

應用文本探勘技術之會計分錄測試方法

孫嘉明* 陳慧巧** 彭宇晨***

摘要：大數據時代來臨，審計作業可結合新興的資訊技術，以分析快速增長的大量資料。如此在原有的審計目標下，可有效加強審計執行情序的效果與效率。長期以來由於會計分錄摘要內容均以簡短文字呈現，故不易使用傳統資料分析工具予以分析；因此本研究應用文本探勘技術，分析會計分錄摘要的非結構內容，辨別及評估其風險，以協助審計人員更有效地找尋異常之會計傳票。

本研究設計採用設計科學研究方法，將文本探勘技術結合資源—事件—人員模型 (Resource-Event-Agent Model, REA Model)，建置會計分錄風險評估系統雛型，以探索文本探勘技術應用於分析會計傳票風險之可行性。其分析程序主要針對單一會計科目的相關分錄摘要內容，應用 REA 模型進行分類、歸納及風險評分；以提供不同風險層級之篩選條件，利於會計師找尋可能誤記或蓄意隱匿之會計分錄。後續並將所設計的系統雛型交由會審專家評估其有效性及可行性，並提出回饋意見與建議。

關鍵詞：會計分錄測試、風險評估、文本探勘、特徵選取

* 國立雲林科技大學會計系副教授

** 國立雲林科技大學會計系副教授

*** 國立雲林科技大學會計碩士

108 年 08 月收稿

109 年 03 月接受

三審接受

DOI: 10.6675/JCA.202111_22(2).02

The Design of a Journal Entry Testing Method Based on Text Mining Techniques

Chia-Ming Sun^{*} Hui-Chiao Chen^{**} Yu-Chen Peng^{***}

Abstract: With the advent of big data, auditing work has combined with emerging information technologies to analyze the increasing volume of data. Without altering original audit objectives, applying new technologies can strengthen the effectiveness and efficiency of the audit process. Since the description of journal entries has been for long time recorded in short text, it is not easy to assess the risk of misrepresentations or intentional omission in journal entries by means of traditional data-analysis techniques. This study applies text mining techniques to analyze unstructured texts to find suspected abnormal journal entries.

Based on the design science research methodology (DSRM) as a research method, we combine text mining techniques and the resource-event-agent (REA) model, to develop a system prototype for risk assessment of journal entries. Its purpose is to explore the feasibility of text mining techniques applying in the analysis of journal entry descriptions. The analysis procedures focus on the descriptions of each account in the journal entries with an application of REA model to classification, generalization, and risk scoring. These provide the filtering condition for different risk levels and assist auditors for locating journal entries caused by unintentional errors or deliberate concealment. The system prototype has further been offered to experienced audit experts to evaluate its effectiveness and feasibility and to provide feedbacks and suggestions.

Keywords: journal entries testing, risk assessment, text mining, feature selection

* Associate Professor, Department of Accounting, National Yunlin University of Science and Technology

** Associate Professor, Department of Accounting, National Yunlin University of Science and Technology

*** Master of Accounting, National Yunlin University of Science and Technology

Submitted August 2019

Accepted March 2020

After 3 rounds of review

DOI: 10.6675/JCA.202111_22(2).02

壹、緒論

一、研究背景及動機

因應大數據應用與人工智慧技術日趨普及，如何結合數據分析應用於審計、稅務等會審專業領域，已成為國內外中大型會計師事務所共同面臨的挑戰。大數據分析在執行審計程序上扮演了突破性的角色，可協助審計人員於相同的時間內得到更高品質的審計證據，也可即時偵查改正審計查核方向（Fukukawa and Mock, 2012; Brown-Liburd and Vasarhelyi, 2015; Alles and Gray, 2016）；因此，審計數據分析（Audit Data Analytics, ADA）隨之成為審計創新與數位轉型的變革重要推力（孫嘉明，2018；Byrnes, Criste, Stewart, and Vasarhelyi, 2014）。

審計數據分析不僅是一種查核工具，更是一套完整的分析程序，其中包含風險評估、內部控制評估的分析過程、證實測試以及詳細測試等連貫過程（Byrnes et al., 2014）。審計數據分析有別於以往傳統審計以人工抽樣做為查核基礎，主要透過高效能的運算設備進行全母體的資料分析並偵查出高風險的會計科目、分錄或交易紀錄，進而取得更精確的查核結果及證據，以改善傳統審計的抽樣風險問題（Fay and Negangard, 2017）。根據英國財務報告委員會（Financial Reporting Council, FRC）之研究報告建議（FRC, 2017）：若將審計數據分析應用得當，可增強審計品質、風險分析、交易及控制測試等，可為管理階層和使用者提供基於審計風險的審計證據及相關資訊。

我國審計準則 49 號公報中指出：證實測試即包含了財務報表與相關會計紀錄核對或調節、檢查重大之日記簿分錄及於編製財務報表過程中所作之其他調整。會計分錄測試為審計人員針對受查單位日常的會計分錄進行測試，從中篩選出含有潛藏舞弊風險特徵的分錄（AICPA, 2002）。周濟群（2016）也提出透過會計分錄測試可以初步偵察出是否具有潛藏舞弊風險的分錄，以便審計人員判斷是否該進行後續的查核程序。然而，目前查核人員仍主要採用抽樣的方式執行查核程序，因而僅能取得合理確信。即使查核工作經過適當規劃及執行，財務報表仍然可能存在有重大不實表達而查核人員並未發現；尤其因為舞弊是受查者故意的行為，管理階層舞弊所導致的重大不實表達風險，可以逾越內部控制，更是不易偵查。

企業前端各大交易循環之財務資訊均依其日記簿傳票或憑單制度逐一過帳彙總至會計總帳系統；尤其目前各大型企業資源規劃系統多採單筆交易個別產生傳票，因此將會產生大量的會計傳票資料（Debreceeny and Gray, 2010）。在多項審計準則公報中也多次強調分錄測試（Journal Entry Testing, JET；或稱日記簿分錄測試）的重要性，被認為是提高審計品質與偵查財報不實風險的重要分析程序。然而，受限於不易取得真實會計分錄資料，先前使用會計分錄進行大量資料分析的研究非常少見。在 Debreceeny and Gray (2010) 的一項探索性研究中，即利用資料探勘（data mining）方法，評估可否利用會計分錄與會計科目的對應關係，偵查財報舞弊的可能性。

