法，將大量的歷史資料進行分析、歸納與整合等工作，找出有興趣之特徵且具有意義的資料。資料探勘之相關定義彙整[7~15]，如表 2-2-1 所示。

綜合以下所有專家學者對於資料探勘的定義，可瞭解到資料探勘吸引人之處，主要原因在於能快速地從資料當中擷取所需要的資訊，亦能有效地分析解決大量與多維度的資料。

二. 資料探勘的功能

資料探勘的功能包含六大項，分別為：資料分類、資料推估、資料預測、資料關聯分組、資料分群及循序樣式採礦，這些功能大多可利用已成熟的統計分析方法來完成，其個別功能說明如下[10][16~17]：

1. 資料分類（Classification）

資料探勘中常見的功能之一，顧名思義即是將分析對象依不同的屬性分類並加以定義，建立不同的類組。資料探勘中的分類是指針對未發生之結果進行預測分類，主要包含歸納和推論兩步驟，其主要目的在於提高分類的準確度，建立分類規則，再評估準則的優劣，常用決策樹進行資料分類。

2. 資料推估（Estimation）

根據不同相關屬性資料的連續性數值，找出各屬性之間的關聯性，以瞭解並獲得某一特定屬性未知的連續性數值，常用的方法包含迴歸分析及類神經網路等統計方法。

3. 資料預測（Prediction）

預測工作的目的在於以其它屬性的值為基礎來預測特定屬性的值。而這個被預測屬性的值通常稱為目標變數或是應變數；而其它屬性則稱為解釋變數或因變數，預測的主要概念在於建立資料當中因變數與自變數間的關係，常用迴歸分析、時間序列分析及類神經網路等方法。

doi:10.6342/NTU201603448

4. 資料關聯分組（Association Rules）

資料關聯分組主要用來發現資料中特徵屬性間具有高度關聯的一種樣式，其所發現的模式通常是用規則來表現。進行資料採礦時，最常應用購物籃分析、GRI（Generalized Rule Induction）、Apriori 演算法[16]、連結分析等分析工具，關聯分組主要處理變數只有二個值的情況，這類資料採礦工具，最重要的關鍵在於有效率地處理大量資料。

5. 資料分群（Clustering）

資料分群主要是利用資料中類似或相同之項目，將同質性較高之資料區隔為不同之集群，集群內資料相似度越高越好；集群間差異度越大越好。在一大群的研究對象當中，根據不同的研究目的必定會有異質化的現象，但異質化的現象可能是幾個同質化的群組所造成，資料分群的主要目的便是將不同之同質化的組別差異找出來，最常使用的方法包括判別分析與集群分析。

6. 循序樣式探勘（Sequential Pattern Mining）

循序樣式探勘是在一序列的資料庫中，找出資料和時間相關之行為模式，並分析此序列之狀態轉變，進而從相關序列中達到預測未來的效果；循序樣式探勘是以時間作為相關項目之區分，因此常應用於股價之預測、行為預測等領域。

三. 資料探勘之流程

資料探勘交叉行業標準過程（Cross-Industry Standard Process for Data Mining，簡稱CRISP-DM）是由歐洲委員會與幾家在資料探勘應用上有經驗的公司共同籌劃的一個特別小組於2000年提出後並加以推廣，其強調完整的資料探勘過程，不能只針對在資料整理、資料呈現、資料分析以及建構模式上，仍需要對企業的需求問題進行了解，以及後期對模式的評價與模式的延伸應用都是一個完整的資料探勘過程不可或缺的要素，CRISP-DM流程圖，如圖2-2-1所示。

其各步驟之敘述說明如下[18]：

1. 定義商業問題（Business Understanding）

資料探勘的中心價值在於商業問題上，所以初步階段必須對組織的問題與需求

doi:10.6342/NTU201603448

深入瞭解，經過不斷與組織討論與確認過後，擬定一個詳盡且可達成的方案。

2. 資料理解（Data Understanding）

定義所需要之資料，並收集完整資料，並對收集的資料做初步分析，包括識別資料的質量問題、找到對資料的基本觀察，除去雜訊或不完整的資料，可提升資料準備的效率，接著並設立假設前提。

3. 資料預處理（Data Preparation）

因為資料來源不同，常會有格式不一致等問題。因此在建立模型之前必須需進行多次的檢查修正，以確保資料得到完整與淨化。

4. 建立模型（Modeling）

根據資料形式，選擇最適合的資料探勘技術並利用不同的資料進行模型測試，以達到預測模型最佳化，模型愈精準，有效性及可靠度愈高，對決策者做出正確之決策愈有利。

5. 評價和解釋（Evaluation）

在測試集中得到之結果，只對該資料有意義，實際應用當中，隨著使用不同的資料集其準確度便會有所差異，因此，此步驟最重要的目的便是瞭解是否有尚未被考慮到的商業問題盲點。

6. 實施（Deployment）

資料探勘流程透過良性循環，最後將整合過後的模型應用於商業上，但模型的完成並非代表整個專案完成，知識的獲得也可以透過組織化、自動化等機制進行預測應用。最後這階段包含部屬計畫、監督、維護、傳承與最後的報告結果，形成整個工作。

四. 資料探勘與文字探勘之關係

資料探勘是一門結合統計學與資訊科學相關理論的方法學，主要藉由各種功能與模式的導入與實踐，使其相關的應用遍及各個領域，成為研究與實務工作者重要的研究方法，是一門兼具問題、理論與方法的學科。同時也可被稱為在資料庫中挖掘知識（Knowledge Mining From Databases）、知識萃取（Knowledge Extraction）、資

doi:10.6342/NTU201603448

料規則分析（Data Pattern Analysis）、資料考古學（Data Archaeology）、資料採集（Data Dredging）等。

資料探勘可在龐大的數據庫中找尋知識，根據不同的依據建立不同的模型，以提供決策時的參考依據。因此，在資料採礦裡如何有效率且正確的從龐大資料庫中汲取有用的資訊是一個很大的挑戰，也隨著非結構化的資料越來越多，處理文字型資料也越來越被人們重視，文字探勘逐漸在資料探勘中成為重要的一環，下節中將仔細介紹何謂文字探勘。

doi:10.6342/NTU201603448

表 2-2-1　資料探勘定義彙整

學者	定義
Frawley [7]	資料探勘是從資料庫中挖掘出明確、前所未知但有用的潛在資訊過程。
Grupe and Owrang [8]	資料探勘乃是從已經存在的資料庫中剖析出新的事實及發現專家仍未知的新關係。
Fayyad [9]	資料探勘是資料庫知識發現的一個部分，而資料庫知識是從大量資料中選取合適之資料，進行資料前處理、資料轉換、解釋評估等工作，再進行資料探勘的一系列過程。
Berry and Linoff [10]	資料探勘為針對大量資料藉由自動或半自動方式進行分析，挖掘出資料中有意義的關係或規則。
Weiss and Indurkhya [11]	資料探勘是從大量資料中挖掘出有價值的資訊。
Kleissner [12]	資料探勘是一種新的且不斷循環的決策支援分析過程，它能夠從組合在一起的資料中，發現隱藏價值的知識，以提供給企業相關人員參考。
Hand [13]	資料探勘從龐大資料庫次級分析的過程，以利找出資料擁有者所關心或獲取有價值的未知關係。
Shaw, Subramaniam, and Tan [14]	資料探勘為尋找和分析資料的過程，其主要之目的是在於找出隱含在資料中的有效資訊。
黃勝崇 [15]	資料探勘為知識發現的核心，是一種自動或半自動的處理，其結果未能預測。
謝邦昌 [16]	資料探勘為找尋隱藏在資料中的有用訊息，如趨勢、特徵及相關性的一種過程，也是從資料當中挖掘出知識。

doi:10.6342/NTU201603448

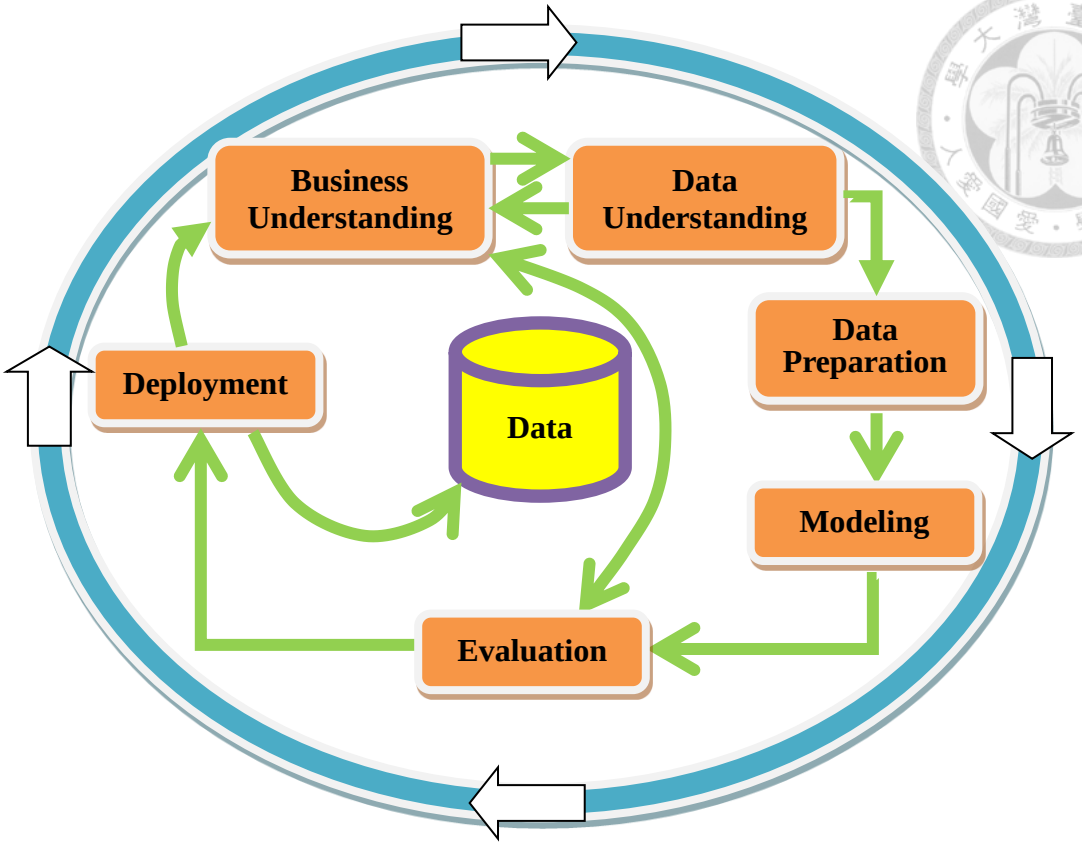


圖 2-2-1　CRISP-DM 流程圖

doi:10.6342/NTU201603448

第三節 文字探勘（Text Mining）

一. 文字探勘的定義

隨著大數據時代的到來，資訊的發展帶來大量的資料，而網站和網頁數量的快速成長，使得現今網路上有各式各樣的電子文件，像是電子書、粉絲專頁、電子新聞、電子郵件等，這些不同種類的文字資料並不像結構化的資料庫的所有資訊已明確定義分類，因此若未經整理，往往過於雜亂無章，無法被有效利用，為了進一步了解文件中更深入的潛藏資訊，因此有了文字探勘的需求。

文件探勘近年來成為興起的文件分析研究課題，文字探勘之相關定義彙整[19~22]，如表 2-3-1 所示，其指的是有效率地從大量文字性資料中整理出有用的資訊。

二. 斷詞的定義

在處理文件資料時，常遇到的首要問題就是大多數的資料不具其結構性，所以無法直接進行分析，因此我們需先對資料做預處理，即為「斷詞」處理。所謂的「詞」，在自然語言處理中是最基本的處理單位，語言學家對其定義為「能夠獨立運用，具有完整語意的最小語言成分。」，其原則如下[23]:

(1)有獨立意義的語法類可依類為一分詞單位。

(2)慣用的語言成分依人的使用習慣切分。

(3)語意失去組合性，或語法起變化，得合為一分詞單位。

(4)有明顯的分隔標記時得切分之。

(5)同形異構的成分依實際語境切分。

在英文裡，每個單字（Word）就可成詞，具有自己的意義，且每個單字之間都以明 顯的空白為分隔，因此沒有斷詞的困擾。反之，在中文裡，詞和詞之間書寫時，並不會以空白區分，因此將正確的詞切分出來，就成為自然語言處理的最基礎工作，斷詞結果的正確性及完整性及效率，其效能優劣或多或少都會影響到後續的處理。

doi:10.6342/NTU201603448

1. 斷詞

對於輸入 N 個連續字的字串（C1 C2 C3……Cn），斷詞的目的就在於找出正確的詞串（W1 W2 W3……Wm），這裡 Wi 可以是單字詞也可以是多字詞。例如:輸入的字串為「我是台灣大學農藝系的學生」， 必須產生出正確的詞串為「我／是／台灣大學／農藝系／的／學生」。

然而，在中文斷詞的過程中，我們並不是利用直接透過人工辨識，而是利用電腦的斷詞系統來做此工作，故在過程中將會遇到的以下三點問題:

(1) 詞的標準不統一

不同的電腦斷詞系統對詞的單位各有不同的要求，這使得斷詞系統之間，難以比較評。例如：台灣的計算語言學會在 1995 年定義「詞是一個具有獨立意義，且具有固定語法功能的字串」，而大陸在 1992 年則定義為「詞是最小的獨立運用的語言單位」。

(2) 歧義性

一個連續中文字串，可能會有多種不同的斷詞組合，斷詞系統必須選出其中最好的一種斷詞方法。因此歧義性可分為兩類：

A.交集型歧義（Overlapping Ambiguity）

令 x、y、z 代表中文字元所組成的字串，若 x、z、xy 與 yz 皆為辭典中的詞，則 xyz 的組合，於不同的文章中，可能會被斷詞成 xy/z 或 x/yz 等兩種不同的結果，其中字元 y 可與字元 x 結合，形成 xy 的詞， 也可與字元 z 結合，產生 yz 的詞，則 xyz 稱為「交集型歧義字串」。

例如：「中國人」三個字，可對應至詞表中的詞有「中、中國、國人、人」，將產生「中／國人」、「中國／人」兩種切分結果。

B.組合型歧義（Covering Ambiguity）

令 x、y 代表中文字元所組成的字串，若 x、y、xy 都是辭典中的詞，xy 的組合中，可在不同的文章中，分別被斷詞成 xy 或 x/y，因為詞 xy 是由 x 與 y 等兩個不同的詞所組成，因此 xy 稱為「組合型歧義字串」。

doi:10.6342/NTU201603448

例如：「才能」二個字，可對應至詞表中的詞有「才、能、才能」，下列句子正確的斷詞為「他／才能／非凡」，「只有／他／才／能／ 勝任」。

(3) 未知詞

由於人類所使用的語言會隨著社會不斷改變，而持續地創造出新的用語，並且詞的衍生現象也非常地普遍，導致新詞會不斷的出現，辭典永遠無法因應新詞產生的速度，所以會出現未知詞問題。因此未知詞指的就是沒有收入在系統辭典裡，但是卻又必須確切分出來的詞，例如:人名、地名、組織名、數字年份等等，這些未知詞往往會是後續應用，關切的重點所在，故斷詞系統必須能夠處理未知詞，才可提高斷詞的正確性。

2. 斷詞的方法

中文斷詞方法共分以下為四種[24~26]：

(1) 詞庫式斷詞法（Word Identification）

利用已建立的詞庫，來比對輸入文件，將文件中出現在詞庫中的片語擷取出來，是最為廣泛被採用的斷詞方法。其缺點是斷詞的正確性取決於所建的詞庫，因此這種斷詞法對於歧義和未登錄詞的切割具有很大的困難[27]。

(2) 統計式斷詞法（Statistical Word Identification）

將原文中任意前後相鄰的兩個字做為一個詞進行出現頻率的統計，出現次數越高，成為一個詞的可能性也就越大，在頻率超過某個預先設定的值時，就將其作為一個詞進行索引。其優點是較不受語文國別與句型的限制，而且可以擷取出未曾被詞庫、語料庫網羅的專業用語、新生詞彙與專有名稱等片語[27]。

(3) 混合式斷詞法（Hybrid Word Identification）

結合前述兩種斷詞法的方法，先利用詞庫式斷詞找出許多不同組合的詞彙，再利用詞彙的統計訊息來找出最佳的斷詞組合。此法目前仍需要大型的語料庫來提供統計資訊。

(4) 專家斷詞法

藉由領域中專家的意見以及用法，加以整理其所需要之詞彙，了解詞彙中之涵義，建立出非通用之詞彙，使詞彙資料庫因不同領域需求而有不同之詞彙，以成為更加完整之詞庫。

doi:10.6342/NTU201603448

3. 常用的斷詞系統

(1) 中研院中文詞知識庫小組（CKIP）-中文斷詞系統

中研院中文詞知識庫小組（Chinese Knowledge Information Processing Group，簡稱 CKIP）是一個跨所合作的中文計算語言研究小組由中央研究院資訊所、語言所於民國七十五年成立，其共同合作建構中文自然語言處理的資源與研究環境，為國內外中文自然語言處理及其相關研究提供基本的研究資料與知識架構。代表性研究成果包括中文詞知識庫、語料庫及中文處理技術等[29]。

此小組所研發的中文斷詞系統目前為國內最為精確的中文斷詞系統[30]，使用者可以線上即時進行斷詞，也可以自行撰寫程式經連線傳送驗證資訊及文本至中文斷詞線上服務伺服器，再由伺服器處理後連線傳回結果，系統使用介面如圖 2-3-1 所示。

這系統包含一個約拾萬詞的詞彙庫及附加詞類、詞頻、詞類頻率、雙連詞類頻率等資料。斷詞依據為此一詞彙庫及定量詞、重疊詞等構詞規律及線上辨識的新詞，並解決分詞歧義問題。除了基本詞彙庫外，使用者可依需要附加領域專屬詞庫。詞類標記為選擇性功能，可附加文本中切分詞的詞類解決詞類歧義並猜測新詞之詞類。分詞系統採用之詞典具可擴充性，使用者可依據不同領域文件，補充以領域詞典做為分詞之用。

(2) 結巴（Jieba）

Jieba 這個中文斷詞程式是由中國百度的一個開發 Sun Junyi 寫的[31]，其為一個開放原始碼（Open Source）的套件（Project），原本僅可支援簡體中文，但在其他的開發者加入開發下，目前已可以支援簡體和繁體中文。而之所以取名為 Jieba 是因為當我們將一句話斷成詞的時候，念起來就會像是結巴一樣[32]。

Jieba 核心程式概觀是利用字典來解決大部份的斷詞，並結合支持最大概率法（Maximum Probability）、隱馬可夫模型（Hidden Markov Model）、索引模型（Query Segment）及混合模型（Mix Segment），共四種核心演算法[33]。Jieba 的斷詞結果也可以具使用者的需求有下列三種選擇:

A. 精確模式:試圖將句子最精確地切開，適合文本分析。

如: 「我來到北京清華大學」=>「我/來到/北京/清華大學」。

B. 全模式:把句子中所有可以組成詞的結果都排列出來，運算速度非常地快，但

doi:10.6342/NTU201603448

是無法解決詞的歧異性。

如:「我來到北京清華大學」=>「我/來到/北京/清華/大學/清華大學/華大/大學」。

C. 搜索引擎模式:在精確模式的基礎上,對較長的詞再次切分,提高召回率,適合用於搜尋引擎分詞。

如:「小明碩士畢業於台灣商學院企管所,後在日本京都大學深造」=>「小明/碩士/畢業/於/台灣商學院/台灣/商學院/商/學院/企管所/企管/所/後/在/日本京都大學/日本/京都大學/京都/大學/深造」。

除了可以自己選擇斷詞結果模式外,Jieba 也提供詞性標註、關鍵詞的篩選、自行添加自定義的辭典等功能。由於前面提過 Jieba 是個開放原始碼的套件,因此有許多人把 Jieba 翻譯成各種程式語言的版本,可供在不同程式語言下進行斷詞,在此整理如下[31]:

A. Python 版本-「jieba」:
https://github.com/fxsjy/jieba

B. Java 版本-「jieba-analysis」:
https://github.com/huaban/jieba-analysis

C. C++版本-「cppjieba」:
https://github.com/yanyiwu/cppjieba

D. Node.JS 版本-「nodejieba」:
https://github.com/yanyiwu/nodejieba

E. iOS 版本-「iosjieba」:
https://github.com/yanyiwu/iosjieba

F. Erlang 版本-「exjieba」:
https://github.com/falood/exjieba

G. R 版本-「jiebaR」:
https://github.com/qinwf/jiebaR

H. PHP 版本-「jieba-php」:
https://github.com/fukuball/jieba-php

doi:10.6342/NTU201603448

三. 文字探勘的應用

對於進行完預處理的資料，擷取適當的資訊後才能進行下一步的分析，在此文字探勘結合一些資訊檢索技術，如：自然語言處理（Natural Language Processing）、統計分析（Statistical Analysis）、機率模式（Probability Models）、機器學習（Machine Learning）等技術。

用來探討概念擷取（Concept Extraction）、文件摘要（Text Summarization）、資訊過濾（Information Filtering）、命名實體的標註或辨別（Named Entities Tagging or Identification）、意見分析（Opinion Analysis）、關係探索（Relation Discovery）、情緒分析（Sentiment Analysis）、文本分類（Text Classification）、文本分群（Text Clustering）、潛在語意分析（Latent Semantic Analysis）等議題。

目前在醫學、法律、商業、工程、電腦等諸多領域已有多種應用被發表，可供作資訊搜尋、訊息過濾、事件關聯、知識萃取、知識管理、決策輔助、病例、論文研究等。

doi:10.6342/NTU201603448

表 2-3-1　文字探勘定義彙整

學者	定義
Hearst [19]	文字探勘是資料探勘的延伸應用，其結合了文字分析及資料探勘的技術，其主要目的是希望透過運算過濾大量的文字內容，並從非結構（Un-structured）或半結構（Semi-structured）的文字中發掘出未知、隱含且有用的資訊，讓需求者能夠有效率的運用。
Dan Sullivan [20]	文字探勘是編輯、組織及分析大量文件的過程，以提供特定使用者（如：決策者、分析師）特定的資訊（如：摘要、關鍵字），及發現某些特定資訊的特性與之間的關聯。
巫啟台 [21]	從非結構化或半結構化的文件資料中，發掘出有價值的片段、模型、方向、趨勢或規則。
曾元顯 [22]	整合傳統資訊檢索技術，包括關鍵字擷取、全文檢索、摘要自動萃取等，讓使用者從文件資料中找出隱含而有價值的資訊。

doi:10.6342/NTU201603448

圖 2-3-1 中研院 CKIP 中文斷詞系統

doi:10.6342/NTU201603448

第四節 台灣大學校務建言之量化研究

在「國立台灣大學校務會議及校務建言系統資料之分析研究」這篇論文中，使用民國94至101年間校務建言系統的量化資料，藉由敘述統計、統計圖表、變方分析、卡方檢定及無母數等統計方法，探討處理效率、點閱次數的關係[2]，其分析的研究結果整理如表2-4-1所示。

此外隨著年份的增加，使用校務建言系統的人數也日益增加。但明顯在十二、一、二月和七、八月的建言筆數較少，原因可能是寒、暑假造成使用校務建言系統的人數減少。

校務建言系統上線使用至今，回覆率平均維持在99%左右。然而對於校務建言系統各項建言分類探討，若進一步探討建議內容，將能更了解建言者的建言需求，校方也才能依不同的建議內容，提升校務建言系統的品質。目前本校已於年度實施行政品質評鑑時，將校務建言系統的執行成果列入評量。

doi:10.6342/NTU201603448

表2-4-1 處理日數及點閱次數之結果

	九組建議類別	六組建議者身分	十七組建議主旨
處理日數 (最長類別/平均天數)	達到統計顯著 (教務27.05天)	達到統計顯著 (短期帳號20.28天)	達到統計顯著 (短期帳號20.28天)
點閱次數 (最多類別/平均筆數)	達到統計顯著 (學生事務62.23筆)	達到統計顯著 (短期帳號76.21筆)	達到統計顯著 (人事120.90筆)

※類別說明

九組建議類別(建言系統預設):

校務、教務、學生事務、總務、計資中心、圖書館、學雜費、體育室、其他。

建議者身分(建言系統預設):

學生、正式教師、正式職員、校友、公務帳號、短期帳號。

建議主旨(自訂):

腳踏車、重大事件、宿舍、汽車、環境、維修、教務、體育館、網路訊號、安全、人事、圖書館、帳目、電腦系統、學生活動、醫院、其他。

doi:10.6342/NTU201603448

第三章 研究方法

第一節　研究流程

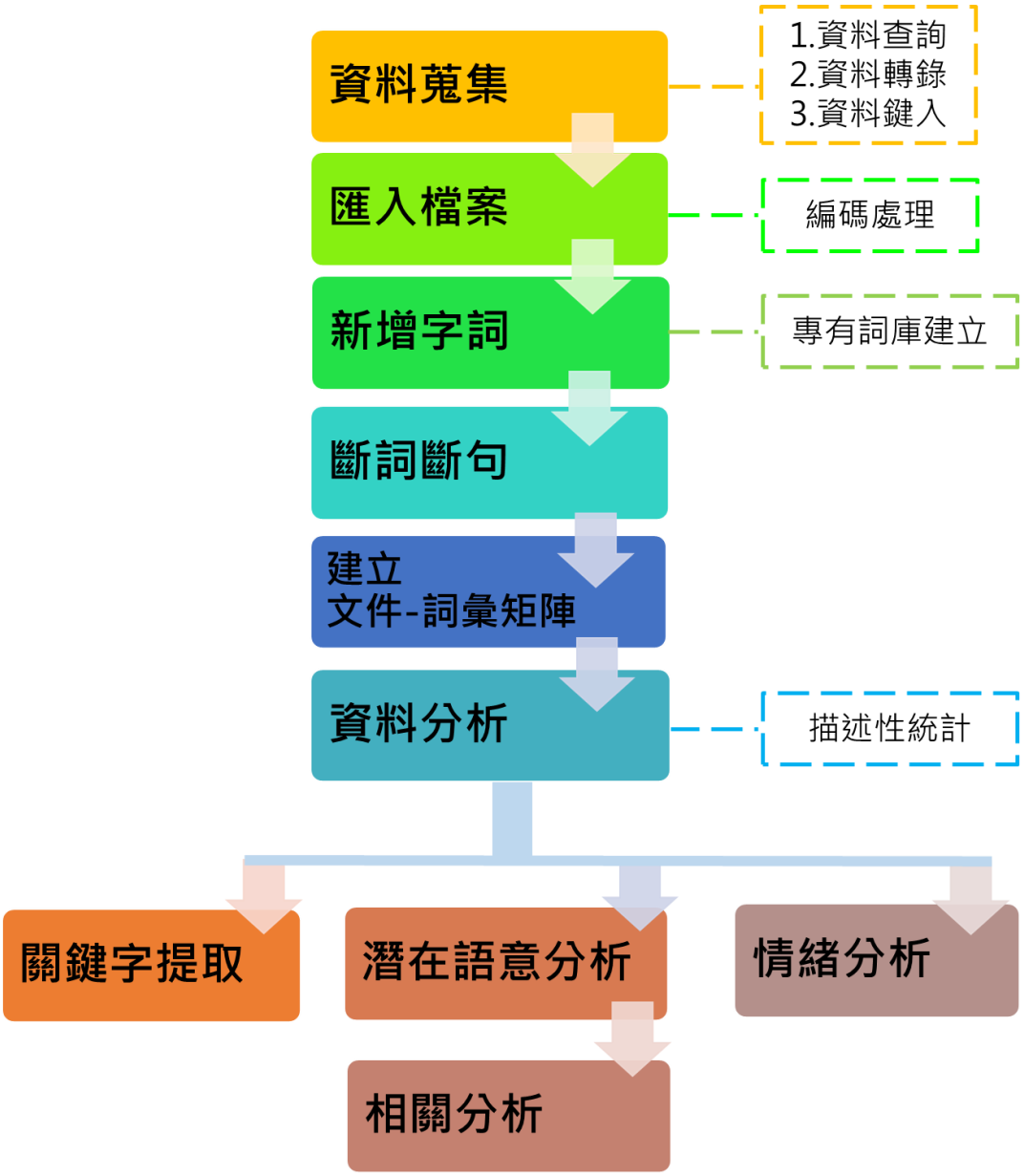


圖 3-1-1 研究流程圖

doi:10.6342/NTU201603448

第二節　研究工具

本研究利用 R 軟體來進行文字探勘，以下將分別闡述所使用的研究工具包含軟體的介紹及選擇此軟體有何優點[34]。

一. R 軟體介紹

R 是一套統計分析及繪圖的語言及軟體，最初是由紐西蘭奧克蘭大學的 Ross Ihaka 及 Robert Gentleman 兩位教授鑒於當時的統計教學實驗室使用的麥金塔電腦，市面上沒有合適的統計軟體，兩人仿 S 語言的架構開發 R 來輔助統計上的教學，經過多年發展後，R 目前已是一套功能強大且廣為使用的自由軟體。