另一方面，文字資料為審計資料中除了數值資料外最常用的資料格式；然而卻也因為傳統上僅能直接分析結構化或已編碼分類的文字類型資料，目前以文本為分析對象的文本探勘 (text mining) 在會審領域逐漸興起 (Fisher, Garnsey, and Hughes, 2016; Sun and Vasarhelyi, 2018)。然而可能受限於企業內部資料取得不易，目前會審領域之文本探勘應用多針對可公開取得之新聞報導、企業財務報告或會計師查核意見進行分析 (如：周濟群與連子杰，2011；王明昌、鄭揚耀與柯建全，2015；Liu and Moffitt, 2016; Tsai and Wang, 2017)。

在會計分錄的組成內容中，除了會計科目與借貸方金額等結構化資料外，分錄摘要為重要的文字說明；因此，一般審計實務均認為摘要內容可作為識別異常之徵兆。在人工自行登錄的會計分錄摘要中，其內容主要由會計人員自行簡述分錄目的或意涵，並沒有標準的記錄方式與原則；因此容易未能加以詳述說明，將易導致與查核人員資訊不對稱之問題，有時不易單由會計分錄得知完整的交易意涵。尤其當分錄摘要若是記載不正確或是過於簡略，其中即可能存在財務報表不實表達之風險。

二、研究目的

以非結構格式儲存的文件資料，隱藏著許多關鍵資訊，其重要性不容小覷，使得文本探勘技術成為近年重要的研究領域之一。中文的句法結構不同於英文語系，文詞間也不似英文單字間以空白明顯分隔，所以需要以詞彙庫及句法分析進行斷詞。中文語料斷詞處理時，為了去除不具代表性的雜訊，需要專家投入或使用系統化方法進行雜訊過濾以找尋有意義的特徵資料。

在會計資訊領域之 REA 模型 (Resource-Event-Agents Model) 中，自 McCarthy 1982 年提出後，即專門應用於會計或經濟相關的決策管理系統之概念性設計，屬於特定領域特徵 (domain-specific pattern) 的塑模工具 (廖紫吟與孫嘉明，2014；McCarthy, 1982; Geerts, 2008)。該模型將會計交易活動相關之本體知識歸類為三大主題，分別為：資源 (resources)、事件 (events)、人員 (agents) (McCarthy, 1982)；故可幫助檢視會計分錄摘要內容是否完整記錄交易內容，其歸類方式可作為文本探勘斷詞結果之特徵篩選原則。

依據前述背景與動機之探討，本研究將文本探勘技術運用於會計分錄測試當中，使查核人員可以快速進行分錄摘要文本資料的前置處理，接著透過 REA 模型將摘要斷詞結果進行分類與篩選，後續並依據風險評估方法進行風險分級及挑選，以協助查核人員分析找尋可能為異常交易之高風險會計分錄。其目的是希望能透過結合文本探勘技術之風險評估程序，補強現有審計會計分錄測試程序不足之處以降低財報舞弊風險。

經由前述之探討，本研究目的簡述如下：

1. 評估將文本探勘技術應用於分析會計分錄之可行性。
2. 以會計總帳循環為例，應用文本探勘技術分析會計分錄中之摘要內容，進行斷詞與篩選其特徵資料。
3. 應用 REA 會計模型對摘要斷詞結果進行分類、篩選及發生頻率統計，發展出風險評量原則以利於挑選異常會計分錄，提升查核活動之查核效果及效率。

貳、文獻探討

一、資料探勘、文本探勘與審計數據分析

隨著大數據的興起，利用各式新興技術應用於審計數據分析成為審計創新浪潮。其中，資料探勘又可稱為資料挖掘或資料採礦，其最主要功能是從大量的資料中尋找出隱藏於內部特別關聯規則的過程 (Frawley, Piatetsky-Shapiro, and Matheus, 1992; Hand, Mannila, and Smyth, 2001)。就會審領域的應用而言，資料探勘技術可就所取得的資料內容作為基礎，依照資料所挖掘出的特徵識別其異常活動或紀錄 (Tan, Steinbach, and Kumar, 2006)，進一步發現可能存在的風險及控制缺失。

先前研究中即提出可透過資料探勘技術來協助查核人員進行資料分析，使其能夠自動地挖掘其資料間潛在的關係 (Lungu and Vătuțiu, 2007)。Ata and Seyrek (2009) 應用資料探勘技術於多家公司的財報，以找出可能存在錯誤及舞弊警訊的財務特徵；另外，張立程 (2015) 則透過分群演算法進行會計分錄異常偵測，將會計分錄的多個文數字欄位進行數值轉換，以快速找尋出與多數會計分錄不同的特殊會計分錄。雖然，造成異常值產生的原因可能為表達方式不同、活動發生頻率低以及人為紀錄不完整等因素，並不全然是內控的重大缺失，但在資料分析範圍及偵查成效上已可大幅提升審計活動的效率。

儘管資料探勘等大數據分析技術已逐漸應用於會審相關研究，但仍然有不足之處，在一篇探討資料探勘運用於財報舞弊偵查的文獻回顧報告中 (Ngai, Hu, Wong, Chen, and Sun, 2011)，歸納發現大多數研究內容為比較各種演算法所能偵測出的舞弊類型，從而判斷各種演算法之預測精確度，進而比較各不同演算法的優劣；然而先前研究並未探討資料探勘技術如何結合應用於審計程序，如何幫助查核人員偵測出內控缺失或舞弊行為，或協助找尋問題癥結，有效出具審計意見。可見該類研究中仍有很大的發展空間。

另外，近年來隨著人工智慧技術的快速發展，自然語言處理 (Natural Language Processing, NLP) 領域由於可對文字或語音進行辨識、解析與擷取，受到各界重視。文本探勘 (text mining) 為適用於文件的資料探勘分析方法，其與一般資料探勘的差異為：一般資料探勘方法用於發掘結構化資料的特徵與規則，而文本探勘則是用於找尋文字中的特徵與規則 (譚家蘭, 2012; Fisher et al., 2016)。