二. R 軟體特色：

1. 免費。
2. 自由、開放源碼。
3. 有效的資料處理及存取能力，可與 C++和 JAVA 等程式連結。
4. 方便的矩陣操作與運算能力。
5. 完整而連貫的資料分析能力，有大量套件程式可免費下載。
6. 強大的繪圖功能。
7. 簡單且發展完善的程式語言環境。
8. 可在 UNIX（含 FreeBSD 與 Linux）、Windows 和 MacOS 執行。
9. 有良好的線上文件說明。

doi:10.6342/NTU201603448

第三節　編碼處理

一. 亂碼問題

在處理中文資料時，最常遇到的首要問題就是收集到的資料要匯入使用的軟體時，竟產生亂碼的問題。因此，對於不熟悉編碼（Encoding）知識的開發者，常需要花費大量的時間在處理資料的編碼問題。

所謂的編碼就是電腦把位元序列（文字資料儲存到電腦後，最終都是 0 和 1 的位元序列）轉譯成人類看得懂的文字的格式，也就是說如果編碼的設定不正確，電腦就沒辦法把資料正確的轉換成文字供人類閱讀。以下舉「我愛你」的例子，並用數種常見的編碼表示其位元序列格式[35]，如表 3-2-1 所示，從此可以看出不同編碼格式的位元序列格式也都不相同。

因此若開啟的檔案或下載網頁有亂碼的情況時，原因通常是發送端與接收編碼設定不同所致。所以在遇到亂碼的問題要處理時，掌握住三個地方就沒什麼問題，第一個是文件的編碼，第二個是作業環境的預設編碼，第三個是軟體的環境編碼，只要確定這三個地方一致就不會產生亂碼的情況。

二. 中文的編碼種類

1. 以作業環境區分

(1) Windows

A.台灣、香港及澳門:系統的字符集都是以 Big5 為基準，無法顯示簡體中文。

B.中國:系統的字符集都是以 GBK 為基準，無法顯示繁體中文[36]。

(2) Mac／Linux

系統的字符集皆採用 UTF-8 為基準，其由萬國碼（Unicode）所延伸出來的編碼方式，但是與萬國碼屬不同的編碼方式，此編碼系統打破所有國家的不同編碼，逐漸成為電子郵件、網頁及其他儲存或傳送文字的應用中，優先採用的編碼。此外，UTF-8 還可通用於繁體中文與簡體中文之間[37]。

doi:10.6342/NTU201603448

2.以軟體環境區分

由於本研究使用的軟體為 RStudio 做為分析工具，因此在此僅介紹 RStudio 中如何更改編碼上的設定。將 R 的環境改為相對應的地區，才可以去處理編碼的問題。

一般來說在台灣地區 R 所設定的預設為 Chinese（Traditional）_Taiwan.950，若想查詢現行環境下的編碼格式可使用 Sys.getlocale( )函數;若要修改現行環境下的編碼格式可使用 Sys.setlocale( )函數。舉例而言，若要處理簡體中文的文件時，則輸 Sys.setlocale(locale='Chinese')函數，即可改變環境編碼為 Chinese(Simplified) _China.936。

另外，RStudio 的使用者，也可以透過點選的方式改變編碼環境[35]，步驟如此 Menu→Tools→Global Options→General→Default text encoding。

doi:10.6342/NTU201603448

表 3-2-1 不同編碼的位元序列格式

中文內容	我愛你
BIG5	A7 DA B7 52 A7 41
GB2312	CE D2 90 DB C4 E3
UCS-2	11 62 1B 61 60 4F
UTF-8	E6 88 91 E6 84 9B E4 BD A0
UTF-16	11 62 1B 61 60 4F
UrlEnc(Big5)	%a7%da%b7R%a7A
UrlEnc(Unicode)	%u6211%u611b%u4f60
UrlEnc(UTF-8)	%e6%88%91%e6%84%9b%e4%bd%a0

doi:10.6342/NTU201603448

第四節　斷詞方法

在第二章的第三節中曾介紹過目前常用的斷詞系統，分別為中研院中文詞知識庫小組（CKIP）的中文斷詞系統及結巴（Jieba）兩種。由於中研院的斷詞系統較不穩定，回傳的資料有可能會有被截斷的情況，且考量到後續的資料分析是使用 R 軟體，為了資料處理的方便性及一致性上，故本研究決定採用 jiebaR 這個套件，進行中文斷詞的部分。

當然在 R 這套軟體中，中文斷詞除了有 jiebaR 這個套件外，還有由中國大陸學者李艦等人所開發出的 Rwordseg 套件。Rwordseg 套件是利用 rJava 去連結 java 分詞工具 ansj （一個開拓 Java 中文分詞工具）來進行斷詞，由於安裝上需變更使用者電腦的眾多設定，加上斷詞效率不佳，經使用比較過後不考慮採用。

在此展示一篇使用 jiebaR 斷詞過後建言結果及詞性的標示，其中英文字母表示此詞的詞性，如表 3-4-1 所示，若想知英文字母所代表的詞性意思請參考附錄一。

doi:10.6342/NTU201603448

表 3-4-1 斷詞後詞性展示

s	z	x	zg	t	d	c
校內	小黑	蚊因	為	夏日	到來	而
a	v	d	v	b	ns	m
嚴重	氾濫	尤其	是	小小	福	四周
c	b	m	n	x	s	x
及	女	九	餐廳	建請	校內	之生
nt	v	x	j	n		
農學院	相關	系所	研發	滅蚊		

doi:10.6342/NTU201603448

第五節　字詞處理-專有詞庫建立

藉由 jiebaR 斷完詞後，可以明顯發現對於 jiebaR 而言出現許多未知詞，可以利用人工判讀的方式大略地做檢查，就舉前一頁的例子而言就會發現很多斷詞斷的並不恰當，然而這些未知詞的產生，絕大部分都跟所分析資料內的專有名詞相關。

本研究是採用台大校務建言系統的建言及回覆，故專有名詞也就是與台大相關的詞彙，如:建築物名稱、系所名稱、各處室名稱及常用簡稱等。這些詞都是未來關切的重點所在，必須確切分出來。因此，本研究按照校方所公布的「校總區及其他校區之主要建築物逐棟編碼地理位置對照」[38]、「院系所課程-台大課程地圖」[39]、「行政組織介紹」[40]及平常同學間在學校習慣使用的名詞簡稱，建立成一本台大專有詞庫，一共收入 588 個詞，可查鑒附錄二。

將專有詞庫建置完成後，新增至 jiebaR 中，並重新進行斷詞，在此使用上頁相同的例子進行重新斷詞，以方便比較結果，如表 3-5-1 所示。結果發現，經過專有詞庫的建置後，原本很多斷詞斷的並不恰當的地方都可明顯改善，達到正確斷詞地效果。

doi:10.6342/NTU201603448

表 3-5-1 專有詞庫斷詞後詞性展示

s	x	x	t	d	c	a
校內	小黑蚊	因為	夏日	到來	而	嚴重
v	d	v	x	m	c	x
氾濫	尤其	是	小小福	四周	及	女九
n	x	s	u	x	v	x
餐廳	建請	校內	之	生農學院	相關	系所
j	n					
研發	滅蚊					

doi:10.6342/NTU201603448

第六節　文件-詞彙矩陣（詞彙-文件矩陣）

文字探勘在做分析前須先建立詞彙與文件的關係矩陣，此關係矩陣可稱為「文件-詞彙矩陣（Document-term Matrix，簡稱 dtm）」或「詞彙-文件矩陣（Term-document Matrix，簡稱 tdm）」。兩者皆是簡單三維矩陣（Simple Triplet Matrix），其建構稀疏矩陣的資料結構和運算元，共三個維度分別為行、列及值。

當資料檔建構成「文件-詞彙詞頻矩陣（dtm）」表示把詞彙出現的頻率設定為行，文件出現的頻率設為列;「詞彙-文件詞頻矩陣（tdm）」表示把把詞彙出現的頻率設定為列，文件出現的頻率設為行。

舉例來說，目前有兩個文件分別為 A 和 B，文件 A 裡的詞彙有「我是台大的學生，台大真漂亮」，文件 B 裡的詞彙有「台大有學生及老師」，經由斷詞後文件 A 與文件 B 其相對應的 Document-term Matrix 及 Term-document Matrix，如表 3-6-1 及表 3-6-2 所示。

doi:10.6342/NTU201603448

表 3-6-1 文件-詞彙詞頻矩陣（Document-term Matrix）

	我	是	台大	的	學生	真	漂亮	有	及	老師
A	1	1	2	1	1	1	1	0	0	0
B	0	0	1	0	1	0	0	1	1	1

表 3-6-2 詞彙-文件詞頻矩陣（Term-document Matrix）

	A	B
我	1	0
是	1	0
台大	2	1
的	1	0
學生	1	1
真	1	0
漂亮	1	0
有	0	1
及	0	1
老師	0	1

doi:10.6342/NTU201603448

第七節　潛在語意分析（Latent Semantic Analysis）

理解生活環境中的語彙意義，是人類相當重要的認知行為之一。人對語彙的理解，不會僅來自單一的字詞，而是從一連串的詞彙而來。句意脈絡或是文章脈絡之所以可以幫助聽者或讀者有效理解單一字詞的意義，其實是因為單一字詞往往有不同的意義，但唯有在該字詞與其它字詞一起出現時，人們才能更正確的理解該字詞的意義。因此，掌握眾多詞彙所隱含錯綜複雜的語意關係，更是人類學習語言歷程中的重要一環[41]。

一. 潛在語意分析與語意空間

潛在語意分析（Latent Semantic Analysis，簡稱 LSA）是一種基於向量空間模型概念的技術，且運用線性代數有效地自動化資訊檢索的方法[42]。目的是探討隱藏在字詞背後的某種關係，這種關係並非以詞典中的定義為基礎，而是參考字詞的使用環境[42]。

潛在語意分析最早是在 1990 年由 Deerwester 等人所提出，其基本概念是先以二維的矩陣空間表徵文字和原始文件間的關係，再利用奇異值分解（Singular Value Decomposition, SVD）的方式，找出字詞對應文件的語意結構，此一技術的特色是將許多原本字面看不到的資訊呈現出來，因此能大幅提昇資訊擷取的有效性，而這也 Deerwester 將此分析方式稱之為「潛在」語意分析的原因。

接著於 1998 年時 Landauer、Foltz 及 Laham 研究提出[48]，若能有一個大型的語料來源能適當的反映人所擁有的語彙知識，即可以藉用 LSA 的技術，透過 SVD 的數學演算方式以及維度化約（Dimension Reduction）方式，就能建置出一個能反應這些語彙知識背後語意關係的語意空間。該語意空間以向量的方式呈現詞彙、段落或是文章在語意空間裡的相對位置，透過兩個文件間的向量角度之餘弦值（cosine value）來評估語意間的相似性，也就是說可藉由比對詞彙在句意脈絡下的對應位置，捕捉到隱藏在不同句子裏兩兩詞彙的語意關聯性。

例如下面的三句話：(1)「學校裏面有很多認真教學的老師」、(2)「老師正一

doi:10.6342/NTU201603448

筆一畫的教學生學習新字的筆順」以及(3)「這星期眼科醫師來學校幫忙做視力檢查」。雖然句子(2)中並沒有出現「學校」這個詞彙，但藉由「老師」這個詞彙的重疊，讀者所建置出的內在語意表徵中，學生和學校會有著一定程度的語意關聯性。同樣的，句子(3)中雖然沒有出現老師和學生，藉由「學校」這個詞彙的聯繫，仍然能讓讀者將「醫師」、「學生」及「老師」的語意關係建置起來。Landauer 認為人捕捉詞彙間語意關係的歷程[42]，其實很類似理解任一城市在此地圖上的意義，是藉由了解城市間彼此對應的位置，例如:地圖上台中在台北的南邊，基隆在台北的北邊，讀者即可以判斷出基隆也在台中的北邊。

LSA 所建置的語意空間，不僅打開一個語意知識表徵的新方法，也提供一種較精確又自動化的方式分析詞與詞、句子與句子、或是文章與文章間的語意關聯程度。另外，LSA 不需要使用文法或事先定義語彙的特性，使得 LSA 的技術可以不受語言環境的限制。也就是說，即是使用者所使用的是非英語的語料庫，只要使用者建立好關鍵詞在文件中出現次數的矩陣，就能利用 LSA 的技術建置出該語系的語意空間。

二. 中文語意空間之建置

使用 LSA 技術建置語意空間，需要以下四個步驟[44]：

1. 建立文件-詞彙矩陣（詞彙-文件矩陣）

從文件中找出關鍵詞來做處理與比對的依據，在本研究所定義的關鍵詞，是在任一文件中曾出現過詞。從建置的文件-詞彙矩陣（詞彙-文件矩陣）可以看出某一個關鍵詞在不同的句子、段落或文件中所出現的情形與每一個句子、段落或文件中有那些重要的關鍵詞。

2. 詞彙權重計算

考慮並非所有的辭彙和所有的文件都有相同的重要性，根據不同的應用研究，關鍵詞彙其重要性也並不完全相同，因此需給予不同的權重。LSA 詞彙權重的調整方式可分為 Local 權重和 Global 權重兩部份。Local 權重是指詞彙在文件中的重

doi:10.6342/NTU201603448

要性，通常以詞彙出現次數代表，出現次數愈多代表愈重要；Global 權重則是詞彙在整個語料庫的重要性，與 Local 權重相反，出現次數愈多，對於文件的重要性相對愈低[45-47] 。

目前常使用建立中文語意空間的詞彙權重方法，分別是：TF-IDF、Log-IDF 及 Log-Entropy。其計算公式分別如下:

(1) TF-IDF

(a) TF（Term Frequency）

$$L(i,j) = tf(i,j) = \frac{n_{i,j}}{\sum_k^N n_{k,j}} \quad (1)$$

L(i,j)表示 local 權重，其中 j 是「某一特定文件」，而 i 是該文件中所使用單詞或單字的「其中一種」，$n_{i,j}$ 就是第 i 個詞彙在第 j 個文件中「出現次數」，$\sum_k^N n_{k,j}$表示在第 j 個文件中所有詞彙的出現次數的總和。 tf(i,j)值愈高，表單詞愈重要。

例如:在兩篇文章中，配篩選出兩個重要名詞，分別為「健康」及「富有」。其分別出現在文章的次數如下表 3-7-1 所示。

表 3-7-1 名詞出現次數

出現次數	第一篇文章	第二篇文章
健康	70	40
富有	30	60

各別計算其 TF 值，如表 3-7-2 所示。

表 3-7-2 TF 值

	第一篇文章	第二篇文章
健康的 TF 值	70/(70+30)=0.7	40/(40+60)=0.4
富有的 TF 值	30/(70+30)=0.3	60/(40+60)=0.6

所以，可以得知「健康」對第一篇文件比較重要，「富有」對第二篇文件比較重要。若搜尋「健康」，那第一篇文件會在較前面的位置；而搜尋「富有」，則第二篇文章會出現在較前面的位置。

doi:10.6342/NTU201603448

(b) IDF（Inverse Document Frequency）

$$G(i) = idf(i) = \log \frac{m}{df(i)} \quad (2)$$

G(i)表示 global 權重，其中 m 表示文件的個數，df(i)表示第 i 詞彙出現在多少個文件中，df(i) 是該單詞在所有文件總數中「出現的文件數」。idf(i)值愈大，表此詞彙的出現機率其普遍性不高，此詞愈重要。

例如: 有 100 篇文章，「健康」出現在 10 篇文章當中，而「富有」出現在 100 篇文章當中，其 IDF 值如下表 3-7-3 所示。

表 3-7-3　IDF 值

健康的 IDF 值	log (100/10) =1
富有的 IDF 值	log (100/100) =0

所以，「健康」出現的機會小，與出現機會很大的「富有」比較起來，便顯得非常重要。

(c) TF-IDF

$$TF - IDF = tf(i,j) \times idf(i) \quad (3)$$

TF-IDF 計算是以某一特定文件內的詞彙頻率，乘上該詞彙在文件總數中的文件頻率。其意義是詞彙相對一份文件的重要性隨著它在該文件中出現的頻率增加，但隨著它在語料庫中出現的頻率減少。也就是說，過濾掉常見的詞彙，保留重要的詞彙。

(2) Log-IDF

(a) Log

$$L(i,j) = \log_2(tf(i,j) + 1)) \quad (4)$$

(b) IDF

$$G(i) = idf(i) = \log \frac{m}{df(i)} \quad (5)$$

(c) Log-IDF

$$Log - IDF = L(i,j) \times G(i) \quad (6)$$

doi:10.6342/NTU201603448

(3) Log-Entropy

(a) Log

$$L(i,j) = \log_2(tf(i,j) + 1)) \quad (7)$$

(b) IDF

$$G(i) = 1 + \sum_j \frac{p(i,j)\log_2(p(i,j))}{\log_2 n} \quad (8)$$

$$p(i,j) = \frac{tf(i,j)}{gf(i)}$$

gf(i)是第 i 個詞彙在所有文件中出現次數的總和。

(c) Log-Entropy

$$Log - IDF = L(i,j) \times G(i) \quad (9)$$

2. 運用 SVD 轉換矩陣

在 LSA 中被使用來分解詞彙－文件矩陣的方法為奇異值分解（SVD），藉由 SVD 的運算過程，可以計算出每個詞彙在對角矩陣中的特徵值，一般來說特徵值愈大的向量，表示具有較大的訊息量，反之則只有微小的訊息量。經過 SVD 轉換後的矩陣，關鍵詞和文件的關係，就不是原本出現次數的關係，取而代之的是表徵關鍵詞在文件中的語意關係。經過 SVD 轉換一個 m×n 的詞彙－文件矩陣 A，其可被拆解成如公式（10）：

$$A = U\sum V^T \quad (10)$$

其中 U 是 m×r 的正交矩陣(Orthogonal Matrix)或稱為左奇異向量(Left Singular Value)，V 為 r×n 的正交矩陣(Orthogonal Matrix)或稱為右奇異向量（Right Singular Value），Σ 為由奇異特徵值組成的 r×r 對角矩陣（Σ=Diagonal (λ_1, λ2,….., λ_r)，其餘元素皆為 0）[46]。而 U 矩陣的列向量稱為詞彙向量（Term Vector），其順序表對應 U 矩陣的字詞順序; V^T矩陣的行向量稱為文件向量（Document Vector）[48] ，其順序表對應V^T矩陣的文件順序。

doi:10.6342/NTU201603448

3. 維度約化（Dimension Reduction)

經由 SVD 後，因為語意空間的矩陣過大或是太多所謂的雜訊，則可以利用維度約化，消除語意空間中不重要之雜訊，其方式是取出 SVD 後前 k 個最大的特徵奇異值，和 U 矩陣、V 矩陣前 k 個行向量(k<r)，並重建矩陣 A 成為$\tilde{A}$，其可被拆解成如公式（11），並如圖 3-8-1 所示[49] [50]。

$$\tilde{A} = U_k \sum_k V_k^T \qquad (11)$$

而經由 SVD 轉換後的矩陣，應該要留下多少個奇異值，也就是說 k 值應為多少，在語意空間建置過程中，相當重要的一個議題。因為在維度約化的過程，如果只用兩到三個維度來表徵每個關鍵詞的向量，則每個關鍵詞之間的相似性會過高；反之，如果保留所有的維度來表徵每個關鍵詞的向量，則每個關鍵詞的相似性就又會幾近於零。

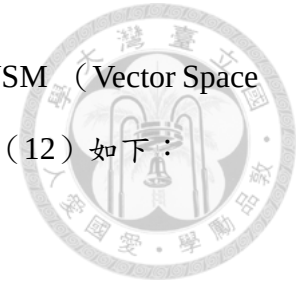
此外，也可以減少計算量與儲存量及去除雜音，其雜音的強度較主要訊息微弱，省略較小的奇異值雖然遺失部分訊息，但遺失的訊息多屬於雜訊，所以用低維近似反而能夠提高資訊檢索的效能[43]。

根據 Landauer 與 Dumais（1997）的研究發現，在維持 100、200 以及 300 個維度的情形下，對同義詞的測試都能有不錯的效果[51]。有鑑於此，本研究在進行 SVD 的矩陣轉換時，即同時建置了含有 100 及 300 個維度的中文語意空間，以達到最理想的語意關聯性評做的效率，建置出一個合理的語意空間。

三. 判斷語意相似性

經由 SVD 重新建置的矩陣$\tilde{A}$，是將詞彙、段落句子或文章以向量形式呈現該詞彙、段落句子或文章在語意空間的相對位置，以兩向量角度的餘弦值表示兩個文件之間的相似度，餘弦值愈大表示兩向量夾角愈小，在語意空間表示兩向量之間的語意愈相似;反之，餘弦值愈大表示兩向量夾角愈大，在語意空間表示兩向量之間的語意愈不相似。所以由 LSA 建置的語意空間，詞彙與詞彙、詞彙與段落、段落與段落間語意相似性都能藉由兩向量之間的餘弦值表示[52]。

doi:10.6342/NTU201603448



假設要判斷第 i 個詞彙和第 j 個詞彙的相似度，則可利用 VSM（Vector Space Model）求兩向量的夾角（餘弦值），即可求得其相似度，公式（12）如下：

$$\cos(t_i, t_j) = \frac{t_i t_j^T}{\|t_i\| \|t_j\|} \quad (12)$$

其中 t_i 表示第 i 個詞彙在語意空間的向量表徵，‖ti‖為 t_i 向量的長度。

假設若要利用 LSA 語意空間判斷兩篇文章（文章一用向量 T_1 表示和文章二用 T_2 向量表示，非語料庫之文章）的相似度，可以利用下述方式，公式（13）計算出：

$T_1=(a_1t_1,a_2t_2,\ldots,a_nt_n)$

$$T_2=(b_1t_1,b_2t_2,\ldots..,b_nt_n) \quad (13)$$

a_i 表示第 i 個詞彙在文章一出現的次數，b_i 表示第 i 個詞彙在文章二出現的次數，t_i 表示第 i 個詞彙在語意空間的向量表徵。基於上述兩篇文章在 LSA 語意空間的向量，其相似程度可由下列公式（14）計算得出，‖Ti‖為 T_i 向量的長度。

$$\cos(T_1, T_2) = \frac{T_1 T_2^T}{\|T_1\| \|T_2\|} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j t_i t_j^T}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j t_i t_j^T}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} b_i b_j t_i t_j^T}} \quad (14)$$

doi:10.6342/NTU201603448

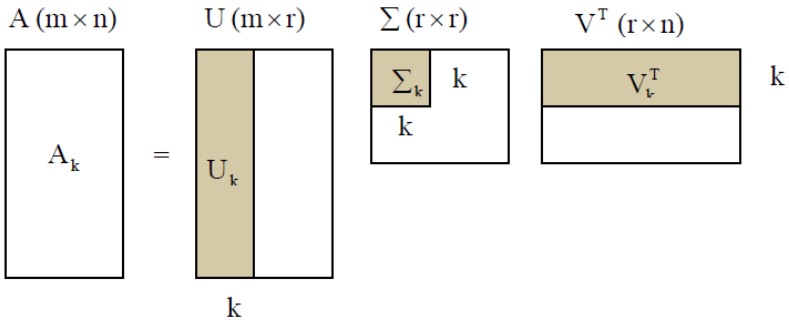


圖 3-8-1 語意空間的矩陣

doi:10.6342/NTU201603448

第八節　相關分析

一. Pearson 相關係數（Pearson Correlation Coefficient）

Pearson 相關係數是用以測量兩連續變數 x 及 y 之間直線關係的強弱[53]。其定義如下:

$$\rho_{xy} = \frac{\mathrm{Cov}\,(x\,,\,y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}} \qquad (15)$$

而用以估計之樣本 Pearson 相關係數則為

$$r_{xy} = \frac{\sum_i \left[\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]}{\sqrt{\sum_i \left(x_i - \bar{x}\right)^2 \sum_i \left(y_i - \bar{y}\right)^2}} \qquad (16)$$

以下說明相關係數如何解讀：

1. r > 0：兩變數 x 及 y 具正相關性，即隨著其中一個變數值增加，另一變數也隨之遞增。
2. r < 0：兩變數 x 及 y 具負相關性，即隨著其中一個變數值增加，另一變數卻遞減。
3. r =±1：兩變數 x 及 y 完全直線相關。
4. -1< r <1：兩變數 x 及 y 具直線相關性，|r|越接近 1 相關性越高，反之|r|越接近 0 相關性越低。
5. r =0：兩變數 x 及 y 不具有直線相關性，並非代表 x 及 y 之間不相關沒有直線關係。

因為在樣本資料中，存在著試驗或抽樣誤差，因此對於所取得的資料，兩變數間是否有關係，也就是相關係數是否為 0，則必須經由 t 值顯著性檢定才能推定，其假設檢定假設檢定為:

$$\begin{cases} H_0: \text{兩變數間無相關} \\ H_1: \text{兩變數間有相關} \end{cases}$$

doi:10.6342/NTU201603448

由相關係數之標準誤差（Standard Error，簡稱 SE），其 $SE(r)=\sqrt{\frac{1-r^2}{n-2}}$，可求得 r 之 t 值如公式（17）：

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = r\sqrt{\frac{n-2}{1-r^2}} \quad (17)$$

t 值服從自由度為 n-2 之 t 分布，當實測 $|t| > t_{\alpha/2,n-2}$ 時，表示兩變數間有相關，至於相關程度有多大，則不得而知，端賴兩變數分布集中程度而定。一般樣本資料之 r 值大小，與樣品點（n）多少有密切關係，通常 n 愈小，r 則越大;反之，n 愈大，則 r 愈小。當 n=2 時，r 必定為 1（兩點必在同一條直線上），因此樣本資料求得之相關係數後，必須再經顯著性檢定才可推論方妥。