目前大多數原始敘述性資訊仍以非結構化或是半結構化的資料型態呈現，文本探勘技術即可用於處理此類型資料；透過演算法或是統計軟體對資料進行拆解與剖析，藉此找出隱藏在文字資料中的規則及架構（Sullivan, 2001）。陳裕菘、謝邦昌、李勝輝與陳郁婷（2014）提及文本探勘技術的成效在於語意分析，透過資料的前置處理程序，辨別出文件的關鍵詞彙；並藉由關鍵詞相關資訊細部分析，可幫助得到有價值的資訊。近年來，由於網路數據大幅增長以及雲端運算普及，文本探勘技術於財務或會計領域的應用需求也快速增高（Loughran and McDonald, 2016）。但是文本探勘相關技術及應用，本身受限於不同社會語言的差異，其應用方法細節上存有差異；尤其中文的斷詞不易特性，成為各專業領域推動上的挑戰（陳世榮，2015）。因此目前的會計研究當中，雖然有部分研究以文本探勘技術應用於財務報告書或會計師查核報告（Kloptchenko, Eklund, Karlsson, Back, Vanharanta, and Visa, 2004; Shirata and Sakagami, 2008）；但可能因為企業內部資料取得不易，仍少有研究將文本探勘應用於會計傳票資料當中（Debreceeny and Gray, 2010）。

二、會計分錄測試與風險評估

我國目前審計準則公報已採納國際審計準則公報原則，將原有之固有風險與控制風險合併評估，共同稱為「重大不實表達風險」。其起因於傳統審計風險模型中雖將不同風險要素獨立分隔，但實務上各風險要素卻難以獨立評估。現今的審計風險模型改為同步考量：經由瞭解受查者及其內部控制環境並考量交易類別、科目餘額與揭露事項辨認，以共同評估不實表達風險。

管理階層舞弊所導致的重大不實表達風險，由於可以逾越內部控制，又經過精密的設計與掩蓋，因此一般的查核程序並不易偵查發現；因此會計分錄測試在現今財報不實風險的查核與風險評估過程中，逐漸日益受到重視（財團法人中華民國會計研究發展基金會，2006，2020；PCAOB, 2002; CAQ, 2008）。

會計分錄測試為審計人員針對受查單位日常的會計分錄進行測試，從中篩選出含有潛藏舞弊風險特徵的分錄（AICPA, 2002）。會計分錄測試為會計事務所為評估財報風險所進行的查核程序。周濟群（2016）提及透過會計分錄測試可以初步偵察出是否具有潛藏舞弊風險的分錄，以便審計人員判斷是否該進行後續的查核程序。本研究將國內外審計實務指引中較常見的高風險分錄特徵及會計分錄測試方法彙整如表 1 所示（財團法人中華民國會計研究發展基金會，2006，2020；顏素英，2018；PCAOB, 2002; CAQ, 2008）。

現在的企業多數採用套裝之企業資源規劃系統（Enterprise Resource Planning, ERP），主要使用自動化程序處理總分類帳及編製財務報表，並包含與各大交易循環作業活動相關的多個功能模組，可在相關交易憑證確認時即自動產製對應之會計分錄，以電子形式拋轉到後端之會計總帳系統。因此審計人員很容易以電腦輔助查核技術辨認出「非標準之人工會計分錄」。雖然「非標準人工會計分錄」發生頻率低，

容易引起會計師關注，然而企業仍可將特殊活動藏於複雜組成的會計分錄 (“mega-combined” entries) 當中，而不易被辨識出來 (Debreceeny and Gray, 2010)。

另一方面，即使會計分錄多數為由 ERP 系統所自動產製之紀錄而極少人工介入，審計人員對相關風險仍無法僅執行少量抽樣之證實程序，即可取得足夠及適切之查核證據。因為該等證據是否足夠及適切，通常有賴於對資訊系統正確性及完整性之有效控制。自動化流程及控制雖然可以降低人為疏失錯誤，但仍無法避免個人不當地踰越資訊系統控制。如果適當控制未能有效執行，則即使是自動化會計分錄仍存在資訊不當產生或遭竄改之可能性。由此可見，以全母體為主之大數據分析方法，可彌補現今採抽樣為主之控制測試或證實測試方法的不足。

表 1 具舞弊風險的會計分錄特徵

項次	說明
1	使用頻率較少的科目或不相關及不尋常之科目。
2	在期末或結帳時記錄的分錄，並未詳述說明其摘要。
3	傳票編製由不常建立者建立或由特定人員建立。
4	金額為整數或是尾數一致的分錄。
5	使用無代碼之會計科目。
6	異常之會計分錄，如借記資產科目，貸記費用科目。
7	於期後立即迴轉之分錄。
8	於假日所建立之分錄。
9	分錄記錄時間超出查核範圍的區間。
10	未記錄登打人員之分錄。
11	編號不連續之傳票。
12	金額重大之傳票。
13	重複之分錄。
14	延遲輸入之分錄。

整理自：財團法人中華民國會計研究發展基金會，2006，2020；PCAOB, 2002; CAQ, 2008。

三、文本特徵選取與 REA 模型

文本探勘技術包含了文句語意擷取分析、關鍵詞分析、文件群集分析等方法 (曾元顯, 2004)，通常需要針對文件進行預先資料處理以助於文本特徵選取 (feature selection)。其程序首先取得關鍵詞 (key words)，再以所萃取的關鍵詞或關鍵詞句 (key phrases) 作搜尋的依據，藉此找出資料間所隱含的關聯及意義。

在關鍵詞產出的過程中，需要選取具適當特徵的資料項目，該特徵為後續分析評估的主要依據。特徵選取演算過程可略分為三個程序，分別為：篩減 (screening)、分級 (ranking) 與選擇 (selecting) (陳裕菘等人, 2014)。「篩減」為在資料斷詞後，將不太重要或無意義的字詞進行排除；「分級」為透過預先制定出的規則所訂定出來之重要程度，將重要性由高到低進行排序；「選擇」為根據重要性的排序並選擇一個特徵字詞庫，而將此字詞庫中的詞彙視為輸入變數，可利用建立後續預測模型。

為了評估會計分錄摘要記錄內容是否符合會計交易紀錄完整性，本研究利用 REA 模型作為特徵篩選之評估原則。REA 模型主要利用組織營運過程所涉及的資源、事件和參與者等三大類主要實體及其關係，來描述企業領域共通的經濟交易活動。其優點是能夠描述經濟活動系統中的所有相關資源、事件與參與者，及其相互之關係。REA 模型為可適用於會計現象的通用語意模型 (generalized semantic representation)，原本希望用於重新設計現有之會計帳務系統；後來有許多學者持續延伸其結構、概念，不但向外擴展到企業價值鏈、向下深入到各種業務事件，成為企業綱目 (enterprise schema) 的概念性塑模 (conceptual modeling) 方法與工具 (Geerts, 2008; Guan, Levitan, and Kuhn Jr., 2013; Dunn, Gerard, and Grabski, 2016)。