二. Spearman 等級相關（Spearman's Rank Correlation Coefficient）

Spearman（1904）之等級相關係數在無母數統計方法中是最早被提出來的，也是今日普遍採用的方法。其資料為序列或有測量單位資料，各變數按觀察值大小給等級（Rank），最小者給等級 1，次小者給等級 2，依此類推，最大者給等級 n（n 為各變數的觀察值個數）。由於此方法不考慮資料的分布型態，計算簡單，效率又高，是很實用的相關性分析方法[53]。其假設檢定為:

$$\begin{cases} H_0: \text{兩變數間無相關} \\ H_1: \text{兩變數間有相關} \end{cases}$$

設兩變數經排列等級後之資料為:

x 變數: $x_1, x_2, \ldots\ldots, x_n$

y 變數: $y_1, y_2, \ldots\ldots, y_n$

兩變數等級差為 $d_i = x_i - y_i$，i=1,2,....,n。當 d_i 很小時，表示兩變數間有相關。

令 $x_i = x_i - \bar{x}$，$y_i = y_i - \bar{y}$，其中 $\bar{x}$ 及 $\bar{y}$ 分別為 x 與 y 變數的算術平均數。由 Pearson 之相關係數公式（18）：

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} \quad (18)$$

doi:10.6342/NTU201603448

今由於 x_i 及 y_i 均為 1,2,.....,n 等級數，故

$$\sum x_i = \frac{n(n+1)}{2} = \sum y_i \quad (19)$$

$$\sum x_i^2 = \frac{n(n+1)(2n+1)}{6} \quad (20)$$

$$\sum x_i^2 = \sum (x_i - \bar{x})^2 = (\sum x_i^2 - (\sum x_i)^2)/n \quad (21)$$

$$= n(n+1)(2n+1)/6 - n(n+1)^2/4 = (n^3 - n)/12$$

同理可得

$$\sum y_i^2 = \frac{n^3-n}{12} \quad (22)$$

由

$$d = x - y \quad (23)$$

$$d^2 = (x-y)^2 = x^2 + y^2 - 2xy \quad (24)$$

$$\sum d^2 = \sum x^2 + \sum y^2 - 2\sum xy \quad (25)$$

故當資料為等級時，其樣本的相關係數r_s為

$$r_s = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = \frac{\sum x^2 + \sum y^2 - \sum d^2}{2\sqrt{\sum x^2 \sum y^2}} = \frac{\frac{n^3-n}{12} + \frac{n^3-n}{12} - \sum d^2}{2\sqrt{(\frac{n^3-n}{12})(\frac{n^3-n}{12})}} \quad (26)$$

$$= 1 - \frac{\sum d^2}{\frac{n^3}{6}} = 1 - \frac{6\sum d^2}{n^3-n}$$

r_s是樣本等級資料的相關係數，也是族群相關係數ρ的估計值，其是否有意義，則需經顯著性程序檢定。檢定統計量為公式（27），其中α為顯著水準，n 為樣本數，Z 服從標準常態分佈:

$$W_\alpha = \frac{Z_\alpha}{\sqrt{n-1}} \quad (27)$$

當$r_s > W_\alpha$時，表示兩變數間有相關，反之則無。須注意此為一漸進分佈，當樣本數足夠大時方才成立。

三. Kendall 等級相關

Kendall（1938）提出另一種與 Spearman 等級相關相似的檢定方法。其以兩變數(x_i, y_i)，i=1,2,.....,n，是否一致以測驗兩變數之間是否有相關[53]。設今有 n 對觀

doi:10.6342/NTU201603448

測值$(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$，$x_i$ 及 y_i 各別依大小給等級後，在依 x_i 之等級有小到大排列，而 y_i 之等級則依 x_i 之等級排於相對位置。在 x_i 等級大小順序下，計算其任意前後兩對之等級大小是否一致，若前一對等級均比後一對等級小，則稱其等級有一致性(Concordance)。反之，若前一對等級均比後一對等極大，則前後兩對 x_i, y_i 之等級沒有一致性(Disconcordance)，全資料所有可能前後對數有$\binom{n}{2} = \frac{n(n-1)}{2}$。

若以 N_c 表示有一致性的對數，N_d 表示沒有一致性的對數，$N_c+N_d=n(n-1)/2$，Kendall 之相關性檢定就是依 N_c 及 N_d 之多少而決定的，其相關性定值以 τ 來表示，其運算公式（28）：

$$\tau = \frac{N_c - N_d}{n(n-1)/2} = \frac{2(N_c - N_d)}{n(n-1)} \qquad (28)$$

若所有對數全為一致性的，則表示$N_c = n(n-1)/2$，$N_d = 0$，即$\tau = 1$；若所有對數皆無一致性的，則表示$N_c = 0$，$N_d = n(n-1)/2$，即$\tau = -1$。故 Kendall 之 τ 值在-1 至 1 之間。其假設檢定為:

$$\begin{cases} H_0: \text{兩變數間無相關} \\ H_1: \text{兩變數間有相關} \end{cases}$$

當實測 $\tau > \tau_{\alpha/2,n}$ 時，表示兩變數間有相關，反之則無。其中 α 表顯著水準，n 為樣本數，$\tau_{\alpha/2,n}$ 表臨界值。

四. Kappa 統計量

Kappa 統計量又可稱為一致性測量、同意度測量，是用來評估不同人(或同人的多次)對於同一事件的判斷結果是否相同。舉例而言，在醫學上不同醫生對同一 X 光片的判斷結果是否相同;在農業上不同茶葉品質品評員之等級評判結果是否一致[53]。

今設 P_{ij} 為一個 $\mathbf{r} \times \mathbf{c}$ 的列聯表中，同一事件第 1 人判斷結果為 i 類而第 2 人判斷結果為 j 類的機率，若大家判斷結果皆相同，其機率和為公式（29），m 為

doi:10.6342/NTU201603448

類別項數:

$$P_0 = \sum^{m} p_{ii} \quad (29)$$

若各類別之觀察值為獨立事件，則其判斷結果皆相同之機率和應為公式（30），$P_{i\cdot}$ 為第 i 類邊際機率和，$P_{\cdot i}$ 為第 j 類邊際機率和。

$$P_e = \sum p_{i\cdot} \times p_{\cdot i} \quad (30)$$

故 $P_0 - P_e$ 為實際與獨立判斷結果機率之差，Cohen(1960)以 Kappa 統計值表示同一事件多次判斷結果同意度如公式（31）:

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (31)$$

當 P_0=1 時，而 K=1，即 r × c 列聯表中離對角線之數值皆為 0，同意度非常好。若 P_0=P_e，即 K=0，同意度不良，其判斷的結果完全是由機會造成的獨立事件，也就是毫無同意度可言。因此，K 值會落在-1 與 1 之間，但通常介於 0 與 1 之間，負值的情況很少出現。

以下說明 K 值如何解讀：

1. K=1：同意度完全一致。
2. K=0：同意度完全不一致。
3. K<0.4：同意度低。
4. 0.4≤K<0.8：中等同意度。
5. 0.8≤K：理想同意度。

我們也可以測試實測 K 值(因是樣本資料)是否為 0。由 K 之變方，如公式(32):

$$V(K) = \frac{1}{n(1-P_e)^2}[P_e + P_e^2 - \sum p_{i\cdot} \times p_{\cdot i}(p_{i\cdot} + p_{\cdot i}) \quad (32)$$

K 之標準化值

$$Z = \frac{K}{\sqrt{V(K)}} \quad (33)$$

若實測 $Z < Z_{\alpha/2}$，則表示 K=0，反之則 K>0。

doi:10.6342/NTU201603448

第九節　情緒分析

文本情緒分析 (Sentiment Analysis)，又可稱為文本情感分析、意見挖掘，指的是用自然語言處理、文本探勘以及計算機語言學等方法來識別和提取原素材中的主觀信息。而其目的是為了找出說話者、作者在某些話題上或者針對一個文本的觀點態度。這個態度或許是他或她的個人判斷或是評估，也許是他當時的情感狀態（就是說，作者在做出這個言論時的情緒狀態），或是作者有意向的情感交流（就是作者想要讀者所體驗的情緒）[55]。

學者 Yang（2007）認為情緒的分類，更能使我們確定對特定事件的個人感情[56]，且 Lin,Chen,Sun 及 Tsai 也認為若有一個系統要能夠識別人類的情感，必須明白人類表達各種情緒的方式 [57]。

以目前情緒辨識研究中，主要是以「表情」、「語音」、「生理資訊」及「文字」四個為主流[58]，因此每種方式都有一定的特徵來代表當下的情緒。如:「表情」的特徵為臉部變化或肢體的變化;「語音」的特徵為聲音的起伏或音調的變化，經由統計的講話音調和強度輪廓作為在講話中表達的情感特徵;「生理」資訊的特徵為生理結構上的數值變化;「文字」的特徵為情緒字彙，利用特徵加上運算將結果以情緒方式呈現。

以下整理這 10 年大部分學者在文本情緒分析上所做的幾個常見的方法[59]:

1.基於文件為基礎的情緒分類（Document-based Sentiment Classification）。

2.以主觀的概念做情緒分析（Subjectivity and Sentiment Classification）。

3.以外觀（屬性）為基礎的情緒分析（Aspect-based Sentiment Analysis）。

4.建立情緒字彙的情緒分析（Lexicon-based Sentiment Analysis）。

5.其他關於情緒分析有趣的議題，包含線上評論的評等預測（Review Rating Prediction）、研究意見比較（Comparative Opinions）、以及意見垃圾偵測（Opinion Spam Detection）等。

因此，本研究透過「建立情緒字彙」的方法來進行情緒分析，利用台灣大學自

doi:10.6342/NTU201603448

然語言處理實驗室由所建立的語意辭典，其將詞彙分為正面及負面情緒，總共一萬多筆詞彙。若假設文章中有一詞彙出現在正面情緒的辭典中記為+1，反之則記為-1。最後計算此文章的總得分數，若分數大於 0 表此文章為正面情緒，等於 0 為中立情緒，小於 0 則代表負面情緒。將情緒簡化至三大類別，分別為正面、中性、負面，以試圖捕捉建言者與回覆者的兩方情緒。

doi:10.6342/NTU201603448

第四章 實證分析

第一節 資料說明

一. 資料蒐集

1. 資料查詢方式

從國立台灣大學網站 myNTU，進入意見交流→校務建言→登入帳號密碼→台灣大學校務建言系統，頁面如圖 4-1-1 所示，即可進行建言或查詢歷年的建言資料。其網址連結如下:

https://my.ntu.edu.tw/sysView.aspx?url=https://mis.cc.ntu.edu.tw/suggest/asp/list.asp

2. 資料轉錄

登入台大校務建言系統後，畫面如圖 4-1-2 所示。其中有金玉集、搜尋、我的建言及提出建言四個功能可以選擇。以下就分別對不同功能做個介紹:

(1) 金玉集:

蒐集從民國 94 年至今的所有建言資料，擁有台大 myNTU 系統帳密的人員皆可以查看，其欄位有編號、類別、建議主旨、提交時間、處理情形和人氣。

點擊任一個金玉集中的建言主旨，即可查看完整的建言內容，包括建議者身分、建議議題類別、主旨、建議內容、處理情形、回覆內容和回覆時間，如圖 4-1-3 所示。

(2) 搜尋:

輸入想要的關鍵字後，即可搜尋與關鍵字相關的主旨、建議內容、回覆內容及提交時間。

(3) 我的建言:

自己所提過的建言內容，可在此功能中被記錄。

doi:10.6342/NTU201603448

(4) 提出建言:

可利用此功能提出建言，需填寫的欄位有建議者身分(此欄系統會利用帳號自行判斷)、建議議題類別(有校務、教務、學生事務、總務、計資中心、圖書館、學雜費、體育室和其他，共有 9 個選項可以選擇)、輸入回覆信箱、主旨及建議內容。

3. 資料鍵入

將國立台灣大學「校務建言系統」民國 94 年 1 月至民國 101 年 12 月校務建言系統，共 4622 筆建言，全部鍵入置 Microsoft Excel 中，如圖 4-1-4 所示，原總檔案共 9716 Kilobyte（KB），詳細資料內容請見附錄三。由於台大校務建言系統，需要帳號密碼才可登入且保密性高，故無法使用爬蟲的方法將資料取得，在此以人工方式一筆筆擷取，因此較為費工且費時。

二. 變數說明

1. 「國立台灣大學校務會議及校務建言系統資料之分析研究」中已有的變數

(1) 建議類別[建言系統預設]: p=1, 2, …, 9

校務、教務、學生事務、總務、計資中心、圖書館、學雜費、體育室和其他。

(2) 建議者身分[建言系統預設]: q=1, 2, …, 6

學生、正式教師、正式職員、校友、公務帳號和短期帳號。

(3) 建議主旨[自訂]: r=1, 2, …., 17

腳踏車、重大事件、宿舍、汽車、環境、維修、教務、體育館、網路訊號、安全、人事、圖書館、帳目、電腦系統、學生活動、醫院和其他。

(4) 建言次數: $\boldsymbol{n_p}$ 、$\boldsymbol{n_q}$ 、$\boldsymbol{n_r}$

建議類別，校務建言系統的總建言次數。

建議者身分，校務建言系統的總建言次數。

建議主旨，校務建言系統的總建言次數。

(5) 平均處理日數: $\frac{\sum_p(\textit{回覆時間}-\textit{提交時間})}{n_p}$ 、$\frac{\sum_q(\textit{回覆時間}-\textit{提交時間})}{n_q}$ 、$\frac{\sum_r(\textit{回覆時間}-\textit{提交時間})}{n_r}$

建議類別，校務建言系統的平均處理日數。

doi:10.6342/NTU201603448

建議者身分，校務建言系統的平均處理日數。

建議主旨，校務建言系統的平均處理日數。

(6) 平均人氣，定義為平均點閱次數: $\frac{\sum_p(\textit{點閱次數})}{n_p}$ 、$\frac{\sum_q(\textit{點閱次數})}{n_q}$ 、$\frac{\sum_r(\textit{點閱次數})}{n_r}$

建議類別，校務建言系統的平均點閱次數。

建議者身分，校務建言系統的平均點閱次數。

建議主旨，校務建言系統的平均點閱次數。

2. 本研究所新增的變數

(1) 建言者字數:

文章的字數，其中包含標點符號。

(2) 回覆者字數:

文章的字數，其中包含標點符號。

(3) 建議內容:

建言者所提出的建言內容。

(4) 回覆內容:

回覆者所提出的建言內容。

(5) 餘弦值_TF-IDF_100 維:

透過潛在語意分析在中文語意空間建置上，詞彙權重計採用 TF-IDF 並降維至 100 維，利用此中文語意空間計算建言者與回覆者文章間的餘弦值。

(6) 餘弦值_TF-IDF_300 維:

透過潛在語意分析在中文語意空間建置上，詞彙權重計採用 TF-IDF 並降維至 300 維，利用此中文語意空間計算建言者與回覆者文章間的餘弦值。

(7) 餘弦值_Log-Entropy_100 維:

透過潛在語意分析在中文語意空間建置上，詞彙權重計採用 Log-Entropy 並降維至 100 維，利用此中文語意空間計算建言者與回覆者文章間的餘弦值。

(8) 餘弦值_Log-Entropy_300 維:

doi:10.6342/NTU201603448

透過潛在語意分析在中文語意空間建置上，詞彙權重計採用 Log-Entropy 並降維至 300 維，利用此中文語意空間計算建言者與回覆者文章間的餘弦值。

(9) 人工標記: answer=1,2,3

利用人工判別的方式，來判斷回覆者是否有回應到建言者的問題。

(10) 建言者情緒:response_s=1,2,3

透過情緒分析計算此文章為正面、中立、負面的哪一種情緒。

(11) 回覆者情緒: opinion_s=1,2,3

透過情緒分析計算此文章為正面、中立、負面的哪一種情緒。

doi:10.6342/NTU201603448

圖 4-1-1 台大校務建言登入頁面

doi:10.6342/NTU201603448

圖 4-1-2 台大校務建言首頁

doi:10.6342/NTU201603448

建議者身份	學生
建議議題類別	圖書館
主旨	公共環境冷氣過冷
建議內容	您好，最近常常到社科院圖書館二樓自習室自習，發現冷氣溫度常常設置在22度，不僅過冷非常不舒服而且又非常浪費能源，希望校方能全面檢視校內所有單位的空調溫度設置，讓環境更舒適且又能節能減碳，謝謝！
處理情形	處理完成
回覆單位	圖書館
回覆內容	同學您好： 您所反應的情況經館員查看，發現C區空調溫控設定被同學擅自調整到22度，其他區域正常。 目前已調回應有的設定，本組也請負責自習室管理同仁在巡視環境時再留意這些分離式空調開關的設定是否有異狀。 若仍有被擅自更改的情況，請向一樓櫃台值班人員反應，我們會立即前往處理。 感謝您的建議，敬祝 學安 社會科學資源服務組 敬覆 聯絡人：林編審 (分機55605)
回覆時間	2016/5/5

圖 4-1-3 建言完整內容

doi:10.6342/NTU201603448

	A	B	C	D	E	F	G	H	I
1	編號	主旨	建言內容	回覆內容	建議類別	建議者身分	提交日期	回覆時間	點閱次數
2	5	經濟系選課問題(教務)	我是經濟四的同學。自從	敬啟者:本案已將建言列印送	教務	學生	2005/9/30	2005/10/4	148
3	13	建議更換校內不符人體工	校內的某種木製座椅,普通	94.10.4 本案已將建言列印送	教務	學生	2005/9/30	2005/10/12	61
4	14	校總區的路牌	近年來校總區興建了許多	您好：您所提之意見，校規	總務	學生	2005/9/30	2005/10/5	49
5	15	新生停車場照明問題	在靠近普通大樓的出口種	敬啟者：本案已通知廠商改	總務	學生	2005/9/30	2005/10/6	19
6	16	關於校內福利社賣菸	印象中校園內應該是全面	同學，您好有關學校福利社	總務	學生	2005/10/1	2005/10/5	96
7	22	雙修輔系(教務)	記得新校長曾說過要降低	1.基於教師教學負擔及各學	教務	學生	2005/10/1	2005/10/14	53
8	23	雙修輔系(教務)	記得新校長曾說過要降低	與第22則建言相同.	教務	學生	2005/10/1	2005/10/12	55
9	25	法學院圖書館二樓的廁所	法學院圖書館二樓的廁所	1.已於9月底接獲訊息，並扌	總務	學生	2005/10/1	2005/10/7	22
10	28	改善現行校車搭載效率	關於現行校車情況，身為	您好：關於本校交通車搭載	總務	學生	2005/10/1	2005/10/13	19
11	30	建議取消填寫教學意見調	個人覺得這樣的類似獎勵	敬啟者:為提昇教學品質,特	教務	學生	2005/10/1	2005/10/4	138
12	31	請問校總區排球場的夜間	相關單位您好！開學到現	您好：感謝您使用本系統，	總務	學生	2005/10/1	2005/10/11	17
13	37	關於合作社以前賣過的那	學校合作社（郵局旁邊）	關於合作社賣過很棒的傘一	總務	學生	2005/10/2	2005/10/5	103
14	42	有關於管理學院選課的問	對於我們這些非管理學院	敬啟者:本案已將建言列印送	教務	學生	2005/10/3	2005/10/4	58
15	43	校園馬路命名	校園馬路命名:原因:校園	您好：您所提的意見，經校	總務	正式教師	2005/10/3	2005/10/5	117
16	48	籃球場夜間燈光	燈光常常沒有開放夜間運	您好：感謝您使用本系統，	總務	學生	2005/10/4	2005/10/5	47

圖 4-1-4 將資料鍵入至 Microsoft Excel

doi:10.6342/NTU201603448

第二節 描述性統計

本研究以台大校務建言的文字資料，即建議內容及回覆內容兩者進行文字探勘，共 4622 篇校務建言。首先，對校方及學生的文章長度做探討，計算每篇內容的使用字數，其中標點符號也包含在字數內，結果如表 4-2-1 及圖 4-2-1 所示。

由表及圖可以得知，校方及學生在文章字數的使用上差異不大，兩者最常使用的字數為 1 至 600 字左右。其中有兩個地方值得注意，第一個，校方有一篇回覆字數為 0 是唯一一篇學校沒有任何回覆的建言，因此對照原始資料後發現，此篇建言於 2006 年 8 月提出，可能校方未熟悉系統故將回覆內容打在處理情形上所導致，內容如圖 4-2-2 所示。

第二個，校方在字數 1801-1900 時篇數竟然高達 74 篇，根據圖 4-2-1 即可明顯看出此異常值。其中字數為 1868 時有共 64 篇、字數為 1870 時共有 9 篇，因此對照原始資料後發現此 64 篇及 9 篇所探討的建言主旨議題相同，皆「反對在永久綠地上開闢永久道路」，而校方對於這個議題的回覆上每篇完全相同，詳細內容如圖 4-2-3 所示，至於 1870 字與 1868 字的差別僅在於此 9 篇有回覆開頭要加上建言同學的姓氏及逗號而已。此舉動引起學生們的不滿，雖然校方在回覆中大篇幅地詳盡說明這個問題，但對於這麼多學生所提的建言，校方每篇的回覆皆為一致，使得學生認為校方只是在官方式的面對此議題。

doi:10.6342/NTU201603448

表 4-2-1 字數長度比較表

字數	學生篇數	學校篇數
0	0	1
1-100	1173	1001
101-200	1434	1459
201-300	788	887
301-400	413	494
401-500	284	299
501-600	163	144
601-700	96	69
701-800	62	70
801-900	54	24
901-1000	33	33
1001-1100	22	21
1101-1200	18	18
1201-1300	14	4
1301-1400	17	4
1401-1500	6	2
1501-1600	6	5
1601-1700	6	0
1701-1800	3	3
1801-1900	4	74
1901-2000	1	1
2001-2100	4	1
2101-2200	5	0
2201-2300	5	2
2301-2400	0	1
2401-2500	0	1
2501-2600	1	1

字數	學生篇數	學校篇數
2601-2700	2	0
2801-2800	1	1
2901-2900	1	0
2901-3000	0	0
3001-3100	1	0
3101-3200	0	0
3201-3300	0	0
3301-3400	1	0
3401-3500	0	0
3501-3600	1	0
3601-3700	0	2
3701-3800	0	0
3801-3900	0	0
3901-4000	0	0
4001-4100	1	0
4101-4200	0	0
…..	0	0
…..	0	0
6301-6400	0	0
6401-6500	1	0
14734	1	0

doi:10.6342/NTU201603448

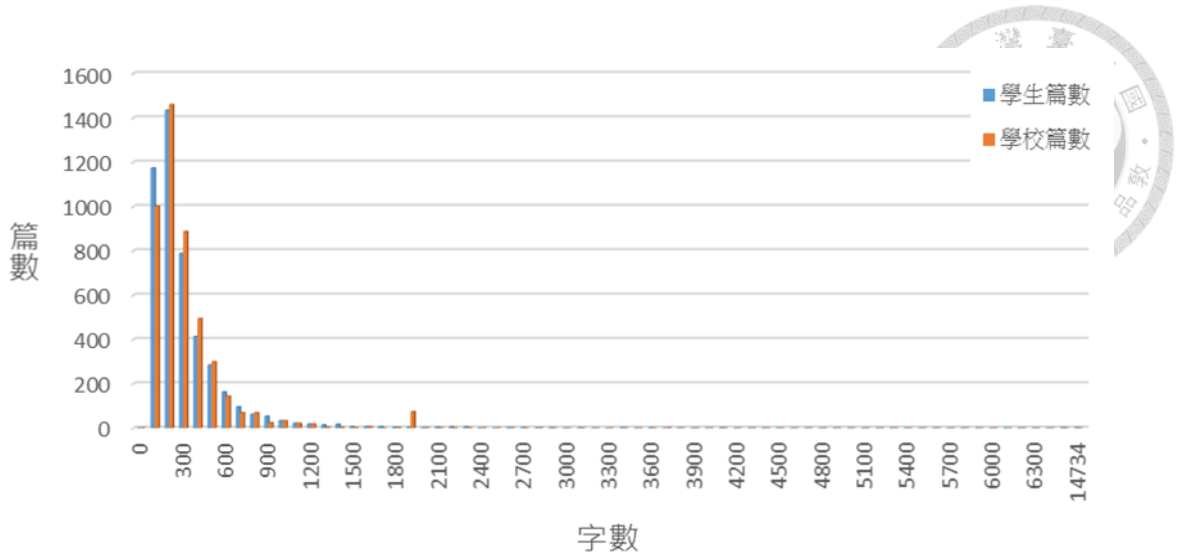


圖 4-2-1 字數長度比較圖

doi:10.6342/NTU201603448

建議者身份	學生
建議議題類別	教務
主旨	關於法律系的轉系辦法，請取消「前學年成績須達全班20%以內」之規定。(第二次建言)
建議內容	一、本人從小就以台大法律為第一志願，但因去年指考前夕太緊張而睡不著，以致沒有考出平日的水準，也未能一償宿願。經過一年的努力，今年轉系放榜，我終於如願以償，轉到法律系司法組了。本人考進外文系的成績算是「吊車尾」，但大一上、下學期學業成績分別拼到第29名與第5名，學年排12名(總算有過20%的第一個門檻)。一年前也不曉得要轉系得先修法律系的課程，但我都有到進修推廣部「旁聽」與憲法與民總有關的法律課程(轉系所指定的考科)。 二、從過來人的經驗(有驚無險)，願挺身而出為今年落榜者或來年有志於轉到法律系的同學請命：我們要的不多，法律系公告轉系的15個名額也不多(本校已放寬為教育部核定名額的20%)，但建議它的門檻可以不要那麼多。 三、有人不以為然，認為今年法律系之轉系，其公告有15名取7名(如連一名同系轉組，共取8名)，其錄取率比同時的轉學考或任何國家考試都高很多？ 四、今年法律系轉學考雖有三百多人報考，但也足額錄取(公告要10名，也取10名)。這二年(94和95年)的轉系考試，其報名人數從以前的三、四百名，驟降到50名以下，就是肇因於「前學年成績須達全班20%」這條緊箍咒。 五、該「緊箍咒」不太合理，因冷門的學系遠比熱門學系容易達到。又它不僅讓考生望而卻步，造成報名人數驟降；而且也變相提高了「到考人數的平均分數」(以往只要前學年學業成績80分即可申請平轉，75分可降轉)。以往報名人數多，水準不會那麼高，平均分數會降低。 六、加以報名雙主修者都是「高年級生」，雙主修與轉系又一起考試，一起算平均數？查該雙主修考生，經過多年的秣馬厲兵，都考得很高分(去年二科平均41分多，今年平均48.34，今年竟有人民總考到83分，二科總分高達141分)。 七、每年招生季節，本校上自校長、註冊組長組長到各系的招生小組(校園博覽會)都大聲疾呼要「選校不選系」，還強調「本校轉系、雙修與輔系已逐年鬆綁」？但查本校財金系與法律系不但沒有提出「配套措施」，甚至「背道而馳」？ 八、請 法律系的教授們能廣開「方便之門」，以「入學從寬，考核從嚴」精神取消「前學年成績須達全班20%以內」之規定。像本校電機系的申請轉系，其「前學年成績只須達75分」而已，該系今年核准轉系高達34名之多。
處理情形	處理完成,本建議案於2006/8/30轉寄法律學系處理,法律學系已於2006/11/1回覆
回覆單位	法律學系
回覆內容	
回覆時間	2006/11/1

圖 4-2-2 未回覆之資料

doi:10.6342/NTU201603448

回覆內容	同學您好： 未來社會科學院與法律學院回歸後，該區將會有近一萬人的活動強度，對於全校三萬名師生的影響為何，這是學校必須要去考量的，也必須從交通（腳踏車、車行及人行）、環境景觀、防災、安全、設施與設備維護、學生活動及與台北市的影響等面向去考量，取得平衡讓學校可以正常運作，而非針對單一議題與單一對象考量。關於道路的設置與綠地保留之問題，總務處秉持一貫立場，並非要設置永久道路，而是要看辛亥門至長興街是否打通，或視社科學院與法律學院回歸後狀況而定回復綠地時程，試想一個人血路如果不通，怎能不生病；同樣地，學校動線也是一樣，學校有服務車輛需求，上下課大量腳踏車通行需求，設施維護需求，電機資工應該也有設備設置或更新需求，若這些需求可以被忽視，那的確可以不用設臨時道路，但是可能嗎？故因此要總務處承諾社科完工後廢除臨時道路，在不確定是否可行的情況下，是有困難的，這個後果也不是總務處可以承擔的。總務處當然可以為避免衝突一口答應社科完工就回復綠地，但這是不負責任的作法。至於是否另開出入口或從椰林大道進出等其他方案，其實都有考慮過，而且各有不可抗力的因素，以下也稍做說明： 1.另開出入口：工地鄰近市區道路，可申請臨時出入口，但也只限施工車輛進出；若要申請一般出入口，依過去交通局的立場是不可能同意的，因為如此會影響市區道路的車輛，區域所有紅綠燈秒差都需調整。本校歷年工程案件經驗，鄰近道路的館舍若另設出入口，都市設計審議都不會通過。 2.施工車輛從其他出入口進出：椰林大道是臺大人共同的記憶，對於外賓或訪客來說，椰林大道是臺大的門面，高聳的椰子樹與平坦的大道也代表著臺大在學術界的崇高地位，因此當初改善完成後，總務處才禁止在椰林大道挖埋管線，施工車輛禁止進入（除非必要，也要申請），如果要開放也是可行，只是椰林大道是柔性路面，經不起重車滾壓，將恢復成以前一樣到處坑坑洞洞，下雨就一灘一灘積水；舟山路為景觀道路，寬度有限，因此就目前來說一四四巷與長興街之出入口並不適合施工車輛通行，因此臨時道路不設，可能會改走獸醫與電機系的道路，反而會更加危險。其實，大型的施工車輛出入校園主要在開挖期，期間約10天，會嚴格要求路口安全導護。 3.一般車輛禁止入校：如果社科停車場完成，且水杉道至復興南路大門之間道路打通，取消校內地上車格，全校教職員工生對禁止一般車輛進入校園有共識，並且不再改建或新建校舍以減少交通量的增加，如此只要處理既有行人、腳踏車、服務車輛與緊急車輛通行的問題，或許可以考慮取消臨時道路。但本校師生人數持續成長，且許多教學研究空間不足或老舊亟需更新，可預期校內仍將陸續進行建築整修、改建、或新建工程，校內交通量仍難調降（即使一般車輛不能進入校園，仍有行人、腳踏車、服務車輛、工程車輛、緊急救災救護車輛等需道路通行），故以目前的情形而言，短期內似乎不易做到。學校的政策是人本交通，逐步進行停車外圍化、地下化，希望盡快做到除了少數例外（經常性的服務車輛、工程車輛、身障者車輛、臨時性的施工車輛、救災救護車輛），校園內沒有動力車輛，把校園還給我們師生員工。那時候所謂道路不再是馬路、車路，而是人的行走與活動場域。 綜觀上述，綠地保留並不是不可行，也不是總務處不願意承諾，而是時機尚未成熟，保留綠地、維護生態環境，原本就是本校教職員工生責無旁貸的事，但是因配合社科大樓興建案而取消了既有道路（語言中心前），為了整體交通動線的考量，不得不尋覓一條替代道路。況且，可預見的是未來東區將是校園人口密集活動的區域，屆時的交通問題能不令人憂心而未雨綢繆？就如同舟山路、德田樓、博理館、化學新館、明達館、法律學院等工程一樣，從規劃到施工完成，皆有令人難以忍受的困擾，但對於學校來說，這是一定要歷經的陣痛期，師生抱怨一定會有，總務處也一定會有需改善的地方，但對不同立場的大家來說，卻都有一樣為了讓臺大變得更好的心，因此對於這樣的問題實在還需要大家的諒解。 各位的意見校方都相當重視，這幾天校長、總務長、資工系主任也因應此議題會談交換意見，希望能解決問題，總務長也與學生會會長及多位同學交換意見。目前結論說明如下：（一）闢設替代道路將考慮教學環境及腳踏車停車，在15米路幅提供行人、腳踏車與車輛通行，會盡量考慮少設置停車位，在靠近房屋教室處，不設置路邊停車位；社科院新大樓完工後就不再設置停車位。（二）15米替代道路將在復興南路大門進來之路打通到水杉道/長興街後（經國發所與新聞所直走，拆掉單身宿舍等）檢討存廢；如不廢除，須提校務會議討論。 最後感謝您的來信，謝謝。 總務處 敬覆

圖 4-2-3 重複回覆之資料

doi:10.6342/NTU201603448

第三節　關鍵字提取

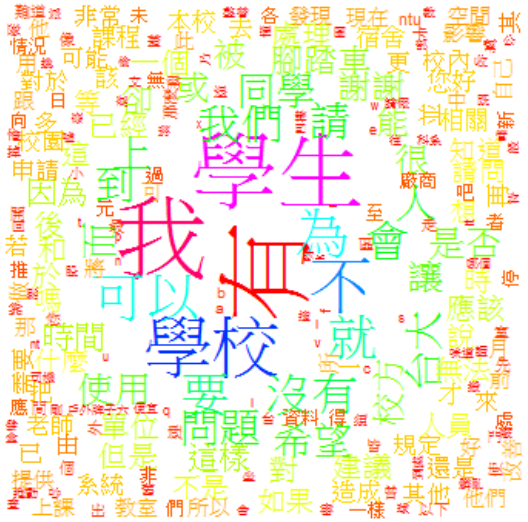
本研究將資料進行文字探勘，先經過中文斷詞及字詞處理後，將所得資料全數轉換成「文件-詞彙矩陣（Document-term Matrix，簡稱 dtm）」的形式，並運用資料視覺化的技術，將資料建構成詞雲（或稱文字雲），透過詞雲的呈現可讓我們迅速瞭解在處理的語料庫中大家所關注的議題及內容為何，也就是哪些詞彙出現次數最多，並以詞彙出現頻率多寡來顯示該詞彙的大小，來衡量詞彙的重要性。

在本研究中以先去除幾個使用頻率高卻沒什麼代表意義的詞，即稱為停字詞（Stop Word）其包含「的、了、是、在、之、也、但、又、都、以、及、與」，並選擇詞彙出現超過五次以上的字，繪製成詞雲以方便之後分析。

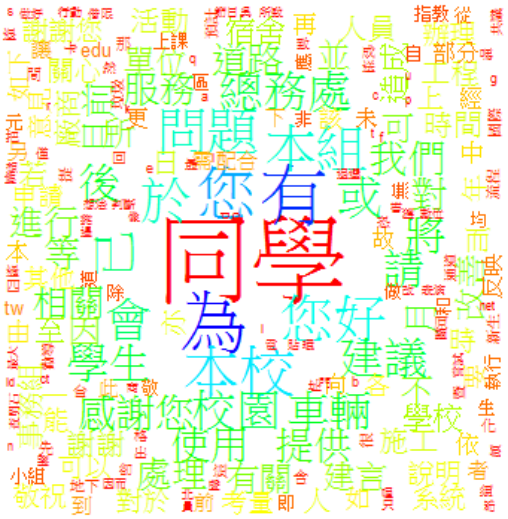
以下本研究將資料整理成三大種形式來分別進行探討:

一. 分為建言內容及回覆內容

1. 建言內容



2.回覆內容



透過詞雲的呈現，可以了解在校務建言系統中，建言者與回覆者之間的用字習慣及文章的內容。在建言者的詞雲中，發現建言者的建言都以「我」字來自稱，訴說「學校」及「學生」的事，而都用「可以」、「有」、「不」、「沒有」來提供建言，從此可以推敲建言者使用校務建言系統內容多為學校「有」哪些地方，「可以」怎樣改進;或者學校哪些地方「不」好及「沒有」注意需改善的地方。而建言的內容

doi:10.6342/NTU201603448

範圍甚廣，包括「腳踏車」、「宿舍」、「課程」、「老師」及「規定」等。

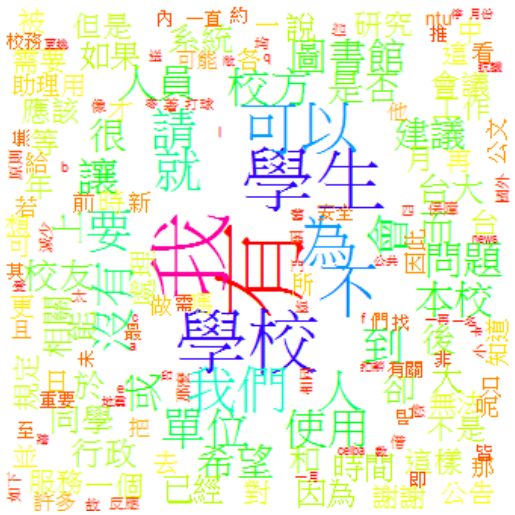
在回覆者的詞雲中，發現建言者的建言都以「您」敬語及「同學」來回稱建言者，由此可見其用詞上是尊敬的，且建言者的身分大多是學生。而回覆者都用「感謝您」、「使用」、「改善」、「提供」、「已」、「會」及「將」來回覆建言，從此可以推敲回覆者回應校務建言系統內容多為「感謝您」「提供」此建言，未來「將」「會」改善;或者此建言「已」處理完畢，可正常「使用」。在回覆的內容範圍甚廣，包括「活動」、「車輛」、「道路」及「宿舍」等。

二. 依照各建議類別，分別探討其建言內容及回覆內容

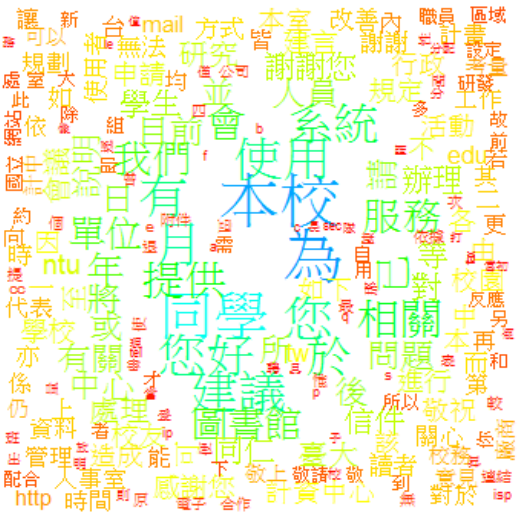
根據「國立台灣大學校務會議及校務建言系統資料之分析研究」此篇論文的量化研究中發現不同建議類別在平均處理日數及點閱次數上有不同，因此本研究想得知在每個類別中建言者最常提出哪些議題的建言，這些議題的內容是否是造成平均處理日數及點閱次數上有不同的原因之一。以下建議類別有校務、教務、學生事務、總務、計資中心、圖書館、學雜費、體育室、其他，共 9 種。

1. 校務:共有 463 筆

(1)建言內容



(2)回覆內容

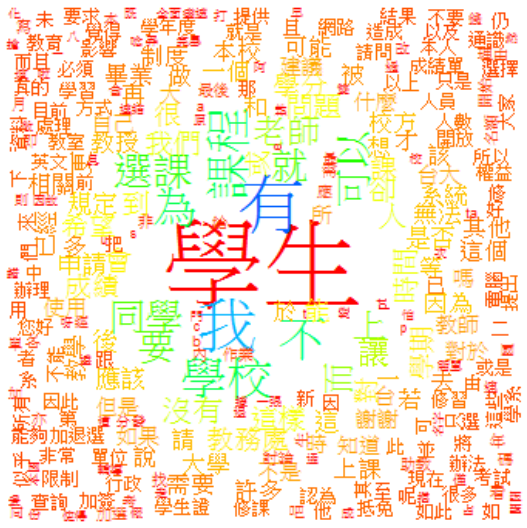


由校務的類別可以得知此類的建言內容議題頗廣，其中有「圖書館」、「校方公文」、「公告」、「系統」、「行政及人員」上的問題。

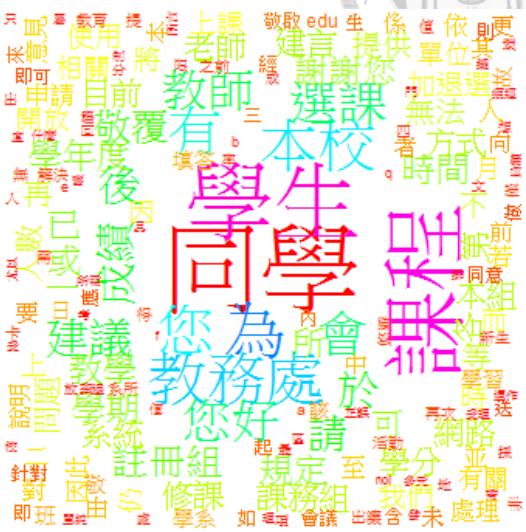
doi:10.6342/NTU201603448

2. 教務:共有 622 筆

(1)建言內容



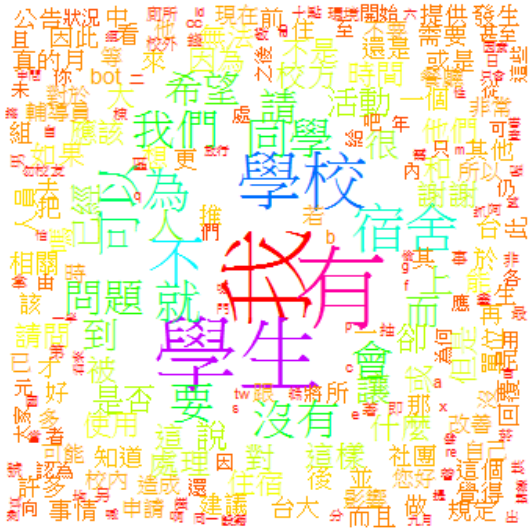
(2)回覆內容



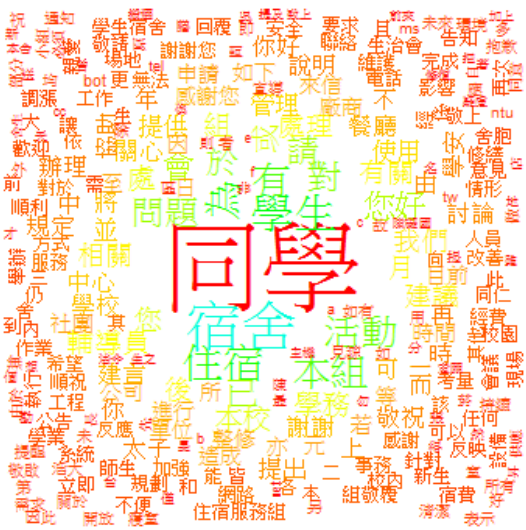
由教務的類別可以得知此類的建言與課程的問題最有關係，有「選課」、「學分」、「英文」、「畢業」等制度規定，其中選課又可細分為「人數開放」、「加退選」、「加簽」及「台大系統」等問題，其次為「老師」或「教授」的問題。

3. 學生事務:共有 748 筆

(1)建言內容



(2)回覆內容

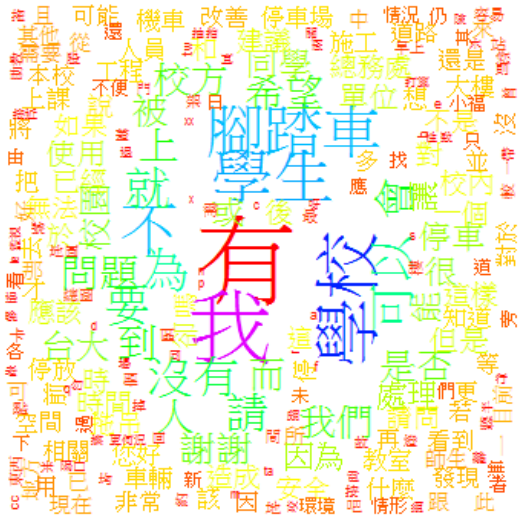


由學生事務的類別可以得知此類的建言最多的為宿舍問題，以校外的「BOT」「宿舍」建言數量最多。除此之外，「社團的活動時間」及「餐廳」等議題也是大家所關切地。

doi:10.6342/NTU201603448

4. 總務:共有 1985 筆

(1)建言內容



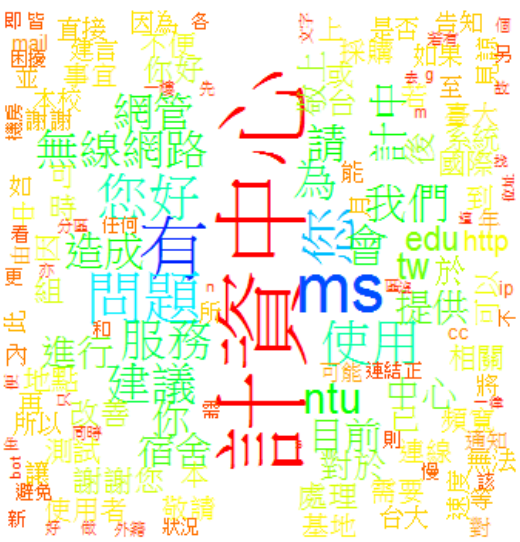
(2)回覆內容

由總務的類別可以得知此類的建言內容以「腳踏車」、「自行車」的問題最多，此外還有「車輛」、「停車場」及「道路」等問題。另外，值得注意總務類的回覆內容上「已」、「將」、「會」這三個詞較其他類別明顯許多，由此可推測總務類的建言相對處理上較為容易或者是總務類的回覆者處理效率較高。

5. 計資中心:共有 132 筆

(1)建言內容

(2)回覆內容



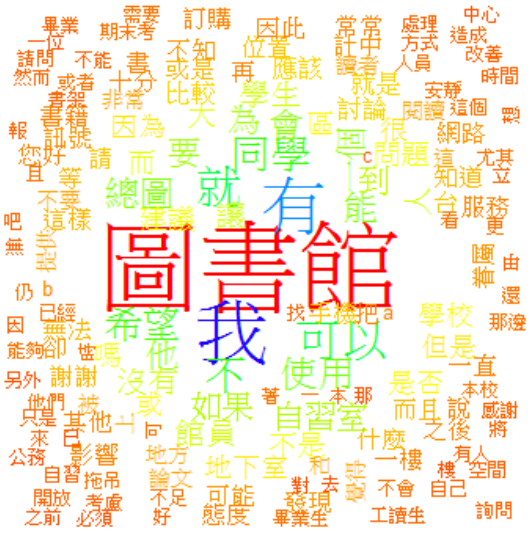
由計資中心的類別可以得知此類的建言內容大部分是「網路」的問題，尤其是「學校」的「無線網路」，可由「ntu」及「peap」兩字看出，因為在台大校園中

doi:10.6342/NTU201603448

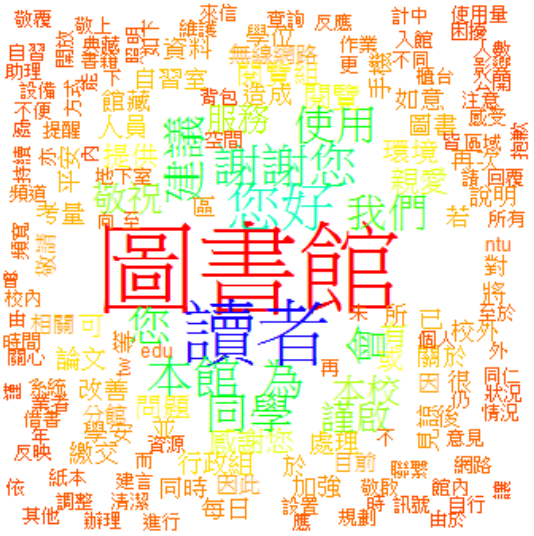
無線網路分別有 ntu_peap 及 NTU 兩個網域可以使用。而網路問題可藉由眾多的否定字「沒有」、「無法」、「不」及「連線」、「訊號」、「速度」等詞，推斷可能是「沒有訊號」、「無法連線」、「不夠速度」的問題。

6. 圖書館:共有 51 筆

(1)建言內容



(2)回覆內容

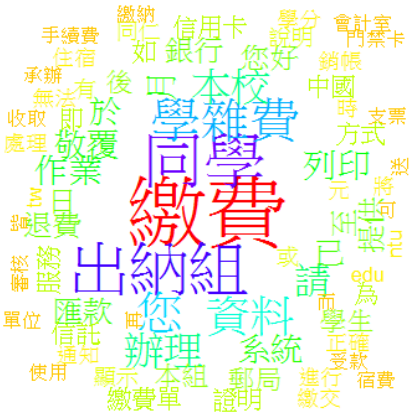


由圖書館的類別可以得知此類的建言地點大多發生在「總圖」，建議內容包含「圖書館」、「自習室」、「書籍」及「館員」等，根據其他詞彙「訂購」、「影響」、「安靜」、「態度」及「服務」，推斷可能是「訂購書籍」、「影響圖書館及自習室的安靜」及「館員的服務態度」的問題

7. 學雜費:共有 14 筆

(1)建言內容

(2)回覆內容



doi:10.6342/NTU201603448

由學雜費的類別可以得知此類的建言內容為「繳費」及「退費」上的問題，而負責此業務的為「出納組」。

8. 體育室:共 54 筆

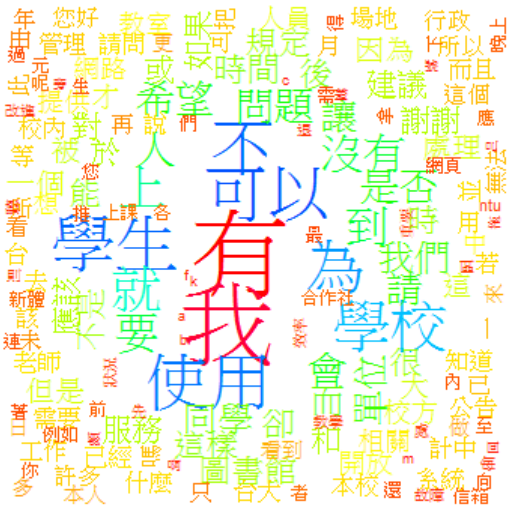
(1)建言內容

(2)回覆內容

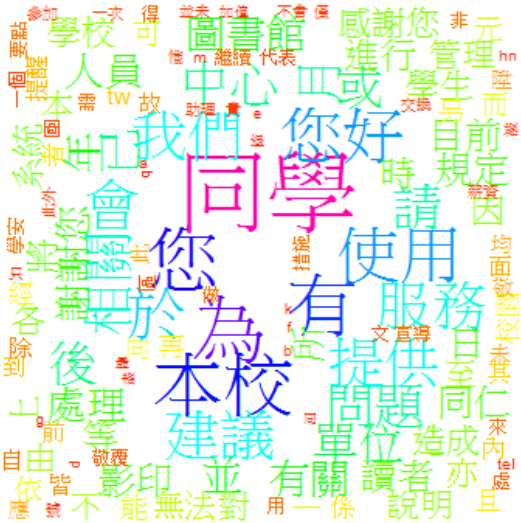
由體育室的類別可以得知此類的建言與「場地」的使用關係最為密切，其中以「新體」及「泳池」使用問題最為嚴重，其次為「人員」的管理及「比賽」的問題。

9. 其他:共有 353 筆

(1)建言內容



(2)回覆內容



由其他的類別可以得知此類的建言內容範圍甚廣，無法明顯得知所主要所關注的議題，有「行政」、「場地」、「圖書館」及「計中」等問題。

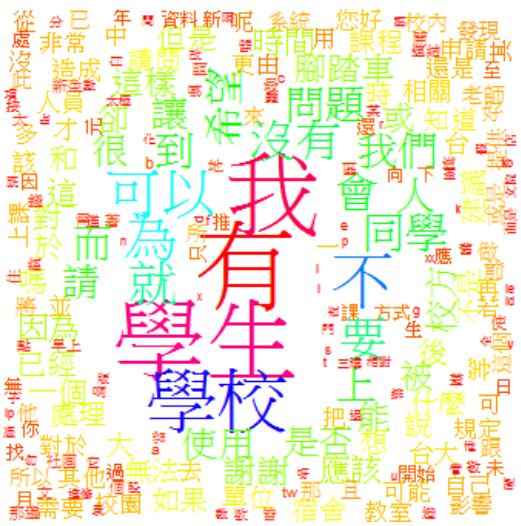
doi:10.6342/NTU201603448

三. 依照各建言者身分，分別探討其建言內容及回覆內容

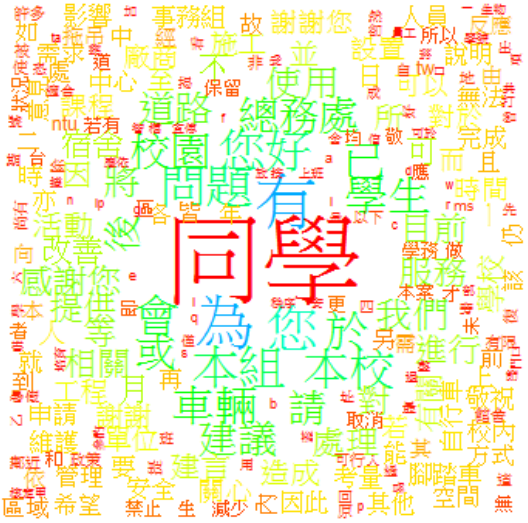
根據「國立台灣大學校務會議及校務建言系統資料之分析研究」此篇論文的量化研究中發現不同建言者身分平均處理日數及點閱次數上有不同，因此本研究想得知在每種不同身分的建言者最常提出哪些議題的建言，這些議題的內容是否是造成平均處理日數及點閱次數上有不同的原因之一。以下建言者身分有學生、正式教師、正式職員、校友、公務帳號及短期帳號，共 6 種。