REA 模型其優點是能夠描述經濟活動系統中的重要實體特徵及其關係，其最初版本的模型組成包含了三類重要會計交易活動相關實體類型 (entity) (McCarthy, 1982)，其定義及示例說明如下：

1. 資源 (resources)：為組織擁有的有形的物品或各種企業可以實際控制、實際擁有的珍貴資源，這些經濟資源會隨著企業的經濟活動增加或減少，例如：現金、原物料等。
2. 事件 (events)：藉由資源變動而產生影響的現象。因為經濟資源將伴隨經濟事件的發生而產生資源變化，可用於記錄影響組織經濟資源變化的交易活動，例如：生產、原物料購入等經濟活動。
3. 參與者 (agents)：乃是參與企業經濟事件的關係人，既包括組織內部的參與者，也包括組織外部的參與者；例如：業務員、倉管、顧客和供應商等。

參、研究設計與分析程序

一、研究設計

本研究針對會計傳票資料進行摘要項目之文本分析，透過文本探勘的技術中可以找出會計人員表達該會計分錄所使用詞彙的習慣，是否有表達不完整的缺失或甚至隱藏交易紀錄所需的基本訊息。其分析程序主要以會計分錄摘要內容作為分析對象，輔以傳統會計分錄測試之會計科目金額與交易頻率分析，透過 REA 模型將會計分錄摘要進行內容篩選與風險分析，以找尋出疑似異常之高風險會計分錄。

Dalal and Zaveri (2011)將文件自動摘要之研究分為 4 大類(1)探索式技術 (heuristic techniques)、(2)語義基礎技術 (semantics-based techniques)、(3)查詢導向技術 (query-oriented techniques)、(4)分群導向技術 (cluster-based techniques)。本研究採用探索式技術，在關鍵詞產出的過程中，選取擁有適當特徵的資料，並進行後續該詞彙使用頻率與該摘要項目敘述完整性的評估。

由於 REA 模型涵蓋了會計領域的交易活動實體類型，因此本研究使用 REA 模型協助文本特徵之選取。於詞彙篩選時，先將詞彙歸類至三大實體類別，以便於後續摘要完整性的評估。系統雛型將應用 REA 模型作為詞彙特徵篩選及風險評估原則，以利於特徵選取及會計分錄之風險分級，並提供檢視功能以利於審計人員進行異常分錄之選擇。

二、研究程序

本研究使用 Peffers, Tuunanen, Rothenberger, and Chatterjee (2007)所提出的「設計科學研究法」(Design Science Research Methodology, DSRM)，其中包含五大步驟：問題定義與動機、定義解決方案的目標、設計與發展、展示、評估與溝通。其研究程序是一個循環檢視的流程，產出結果包括模型、方法、概念及案例，可進一步評估其成效，幫助逐步建構可行的研究成果。因此，對於新興科技應用的相關研究議題，利用設計科學研究法可以幫助研究者對研究主題逐步釐清問題及研究方向，探索研究對象所表現出的特性，歸納並發展出一套可行的解決方案。

為利於會計分錄風險評估系統雛型之可行性評估，本研究使用某家公司單一年度之會計分錄資料，進行分析與雛型設計之測試實例對象。所使用的資料為 2017 年之會計分錄資料，共有 19,102 張會計分錄，共有 93,266 筆資料。在多數企業中會計科目：「其他費用」，屬於非例行交易多採用人工登打分錄，其記錄品質可能變動較大，因此本研究選擇該科目相關之會計分錄作為雛型展示樣本來源。以本個案公司而言，共計有 685 筆資料。分析結果所展示的圖例，考量對該公司營業資訊的保護，已針對特定公司名稱及人員姓名進行去識別化處理。

基於設計科學研究法，本研究依其主要步驟作為研究流程，以下簡要說明研究程序及相關研究設計考量因素。

(一)確認問題與動機

此步驟為定義明確的研究問題與動機，並且提出有價值的解決方案。此階段需要理解問題的內容以及評估解決問題的重要性 (Peffers et al., 2007)。

根據我國審計公報第四十三號第七十五條，查核人員為因應管理階層踰越控制之風險，應設計及執行查核程序以測試普通日記簿分錄及編製財務報表所作其他調整之適當性。第七十七條中亦提到：「在電腦資訊系統環境下，普通日記簿分錄及其他調整分錄則可能僅儲存於電腦檔案中」；尤其現今 ERP 系統多依個別交易逐一自

動產製日記簿分錄，將導致審計人員逐一檢視會計分錄內容適當性並不容易。另外，分錄摘要內容為非結構化之文字敘述，一般審計數據分析軟體不易分析比對，故常被忽略而未能善用其資訊價值。

(二)定義解決方案的目標

根據步驟一所定義問題，推論出可行的解決方案目標。而解決方案目標可採量化或質化目標。量化目標可用於比較此解決方案是否比現存方案更具效果或效率；質化目標為描述新的產出如何解決當前尚未解決問題 (Peffer et al., 2007)。

本研究為改善原有會計分錄測試之分析限制，以會計傳票資料之摘要內容為分析標的；因為若會計傳票中摘要表達不完整或不適當，除了為人員疏失外，也可能存在財務報表不實表達之疑慮。本研究將以摘要的內容結合會計科目、金額大小等特徵，識別是否可能異常的交易，希望協助會審人員快速發覺具有潛在風險的會計傳票，以達到降低財報查核風險及企業內控風險。

雛型設計目的在於可行性評估，因此避免大量資料整理跟分析過於耗時，影響雛型開發與修正之效率；本研究透過少量資料反覆評估、測試及修改其設計規則以驗證系統的可行性。

(三)設計與發展

在設計發展階段，設計科學研究法也可以透過某些理論或原理使其設計更有根據及說服力。此階段的產出可能為構面、模式、方法、案例或任何具有研究貢獻的設計成果 (Peffer et al., 2007)。

本研究使用某企業的年度會計傳票資料作為分析樣本，建構一個會計分錄風險評估系統雛型。首先針對傳票的摘要進行斷詞處理，保留具有意義的中文詞彙後，將詞彙依照 REA 會計模型之三大類別特性進行分類，並針對各分錄的資料特徵，依據所歸納之審計原則及資料特性進行風險等級歸類。風險評估之分析規則與雛型設計細節將詳述於下節。

(四)展示

本階段將展示研究設計之產出如何解決前面階段所定義的問題，且透過一個或多個的實際案例進行驗證。其作法可能使用實驗、模擬、案例研究、證明或其它有效的方法 (Peffer et al., 2007)。

本研究為確認會計分錄風險評估雛型是否實現其設計目標，使用微軟 Power BI 商業智慧視覺化分析工具，以風險矩陣呈現不同風險等級會計分錄之分布情形，並可點選項目即時呈現其分錄內容，以供會審人員評估其分析成果有效性。

(五)評估與溝通

此階段目的在於觀察及評估解決方案的有效性。將解決方案目標與展示結果進行比較，評估系統是否可行且功能是否達成目標。研究成果需要進行使用者之回饋，可以透過與相關領域專家或使用者說明及溝通研究成果，討論解決方案是否符合實際及未來的需求，以提升解決方案的可用性 (Peffer et al., 2007)。

本研究的風險評估雛型初步完成後，為了確認風險評估結果是否可行有用，故將分析結果之系統雛型提供給三位實務工作經驗均超過十年以上之會審專家進行訪談及討論，以評估判斷是否符合實際情境。如有不符合之情形，則重新修正本研究之設計規則，以確認其有用性。

三、中文斷詞處理程序

本研究採用中央研究院數位典藏國家型科技計畫所研發之中文斷詞系統 (Chinese Knowledge and Information Processing, CKIP) 作為詞彙篩選工具，藉由該系統所累積的大量中文詞庫輔助辨識分錄摘要內容的詞彙。該系統的語料庫之分詞原則為依據中央標準局的「資訊處理用中文分詞規範」。此系統藉由與詞庫比對之外，亦可針對詞庫中未存在的未知詞加以剖析，透過文法剖析規則來新增詞彙。

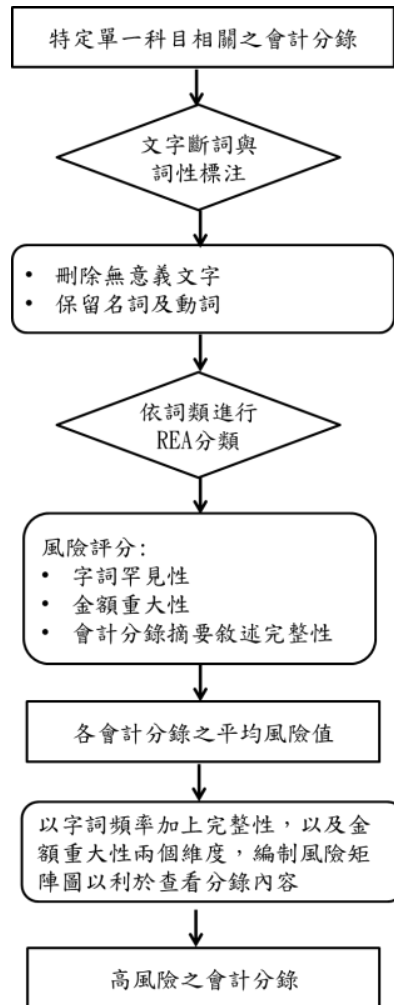
本研究針對會計傳票的摘要，應用 CKIP 中文斷詞系統線上服務進行斷詞及詞性標記，以取得每筆會計分錄摘要的組成文詞及所對應的詞性；並利用詞性作為詞彙統整及分類之依據。詞類標記乃依據中研院平衡語料庫詞類標記集，所傳回之精簡詞類標記。為避免過多無效詞彙影響分析效益，因此依據 REA 模型的實體分類篩選原則，將完成的斷詞進行過濾，僅保留斷詞結果標示詞性為名詞及動詞之字詞進行後續分析。

會計傳票摘要內容多以簡單扼要的方式說明，因此摘要內容可能句法不完整或是存在特殊交易對象名稱。該系統透過文字拆解的規則及針對語法的結構來分析字詞，並識別字詞的詞性，對於某些無法辨認之文字仍可歸類至特定類別，如姓名、新詞等詞庫中未建立之詞彙，將依語法結構歸類於「普通名詞 (Na)」；如此便無須擔心字詞遺失的情況，仍可進行完整分析。

肆、解決方案設計與雛型展示

一、解決方案設計

本研究使用某企業年度會計傳票資料作為分析樣本，應用文本探勘技術於會計分錄摘要，將摘要斷詞並依照詞性進行分類，及整體風險評估以篩選出高風險傳票。系統雛型之設計流程 (如圖 1 所示)，主要分為三大步驟，以下將逐一詳加說明，並後續以案例說明雛型設計過程：



資料來源：本研究整理

圖 1 會計分錄風險評估之分析流程

(一)摘要斷詞與詞性標記

本研究選取原始資料中科目名稱為「其他費用」之會計分錄作為分析樣本。首先經過 CKIP 斷詞系統將摘要內容初步拆解，依斷詞系統所標記的詞性，剔除不具意義的詞彙，保留後續可供 REA 分類的重要詞彙。詞彙挑選原則如下：

1. 刪除無意義詞彙，包含標點、特殊符號、注音符號、數值、英文等。
2. 刪除單一詞彙，對於中文字而言，斷詞結果詞彙長度為一個文字的資料視為單一詞彙。
3. 考量 REA 三大實體之詞性應為動詞或名詞，故保留詞性為動詞跟名詞之詞彙，刪除形容詞、副詞、助動詞等詞彙。

(二)應用 REA 模型進行詞彙分類

本研究透過 REA 會計模型將詞彙加以分類，由於經濟活動中包含資源、事件、人員，可分別對應至不同的詞性；因此可將標記為這些詞性的字詞作為 REA 之分類規則。其中，「事件」類別實體主要對應到動詞（含及物動詞、不及物動詞，例如：「租用」、「代墊」、「沖」）；「資源」類別實體主要對應至普通名詞（例如：「服務費」、「產品」、「車險」）；「人員」類別實體則對應至專有名詞及地方名詞（例如：「紡拓會」、「德國」、「林○玉」）。後續並將所挑選的詞彙之出現頻率進行累計，以得到詞頻表供作後續風險評量。

(三)風險評量

此步驟針對 REA 分類後的字詞頻率、對應金額、REA 完整性進行風險評量。所考量的風險因素與評量原則說明如下：

1. 風險因素—字詞的頻率

本研究將風險程度分為「極高」、「高」、「中」、「低」、「極低」，共五個等級。考量在會計帳務內控為適當有效性的前提下，高風險的分錄發生頻率較低；另外，特定的異常活動如果發生頻率或記錄頻率高，將容易曝露被發覺，故傾向隱匿而不完整揭露。故此階段將各字詞之頻率依不同數值範圍做風險等級，依字詞頻率“由小至大”分為五個組距，依序將不同組距範圍內的字詞，依頻率越低則風險越高的原則，分別給予高低不同的風險等級。