1. 學生:共有 3687 筆

(1)建言內容



(2)回覆內容

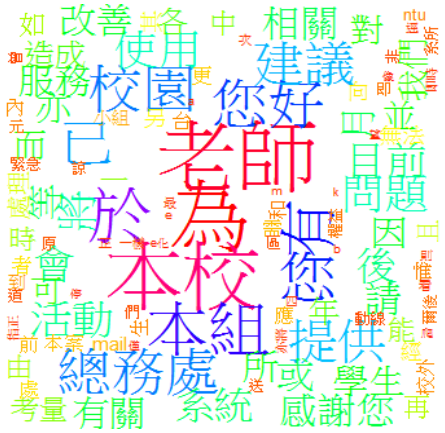


由身分學生的類別可以得知此類的建言內容範圍甚廣，學生們主要所關注的議題，有「腳踏車」、「課程」、「教室」及「宿舍」等問題。

2. 正式教師:共有 252 筆

(1)建言內容

(2)回覆內容

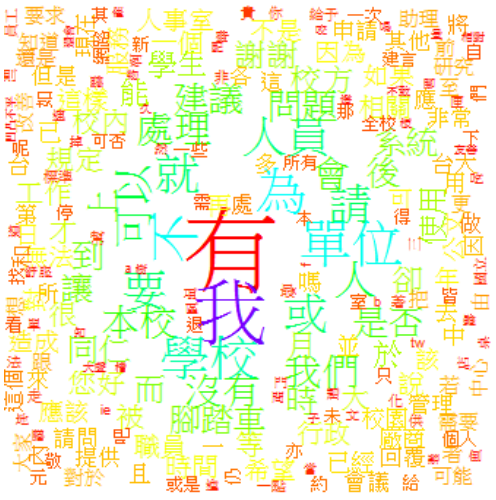


doi:10.6342/NTU201603448

由身分正式教師的類別可以得知此類的建言內容沒有特定議題，老師們所關注的議題，不外乎有學生們也關切的「腳踏車」問題，也有「活動」、「研究」及「停車場」等問題。

3. 正式職員:共有 508 筆

(1)建言內容

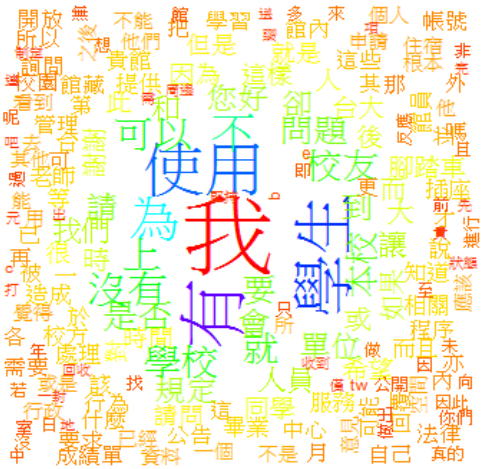


(2)回覆內容

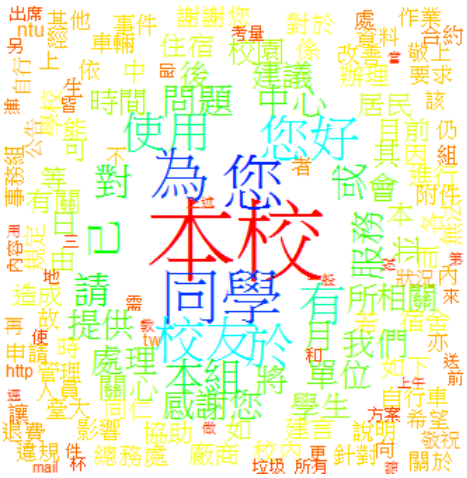
由身分正式職員的類別可以得知此類的建言內容大多與「人(人員)」及「系統」相關，從回覆的內容也可看見來自「人事處」及「事務組」的回覆。

4. 校友:共有 91 筆

(1)建言內容



(2)回覆內容



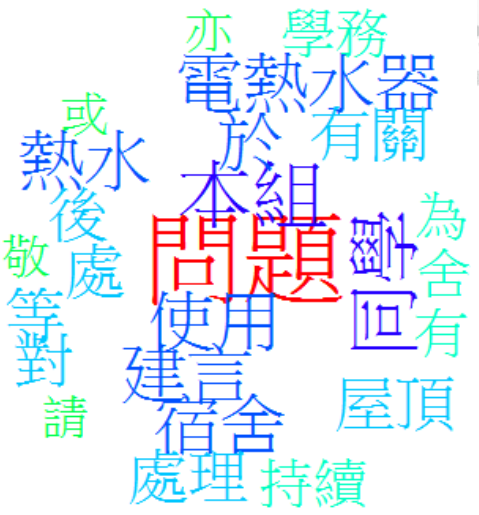
由身分校友的類別可以得知此類的建言內容沒有特定議題，其包含「畢業」後申請「成績單」、使用圖書館「館藏」及校園中「腳踏車」等可以改進的問題。

doi:10.6342/NTU201603448

5. 公務帳號:共有 12 筆

(1)建言內容

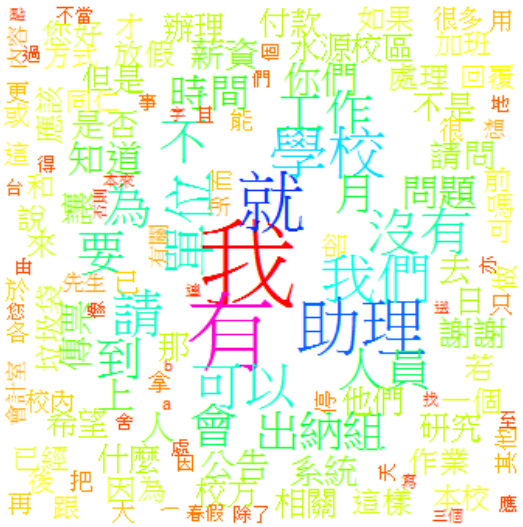
(2)回覆內容



由身分公務帳號的類別可以得知此類的建言者為問題明顯，其原因是此資料筆數稀少所導致，內容包含「熱水」、「路面突出物」及「屋頂雨水」滲入等問題。

6. 短期帳號:共有 72 筆

(1)建言內容



(2)回覆內容

由身分短期帳號的類別可以得知此類的建言者為「助理」居多，且與「工作」相關，其中包含「薪資」、「放假時間」及「系統公告」等的問題。

doi:10.6342/NTU201603448

第四節　建言與回覆之關聯性

本研究欲探討學生所表達的意見，校方是否有真實回答學生所需;換句話說想探討建言者所表達的意見，回覆者是否有真實回答建言者的所需。因此，本研究先透過潛在語意分析的方法來判斷建言與回覆之間的語意相似程度，並透過人工方式判讀回覆者是否有回應到建言者。

將潛在語意分析及人工方式判讀的結果，進行相關分析，探討是否可使用潛在語意分析的結果，來做為校方是否有真實回覆到建言者的問題。

一. 潛在語意分析 Latent semantic analysis

透過潛在語意分析來進行文章與文章的相似度，若回覆者有真正回答建言者的問題時，兩者探討的議題內容可能是相同的，換句話說內容可能會是相似的。反之，若回覆者避而不答或文不對題的狀況，內容可能會是不相似的。

本研究使用之語料庫，原為中央研究院現代漢語標記語料庫 4.0 版，共有 19,247 份文件，11,245,330 詞，其主題涵蓋文學、生活、社會、科學、哲學及藝術。但由於本研究內容較為特殊，關鍵詞彙多為台大校內的專有名詞，故使用中央研究院建置的現代漢語平衡語料庫作為語料庫並不恰當。

因此，本研究使用資料探勘建模時把資料切割成訓練資料（Training Data）和測試資料（Testing Data）的概念，將校務建言資料共 4622 筆資料，依「80:20」比例，利用隨機亂數的方法，分割為「訓練資料:測試資料」，也就是各「3697:925」筆，其中把訓練資料當成本研究的語料庫。

至於本研究的語料庫建置步驟為「建立文件-詞彙矩陣（詞彙-文件矩陣）」、「詞彙權重計算」、「運用 SVD 轉換矩陣」及「維度約化」，詳細的步驟說明請見第三章研究方法的第七節潛在語意分析，在此研究中分別建置四種語意空間分別為詞彙權重計採用 TF-IDF 並降維至 100 維、詞彙權重計採用 TF-IDF 並降維至 300 維、詞彙權重計採用 Log-Entropy 並降維至 100 維、詞彙權重計採用 Log-Entropy 並降維至 300 維。

doi:10.6342/NTU201603448

接著，利用 VSM （Vector Space Model）求兩向量的夾角（餘弦值），即可求得建言與回覆間的相似度，計算結果可見如附錄四，其值皆取到小數第二位，第三位四捨五入。

根據郭伯臣「應用潛在語意分析探究詞彙對語料庫之重要性」研究，指出詞彙權重方法的比較後發現 Log-Entropy 詞彙權重方法在各指標與詞頻、詞彙出現文件數之相關性較高[43]。因此，本研究採用 Log-Entropy 詞彙權重方法，作為詞彙對於語料庫的重要性程度之參考，故使用語意空間為 100 維 Log-Entropy 及 300 維 Log-Entropy 的餘弦值做為結果。

二. 相關分析

本研究將利用潛在語意分析，將語意空間為 100 維 Log-Entropy 及 300 維 Log-Entropy 所得的餘弦值，依據兩種標準各分為三個等級。種類一的標準較為寬鬆，若餘弦值是 0.0-0.2 則為等級 1 其所對應的人工判斷標準為完全離題，0.2-0.4 則為等級 2 其所對應的人工判斷標準為含糊不清回應，0.4-0.1 則為等級 3 其所對應的人工判斷標準為有清楚回應; 種類二的標準較為嚴格，若餘弦值是 0.0-0.25 則為等級 1 其所對應的人工判斷標準為完全離題，0.25-0.5 則為等級 2 其所對應的人工判斷標準為含糊不清回應，0.5-0.1 則為等級 3 其所對應的人工判斷標準為有清楚回應，其等級劃分規則整理如表 4-4-2 所示。

從 925 篇的測試資料中，取前 432 篇做人工標記，其結果請詳見附錄五。將 LSA 所得的餘弦值經由等級劃分後，並與人工判斷的結果，進行相關分析，分別利用 Pearson 相關係數、Spearman 等級相關、Kendall 等級相關及 Kappa 統計量，用以判斷 LSA 的餘弦值與人工判斷的結果是否有關，檢定結果如表 4-4-3。

由結果得知，不論在維度為 100 維或是 300 維，餘弦種類一或是二的情況下，所有相關分析的指標，在 95%信賴區間下其值皆不包含 0，也就是拒絕虛無假設。因此，可知 LSA 的餘弦值與人工判斷的結果是有相關的。LSA 的餘弦值可做為日後評斷回覆者是否有回覆到建言者的建言內容之重要指標。

doi:10.6342/NTU201603448

表 4-4-2 餘弦值等級劃分

等級	餘弦值種類一	餘弦值種類二	人工判斷標準
1	0.0-0.2	0.0-0.25	完全離題
2	0.2-0.4	0.25-0.5	含糊不清回應
3	0.4-1.0	0.5-1	有清楚回應

doi:10.6342/NTU201603448

表 4-4-3 相關分析之結果

維度及種類	相關指標	指標值	95%信賴區間
100 維_餘弦值種類一	Pearson	0.2055	(0.0708,0.3403)
	Spearman	0.1392	(0.0231,0.2553)
	Kendall	0.1357	(0.0229,02485)
	Kappa	0.1243	(0.0202,0.2284)
100 維_餘弦值種類二	Pearson	0.1843	(0.0639,0.3048)
	Spearman	0.1387	(0.0322,0.2452)
	Kendall	0.1345	(0.0314,0.2377)
	Kappa	0.1169	(0.0272,0.2066)
300 維_餘弦值種類一	Pearson	0.1918	(0.0589,0.3247)
	Spearman	0.1313	(0.0187,0.2439)
	Kendall	0.1277	(0.1200,0.2370)
	Kappa	0.1234	(0.0207,0.2261)
300 維_餘弦值種類二	Pearson	0.1496	(0.0314,0.2677)
	Spearman	0.1044	(0.0011,0.2077)
	Kendall	0.1007	(0.0010,0.2003)
	Kappa	0.0899	(0.0069,0.1729)

doi:10.6342/NTU201603448

第五節　情緒分析

透過情緒分析來判斷建言者（學生）與回覆者（校方）兩者在溝通的過程中，兩方的情緒是否達到理性的溝通。

本研究透過「建立情緒字彙」的方法來進行情緒分析，利用台灣大學自然語言處理實驗室由所建立的語意辭典，其將詞彙分為正面及負面情緒，總共一萬多筆詞彙。若假設文章中有一詞彙出現在正面情緒的辭典中記為+1，反之則記為-1。最後計算此文章的總得分數，若分數大於 0 表此文章為正面情緒，等於 0 為中立情緒，小於 0 則代表負面情緒。將情緒簡化至三大類別，分別為正面、中性、負面，以試圖捕捉建言者與回覆者的兩方情緒。

研究結果如圖 4-5-1 所示，共分析 4622 篇建言內容及回覆內容。從圖 4-5-1 中明顯得知，學生在提出建言時情緒較為負面，其比例為 52%；而學校在回答學生的建言時，多以正面的情緒回應學生，其比例高達 75%。

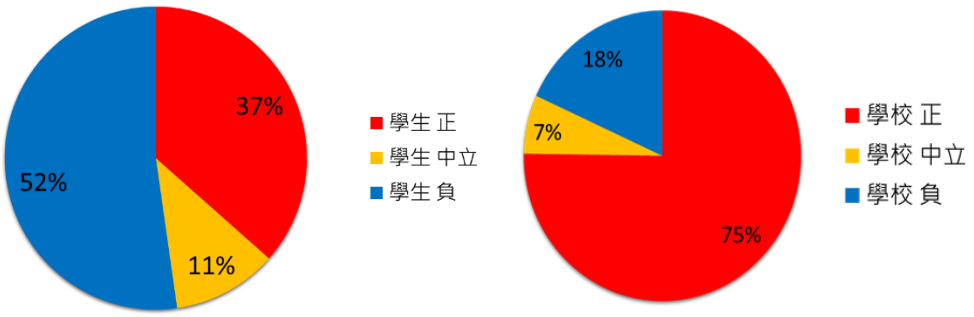


圖 4-5-1　情緒分析結果

接著，本研究進行對等性檢定，針對同一篇建言，建言者與回覆者間的情緒是否是一致的，結果如表 4-5-1 所示。結果其 P-value<0.0001，故兩者是不對稱的，情緒上是不一致的。

doi:10.6342/NTU201603448

表 4-5-1 情緒分析之 3×3 列聯表

學生 / 學校	正	中立	負	和
正	1688	519	1270	1270 (75.22%)
中立	0	0	303	303 (6.56%)
負	0	0	842	842 (18.22%)
和	1688 (36.52%)	519 (11.23%)	2415 (52.25%)	4622

統計值	對稱性檢定
P-value	<0.0001

doi:10.6342/NTU201603448

第五章 結論與建議

第一節　結論

本研究利用西元2005至2012年台灣大學校務建言系統的文字資料，進行文字探勘（Text Mining）分析。首先，利用建言者（學生）及回覆者（校方）的使用字數上進行描述性統計，可以得知兩者的在文長方面沒有太大的差異，除少數極端值以外，在極端的部分明顯發現回覆者（校方）面對相同議題的建言，皆採取同樣的內容做為回答，回應方式十分的官方且制式。

再者，藉由文字雲的呈現可快速了解如此多篇的建言內容所關切的議題為何。建言者大多關注於腳踏車、宿舍、課程、老師及規定等方面，若在細分建言者身分及建議類別，可得知在不同建議者身分及建議類別上，其所關注的議題、事物皆不相同。其中最常被提出建言的行政單位為總務類，其次為學生事務及教務類，在總務方面以腳踏車、車輛、停車場居多；在學生事務方面以宿舍（BOT）、社團活動時間、餐廳居多；在教務方面以選課、學分、英文、畢業居多。而建言者（學生）與回覆者（校方）在用字習慣上，回覆者（校方）用詞是較為禮貌的，其以「您」來做為稱呼，且在文末會回覆「感謝您」。

另外，本研究透過潛在語意分析（Latent Semantic Analysis）計算建言者與回覆者之間的文章相似度（餘弦值），並結合人工判斷結果，利用Pearson相關係數、Spearman等級相關、Kendall等級相關及Kappa統計量進行相關分析，其在統計上達皆顯注差異，因此潛在語意分析（Latent Semantic Analysis）之餘弦值，可用來做為是否有回答的重要依據。最後，利用情緒分析（Sentiment Analysis）可得知校方正面情緒高達75%，學生負面情緒為52%，故可知校方情緒上大致為正面，學生情緒上大都呈現負面。兩者在溝通的過程中情緒的反應是不一致的，其在統計上也達顯著差異。

doi:10.6342/NTU201603448

第二節　研究建議與未來展望

一. 研究建議

在本研究發現可以將潛在語意分析（Latent Semantic Analysis）之餘弦值用來做為回覆者是否有回應建言者的重要依據指標，其結果在統計上皆達顯注差異，但是所得的指標值皆較小，在實務上可能不具其意義。

然而導致此狀況的原因，可能因研究限制上所造成，在人工判斷是否有回應的情況下，由於在金錢及時間的考量下只採用研究者本人來判斷，其可能會有偏差畢竟每個人的判斷標準嚴格程度不同，因此在未來的後續研究者可以找同學或者是師長多人一起來做人工判斷，將人為的干擾因素降低。

另外，在文字斷詞斷句及進行潛在語意分析（Latent Semantic Analysis）時 SVD 所降維的維度也可能影響結果。本研究採用語意空間為 100 維 Log-Entropy 及 300 維 Log-Entropy 所得的餘弦值，其中在語意空間為 100 維的情況下，100 個主成分占所有變異之比例為 111.1721/496.3637=22.4%，表取到 100 維時可解釋全部變異的 22.4%；在語意空間為 300 維的情況下，300 個主成分占所有變異之比例為 221.3605/496.3637=44.6%，表取到 300 維時可解釋全部變異的 44.6%。

因此，可能因為解釋變異的比例不高，導致結果不盡理想，未來的研究者可以在嘗試在不同維度下進行語意空間建置。

二. 未來展望

未來文字探勘透過文字雲的呈現，可幫決策者或分析者快速的了解龐大的文字內容，其所要表達及關切的議題訊息。此外，文字探勘除了使用傳統的自然語言處理技術外，將來可透過深度學習，如:類神經網路（Neural Network）捕捉文本的語境以及詞彙的前後關係，來可以減少對語言知識的依賴，讓系統可以從文本裡學習，不需要或只需使用一點點的處理。這可以幫助我們擴增到各種語言，而且花最少的工程力氣。

doi:10.6342/NTU201603448

在現今社會裡無論是在政府機關、金融業、社群經營上等，皆重視個人化服務及與客戶溝通互動上的滿意度，經本研究結果發現可以透過潛在語意分析（Latent Semantic Analysis）的運用，利用其餘弦值用來做為回覆者是否有回應建言者的重要依據指標，可以幫助各企業及機關來評斷客服部門及社群管理者，是否有以積極的誠意及態度且詳細的回覆客戶所提出的建議及問題。再者，利用情緒分析（Sentiment Analysis）可得知管理者與客戶間是否具有良性的互動溝通。

透過以上種種的文字探勘分析，從質化資料中挖掘可用的資訊。最後，再結合上量化資料，利用數字及文字資料間彼此的互動，如在本研究中就可以結合情緒分析與提交時間，探討夜深時學生們的脾氣是否比較不佳等等。這將使得我們未來在挖掘企業或機關在處理與客戶互動關係中，不僅可以分析數值型資料外，還可透過解了大量的文件資料，從中獲取與客戶相關的知識及需求，這將使得各企業及機關運作上有更好的發展。

doi:10.6342/NTU201603448

參考文獻

[英文參考文獻]

1. Gantz J. & Reinsel D. (2012). The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. IDC Report. Published by International Data Corporation, sponsored by EMC Corporation.

7. Frawley,W. J., Paitetsky-Shapiro, G., & Matheus, C.J. (1991). Knowledge Discovery in Databases: An overview. Communication of the ACM, 39, 1-34.

8. Grupe, F. H., & Owrang, M. M. (1995). Database mining discovering new knowledge and cooperative advantage. Information systems management, 12, 26-31.

9. Fayyad, U. M., Piatetsky,S. G. & Padhraic, S. (1996). From data mining to knowledge discovery in databases. American Association for Artificial Intelligence, 11(5), 20-25.

10. Berry, M.J.A., & Linoff, G.S. (1997).Data mining techniques: For marketing, sales, and customer support. John Wiley & Sons, Inc. New York, NY, USA.

11. Sholom M. Weiss & Nitin Indurkhya (1998). Predictive data mining: a practical guide. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.

12. Kleissner, C.(1998). Data Mining for the Enterprise. Proceedings of the 31st Annual Hawaii International Conference On System Sciences, 295-304.

13. Hand , D. J., Blunt, G., Kelly, M. G., & Adams, N. M. (2000). Data mining for fun and profit. Statistical Sci., 15,111-131.

14. Shaw,M. J., Subramaniam, C., Tan, G. W. E.(2001).Knowledge management and data mining for marketing. Decision Support System, 31(1),127-137.

17. Burges, C.J.C. (1998).A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2, 121-167.

doi:10.6342/NTU201603448

18. Berson, A., Smith, S., & Therling, K. (1999). Building Data Mining application for CRM. McGraw-Hill Companies, New York, NY, USA.

19. Hearst, M.A.(1999).Untangling text data mining. Proceeding ACL '99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics Pages 3-10. College Park, Maryland.

20. Dan Sullivan(2001).Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales. John Wiley & Sons, Inc. New York, NY, USA.

22. Yuen-Hsien Tseng, Yeong-Ming Wang, Dai-Wei Juang, Chi-Jen Lin(2005). Text Mining for Patent Map Analysis. Proceedings of IACIS Pacific 2005 Conference. Taipei, Taiwan.

28. Sproat,R. and Shih,C. (1990). A statistical method for finding word boundaries in chinese text. Computer Processing of Chinese and Oriental Languages, Vol. 4No. 4, 336-351.

42. Martin, D.I., & Berry, M. W.(2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), Handbook of Latent Semantic Analysis. (pp. 35-55). Mahwah, NJ: Lawrence Erlbaum Associate.

45. Berry, M.W., & Browne, M. (2005). Understanding search engines: Mathematical Modeling and Text Retrieval. Philadelphia: SIAM, 2,12-14.

46. Dumais, S. (1991). Improving the retrieval of information from external sources. Behavior Research Methods, Instruments, and Computers, 23, 229–236.

47. Letsche, T., & Berry, M. W. (1997). Large-scale information retrieval with latent semantic indexing. Information Sciences, 100, 105–137.

doi:10.6342/NTU201603448

48. Landauer, T.K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25, 259-284.

49. Berry, M.W., Dumais, S., & O'Brien, G. (1995). Using linear algebra for intelligent information retrieval. SIAM Review, 37, 573–595.

50. Witter, D., & Berry, M. W. (1998). Downdating the latent semantic indexing model for conceptual information retrieval. The Computer Journal, 41, 589–6

51. Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211–240.

52. Cooley, W. W. & Lohnes, P. R. (1971) .Multivariate Data Analysis. Wiley, New York, NY.

56. Yang, Changhua, Lin, Kevin Hsin-Yih, & Chen, Hsin-Hsi. (2007). Emotion Classification Using Web Blog Corpora. Proceeding WI '07 Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence , 275-278. Washington, DC, USA.

57. Lin, Hao-Chiang Koong, Chen, Nian-Shing, Sun, Rui-Ting, & Tsai, I. Hen.(2012). Usability of affective interfaces for a digital arts tutoring system. Behaviour & Information Technology, Volume 33, Issue2, 1-12.

[中文參考文獻]

2. 姚力維(2014)。國立台灣大學校務會議及校務建言系統資料之分析研究。國立台灣大學農藝學系未發表碩士論文。臺北，臺灣。

4. 胡世忠(2013)。雲端時代的殺手級應用：Big Data 海量資料分析。臺北市:天下雜誌。

15. 黃勝崇(2000)。資料探勘應用於醫療院所輔助病患看診指引之研究。南華大學資訊管理研究所未發表碩士論文。嘉義，臺灣。

doi:10.6342/NTU201603448

16. 謝邦昌(2001)。資料採礦入門及應用－統計技術看資料採礦。臺北市：資商訊息顧問股份有限公司。

21. 巫啟台(2001)。文件之關聯資訊萃取及其概念圖自動建構。國立成功大學資訊工程學系未發表碩士論文。臺南，臺灣。

23. 朱怡霖(2002)。中文斷詞與專有名詞辨識之研究。國立臺灣大學資訊工程學系未發表碩士論文。臺北，臺灣。

24. 曾元顯(1997)。關鍵詞自動擷取技術與相關詞回饋。中國圖書館學會會報 59 期 Pages 59-64。

25. 喻欣凱(2008)。運用支援向量機與文字探勘於股價漲跌趨勢之預測。輔仁大學資訊管理學系未發表碩士論文。臺北，臺灣。

26. 林千翔、張嘉惠、陳貞伶(2010)。結合長詞優先與序列標記之中文斷詞研究。中文計算語言學期刊 15 卷 3-4 期 Pages 161 -179。

27. 陳克建、陳正佳、林隆基(1986)。中文語句分析的研究-斷詞與構詞。中央研究院資訊所技術報告 TR86-004。

29. 詞庫小組(1998)。中央研究院平衡語料庫的內容與說明(修訂版)。臺北市:中央研究院資訊科學研究所中文詞知識庫小組。

41. 陳明蕾、王學誠、柯華葳(2009)。中文語意空間建置及心理效度驗證：以潛在語意分析技術爲基礎。中華心理學刊 51 卷 4 期 Pages 415 – 435。

44. 白鎧誌、李政軒、郭伯臣、廖晨惠（2011）。應用潛在語意分析探究詞彙對語料庫之重要性。2011 資訊科技國際研討會，朝陽科技大學。

53. 沈明來(2011)。統計分析與 SAS 應用。臺北市:九州圖書文物有限公司。

54. 沈明來(2007)。實用無母數統計學。臺北市:九州圖書文物有限公司。

58. 林豪鏘(2013)。以 FACEBOOK 塗鴉牆文本分析情緒文字的關係。國立台南大學數位科技學習系未發表碩士論文。臺南，臺灣。

doi:10.6342/NTU201603448

［網頁］

3. Here’s What Happens in 60 Seconds on the Internet。Accessed date: March 05,2016.

http://smallbiztrends.com/2015/12/60-seconds-on-the-internet.html

5. 劃時代的掏金術 Big Data。Accessed date: March 05,2016.

http://www.moneydj.com/topics/bigdata/

6. IBM 海量資料的淘金術。Accessed date: March 20,2016.

http://www-07.ibm.com/tw/blueview/2012oct/8.html

30. 中研院中文詞知識庫小組(CKIP)-中文斷詞系統。Accessed date: March 22,2015.

http://ckipsvr.iis.sinica.edu.tw/

31. GitHub - fxsjy/jieba: 結巴中文分詞。Accessed date: March 30,2015.

https://github.com/fxsjy/jieba

32. JIEBA 結巴中文斷詞。Accessed date: March 30,2015.

https://speakerdeck.com/fukuball/jieba-jie-ba-zhong-wen-duan-ci

33. jiebaR 中文分词。Accessed date: March 30,2015.

http://doc.qinwf.com/jiebaR_v0_7/index.html

34. 國立台灣大學統計教學中心-統計軟體介紹。Accessed date: April 03,2016.

http://www.statedu.ntu.edu.tw/lab/%E7%B5%B1%E8%A8%88%E8%BB%9F%E9%AB%94%E7%B0%A1%E4%BB%8B.asp

35. R 講題分享－SpideR--用 R 自製網路爬蟲收集資料。Accessed date: April 05,2015.

http://programmermagazine.github.io/201311/htm/article6.html

36. 維基百科:大五碼（Big5）。Accessed date: April 10,2015.

https://zh.wikipedia.org/wiki/%E5%A4%A7%E4%BA%94%E7%A2%BC

37. 維基百科: UTF-8。Accessed date: April 10,2015.

doi:10.6342/NTU201603448

https://zh.wikipedia.org/wiki/UTF-8

38. 台大-校總區及其他校區之主要建築物逐棟編碼地理位置對照。Accessed date: April 15,2015. https://www.google.com.tw/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiJ8r2Li9HMAhXFjJQKHe0hDHoQFggbMAA&url=http%3A%2F%2Fhomepage.ntu.edu.tw%2F~cpo%2Fenactment%2F991102.pdf&usg=AFQjCNGIg1iMoavfKXirggQyklJloRreSw&sig2=rJ2f0efNEnceGsU5TxCDVA

39. 院系所課程-台大課程地圖。Accessed date: April 15,2015. http://coursemap.aca.ntu.edu.tw/course_map_all/map.php.htm

40. 國立臺灣大學-行政組織。Accessed date: April 15,2015. http://www.ntu.edu.tw/administration/administration.html

41. 線代啟示錄- SVD 於資訊檢索與文本搜尋的應用。Accessed date: July 08,2015. https://ccjou.wordpress.com/2009/11/04/svd-%E6%96%BC%E8%B3%87%E8%A8%8A%E6%AA%A2%E7%B4%A2%E8%88%87%E6%96%87%E6%9C%AC%E6%90%9C%E5%B0%8B%E7%9A%84%E6%87%89%E7%94%A8/

55. 維基百科:文本情感分析。Accessed date: February 16,2016. https://zh.wikipedia.org/wiki/%E6%96%87%E6%9C%AC%E6%83%85%E6%84%9F%E5%88%86%E6%9E%90

59. 資料科學實驗室: 情緒分析(Sentiment Analysis)的作法與商業價值。Accessed date: February 16,2016. http://dataology.blogspot.tw/2015/04/sentiment-analysis.html

doi:10.6342/NTU201603448

附錄一、詞性對照表

本附錄提供 jiebaR 的詞性標示及中研院平衡語料庫詞類標記集，做為英文字母對照之中文詞性類別。

1. jiebaR

符	詞性	符	詞性	符	詞性
Ag	形語素	j	簡稱略	r	代詞
a	形容詞	k	後接成	s	處所詞
ad	副形詞	l	習用語	Tg	時語素
an	名形詞	m	數詞	t	時間詞
b	區別詞	Ng	名語素	u	助詞
c	連詞	n	名詞	Vg	動語素
Dg	副語素	nr	人名	v	動詞
d	副詞	ns	地名	vd	副動詞
e	嘆詞	nt	機構團	vn	名動詞
f	方位詞	nz	其他專	w	標點符
g	語素	o	擬聲詞	x	非語素
h	前接成	p	介詞	y	語氣詞
i	成語	q	量詞	z	狀態詞

2. 中研院平衡語料庫詞類標記集

精簡詞類	簡化標記	對應的CKIP詞類標記	
A	A	A	/*非謂形容詞*/
C	Caa	Caa	/*對等連接詞，如：和、跟*/
POST	Cab	Cab	/*連接詞，如：等等*/
POST	Cba	Cbab	/*連接詞，如：的話*/
C	Cbb	Cbaa, Cbba, Cbbb, Cbca, Cbcb	/*關聯連接詞*/
ADV	Da	*Daa*	/*數量副詞*/
ADV	Dfa	Dfa	/*動詞前程度副詞*/
ADV	Dfb	Dfb	/*動詞後程度副詞*/
ASP	Di	Di	/*時態標記*/
ADV	Dk	Dk	/*句副詞*/
ADV	D	*Dab*, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	/*副詞*/
N	Na	Naa, Nab, Nac, Nad, Naea, Naeb	/*普通名詞*/
N	Nb	Nba, Nbc	/*專有名稱*/
N	Nc	Nca, Ncb, Ncc, Nce	/*地方詞*/

doi:10.6342/NTU201603448

N	Ncd	Ncda, Ncdb	/*位置詞*/
N	Nd	Ndaa, Ndab, Ndc, Ndd	/*時間詞*/
DET	Neu	*Neu*	/*數詞定詞*/.
DET	Nes	*Nes*	/*特指定詞*/
DET	Nep	*Nep*	/*指代定詞*/
DET	Neqa	*Neqa*	/*數量定詞*/
POST	Neqb	*Neqb*	/*後置數量定詞*/
M	Nf	Nfa, Nfb, Nfc, Nfd, Nfe, Nfg, Nfh, Nfi	/*量詞*/
POST	Ng	Ng	/*後置詞*/
N	Nh	Nhaa, Nhab, Nhac, Nhb, Nhc	/*代名詞*/
Nv	Nv	Nv1,Nv2,Nv3,Nv4	/*名物化動詞*/
T	I	I	/*感嘆詞*/
P	P	P*	/*介詞*/
T	T	Ta, Tb, Tc, Td	/*語助詞*/
Vi	VA	VA11,12,13,VA3,VA4	/*動作不及物動詞*/
Vt	VAC	VA2	/*動作使動動詞*/
Vi	VB	VB11,12,VB2	/*動作類及物動詞*/
Vt	VC	VC2, VC31,32,33	/*動作及物動詞*/
Vt	VCL	VC1	/*動作接地方賓語動詞*/
Vt	VD	VD1, VD2	/*雙賓動詞*/
Vt	VE	VE11, VE12, VE2	/*動作句賓動詞*/
Vt	VF	VF1, VF2	/*動作謂賓動詞*/
Vt	VG	VG1, VG2	/*分類動詞*/
Vi	VH	VH11,12,13,14,15,17,VH21	/*狀態不及物動詞*/
Vt	VHC	VH16, VH22	/*狀態使動動詞/
Vi	VI	VI1,2,3	/*狀態類及物動詞*/
Vt	VJ	VJ1,2,3	/*狀態及物動詞*/
Vt	VK	VK1,2	/*狀態句賓動詞*/
Vt	VL	VL1,2,3,4	/*狀態謂賓動詞*/
Vt	V_2	V_2	/*有*/
T	DE	/*的，之，得，地*/	
Vt	SHI	/*是*/	
FW	FW	/*外文標記*/	
COLONCATEGORY			/* 冒號 */
COMMACATEGORY			/* 逗號 */
DASHCATEGORY			/* 破折號 */
ETCCATEGORY			/* 刪節號 */

doi:10.6342/NTU201603448

EXCLAMATIONCATEGORY	/* 驚嘆號 */
PARENTHESISCATEGORY	/* 括弧 */
PAUSECATEGORY	/* 頓號 */
PERIODCATEGORY	/* 句號 */
QUESTIONCATEGORY	/* 問號 */
SEMICOLONCATEGORY	/* 分號 */
SPCHANGECATEGORY	/* 雙直線 */

doi:10.6342/NTU201603448

附錄二、專有詞庫表

本附錄為本研究的專有詞庫，共有 588 個詞。其內容分為:「校總區及其他校區之主要建築物逐棟編碼地理位置對照」、「院系所課程-台大課程地圖」、「行政組織介紹」及「平常同學間在學校習慣使用的名詞簡稱」

新月台	女八	研一舍
綜合體育館	心理系南館	女一
綜合新體育館	心理系北館	女三
普通教學館	應用力學館	女二
小福樓	漁業科學館	女五
博雅教學館	水產養殖池	大一女
體育館	視聽教育館	研一
游泳池	語言大樓	二號館
原子與分子科學研究所館	計算機及資訊網路中心	三號館
電機一館	人文大樓	農化新館
化學系館	農業陳列館	第二行政大樓
生化科學研究所館	舊圖書館	性別平等教育委員會館
新生教學館	樂學館	行政大樓西側平房
數學研究中心	文學院館	望樂樓
數學館	文學院研究大樓	行政大樓
思亮館	土木系館	進修教育大樓
全球變遷中心	化學工程系館	小小福
海洋研究所館	圖書資訊館	保管組倉庫
天文數學館	綜合教學館	四號館
物理學系	第一學生活動中心	五號館
凝態科學研究中心	一活	農業綜合大樓
凝態中心	活大	共同教學館
凝態館	圖書館	農產品展售中心
梁次震與國家理論中心	農藝館	森林環境暨資源系館
機械工程館	一號館	水工試驗所館
志鴻館	植物標本館	花卉館
工學院綜合大樓	女一舍	水工試驗大樓
女九舍餐廳	女三舍	衛生保健及醫療中心
女九舍	女二舍	航空測量館
女八舍	女五舍	林產館
女九	大一女舍	推廣教育大樓

doi:10.6342/NTU201603448

建築與城鄉研究所館	育成中心 A 棟	男六舍
管理學院教研館	育成中心 B 棟	男一
雅頌坊	育成中心 C 棟	男三
第二學生活動中心	澄思樓	男五
二活	思源樓	男七
管理學院二號館	飲水樓	男八
管理學院一號館	卓越研究大樓	男六
管一	獸醫館	太子學舍
管二	獸醫系三館	長興舍區
尊賢館	電機二館	中華經濟研究院
尊賢會館	德田館	生物技術研究中心
展書樓	資訊工程館	土木研究大樓
地質系館	博理館	國家地震工程研究中心
鹿鳴堂	機械系臨時工廠	環境研究大樓
鹿鳴雅舍	新聞研究所館	全校性實驗動物研究中心
台大附設幼稚園	社會及社工館	資源回收場
生命科學館	國家發展所大樓	園藝教學管理研究室
地理系館	社科院大樓	生醫工程館
浩瀚樓	農機館	芳蘭大厝
氣象館	知武館	農業昆蟲館
大氣科學館	生機二號館	臺大醫院西址
精密溫室	中菲大樓	臺大兒童醫院
自動控溫管理室	霖澤館	醫學人文館
轉殖溫室	國青大樓	基礎醫學大樓
食品科技館	萬才館	聯合教學館
食品研發大樓	工程科學及海洋工程學系館	圓形小劇場
園產品加工館	環境工程館	會議中心暨醫學研究大樓
造園館	嚴慶齡工業研究中心	體育館
綠房子	明達館	醫學院男二女六舍
人工氣候室	動物科學技術學系館	藥學系館
種子研究室	動物醫院	六號館
農業試驗場辦公室	臺大癌醫中心醫院	護理系館
昆蟲館	男一舍	臺大醫院東址
太子學舍	男三舍	公共衛生學院
修齊會館	男五舍	公衛學院教學後棟館
水源舍區	男七舍	女七舍餐廳
行政大樓	男八舍	女四舍

doi:10.6342/NTU201603448

男四舍	萬霖館	生農學院
女四	醉月湖	管理學院
男四	生態湖	管院
藥物科技大樓	水源池	公共衛生學院
藥學大樓	普通大樓	公衛學院
行政大樓	普通教室	電機資訊學院
社科院學生活動中心	共同教室	電資學院
綜合大樓	共同大樓	法律學院
社科院圖書分館	新生大樓	法學院
前排教室	博雅館	生命科學院
後排教室	多功能生活廳	牙醫專業學院
經研大樓	鹿鳴廣場	中國文學系
研究大樓	小福	中文系
校總區	小巨蛋	外國語文學系
社科院校區	新體	外語系
城中校區	舊體	歷史學系
水源校區	總圖	歷史系
徐州路校區	視聽教育館	哲學系
建國校區	DVD	人類學系
竹北分部	PTT	圖書資訊學系
雲林分部	ptt	圖資系
汀洲路	NTU	日本語文學系
舟山路	Webmail	日語系
徐州路	Ceiba	戲劇學系
基隆路	ceiba	戲劇系
新生南路	CEIBA	藝術史研究所
羅斯福路	ceiba	藝術所
辛亥路	BOT	語言學研究所
思源街	bot	語言所
長興街	Bot	音樂學研究所
大椰林	文學院	音樂所
小椰林	理學院	臺灣文學研究所
大小椰林	社會科學院	華語教學碩士學位學程
大椰林道	社科院	獸醫專業學院
小椰林道	醫學院	數學系
大小椰林道	工學院	物理學系
椰林大道	生物資源暨農學院	物理系

doi:10.6342/NTU201603448

化學系	分子醫學研究所	動物科學技術學系
地質科學系	免疫學研究所	畜產學系
地質系	口腔生物科學研究所	動科系
心理學系	臨床藥學研究所	農業經濟學系
心理系	法醫學研究所	農經系
地理環境資源學系	腫瘤醫學研究所	園藝暨景觀學系
大氣科學系	腦與心智科學研究所	園藝系
大氣系	臨床基因醫學研究所	獸醫學系
海洋研究所	轉譯醫學博士學位學程	獸醫系
天文物理研究所	新聞研究所	生物產業傳播暨發展學系
應用物理學研究所	新聞所	生傳系
政治學系	土木工程學系	生物產業機電工程學系
政治系	土木系	生機系
經濟學系	機械工程學系	昆蟲學系
經濟系	機械系	昆蟲系
社會學系	化學工程學系	植物病理與微生物學系
社會系	化工系	植微系
社會工作學系	工程科學及海洋工程學系	食品科技研究所
國家發展研究所	材料科學與工程學系	食科所
國發所	環境工程學研究所	生物科技研究所
醫學系	環工所	生科所
牙醫學系	應用力學研究所	臨床動物醫學研究所
牙醫系	力學所	分子暨比較病理生物學研究所
藥學系	建築與城鄉研究所	植物醫學碩士學位學程
醫學檢驗暨生物技術學系	工業工程學研究所	工商管理學系
護理學系	工工所	會計學系
物理治療學系	醫學工程學研究所	會計系
職能治療學系	醫工所	財務金融學系
臨床醫學研究所	高分子科學與工程學研究所	財金系
臨床牙醫學研究所	農藝學系	國際企業學系
生理學研究所	農藝系	國企系
生化學研究所	生物環境系統工程學系	資訊管理學系
藥理學研究所	生工系	資管系
病理學研究所	農業化學系	高階管理碩士專班
微生物學研究所	農化系	EMBA
解剖學研究所	森林環境暨資源學系	商學研究所
毒理學研究所	森林系	商研所

doi:10.6342/NTU201603448

管理學院高階公共管理組	基因體與系統生物學學位學程	經營管理組
管理學院會計與管理決策組	校長室	教職員住宿服組
管理學院財務金融組	副校長室	駐警隊
管理學院國際企業管理組	秘書室	醫學院總務分處
管理學院資訊管理組	教務處	生活輔導組
管理學院商學組	總務處	住宿服務組
管理學院企業管理碩士專班	學生事務處	課外活動組
復旦 EMBA	研究發展處	保健中心
公共衛生學系	國際事務處	健康中心
公衛系	財務管理處	醫學院學務分處
職業醫學與工業衛生研究所	主計室	職涯中心僑生及陸生輔導組
職工所	人事室	軍訓室
環境衛生研究所	計算機及資訊網路中心	活動中心
衛生政策與管理研究所	計算機中心	心輔中心
公共衛生碩士學程	計資中心	企劃組
健康政策與管理研究所	計中	研究計畫服務組
流行病學與預防醫學研究所	圖書館	產學合作總中心
流預所	出版中心	研究倫理中心
生命科學系	環保暨安衛中心	醫學院研發分處
生科系	環安中心	資源發展組
生化科技學系	教務長	財務管理組
生化系	教務室	新事業發展組
動物學研究所	教務處秘書室	主任
植物科學研究所	招生辦公室	歲計組
分子與細胞生物學研究所	註冊組	會計組
生態學與演化生物學研究所	課務組	基金組
漁業科學研究所	研究生教務組	審核組
生化科學研究所	資訊組	綜合業務組
牙醫學系	醫學院教務分處	醫學院會計組
牙醫系	教學發展中心	任免組
臨床牙醫學研究所	總務長室暨總務處秘書室	考訓組
口腔生物科學研究所	文書組	退撫保險組
微生物與生化學研究所	事務組	綜合業務組
獸醫學系	保管組	醫學院
獸醫系	營繕組	主任
臨床動物醫學研究所	出納組	行政室
分子暨比較病理生物學研究所	採購組	教學研究組

doi:10.6342/NTU201603448

作業管理組	博碩士
資訊網路組	碩博士
程式設計組	雙主修
醫學院資訊組	台灣大學
館長室	台科大
校史館營運組	台大
館藏徵集組	三總
閱覽組	麵包
學科服務組	資源回收
書目服務組	羽球
推廣服務組	e 化
多媒體服務組	教學意見
特藏組	期末意見
系統資訊組	期末教學意見
行政組	
社會科學資源服務組	
醫分館成員	
行政組	
銷售發行組	
編輯出版組	
垃圾筒	
停車格	
在學證明	
助教費	
工讀金	
校務建言	
差勤系統	
加退選	
便當	
上下課	
小黑蚊	
冷氣	
籃球框	
糞便	

doi:10.6342/NTU201603448

附錄三、校務建言原始資料

本研究使用「國立台灣大學校務建言系統」民國 94 年 1 月至民國 101 年 12 月校務建言系統，共 4622 筆建言資料。原總檔案共 9716 Kilobyte（KB），全部鍵入置 Microsoft Excel 中，詳細資料可見光碟。

doi:10.6342/NTU201603448

附錄四、測試資料之相似結果

本附錄為計算 925 篇的測試資料，在四種不同語意空間，其分別為詞彙權重計採用 TF-IDF 並降維至 100 維、詞彙權重計採用 TF-IDF 並降維至 300 維、詞彙權重計採用 Log-Entropy 並降維至 100 維、詞彙權重計採用 Log-Entropy 並降維至 300 維。計算建言與回覆間的相似性（餘弦值）的結果，所有值皆取到小數第二位，第三位四捨五入。

篇數	金玉集編號	100 維 TF-IDF	300 維 TF-IDF	100 維 Log-Entropy	300 維 Log-Entropy
1	13	0.56	0.53	0.59	0.54
2	31	0.67	0.63	0.71	0.7
3	50	0.66	0.58	0.66	0.6
4	53	0.68	0.65	0.7	0.67
5	66	0.87	0.77	0.75	0.69
6	88	0.07	0.07	0.08	0.07
7	99	0.57	0.47	0.54	0.46
8	119	0.43	0.35	0.31	0.29
9	129	0.75	0.63	0.68	0.62
10	134	0.78	0.73	0.73	0.71
11	136	0.58	0.51	0.47	0.45
12	139	0.81	0.72	0.75	0.69
13	159	0.77	0.69	0.77	0.71
14	162	0.83	0.75	0.76	0.7
15	173	0.6	0.49	0.63	0.43
16	177	0.83	0.76	0.61	0.53
17	194	0.82	0.75	0.74	0.71
18	197	0.73	0.65	0.62	0.59
19	201	0.83	0.77	0.74	0.71
20	202	0.83	0.68	0.63	0.59
21	208	0.81	0.78	0.71	0.7
22	214	0.9	0.85	0.9	0.87
23	218	0.46	0.44	0.58	0.56
24	235	0.84	0.82	0.78	0.77
25	237	0.63	0.6	0.63	0.63
26	238	0.41	0.37	0.46	0.45

doi:10.6342/NTU201603448

27	248	0.38	0.27	0.33	0.22
28	252	0.94	0.92	0.97	0.96
29	259	0.64	0.57	0.48	0.47
30	263	0.74	0.68	0.67	0.64
31	277	0.92	0.91	0.93	0.92
32	295	0.87	0.78	0.75	0.73
33	298	0.86	0.8	0.77	0.74
34	299	0.75	0.66	0.61	0.57
35	300	0.68	0.61	0.65	0.62
36	301	0.87	0.83	0.83	0.82
37	302	0.8	0.64	0.67	0.58
38	306	0.79	0.75	0.78	0.75
39	313	0.79	0.7	0.75	0.71
40	321	0.62	0.6	0.64	0.63
41	330	0.4	0.35	0.38	0.36
42	334	0.75	0.7	0.74	0.72
43	335	0.66	0.62	0.65	0.63
44	388	0.75	0.69	0.69	0.66
45	391	0.47	0.45	0.55	0.53
46	399	0.31	0.26	0.36	0.31
47	403	0.46	0.42	0.49	0.48
48	411	0.57	0.55	0.61	0.59
49	412	0.38	0.36	0.48	0.46
50	419	0.55	0.53	0.62	0.6
51	422	0	0	0.05	0
52	423	0.85	0.81	0.87	0.84
53	424	0	0	0.05	0
54	425	0	0	0.05	0
55	432	0.35	0.33	0.44	0.42
56	437	0.82	0.76	0.76	0.73
57	450	0.74	0.71	0.76	0.72
58	453	0.61	0.53	0.56	0.52
59	474	0.11	0.11	0.12	0.12
60	475	0.51	0.48	0.44	0.43
61	489	0.86	0.82	0.78	0.76
62	499	0.28	0.27	0.35	0.34
63	503	0.55	0.5	0.41	0.43

doi:10.6342/NTU201603448

64	509	0.7	0.63	0.73	0.7
65	514	0.71	0.65	0.56	0.54
66	534	0.8	0.74	0.7	0.67
67	535	0.9	0.88	0.87	0.87
68	537	0.89	0.86	0.82	0.8
69	539	0.65	0.51	0.54	0.43
70	546	0.57	0.5	0.56	0.53
71	555	0.77	0.7	0.81	0.77
72	578	0.66	0.6	0.64	0.59
73	579	0.6	0.56	0.61	0.58
74	584	0.76	0.69	0.78	0.67
75	587	0.92	0.88	0.88	0.87
76	611	0.75	0.66	0.61	0.58
77	618	0.17	0.16	0.2	0.19
78	619	0.6	0.56	0.56	0.53
79	622	0.84	0.8	0.8	0.77
80	624	0.71	0.67	0.65	0.64
81	639	0.77	0.66	0.73	0.68
82	642	0.23	0.19	0.23	0.18
83	645	0.82	0.69	0.68	0.64
84	668	0.8	0.73	0.78	0.75
85	675	0.66	0.59	0.64	0.57
86	682	0.71	0.62	0.63	0.59
87	689	0.81	0.76	0.78	0.76
88	700	0.78	0.68	0.75	0.68
89	707	0.72	0.66	0.62	0.61
90	715	0.76	0.69	0.7	0.66
91	728	0.86	0.76	0.83	0.75
92	733	0.62	0.44	0.56	0.44
93	737	0.67	0.63	0.69	0.68
94	738	0.84	0.77	0.75	0.7
95	740	0.6	0.53	0.58	0.54
96	742	0.39	0.3	0.42	0.37
97	748	0.1	0.09	0.11	0.11
98	750	0.64	0.53	0.61	0.58
99	753	0.24	0.1	0.22	0.12
100	757	0.65	0.58	0.52	0.51

doi:10.6342/NTU201603448

101	758	0.46	0.42	0.38	0.38
102	761	0.7	0.66	0.61	0.56
103	775	0.6	0.53	0.46	0.45
104	776	0.3	0.26	0.22	0.16
105	780	0.72	0.67	0.72	0.7
106	781	0.74	0.66	0.67	0.64
107	785	0.65	0.6	0.56	0.55
108	795	0.43	0.38	0.38	0.37
109	796	0.62	0.52	0.5	0.44
110	804	0.6	0.49	0.42	0.4
111	806	0.65	0.58	0.67	0.61
112	808	0.31	0.23	0.28	0.23
113	810	0.62	0.58	0.51	0.49
114	821	0.75	0.65	0.63	0.58
115	828	0.61	0.51	0.43	0.41
116	833	0.48	0.45	0.4	0.39
117	847	0.82	0.78	0.68	0.67
118	858	0.65	0.64	0.66	0.65
119	859	0.83	0.75	0.72	0.7
120	862	0.43	0.36	0.42	0.37
121	865	0.43	0.26	0.31	0.2
122	868	0.86	0.81	0.78	0.76
123	871	0.85	0.8	0.74	0.74
124	873	0.76	0.76	0.64	0.65
125	879	0.77	0.73	0.71	0.69
126	892	0.7	0.63	0.69	0.65
127	909	0.79	0.71	0.75	0.69
128	923	0.88	0.78	0.82	0.78
129	926	0.79	0.76	0.76	0.74
130	931	0.69	0.63	0.76	0.72
131	937	0.5	0.44	0.47	0.45
132	938	0.75	0.64	0.72	0.69
133	940	0.01	0.01	0.11	0.04
134	948	0.9	0.87	0.85	0.83
135	951	0.73	0.67	0.65	0.6
136	956	0.2	0.19	0.25	0.24
137	962	0.92	0.86	0.84	0.82

doi:10.6342/NTU201603448

138	969	0.67	0.54	0.44	0.43
139	972	0.51	0.45	0.4	0.38
140	982	0.68	0.64	0.52	0.52
141	999	0.45	0.4	0.58	0.52
142	1002	0.49	0.41	0.37	0.35
143	1007	0.03	0.02	0.09	0.07
144	1013	0.69	0.64	0.68	0.65
145	1017	0.64	0.57	0.54	0.52
146	1025	0.86	0.8	0.82	0.79
147	1034	0.61	0.58	0.61	0.59
148	1043	0.45	0.37	0.36	0.34
149	1045	0.44	0.4	0.35	0.36
150	1046	0.86	0.85	0.85	0.85
151	1048	0.61	0.48	0.46	0.41
152	1051	0.79	0.7	0.68	0.63
153	1055	0.51	0.11	0.43	0.21
154	1059	0.48	0.46	0.52	0.51
155	1086	0.7	0.65	0.7	0.67
156	1093	0.84	0.77	0.76	0.72
157	1095	0.81	0.72	0.67	0.64
158	1096	0.69	0.64	0.61	0.57
159	1099	0.6	0.55	0.61	0.59
160	1104	0.79	0.73	0.66	0.65
161	1122	0.82	0.77	0.78	0.74
162	1132	0.78	0.74	0.77	0.72
163	1134	0.8	0.68	0.74	0.65
164	1135	0.81	0.77	0.81	0.76
165	1144	0.37	0.27	0.43	0.26
166	1149	0.88	0.8	0.9	0.82
167	1157	0.83	0.77	0.79	0.74
168	1158	0.05	0.05	0.07	0.07
169	1159	0.05	0.05	0.07	0.07
170	1166	0.86	0.84	0.88	0.83
171	1174	0.8	0.75	0.84	0.78
172	1178	0.72	0.64	0.66	0.63
173	1190	0.78	0.58	0.7	0.55
174	1195	0.84	0.76	0.77	0.72