2. 風險因素—字詞對應交易的金額

此項風險因素考慮金額的重大性，當交易金額越高，則對於組織的資產或收支造成影響越大。本研究將各字詞對應的交易金額依不同數值範圍進行風險分級，“由小至大”分為五個組距，以交易金額最小的區間為第一組距，將各組距所對應的風險分為極低、低、中、高、極高。

3. 風險因素—摘要敘述完整性

REA 模型可確認經濟活動中的所有相關資源、事件與代理人及其相互之關係是否完整描述。本研究將觀察會計傳票中的摘要內容是否有完整的敘述，若是會計分錄中摘要內容缺乏完整 REA 分類者，可判斷摘要內容記載不完整，財務報表不實表達之風險相對較高。

二、雛型展示與評估

為提升分析程序反覆測試之時效性，本研究透過單一科目「其他費用」的相關分錄資料進行測試及評估，以驗證系統的可行性，並進行系統雛型設計及展示。

(一)應用文本探勘技術進行摘要分析

首先經過 CKIP 斷詞系統將分錄摘要進行斷詞分析，總共拆解出 4,187 個字詞（包含重覆的字詞）。由於拆解出的字詞包含許多無意義的詞彙或是標點符號，因此進行詞彙挑選。過程如下：

1. 刪除無意義詞彙，包含標點符號、注音符號、數值、英文符號；例如：「10508/16-106/08/16」、「6%」、「SMS-SHI1011-JP」，共刪除 1524 個字詞。
2. 刪除單一無意義詞彙：例如：「共」、「些」、「及」、「後」，共刪除 104 個單一字詞。
3. 保留詞性為「普通名詞 (Na)」、「專有名詞 (Nb)」、「地方名詞 (Nc)」、「及物動詞 (VA)」、「不及物動詞 (VC)」等字詞（如圖 2），刪除形容詞、副詞、助動詞者等詞性字詞，共刪除 9 個字詞。

經過資料前置處理過後，最終得到 148 個具有意義之詞彙。

1. 會計傳票

#	A	B	C	摘要
1	傳單編號	科目編號	科目名稱	摘要
2	9115-1707250002	1101001	庫存現金-台幣	楊○慧預支台幣換日幣退回PY500*0.2743與○庭保管
3	9120-1701200004	1101001	庫存現金-台幣	AHIG-170120001 1030006 楊○慧 MISC
4	9120-1702130001	1101001	庫存現金-台幣	AHIG-170213001 1030006 楊○慧 MISC
5	9120-1703160001	1101001	庫存現金-台幣	S9977 00薪資 1703/16
6	9120-1703160005	1101001	庫存現金-台幣	AHIG-170316001 1030006 楊○慧 MISC
7	9120-1703210002	1101001	庫存現金-台幣	AHIG-170321001 1030006 楊○慧 MISC
8	9120-1703230001	1101001	庫存現金-台幣	EMPL 楊○慧 1703/23
9	9120-1704070002	1101001	庫存現金-台幣	AHIG-17
10	9120-1705190005	1101001	庫存現金-台幣	AHIG-17
11	9120-1706210004	1101001	庫存現金-台幣	AHIG-17
12	9120-1707250005	1101001	庫存現金-台幣	AHIG-17
13	9120-1707250015	1101001	庫存現金-台幣	EMPL 楊○
14	9120-1707280005	1101001	庫存現金-台幣	AHIG-17
15	9120-1708180001	1101001	庫存現金-台幣	AHIG-17
16	9120-1708290004	1101001	庫存現金-台幣	AHIG-17
17	9120-1709270001	1101001	庫存現金-台幣	AHIG-17
18	9120-1711130002	1101001	庫存現金-台幣	AHIG-17
19	9120-1711150004	1101001	庫存現金-台幣	EMPL 程○
20	9120-1711170005	1101001	庫存現金-台幣	AHIG-17
21	9120-1701170003	1101001	庫存現金-台幣	AHIP-17
22	9120-1701230004	1101001	庫存現金-台幣	AHIP-17
23	9120-1701230005	1101001	庫存現金-台幣	AHIP-17
24	9120-170230001	1101001	庫存現金-台幣	紹○500具
25	9120-1702200004	1101001	庫存現金-台幣	AHIP-17
26	9120-1702210007	1101001	庫存現金-台幣	AHIP-17
27	9120-1702210008	1101001	庫存現金-台幣	AHIP-17
28	9120-1702240016	1101001	庫存現金-台幣	AHIP-17

2. 中文斷詞

詞是最小有意義且可以自由使用的語言單位。任本中的詞才能進行進一步的處理，例如機器翻譯、語中文自動分詞的工作成了語言處理不可或缺的技术。詞和文本做比對，找出可能包含的詞，由於存在歧義式多討論如何解決分詞歧義的問題，而較少討論如何（新詞如何辨認）。

3. 斷詞結果

由 楊(Nb) 0(FW) 慧(Na) 預支(VC) 台幣(Na) 換(VC) 當處理 2 7 4 3 (Neu) 吳(Nb) 0(FW) 庭(Nc) 保管(VC) 詞彙的 慧(Na) MISCAH(FW) 1 (Neu) G-(FW) 1 7 0 2 1 3 彙的裡 OO(FW) 薪資(Na) 1 7 (Neu) / (FW) 0 3 (Neu) / (FW) 出現在 楊(Nb) 0(FW) 慧(Na) MISCAH(FW) 1 (Neu) G-(FW) 詞彙比 (FW) 楊(Nb) 0(FW) 慧(Na) 1 7 (Neu) / (FW) 0 3 及詞彙 0 0 6 (Neu) 楊(Nb) 0(FW) 慧(Na) MISCAH(FW) 1 建立領 MISCAH(FW) 1 (Neu) G-(FW) 1 7 0 6 2 1 0 0 1 1 0 之中文 7 2 5 0 0 1 1 0 3 0 0 0 6 (Neu) 楊(Nb) 0(FW) 慧 連詞類 (PARENTHESISCATEGORY) EMPL(FW) 楊(Nb) 0(FW) 慧 詞，並 1 7 0 7 2 8 0 0 1 1 0 3 0 0 0 6 (Neu) 楊(Nb) 0 標記為 (Neu) 楊(Nb) 0(FW) 慧(Na) MISCAH(FW) 1 (Neu) 被拆成 MISCAH(FW) 1 (Neu) G-(FW) 1 7 0 9 2 7 0 0 1 1 0 1 3 0 0 2 1 0 3 0 0 0 6 (Neu) 楊(Nb) 0(FW) 慧 1 5 (Neu) AH(FW) 1 (Neu) G-(FW) 1 (Neu)

7 1 1 1 7 0 0 1 1 0 3 0 0 0 6 (Neu) 楊(Nb) 0(FW) (Neu) 李(Nb) 0(FW) 蘭(Na) MISC(FW) 黃(Nb) 0(FW) 1 7 0 1 2 3 0 0 1 1 0 1 0 0 0 8 (Neu) 李(Nb) (Neu) AH(FW) 1 (Neu) P-(FW) 1 7 0 1 2 3 0 0 2 (Neu) +(FW) 其他(Neqa) 6 6 8 0 (Neu) +(FW) 江 1 5 0 0 (Neu) 江(Na) 0(FW) 祥和(VH) 6 3 4 6 0 (FW) 蘭(Na) MISCAH(FW) 1 (Neu) P-(FW) 1 7 P-(FW) 1 7 0 2 2 1 0 0 1 1 0 1 0 0 0 8 (Neu) 等 0 0 8 (Neu) 李(Nb) 0(FW) 蘭(Na) MISCAH(FW) 1

圖 2 中文斷詞處理程序

(二)應用 REA 模型詞彙分類

透過 REA 模型分類原則萃具有意義之字詞。各字詞與會計分錄編號進行串聯與統計各字詞頻率。表 2 為透過詞性為普通名詞(Na)，故被歸類為「資源」(Resource: R) 類別所包含之字詞摘錄列表。由該表可快速檢視該公司其他費用科目的實際用

途頻率以及佔科目總金額比例。部分頻率偏低但總金額佔比相對較高的項目（如：退佣、經驗）較容易引起關注。另外，如結合字詞頻率與總金額，可發現「服務費」、「寢具」以其規律性（約每月一次）但總金額佔比達到 60% 以上且平均金額也最高，最容易引起關注。相對而言，「職災險」、「誠實險」、「團險」，雖然總金額佔比超出 15%，卻因為字詞頻率最高，平均金額相對偏低，而降低了可能的風險。

表 2 資源類別字詞分析（摘錄）

編號	字詞	次數	交易平均金額	交易加總金額	佔科目總金額之比例
1	退佣	1	63,389.00	63,389	4.09%
2	經驗	1	44,048.00	44,048	2.84%
3	網路	1	1,500.00	1,500	0.10%
7	物流	2	8,201.50	16,403	1.06%
8	補助費	2	8,201.50	16,403	1.06%
10	旅平險	4	4,386.00	17,544	1.13%
11	保險費	4	1,200.25	4,801	0.31%
14	商品	8	1,178.38	9,427	0.61%
15	行銷	9	1,994.11	17,947	1.16%
R	工務	10	1,458.20	14,582	0.94%
(資源)	汽車材	10	991.60	9,916	0.64%
19	設備	11	618.18	6,800	0.44%
20	生技	11	124.18	1,366	0.09%
21	服務費	12	80,000.00	960,000	61.87%
22	寢具	13	74,173.85	964,260	62.14%
23	責任險	14	10,610.00	148,540	9.57%
24	業務	15	1,760.40	26,406	1.70%
26	職災險	350	741.27	259,444	16.72%
27	誠實險	362	841.44	304,600	19.63%
28	團險	385	799.43	307,780	19.84%

1. 上表為各字詞所對應之交易總金額，計算佔「其他費用」科目總金額之比例，可與將金額之重大性與字詞頻率併同考量風險程度。
2. 資料來源：本研究整理

上述字詞頻率與金額之分布情形如以散佈圖表達，可明顯約略分為三個不同資料群組如圖 3 所示。分別為左上角（次數低且金額較高）、左下角（次數低且金額較低）以及右下角（次數最高但平均金額很低）。由風險觀點，左上角之資料項目因交易平均金額較高且發生次數較低故最容易引起關注，屬於高風險區域。



圖 3 資源類別字詞頻率與交易平均金額之散佈圖

(三)分錄整體風險評量與風險構面

由於一個會計分錄由多筆資料項目所組成，每個項目都有摘要備註，分別對應了不同的科目與借貸方金額；因此為了綜合評估每一個會計分錄所組成項目的整體風險，本研究將前述之二項風險因素：REA 字詞頻率與完整性整併為摘要評分構面；另一構面則為傳票金額重大性，分別予以分級作為風險評量的標準。分級標準說明如下：

1. 風險構面—摘要評分（字詞頻率與完整性）

每個類別（分為資源、事件、人員等三大類別），依字詞發生次數做風險評估，“由小至大”以五分位數分為五個組距，以頻率次數極低區間為第一組距，將風險分為極高、高、中、低、極低等五個等級。表 3 為測試資料字詞頻率的分組評分標準。

將每個會計分錄單一項目的摘要敘述進行 REA 分類後，即可依字詞頻率得知該分錄在 REA 三大分類的個別等級評分。如果某個分錄項目缺乏某個類別字詞，則因說明不完整而將該類風險等級標示最高為 6；之後再分別先得到整個分錄的各類別

平均值，以及將三個類別的平均值再次平均後作為該分錄的 REA 完整性風險評估值，示例如表 5。

表 3 字詞頻率之風險等級標準（依五分位數之分布）

組距	次數	評分	風險等級
組距一	1~5	5	極高
組距二	6~12	4	高
組距三	13~35	3	中
組距四	36~50	2	低
組距五	51~429	1	極低

資料來源：本研究整理

2. 風險構面—金額重大性

由於每個分錄的項目數量多寡不一且各對應不同金額，為綜合評估分錄的重大性，依各項目金額“由小至大”以五分位數分為五個組距，以金額極低區間為第一組距，將風險分為極低、低、中、高、極高，表 4 為字詞的金額大小的評分標準。最後並計算各個分錄金額重大性的多個項目評分之平均值，以利於彙整評量（示例如表 5）。