doi:10.6342/NTU201603448

175	1211	0.76	0.67	0.72	0.69
176	1214	0.81	0.77	0.83	0.81
177	1215	0.74	0.68	0.73	0.66
178	1217	0.64	0.58	0.57	0.54
179	1218	0.71	0.66	0.69	0.67
180	1227	0.91	0.88	0.91	0.88
181	1229	0.92	0.88	0.9	0.88
182	1230	0.6	0.5	0.52	0.51
183	1232	0.41	0.37	0.36	0.33
184	1243	0.84	0.72	0.8	0.72
185	1245	0.73	0.68	0.62	0.61
186	1252	0.68	0.6	0.68	0.64
187	1258	0.05	0.04	0.09	0.08
188	1263	0.93	0.88	0.88	0.87
189	1266	0.92	0.85	0.84	0.81
190	1268	0.58	0.46	0.53	0.44
191	1275	0.68	0.56	0.56	0.45
192	1295	0.73	0.66	0.71	0.67
193	1298	0.9	0.83	0.86	0.81
194	1300	0.92	0.87	0.94	0.9
195	1307	0.52	0.46	0.38	0.37
196	1308	0.79	0.71	0.66	0.62
197	1314	0.45	0.39	0.37	0.35
198	1322	0.59	0.5	0.48	0.46
199	1327	0.76	0.69	0.76	0.7
200	1339	0.35	0.39	0.28	0.27
201	1350	0.77	0.69	0.7	0.67
202	1367	0.92	0.84	0.8	0.78
203	1377	0.59	0.55	0.53	0.5
204	1379	0.22	0.21	0.25	0.24
205	1390	0.89	0.82	0.82	0.79
206	1407	0.74	0.69	0.67	0.63
207	1409	0.34	0.27	0.33	0.28
208	1415	0.39	0.34	0.36	0.33
209	1417	0.69	0.55	0.52	0.49
210	1430	0.16	0.12	0.17	0.16
211	1445	0.66	0.62	0.7	0.66

doi:10.6342/NTU201603448

212	1469	0.64	0.59	0.66	0.64
213	1478	0.7	0.67	0.67	0.64
214	1486	0.74	0.67	0.71	0.7
215	1487	0.78	0.68	0.68	0.65
216	1501	0.8	0.76	0.78	0.74
217	1505	0.47	0.32	0.31	0.25
218	1509	0.6	0.49	0.41	0.38
219	1516	0.7	0.66	0.66	0.64
220	1523	0.64	0.57	0.51	0.48
221	1528	0.88	0.86	0.9	0.89
222	1541	0.36	0.31	0.31	0.3
223	1542	0.36	0.31	0.31	0.3
224	1545	0.73	0.67	0.65	0.61
225	1547	0.71	0.67	0.74	0.71
226	1560	0.6	0.52	0.48	0.45
227	1567	0.71	0.68	0.69	0.68
228	1568	0.55	0.51	0.44	0.43
229	1569	0.68	0.64	0.6	0.57
230	1578	0.62	0.57	0.73	0.71
231	1579	0.47	0.41	0.44	0.38
232	1581	0.64	0.56	0.56	0.51
233	1599	0.51	0.47	0.54	0.5
234	1632	0.64	0.58	0.48	0.47
235	1635	0.53	0.39	0.44	0.4
236	1636	0.85	0.75	0.88	0.84
237	1643	0.82	0.77	0.74	0.72
238	1652	0.89	0.84	0.87	0.8
239	1659	0.59	0.53	0.49	0.47
240	1660	0.81	0.75	0.74	0.69
241	1666	0.68	0.62	0.7	0.65
242	1688	0.76	0.7	0.71	0.67
243	1689	0.56	0.43	0.54	0.45
244	1691	0.69	0.63	0.62	0.6
245	1696	0.47	0.38	0.34	0.33
246	1713	0.89	0.84	0.88	0.85
247	1716	0.45	0.35	0.35	0.31
248	1728	0.75	0.72	0.68	0.66

doi:10.6342/NTU201603448

249	1730	0.9	0.83	0.88	0.84
250	1731	0.84	0.78	0.75	0.71
251	1732	0.8	0.75	0.85	0.83
252	1748	0.58	0.53	0.68	0.59
253	1753	0.7	0.63	0.54	0.52
254	1755	0.28	0.28	0.33	0.33
255	1762	0.32	0.25	0.28	0.23
256	1772	0.45	0.4	0.33	0.3
257	1783	0.59	0.53	0.51	0.47
258	1785	0.58	0.57	0.64	0.64
259	1788	0.75	0.7	0.66	0.62
260	1789	0.66	0.6	0.49	0.49
261	1793	0.86	0.82	0.69	0.69
262	1799	0.39	0.35	0.45	0.44
263	1801	0.75	0.67	0.72	0.67
264	1808	0.8	0.75	0.85	0.77
265	1811	0.59	0.54	0.54	0.52
266	1816	0.57	0.53	0.51	0.45
267	1820	0.65	0.6	0.52	0.53
268	1822	0.9	0.87	0.88	0.85
269	1833	0.65	0.57	0.57	0.54
270	1837	0.81	0.75	0.71	0.66
271	1846	0.5	0.5	0.57	0.53
272	1854	0.5	0.47	0.48	0.47
273	1856	0.81	0.74	0.72	0.68
274	1859	0.53	0.48	0.44	0.41
275	1860	0.9	0.87	0.88	0.86
276	1866	0.39	0.28	0.32	0.26
277	1872	0.76	0.69	0.68	0.62
278	1880	0.65	0.61	0.64	0.61
279	1881	0.65	0.59	0.47	0.46
280	1888	0.76	0.65	0.68	0.61
281	1890	0.75	0.67	0.58	0.55
282	1893	0.67	0.48	0.48	0.46
283	1898	0.52	0.46	0.45	0.43
284	1909	0.24	0.22	0.24	0.22
285	1918	0.71	0.64	0.53	0.5

doi:10.6342/NTU201603448

286	1931	0.39	0.35	0.35	0.32
287	1934	0.71	0.61	0.73	0.7
288	1936	0.52	0.45	0.52	0.49
289	1953	0.38	0.32	0.28	0.25
290	1960	0.77	0.72	0.71	0.69
291	1965	0.66	0.59	0.71	0.69
292	1976	0.77	0.7	0.6	0.56
293	1977	0.75	0.69	0.73	0.67
294	1987	0.56	0.52	0.59	0.52
295	1992	0.67	0.62	0.59	0.58
296	2004	0.73	0.62	0.61	0.53
297	2008	0.61	0.54	0.56	0.52
298	2017	0.16	0.13	0.18	0.15
299	2023	0.74	0.67	0.79	0.73
300	2045	0.84	0.73	0.7	0.69
301	2051	0.63	0.6	0.55	0.55
302	2059	0.77	0.73	0.83	0.77
303	2068	0.77	0.75	0.74	0.73
304	2069	0.43	0.31	0.42	0.25
305	2070	0.68	0.59	0.64	0.56
306	2071	0.82	0.76	0.8	0.74
307	2076	0.78	0.66	0.77	0.72
308	2079	0.82	0.74	0.77	0.72
309	2080	0.3	0.29	0.31	0.31
310	2083	0.55	0.49	0.49	0.47
311	2092	0.85	0.7	0.73	0.64
312	2093	0.76	0.68	0.68	0.61
313	2099	0.82	0.79	0.8	0.78
314	2109	0.59	0.55	0.54	0.53
315	2129	0.58	0.5	0.42	0.42
316	2132	0.29	0.23	0.28	0.21
317	2150	0.63	0.53	0.55	0.51
318	2159	0.51	0.38	0.47	0.38
319	2164	0.51	0.38	0.47	0.38
320	2189	0.63	0.6	0.59	0.51
321	2191	0.82	0.77	0.65	0.62
322	2200	0.65	0.56	0.55	0.52

doi:10.6342/NTU201603448

323	2214	0.44	0.3	0.36	0.32
324	2215	0.69	0.55	0.58	0.53
325	2218	0.71	0.58	0.56	0.52
326	2222	0.57	0.49	0.54	0.5
327	2233	0.58	0.54	0.51	0.49
328	2234	0.02	0.01	0.21	0.05
329	2235	0.62	0.61	0.63	0.64
330	2237	0.69	0.64	0.65	0.61
331	2246	0.61	0.56	0.51	0.48
332	2247	0.28	0.22	0.17	0.16
333	2248	0.67	0.61	0.58	0.56
334	2262	0.91	0.86	0.88	0.87
335	2265	0.75	0.66	0.7	0.65
336	2274	0.76	0.72	0.76	0.74
337	2276	0.74	0.72	0.7	0.63
338	2284	0.72	0.63	0.7	0.64
339	2289	0.71	0.61	0.53	0.51
340	2290	0.66	0.6	0.71	0.68
341	2295	0.78	0.72	0.71	0.69
342	2296	0.74	0.66	0.67	0.64
343	2297	0.45	0.44	0.52	0.48
344	2302	0.51	0.46	0.44	0.43
345	2304	0.66	0.61	0.63	0.6
346	2338	0.61	0.55	0.5	0.47
347	2346	0.5	0.43	0.39	0.37
348	2355	0.53	0.49	0.56	0.54
349	2356	0.79	0.67	0.68	0.64
350	2381	0.85	0.83	0.86	0.84
351	2386	0.75	0.69	0.76	0.71
352	2395	0.77	0.73	0.67	0.65
353	2401	0.36	0.34	0.54	0.45
354	2404	0.72	0.6	0.59	0.54
355	2413	0.44	0.38	0.42	0.39
356	2416	0.21	0.17	0.26	0.21
357	2417	0.68	0.64	0.65	0.63
358	2439	0.72	0.63	0.66	0.63
359	2441	0.54	0.51	0.45	0.44

doi:10.6342/NTU201603448

360	2448	0.71	0.61	0.77	0.72
361	2450	0.74	0.68	0.56	0.53
362	2458	0.6	0.53	0.45	0.45
363	2459	0.75	0.69	0.6	0.58
364	2462	0.62	0.62	0.5	0.48
365	2468	0.87	0.8	0.81	0.78
366	2470	0.4	0.32	0.44	0.38
367	2475	0.59	0.53	0.44	0.43
368	2478	0.18	0.17	0.2	0.19
369	2479	0.29	0.23	0.23	0.21
370	2482	0.87	0.84	0.84	0.82
371	2483	0.76	0.65	0.63	0.54
372	2484	0.76	0.74	0.71	0.68
373	2487	0.76	0.63	0.7	0.65
374	2488	0.91	0.88	0.94	0.92
375	2491	0.86	0.76	0.83	0.72
376	2492	0.47	0.4	0.36	0.34
377	2506	0.49	0.42	0.44	0.42
378	2510	0.73	0.67	0.64	0.61
379	2524	0.66	0.6	0.59	0.56
380	2533	0.85	0.81	0.83	0.81
381	2535	0.7	0.64	0.72	0.67
382	2545	0.71	0.62	0.63	0.59
383	2554	0.59	0.52	0.52	0.49
384	2564	0.77	0.72	0.72	0.7
385	2565	0.86	0.73	0.7	0.64
386	2576	0.44	0.38	0.4	0.36
387	2580	0.56	0.53	0.46	0.45
388	2586	0.49	0.35	0.34	0.31
389	2594	0.71	0.61	0.76	0.67
390	2598	0.75	0.72	0.74	0.71
391	2611	0.69	0.59	0.66	0.6
392	2617	0.38	0.34	0.35	0.32
393	2642	0.71	0.66	0.59	0.58
394	2654	0.78	0.74	0.72	0.69
395	2660	0.57	0.53	0.49	0.46
396	2666	0.66	0.6	0.58	0.54

doi:10.6342/NTU201603448

397	2671	0.78	0.74	0.72	0.69
398	2676	0.57	0.52	0.49	0.46
399	2679	0.81	0.75	0.72	0.69
400	2690	0.66	0.61	0.6	0.58
401	2693	0.49	0.43	0.51	0.46
402	2699	0.76	0.72	0.7	0.68
403	2707	0.69	0.65	0.65	0.64
404	2716	0.87	0.84	0.75	0.74
405	2717	0.87	0.82	0.87	0.85
406	2719	0.81	0.72	0.62	0.6
407	2721	0.41	0.37	0.33	0.31
408	2728	0.54	0.49	0.41	0.41
409	2732	0.72	0.67	0.7	0.68
410	2734	0.2	0.19	0.21	0.21
411	2736	0.76	0.74	0.69	0.68
412	2738	0.78	0.71	0.8	0.72
413	2747	0.71	0.67	0.73	0.71
414	2752	0.69	0.56	0.55	0.52
415	2757	0.82	0.77	0.78	0.76
416	2760	0.78	0.73	0.66	0.6
417	2761	0.49	0.47	0.32	0.36
418	2762	0.79	0.76	0.76	0.75
419	2766	0.69	0.49	0.66	0.57
420	2777	0.25	0.24	0.33	0.33
421	2783	0.42	0.39	0.27	0.27
422	2786	0.74	0.69	0.58	0.57
423	2790	0.53	0.46	0.61	0.54
424	2804	0.82	0.77	0.82	0.79
425	2822	0.49	0.23	0.25	0.23
426	2843	0.68	0.64	0.67	0.66
427	2847	0.62	0.51	0.46	0.43
428	2875	0.83	0.79	0.79	0.77
429	2877	0.36	0.34	0.45	0.43
430	2878	0.83	0.77	0.82	0.8
431	2879	0.63	0.59	0.62	0.59
432	2881	0.27	0.24	0.25	0.25
433	2882	0.61	0.59	0.66	0.63

doi:10.6342/NTU201603448

434	2884	0.71	0.65	0.6	0.57
435	2886	0.88	0.85	0.84	0.83
436	2896	0.87	0.82	0.86	0.84
437	2901	0.87	0.78	0.76	0.74
438	2907	0.54	0.44	0.42	0.39
439	2909	0.53	0.45	0.46	0.4
440	2910	0.64	0.59	0.62	0.57
441	2920	0.4	0.32	0.27	0.25
442	2929	0.72	0.64	0.65	0.61
443	2938	0.56	0.5	0.56	0.53
444	2939	0.58	0.51	0.48	0.45
445	2945	0.8	0.74	0.72	0.7
446	2948	0.79	0.66	0.62	0.56
447	2950	0.61	0.57	0.45	0.45
448	2959	0.83	0.76	0.74	0.7
449	2966	0.42	0.36	0.4	0.35
450	2970	0.78	0.75	0.68	0.66
451	2984	0.89	0.85	0.85	0.83
452	2987	0.11	0.1	0.13	0.12
453	2993	0.44	0.36	0.37	0.33
454	2994	0.52	0.47	0.42	0.42
455	2995	0.85	0.82	0.8	0.79
456	3000	0.84	0.81	0.8	0.79
457	3002	0.62	0.59	0.56	0.54
458	3006	0.75	0.71	0.75	0.71
459	3008	0.69	0.61	0.64	0.55
460	3012	0.77	0.66	0.73	0.7
461	3013	0.25	0.22	0.19	0.18
462	3016	0.86	0.82	0.74	0.72
463	3019	0.72	0.66	0.59	0.57
464	3021	0.59	0.5	0.49	0.46
465	3024	0.64	0.59	0.61	0.6
466	3025	0.84	0.76	0.79	0.72
467	3032	0.74	0.67	0.71	0.68
468	3033	0.88	0.85	0.78	0.77
469	3037	0.64	0.6	0.58	0.58
470	3044	0.77	0.69	0.64	0.59

doi:10.6342/NTU201603448

471	3048	0.54	0.47	0.46	0.43
472	3054	0.79	0.74	0.78	0.74
473	3057	0.85	0.8	0.78	0.76
474	3070	0.76	0.68	0.66	0.62
475	3082	0.84	0.77	0.74	0.71
476	3092	0.65	0.6	0.56	0.54
477	3114	0.4	0.34	0.31	0.29
478	3125	0.92	0.87	0.87	0.86
479	3129	0.79	0.76	0.67	0.65
480	3138	0.7	0.61	0.55	0.52
481	3147	0.56	0.45	0.44	0.4
482	3150	0.89	0.84	0.84	0.81
483	3158	0.47	0.4	0.35	0.34
484	3164	0.8	0.76	0.71	0.69
485	3199	0.88	0.86	0.89	0.88
486	3201	0.91	0.88	0.9	0.87
487	3208	0.75	0.69	0.7	0.67
488	3217	0.65	0.58	0.59	0.55
489	3218	0.59	0.47	0.51	0.49
490	3225	0.6	0.49	0.52	0.47
491	3247	0.68	0.53	0.57	0.53
492	3250	0.68	0.57	0.73	0.67
493	3252	0.27	0.23	0.31	0.27
494	3257	0.82	0.76	0.75	0.7
495	3270	0.8	0.72	0.79	0.75
496	3272	0.55	0.48	0.48	0.43
497	3276	0.75	0.7	0.7	0.68
498	3279	0.68	0.64	0.7	0.68
499	3282	0.47	0.4	0.42	0.36
500	3283	0.87	0.81	0.85	0.82
501	3287	0.9	0.88	0.77	0.77
502	3296	0.74	0.7	0.68	0.67
503	3298	0.6	0.58	0.59	0.55
504	3306	0.59	0.58	0.62	0.59
505	3308	0.87	0.82	0.89	0.86
506	3313	0.4	0.38	0.41	0.4
507	3331	0.44	0.35	0.38	0.35

doi:10.6342/NTU201603448

508	3333	0.59	0.52	0.46	0.43
509	3342	0.71	0.61	0.68	0.55
510	3343	0.69	0.55	0.48	0.47
511	3348	0.65	0.59	0.55	0.52
512	3354	0.77	0.67	0.58	0.55
513	3355	0.49	0.44	0.48	0.46
514	3360	0.76	0.7	0.71	0.69
515	3361	0.25	0.15	0.22	0.15
516	3365	0.65	0.59	0.44	0.44
517	3389	0.86	0.74	0.81	0.77
518	3391	0.92	0.88	0.9	0.87
519	3399	0.7	0.64	0.61	0.58
520	3421	0.72	0.68	0.7	0.65
521	3425	0.48	0.38	0.52	0.39
522	3426	0.87	0.8	0.72	0.69
523	3428	0.79	0.75	0.69	0.67
524	3434	0.48	0.47	0.57	0.55
525	3446	0.82	0.77	0.72	0.7
526	3449	0.62	0.56	0.5	0.46
527	3459	0.67	0.57	0.57	0.52
528	3462	0.83	0.76	0.81	0.78
529	3467	0.65	0.56	0.58	0.53
530	3468	0.82	0.75	0.78	0.75
531	3469	0.35	0.27	0.38	0.3
532	3480	0.61	0.49	0.41	0.39
533	3486	0.35	0.34	0.41	0.41
534	3496	0.86	0.82	0.79	0.78
535	3510	0.63	0.53	0.47	0.44
536	3520	0.43	0.34	0.39	0.32
537	3521	0.37	0.32	0.34	0.32
538	3532	0.63	0.51	0.52	0.46
539	3533	0.83	0.73	0.79	0.71
540	3538	0.86	0.83	0.87	0.85
541	3541	0.87	0.85	0.86	0.85
542	3545	0.86	0.72	0.65	0.55
543	3550	0.57	0.51	0.61	0.57
544	3555	0.78	0.68	0.66	0.62

doi:10.6342/NTU201603448

545	3563	0.83	0.76	0.8	0.77
546	3564	0.78	0.73	0.73	0.72
547	3565	0.45	0.41	0.45	0.43
548	3566	0.81	0.67	0.72	0.5
549	3580	0.43	0.41	0.47	0.4
550	3583	0.68	0.58	0.57	0.53
551	3603	0.48	0.33	0.33	0.27
552	3610	0.55	0.47	0.43	0.41
553	3618	0.31	0.28	0.33	0.3
554	3623	0.65	0.61	0.66	0.63
555	3624	0.83	0.71	0.69	0.63
556	3627	0.91	0.85	0.87	0.84
557	3632	0.79	0.69	0.9	0.82
558	3637	0.86	0.84	0.75	0.75
559	3642	0.86	0.82	0.82	0.8
560	3644	0.81	0.69	0.76	0.68
561	3656	0.52	0.44	0.41	0.39
562	3666	0.78	0.74	0.65	0.65
563	3672	0.93	0.89	0.85	0.83
564	3693	0.83	0.78	0.75	0.73
565	3696	0.7	0.69	0.69	0.68
566	3698	0.63	0.56	0.57	0.53
567	3703	0.71	0.61	0.58	0.56
568	3706	0.83	0.74	0.83	0.8
569	3708	0.7	0.64	0.75	0.7
570	3715	0.83	0.73	0.69	0.64
571	3718	0.6	0.48	0.48	0.46
572	3719	0.35	0.29	0.29	0.26
573	3723	0.62	0.54	0.62	0.59
574	3730	0.68	0.61	0.68	0.6
575	3736	0.81	0.75	0.72	0.68
576	3739	0.76	0.61	0.65	0.63
577	3750	0.54	0.44	0.39	0.34
578	3751	0.63	0.58	0.59	0.55
579	3752	0.76	0.66	0.67	0.62
580	3755	0.61	0.52	0.42	0.41
581	3756	0.67	0.64	0.71	0.68

doi:10.6342/NTU201603448

582	3758	0.68	0.62	0.66	0.61
583	3759	0.51	0.44	0.42	0.4
584	3760	0.7	0.66	0.75	0.73
585	3761	0.54	0.47	0.49	0.46
586	3765	0.83	0.77	0.8	0.77
587	3774	0.84	0.8	0.77	0.76
588	3790	0.71	0.62	0.6	0.53
589	3792	0.71	0.63	0.6	0.53
590	3793	0.56	0.5	0.54	0.51
591	3794	0.8	0.77	0.75	0.74
592	3795	0.49	0.42	0.41	0.39
593	3816	0.53	0.49	0.5	0.47
594	3820	0.43	0.36	0.39	0.38
595	3823	0.78	0.7	0.8	0.74
596	3833	0.5	0.43	0.45	0.41
597	3834	0.69	0.59	0.62	0.54
598	3835	0.56	0.47	0.45	0.41
599	3853	0.55	0.52	0.58	0.56
600	3854	0.83	0.75	0.71	0.69
601	3870	0.46	0.38	0.44	0.37
602	3875	0.84	0.76	0.72	0.7
603	3889	0.84	0.79	0.75	0.71
604	3898	0.77	0.7	0.75	0.71
605	3899	0.64	0.58	0.65	0.59
606	3909	0.44	0.39	0.34	0.33
607	3917	0.64	0.59	0.63	0.59
608	3925	0.56	0.47	0.5	0.46
609	3929	0.84	0.79	0.84	0.82
610	3930	0.72	0.64	0.59	0.57
611	3931	0.53	0.48	0.37	0.38
612	3937	0.69	0.53	0.71	0.61
613	3938	0.8	0.65	0.77	0.69
614	3942	0.45	0.35	0.37	0.34
615	3946	0.68	0.63	0.63	0.61
616	3949	0.68	0.62	0.66	0.64
617	3950	0.81	0.76	0.81	0.78
618	3962	0.77	0.68	0.68	0.67

doi:10.6342/NTU201603448

619	3964	0.82	0.8	0.88	0.85
620	3970	0.73	0.63	0.47	0.46
621	3971	0.19	0.17	0.23	0.21
622	3989	0.87	0.79	0.81	0.77
623	3990	0.7	0.62	0.6	0.57
624	3995	0.66	0.5	0.77	0.67
625	4011	0.68	0.54	0.57	0.53
626	4017	0.66	0.61	0.66	0.64
627	4033	0.12	0.08	0.14	0.1
628	4035	0.25	0.22	0.28	0.24
629	4045	0.55	0.47	0.44	0.38
630	4046	0.62	0.52	0.55	0.5
631	4067	0.83	0.77	0.78	0.73
632	4071	0.42	0.43	0.29	0.3
633	4081	0.69	0.65	0.56	0.54
634	4083	0.3	0.18	0.31	0.18
635	4084	0.69	0.62	0.59	0.56
636	4089	0.59	0.48	0.52	0.47
637	4093	0.45	0.42	0.45	0.43
638	4094	0.59	0.57	0.57	0.57
639	4104	0.77	0.68	0.71	0.65
640	4129	0.65	0.6	0.57	0.53
641	4131	0.94	0.89	0.95	0.92
642	4134	0.77	0.65	0.8	0.71
643	4135	0.85	0.82	0.85	0.83
644	4172	0.61	0.52	0.56	0.51
645	4173	0.7	0.62	0.6	0.58
646	4174	0.63	0.39	0.49	0.29
647	4181	0.85	0.83	0.8	0.8
648	4185	0.7	0.63	0.66	0.64
649	4187	0.4	0.37	0.35	0.34
650	4194	0.66	0.52	0.56	0.47
651	4196	0.34	0.33	0.38	0.39
652	4198	0.77	0.64	0.69	0.63
653	4203	0.56	0.54	0.6	0.59
654	4204	0.54	0.48	0.53	0.51
655	4210	0.39	0.36	0.4	0.38

doi:10.6342/NTU201603448

656	4214	0.68	0.61	0.65	0.61
657	4218	0.51	0.48	0.39	0.36
658	4222	0.67	0.59	0.66	0.64
659	4230	0.67	0.57	0.53	0.47
660	4238	0.49	0.39	0.44	0.38
661	4241	0.88	0.82	0.85	0.82
662	4246	0.89	0.87	0.84	0.82
663	4247	0.54	0.51	0.56	0.52
664	4248	0.79	0.74	0.66	0.63
665	4250	0.72	0.58	0.52	0.5
666	4252	0.7	0.64	0.6	0.58
667	4253	0.78	0.74	0.72	0.69
668	4255	0.52	0.45	0.43	0.4
669	4259	0.77	0.7	0.71	0.69
670	4267	0.7	0.54	0.53	0.47
671	4274	0.78	0.74	0.72	0.7
672	4275	0.75	0.58	0.78	0.62
673	4303	0.4	0.31	0.41	0.39
674	4305	0.9	0.86	0.89	0.87
675	4310	0.44	0.41	0.36	0.35
676	4320	0.71	0.58	0.54	0.48
677	4323	0.52	0.46	0.47	0.46
678	4327	0.75	0.71	0.72	0.7
679	4328	0.7	0.63	0.7	0.66
680	4331	0.7	0.62	0.71	0.65
681	4334	0.78	0.67	0.63	0.59
682	4342	0.54	0.52	0.75	0.72
683	4350	0.73	0.62	0.76	0.73
684	4361	0.81	0.76	0.77	0.75
685	4370	0.67	0.61	0.56	0.54
686	4390	0.91	0.84	0.84	0.8
687	4391	0.84	0.77	0.79	0.74
688	4407	0.94	0.88	0.86	0.81
689	4413	0.75	0.72	0.82	0.8
690	4417	0.76	0.68	0.75	0.7
691	4427	0.85	0.79	0.84	0.78
692	4433	0.8	0.77	0.88	0.86

doi:10.6342/NTU201603448

693	4435	0.55	0.43	0.59	0.42
694	4436	0.52	0.39	0.47	0.37
695	4440	0.74	0.68	0.76	0.7
696	4447	0.3	0.24	0.31	0.22
697	4454	0.63	0.56	0.5	0.47
698	4458	0.77	0.72	0.75	0.71
699	4464	0.57	0.47	0.61	0.45
700	4467	0.76	0.69	0.67	0.65
701	4472	0.76	0.7	0.67	0.64
702	4485	0.85	0.8	0.78	0.75
703	4489	0.13	0.11	0.12	0.12
704	4497	0.39	0.36	0.32	0.3
705	4509	0.88	0.81	0.82	0.76
706	4517	0.87	0.77	0.81	0.74
707	4522	0.69	0.6	0.56	0.53
708	4526	0.52	0.47	0.46	0.44
709	4547	0.59	0.5	0.52	0.49
710	4549	0.76	0.65	0.65	0.62
711	4557	0.66	0.59	0.5	0.47
712	4560	0.81	0.72	0.71	0.68
713	4566	0.75	0.7	0.77	0.7
714	4567	0.54	0.46	0.45	0.39
715	4585	0.71	0.67	0.57	0.56
716	4596	0.69	0.64	0.68	0.65
717	4626	0.68	0.59	0.7	0.62
718	4635	0.75	0.68	0.7	0.64
719	4637	0.29	0.23	0.26	0.19
720	4638	0.59	0.52	0.5	0.44
721	4639	0.88	0.85	0.89	0.87
722	4643	0.65	0.6	0.56	0.55
723	4650	0.65	0.6	0.49	0.46
724	4653	0.62	0.53	0.65	0.61
725	4656	0.62	0.53	0.43	0.41
726	4658	0.8	0.69	0.7	0.64
727	4664	0.72	0.56	0.59	0.48
728	4668	0.74	0.62	0.59	0.55
729	4672	0.61	0.55	0.63	0.57

doi:10.6342/NTU201603448

730	4676	0.35	0.35	0.28	0.27
731	4684	0.29	0.27	0.31	0.3
732	4694	0.63	0.55	0.57	0.55
733	4695	0.82	0.78	0.7	0.68
734	4696	0.49	0.41	0.41	0.38
735	4707	0.88	0.81	0.82	0.8
736	4709	0.74	0.68	0.72	0.68
737	4728	0.85	0.77	0.86	0.79
738	4729	0.61	0.44	0.53	0.44
739	4734	0.26	0.21	0.22	0.2
740	4738	0.63	0.54	0.41	0.38
741	4747	0.62	0.55	0.59	0.58
742	4752	0.93	0.9	0.79	0.85
743	4758	0.83	0.78	0.85	0.81
744	4759	0.38	0.31	0.42	0.26
745	4761	0.73	0.69	0.7	0.68
746	4763	0.56	0.5	0.48	0.44
747	4766	0.9	0.85	0.87	0.85
748	4768	0.77	0.7	0.78	0.75
749	4771	0.57	0.41	0.46	0.4
750	4773	0.8	0.75	0.69	0.65
751	4780	0.54	0.48	0.52	0.49
752	4781	0.82	0.74	0.77	0.74
753	4784	0.74	0.68	0.71	0.67
754	4786	0.49	0.43	0.56	0.41
755	4792	0.66	0.62	0.64	0.62
756	4799	0.69	0.61	0.6	0.55
757	4812	0.69	0.57	0.49	0.47
758	4814	0.86	0.78	0.83	0.77
759	4818	0.63	0.57	0.59	0.54
760	4821	0.6	0.56	0.56	0.54
761	4830	0.4	0.32	0.35	0.32
762	4836	0.78	0.71	0.67	0.59
763	4841	0.69	0.64	0.6	0.58
764	4843	0.56	0.52	0.57	0.52
765	4847	0.89	0.87	0.81	0.8
766	4849	0.69	0.65	0.59	0.6

doi:10.6342/NTU201603448

767	4850	0.36	0.31	0.29	0.27
768	4851	0.56	0.51	0.52	0.5
769	4856	0.77	0.71	0.71	0.68
770	4857	0.67	0.63	0.57	0.56
771	4873	0.63	0.57	0.54	0.52
772	4875	0.73	0.67	0.68	0.66
773	4886	0.85	0.78	0.74	0.72
774	4906	0.56	0.48	0.49	0.45
775	4913	0.73	0.69	0.77	0.73
776	4924	0.81	0.75	0.85	0.81
777	4934	0.65	0.59	0.64	0.62
778	4938	0.77	0.71	0.75	0.72
779	4947	0.68	0.62	0.64	0.6
780	4971	0.31	0.15	0.45	0.38
781	4991	0.68	0.48	0.54	0.42
782	4994	0.51	0.43	0.44	0.36
783	4997	0.66	0.56	0.59	0.56
784	5002	0.63	0.56	0.59	0.55
785	5008	0.63	0.57	0.6	0.53
786	5010	0.76	0.7	0.69	0.67
787	5016	0.61	0.57	0.64	0.61
788	5018	0.56	0.48	0.44	0.43
789	5020	0.82	0.78	0.83	0.8
790	5030	0.59	0.47	0.53	0.46
791	5032	0.86	0.79	0.84	0.79
792	5034	0.53	0.46	0.48	0.44
793	5035	0.78	0.71	0.75	0.69
794	5037	0.56	0.5	0.49	0.48
795	5040	0.73	0.7	0.75	0.72
796	5045	0.84	0.76	0.74	0.71
797	5049	0.69	0.57	0.73	0.67
798	5055	0.78	0.71	0.59	0.58
799	5058	0.35	0.3	0.37	0.36
800	5061	0.82	0.78	0.84	0.8
801	5068	0.75	0.65	0.66	0.61
802	5071	0.66	0.57	0.7	0.62
803	5075	0.72	0.67	0.65	0.61

doi:10.6342/NTU201603448

804	5076	0.23	0.16	0.17	0.16
805	5084	0.34	0.3	0.31	0.29
806	5086	0.65	0.6	0.59	0.54
807	5088	0.85	0.79	0.86	0.84
808	5099	0.56	0.53	0.58	0.53
809	5104	0.68	0.62	0.53	0.51
810	5113	0.58	0.5	0.48	0.44
811	5116	0.78	0.7	0.62	0.59
812	5117	0.84	0.77	0.74	0.7
813	5123	0.64	0.56	0.57	0.53
814	5128	0.39	0.35	0.32	0.31
815	5136	0.88	0.83	0.81	0.78
816	5146	0.73	0.68	0.66	0.65
817	5155	0.79	0.73	0.76	0.72
818	5156	0.39	0.39	0.35	0.35
819	5163	0.71	0.6	0.62	0.51
820	5165	0.82	0.78	0.8	0.78
821	5170	0.49	0.31	0.42	0.27
822	5173	0.72	0.66	0.78	0.74
823	5175	0.62	0.57	0.52	0.52
824	5177	0.8	0.77	0.68	0.7
825	5182	0.12	0.1	0.14	0.1
826	5185	0.45	0.33	0.54	0.33
827	5191	0.66	0.6	0.54	0.53
828	5199	0.68	0.61	0.6	0.59
829	5202	0.4	0.42	0.45	0.45
830	5210	0.87	0.79	0.87	0.82
831	5221	0.9	0.84	0.86	0.8
832	5250	0.84	0.8	0.7	0.7
833	5252	0.42	0.37	0.48	0.43
834	5260	0.5	0.48	0.54	0.52
835	5271	0.68	0.55	0.66	0.59
836	5276	0.75	0.65	0.7	0.64
837	5288	0.77	0.7	0.71	0.68
838	5290	0.7	0.64	0.56	0.53
839	5291	0.59	0.5	0.62	0.57
840	5292	0.76	0.73	0.79	0.78

doi:10.6342/NTU201603448

841	5303	0.75	0.7	0.73	0.7
842	5307	0.79	0.74	0.77	0.74
843	5310	0.85	0.81	0.8	0.78
844	5317	0.78	0.69	0.73	0.68
845	5325	0.76	0.7	0.66	0.65
846	5326	0.63	0.5	0.75	0.68
847	5340	0.71	0.66	0.67	0.64
848	5347	0.73	0.69	0.59	0.58
849	5351	0.68	0.64	0.67	0.65
850	5353	0.57	0.51	0.44	0.43
851	5363	0.72	0.69	0.77	0.76
852	5368	0.79	0.73	0.8	0.77
853	5370	0.6	0.5	0.4	0.38
854	5372	0.52	0.42	0.41	0.36
855	5381	0.84	0.78	0.71	0.69
856	5382	0.56	0.52	0.49	0.47
857	5384	0.48	0.41	0.4	0.35
858	5397	0.76	0.71	0.72	0.7
859	5403	0.72	0.65	0.69	0.66
860	5407	0.49	0.4	0.43	0.4
861	5409	0.66	0.62	0.53	0.54
862	5410	0.71	0.66	0.73	0.69
863	5416	0.84	0.75	0.69	0.67
864	5421	0.49	0.45	0.39	0.38
865	5422	0.7	0.63	0.75	0.68
866	5423	0.77	0.67	0.72	0.64
867	5428	0.47	0.48	0.36	0.37
868	5429	0.69	0.59	0.54	0.49
869	5430	0.75	0.7	0.67	0.64
870	5461	0.78	0.68	0.71	0.67
871	5465	0.69	0.66	0.62	0.6
872	5467	0.75	0.62	0.67	0.63
873	5478	0.67	0.65	0.54	0.54
874	5483	0.91	0.86	0.87	0.84
875	5489	0.56	0.47	0.47	0.43
876	5508	0.67	0.62	0.66	0.64
877	5513	0.49	0.43	0.49	0.44

doi:10.6342/NTU201603448

878	5515	0.34	0.28	0.36	0.32
879	5517	0.67	0.59	0.63	0.57
880	5518	0.76	0.67	0.68	0.63
881	5520	0.8	0.76	0.85	0.83
882	5525	0.88	0.85	0.82	0.81
883	5530	0.62	0.59	0.48	0.47
884	5532	0.71	0.61	0.6	0.55
885	5536	0.88	0.85	0.92	0.9
886	5543	0.61	0.58	0.68	0.64
887	5544	0.45	0.36	0.47	0.38
888	5550	0.67	0.62	0.63	0.61
889	5551	0.77	0.68	0.67	0.64
890	5554	0.68	0.61	0.59	0.58
891	5567	0.74	0.67	0.72	0.68
892	5573	0.59	0.54	0.53	0.52
893	5579	0.71	0.62	0.58	0.54
894	5582	0.73	0.68	0.71	0.68
895	5583	0.51	0.38	0.46	0.4
896	5585	0.78	0.67	0.6	0.58
897	5586	0.65	0.56	0.52	0.47
898	5589	0.52	0.44	0.51	0.46
899	5600	0.78	0.71	0.78	0.72
900	5619	0.86	0.76	0.74	0.7
901	5646	0.51	0.44	0.39	0.38
902	5655	0.85	0.76	0.87	0.81
903	5661	0.81	0.78	0.84	0.81
904	5674	0.71	0.6	0.82	0.74
905	5680	0.69	0.63	0.68	0.64
906	5686	0.46	0.4	0.38	0.35
907	5692	0.55	0.51	0.57	0.54
908	5693	0.92	0.87	0.9	0.87
909	5698	0.67	0.61	0.74	0.7
910	5699	0.6	0.57	0.58	0.56
911	5702	0.63	0.58	0.62	0.6
912	5707	0.5	0.43	0.49	0.43
913	5708	0.76	0.65	0.71	0.67
914	5710	0.72	0.63	0.62	0.57

doi:10.6342/NTU201603448

915	5716	0.82	0.74	0.83	0.79
916	5717	0.84	0.78	0.74	0.71
917	5719	0.66	0.61	0.46	0.48
918	5726	0.7	0.68	0.59	0.58
919	5732	0.71	0.65	0.62	0.57
920	5733	0.49	0.44	0.4	0.39
921	5750	0.66	0.6	0.62	0.58
922	5755	0.72	0.66	0.62	0.6
923	5756	0.62	0.55	0.55	0.52
924	5775	0.57	0.45	0.5	0.44
925	5777	0.66	0.64	0.72	0.69

doi:10.6342/NTU201603448

附錄五、LSA 與人工標記結果

本附錄從 925 篇的測試資料中，取前 432 篇做人工標記，且列出其對應的 100 維 Log-Entropy 種類一、100 維 Log-Entropy 種類二、300 維 Log-Entropy 種類一、300 維 Log-Entropy 種類二的所得餘弦值，其依等級畫分。

篇數	金玉集編號	人工標記	100 維 Log-Entropy 種類一	100 維 Log-Entropy 種類二	300 維 Log-Entropy 種類一	300 維 Log-Entropy 種類二
1	13	3	3	3	3	3
2	31	3	3	3	3	3
3	50	3	3	3	3	3
4	53	3	3	3	3	3
5	66	3	3	3	3	3
6	88	3	1	1	1	1
7	99	3	3	3	2	3
8	119	3	2	2	2	2
9	129	3	3	3	3	3
10	134	3	3	3	3	3
11	136	3	2	3	2	3
12	139	3	3	3	3	3
13	159	3	3	3	3	3
14	162	3	3	3	3	3
15	173	3	3	3	2	3
16	177	3	3	3	3	3
17	194	1	3	3	3	3
18	197	3	3	3	3	3
19	201	3	3	3	3	3
20	202	3	3	3	3	3
21	208	3	3	3	3	3
22	214	3	3	3	3	3
23	218	3	3	3	3	3
24	235	3	3	3	3	3
25	237	3	3	3	3	3
26	238	1	2	3	2	3
27	248	1	2	2	1	2
28	252	3	3	3	3	3

doi:10.6342/NTU201603448

29	259	3	2	3	2	3
30	263	3	3	3	3	3
31	277	3	3	3	3	3
32	295	2	3	3	3	3
33	298	3	3	3	3	3
34	299	3	3	3	3	3
35	300	3	3	3	3	3
36	301	2	3	3	3	3
37	302	3	3	3	3	3
38	306	2	3	3	3	3
39	313	3	3	3	3	3
40	321	3	3	3	3	3
41	330	3	2	2	2	2
42	334	3	3	3	3	3
43	335	2	3	3	3	3
44	388	3	3	3	3	3
45	391	2	3	3	3	3
46	399	2	2	2	2	2
47	403	2	2	3	2	3
48	411	2	3	3	3	3
49	412	2	2	3	2	3
50	419	2	3	3	3	3
51	422	2	1	1	1	1
52	423	3	3	3	3	3
53	424	2	1	1	1	1
54	425	2	1	1	1	1
55	432	2	2	3	2	3
56	437	3	3	3	3	3
57	450	3	3	3	3	3
58	453	2	3	3	3	3
59	474	3	1	1	1	1
60	475	3	2	3	2	3
61	489	3	3	3	3	3
62	499	3	2	2	2	2
63	503	3	2	3	2	3
64	509	3	3	3	3	3
65	514	3	3	3	3	3

doi:10.6342/NTU201603448

66	534	3	3	3	3	3
67	535	3	3	3	3	3
68	537	3	3	3	3	3
69	539	3	3	3	2	3
70	546	3	3	3	3	3
71	555	3	3	3	3	3
72	578	3	3	3	3	3
73	579	3	3	3	3	3
74	584	3	3	3	3	3
75	587	3	3	3	3	3
76	611	3	3	3	3	3
77	618	3	1	2	1	1
78	619	3	3	3	3	3
79	622	3	3	3	3	3
80	624	3	3	3	3	3
81	639	3	3	3	3	3
82	642	3	1	2	1	1
83	645	3	3	3	3	3
84	668	3	3	3	3	3
85	675	3	3	3	3	3
86	682	3	3	3	3	3
87	689	3	3	3	3	3
88	700	3	3	3	3	3
89	707	3	3	3	3	3
90	715	3	3	3	3	3
91	728	3	3	3	3	3
92	733	3	3	3	2	3
93	737	3	3	3	3	3
94	738	3	3	3	3	3
95	740	3	3	3	3	3
96	742	2	2	3	2	2
97	748	3	1	1	1	1
98	750	2	3	3	3	3
99	753	1	1	2	1	1
100	757	3	3	3	3	3
101	758	3	2	2	2	2
102	761	3	3	3	3	3

doi:10.6342/NTU201603448

103	775	3	2	3	2	3
104	776	3	1	2	1	1
105	780	3	3	3	3	3
106	781	3	3	3	3	3
107	785	3	3	3	3	3
108	795	3	2	2	2	2
109	796	3	3	3	2	3
110	804	3	2	3	2	3
111	806	3	3	3	3	3
112	808	3	2	2	1	2
113	810	3	3	3	2	3
114	821	3	3	3	3	3
115	828	3	2	3	2	3
116	833	3	2	3	2	2
117	847	3	3	3	3	3
118	858	3	3	3	3	3
119	859	3	3	3	3	3
120	862	3	2	3	2	2
121	865	3	2	2	1	2
122	868	3	3	3	3	3
123	871	3	3	3	3	3
124	873	3	3	3	3	3
125	879	3	3	3	3	3
126	892	3	3	3	3	3
127	909	3	3	3	3	3
128	923	3	3	3	3	3
129	926	3	3	3	3	3
130	931	3	3	3	3	3
131	937	3	2	3	2	3
132	938	3	3	3	3	3
133	940	3	1	1	1	1
134	948	3	3	3	3	3
135	951	3	3	3	3	3
136	956	3	2	2	1	2
137	962	3	3	3	3	3
138	969	3	2	3	2	3
139	972	3	2	3	2	2

doi:10.6342/NTU201603448

140	982	3	3	3	3	3
141	999	3	3	3	3	3
142	1002	3	2	2	2	2
143	1007	3	1	1	1	1
144	1013	3	3	3	3	3
145	1017	3	3	3	3	3
146	1025	3	3	3	3	3
147	1034	3	3	3	3	3
148	1043	3	2	2	2	2
149	1045	3	2	2	2	2
150	1046	3	3	3	3	3
151	1048	3	2	3	2	3
152	1051	3	3	3	3	3
153	1055	3	2	3	1	2
154	1059	3	3	3	3	3
155	1086	3	3	3	3	3
156	1093	3	3	3	3	3
157	1095	3	3	3	3	3
158	1096	3	3	3	3	3
159	1099	3	3	3	3	3
160	1104	3	3	3	3	3
161	1122	3	3	3	3	3
162	1132	2	3	3	3	3
163	1134	2	3	3	3	3
164	1135	2	3	3	3	3
165	1144	2	2	3	2	2
166	1149	2	3	3	3	3
167	1157	2	3	3	3	3
168	1158	2	1	1	1	1
169	1159	2	1	1	1	1
170	1166	2	3	3	3	3
171	1174	2	3	3	3	3
172	1178	2	3	3	3	3
173	1190	2	3	3	3	3
174	1195	2	3	3	3	3
175	1211	3	3	3	3	3
176	1214	3	3	3	3	3

doi:10.6342/NTU201603448

177	1215	2	3	3	3	3
178	1217	3	3	3	3	3
179	1218	3	3	3	3	3
180	1227	2	3	3	3	3
181	1229	3	3	3	3	3
182	1230	3	3	3	3	3
183	1232	3	2	2	2	2
184	1243	3	3	3	3	3
185	1245	2	3	3	3	3
186	1252	3	3	3	3	3
187	1258	1	1	1	1	1
188	1263	3	3	3	3	3
189	1266	3	3	3	3	3
190	1268	3	3	3	2	3
191	1275	3	3	3	2	3
192	1295	3	3	3	3	3
193	1298	3	3	3	3	3
194	1300	3	3	3	3	3
195	1307	3	2	2	2	2
196	1308	3	3	3	3	3
197	1314	3	2	2	2	2
198	1322	3	2	3	2	3
199	1327	3	3	3	3	3
200	1339	3	2	2	2	2
201	1350	3	3	3	3	3
202	1367	3	3	3	3	3
203	1377	3	3	3	3	3
204	1379	3	2	2	1	2
205	1390	3	3	3	3	3
206	1407	3	3	3	3	3
207	1409	3	2	2	2	2
208	1415	3	2	2	2	2
209	1417	2	3	3	2	3
210	1430	1	1	1	1	1
211	1445	3	3	3	3	3
212	1469	3	3	3	3	3
213	1478	3	3	3	3	3

doi:10.6342/NTU201603448

214	1486	3	3	3	3	3
215	1487	3	3	3	3	3
216	1501	3	3	3	3	3
217	1505	3	2	2	2	2
218	1509	3	2	3	2	2
219	1516	3	3	3	3	3
220	1523	3	3	3	2	3
221	1528	3	3	3	3	3
222	1541	2	2	2	2	2
223	1542	2	2	2	2	2
224	1545	3	3	3	3	3
225	1547	3	3	3	3	3
226	1560	3	2	3	2	3
227	1567	3	3	3	3	3
228	1568	3	2	3	2	3
229	1569	3	3	3	3	3
230	1578	3	3	3	3	3
231	1579	3	2	3	2	2
232	1581	3	3	3	3	3
233	1599	3	3	3	3	3
234	1632	3	2	3	2	3
235	1635	2	2	3	2	3
236	1636	3	3	3	3	3
237	1643	3	3	3	3	3
238	1652	3	3	3	3	3
239	1659	3	2	3	2	3
240	1660	3	3	3	3	3
241	1666	3	3	3	3	3
242	1688	3	3	3	3	3
243	1689	3	3	3	2	3
244	1691	3	3	3	3	3
245	1696	3	2	2	2	2
246	1713	3	3	3	3	3
247	1716	3	2	2	2	2
248	1728	3	3	3	3	3
249	1730	3	3	3	3	3
250	1731	3	3	3	3	3

doi:10.6342/NTU201603448

251	1732	3	3	3	3	3
252	1748	3	3	3	3	3
253	1753	3	3	3	3	3
254	1755	3	2	2	2	2
255	1762	3	2	2	1	2
256	1772	3	2	2	2	2
257	1783	3	3	3	2	3
258	1785	3	3	3	3	3
259	1788	3	3	3	3	3
260	1789	3	2	3	2	3
261	1793	3	3	3	3	3
262	1799	3	2	3	2	3
263	1801	3	3	3	3	3
264	1808	3	3	3	3	3
265	1811	3	3	3	3	3
266	1816	3	3	3	2	3
267	1820	3	3	3	3	3
268	1822	3	3	3	3	3
269	1833	3	3	3	3	3
270	1837	3	3	3	3	3
271	1846	3	3	3	3	3
272	1854	2	2	3	2	3
273	1856	3	3	3	3	3
274	1859	3	2	3	2	3
275	1860	3	3	3	3	3
276	1866	2	2	2	2	2
277	1872	3	3	3	3	3
278	1880	3	3	3	3	3
279	1881	3	2	3	2	3
280	1888	3	3	3	3	3
281	1890	3	3	3	3	3
282	1893	3	2	3	2	3
283	1898	3	2	3	2	3
284	1909	3	1	2	1	2
285	1918	3	3	3	3	3
286	1931	3	2	2	2	2
287	1934	1	3	3	3	3

doi:10.6342/NTU201603448

288	1936	3	3	3	2	3
289	1953	3	2	2	2	2
290	1960	2	3	3	3	3
291	1965	3	3	3	3	3
292	1976	3	3	3	3	3
293	1977	3	3	3	3	3
294	1987	3	3	3	3	3
295	1992	3	3	3	3	3
296	2004	3	3	3	3	3
297	2008	3	3	3	3	3
298	2017	3	1	1	1	1
299	2023	3	3	3	3	3
300	2045	3	3	3	3	3
301	2051	1	3	3	3	3
302	2059	3	3	3	3	3
303	2068	3	3	3	3	3
304	2069	3	2	3	2	2
305	2070	3	3	3	3	3
306	2071	3	3	3	3	3
307	2076	3	3	3	3	3
308	2079	3	3	3	3	3
309	2080	2	2	2	2	2
310	2083	3	2	3	2	3
311	2092	3	3	3	3	3
312	2093	3	3	3	3	3
313	2099	3	3	3	3	3
314	2109	3	3	3	3	3
315	2129	2	2	3	2	3
316	2132	3	2	2	1	2
317	2150	3	3	3	3	3
318	2159	3	2	3	2	2
319	2164	3	2	3	2	2
320	2189	3	3	3	3	3
321	2191	3	3	3	3	3
322	2200	3	3	3	3	3
323	2214	3	2	2	2	2
324	2215	3	3	3	3	3

doi:10.6342/NTU201603448

325	2218	2	3	3	3	3
326	2222	3	3	3	3	3
327	2233	3	3	3	2	3
328	2234	3	1	2	1	1
329	2235	2	3	3	3	3
330	2237	3	3	3	3	3
331	2246	3	3	3	2	3
332	2247	3	1	1	1	1
333	2248	3	3	3	3	3
334	2262	3	3	3	3	3
335	2265	3	3	3	3	3
336	2274	3	3	3	3	3
337	2276	3	3	3	3	3
338	2284	3	3	3	3	3
339	2289	3	3	3	3	3
340	2290	3	3	3	3	3
341	2295	3	3	3	3	3
342	2296	3	3	3	3	3
343	2297	3	3	3	2	3
344	2302	3	2	3	2	3
345	2304	3	3	3	3	3
346	2338	3	3	3	2	3
347	2346	3	2	2	2	2
348	2355	3	3	3	3	3
349	2356	3	3	3	3	3
350	2381	3	3	3	3	3
351	2386	3	3	3	3	3
352	2395	3	3	3	3	3
353	2401	3	3	3	2	3
354	2404	3	3	3	3	3
355	2413	3	2	3	2	2
356	2416	3	2	2	1	2
357	2417	3	3	3	3	3
358	2439	3	3	3	3	3
359	2441	3	2	3	2	3
360	2448	3	3	3	3	3
361	2450	3	3	3	3	3

doi:10.6342/NTU201603448

362	2458	3	2	3	2	3
363	2459	3	3	3	3	3
364	2462	3	3	3	2	3
365	2468	3	3	3	3	3
366	2470	3	2	3	2	2
367	2475	3	2	3	2	3
368	2478	3	1	2	1	1
369	2479	3	1	2	1	2
370	2482	3	3	3	3	3
371	2483	3	3	3	3	3
372	2484	3	3	3	3	3
373	2487	3	3	3	3	3
374	2488	3	3	3	3	3
375	2491	3	3	3	3	3
376	2492	3	2	2	2	2
377	2506	3	2	3	2	3
378	2510	3	3	3	3	3
379	2524	3	3	3	3	3
380	2533	3	3	3	3	3
381	2535	3	3	3	3	3
382	2545	3	3	3	3	3
383	2554	3	3	3	2	3
384	2564	3	3	3	3	3
385	2565	3	3	3	3	3
386	2576	3	2	3	2	2
387	2580	3	2	3	2	3
388	2586	3	2	2	2	2
389	2594	2	3	3	3	3
390	2598	3	3	3	3	3
391	2611	3	3	3	3	3
392	2617	3	2	2	2	2
393	2642	3	3	3	3	3
394	2654	3	3	3	3	3
395	2660	3	2	3	2	3
396	2666	3	3	3	3	3
397	2671	3	3	3	3	3
398	2676	3	2	3	2	3

doi:10.6342/NTU201603448

399	2679	3	3	3	3	3
400	2690	3	3	3	3	3
401	2693	3	3	3	2	3
402	2699	3	3	3	3	3
403	2707	2	3	3	3	3
404	2716	3	3	3	3	3
405	2717	3	3	3	3	3
406	2719	3	3	3	3	3
407	2721	1	2	2	2	2
408	2728	3	2	3	2	3
409	2732	3	3	3	3	3
410	2734	3	1	2	1	2
411	2736	3	3	3	3	3
412	2738	3	3	3	3	3
413	2747	2	3	3	3	3
414	2752	3	3	3	3	3
415	2757	3	3	3	3	3
416	2760	3	3	3	3	3
417	2761	3	2	2	2	2
418	2762	3	3	3	3	3
419	2766	2	3	3	3	3
420	2777	1	2	2	2	2
421	2783	3	2	2	2	2
422	2786	3	3	3	3	3
423	2790	2	3	3	3	3
424	2804	2	3	3	3	3
425	2822	3	2	2	1	2
426	2843	3	3	3	3	3
427	2847	3	2	3	2	3
428	2875	3	3	3	3	3
429	2877	3	2	3	2	3
430	2878	3	3	3	3	3
431	2879	3	3	3	3	3
432	2881	3	2	2	2	2

doi:10.6342/NTU201603448