表 4 金額重大性之風險等級標準（依五分位數之分布）

組距	金額	評分	風險等級
組距一	1.00~ 1,410.00	1	極低
組距二	1,411.00 ~ 3,469.00	2	低
組距三	3,470.00 ~ 10,700.55	3	中
組距四	10,700.56 ~ 33,105.25	4	高
組距五	33,105.26 ~ 199,047.50	5	極高

資料來源：本研究整理

表 5 單一會計分錄整體風險評量計算示例

傳票編號	科目名稱	摘要	借貸	R 類別	E 類別	A 類別	金額	摘要平
			金額	評分	評分	評分	評分	均評分
XXX-1712	銷—其他保險費-林○○.葉○○—出差歐洲	2392	3.000	3.000	5.000	2.250		
010003	保險費 1105-1112-共 7 天							
	管—其他保險費-鐘○○.詹○○—出差美國	2992	3.000	3.000	4.500	2.500		
	保險費 0129-0212-共 14 天							
	應付費用 SFB05 保險費 17/12/20	-5384	3.000	6.000	6.000	3.000		
會計分錄整體風險各字詞類別平均評分				3.000	4.000	5.167	2.583	
								4.056

最後本研究依據每個會計分錄於摘要評分（REA 字詞頻率及完整性）與金額重大性兩個構面，組成風險矩陣以利於快速識別各不同風險之會計分錄，並以散佈圖表達其及分布情形。審計人員可以利用 Power BI 商業智慧分析工具所設計之系統離型介面，依據風險程度優先針對高風險區域之傳票進行查核相關傳票明細內容，其示例如圖 4 所示。由圖中審計人員可採用風險圖表的相對分佈位置查看不同特徵的傳票內容；由於金額越大越具有重大性加上摘要評分越高代表摘要的 REA 字詞出現頻率較低或是缺乏不完整，因此越偏右上角的圓點代表風險越大的傳票。另外，為了兼俱表達同一張傳票多個項目的平均值與傳票的借貸方總額兩者不同的特性，圖表的 Y 軸是依各傳票組成項目的金額平均值分組採 1 至 5 分表達；而各圓點的大小則是表達單張傳票的借貸方金額大小。使用者可依風險考量點選圓點後，下方顯示對應之傳票內容。

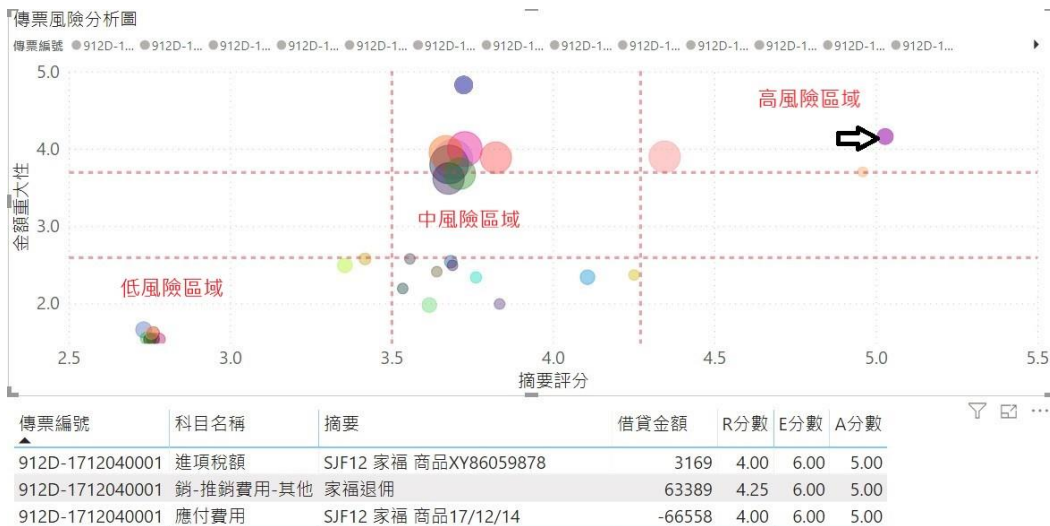


圖 4 以互動圖表呈現會計傳票之風險分布情形

三、評估解決方案

本研究目的為應用文本探勘技術進行會計分錄測試，建立一套風險評估之離型，藉由字詞頻率、交易金額與摘要完整性等風險因素進行風險評量，以利於快速呈現不同風險的會計傳票。為使本研究建立之風險評估模型與實務相符，因此以三位實務工作年資均超過十年之會審專家為模擬使用者進行訪談，以確認風險評估機制的可行性及有效性。評估方式以實地離型展示與訪談為主；說明研究問題及解決方案，並展示透過風險評估離型所發現的高風險會計傳票內容。以下為由質性訪談紀錄所歸納之受訪者回饋意見及建議：

(一)本研究風險評估離型之優點

1. 透過各字詞之發生頻率與對應金額(如表 2 與圖 3)，可以快速瞭解該科目會計分錄交易之分布類型、頻率與金額大小，以評估其合理性及風險。相較於傳統交易金額篩選方法，傳統方法偏向僅能就單張傳票金額較大者進行抽樣，在查核時間的限制下，通常選擇金額最大者的前幾項傳票進行審視。結合文本探勘方法，則將可就金額相對較高但發生字詞次數偏低的傳票進行分析；或是字詞頻率高且同一字詞相關傳票彙總金額較高的傳票亦進行審查。如此可減少受查者為規避抽查，針對審計人員的抽樣金額門檻，將重大交易拆解為多項紀錄的偵查風險。
2. 可針對會計分錄的摘要分析出字詞頻率較低或敘述不完整的內容特徵，輔以金額重大性原則，迅速有效地篩選出高風險的傳票。尤其針對摘要不完整敘述的傳票，審查人員也可能就所擷取字詞與會計科目的關連性與合理性進行評估。
3. 由於相較於傳統交易金額篩選方法，本研究之風險評估方法，增加了摘要字詞頻率與摘要完整性等構面，資訊量較多，使用視覺化圖表呈現，將更容易一覽某一科目相關字詞與金額的分佈情形。透過風險矩陣視覺化分析介面(如圖 4)，可將不同類型傳票的發生頻率(點數多寡)、金額大小(圓點大小)與字詞頻率及完整性等分錄摘要品質等，得到直觀的整體分佈概況。並可快速逐一檢視高風險區域所涵蓋傳票的分錄內容，而這些往往是審計人員僅憑以往依金額大小之審計抽樣方法所不易進一步偵查的異常資料(傳統作法僅能輔以審計人員針對分錄清冊快速大量瀏覽審視，在大量傳票情境下，較不易謹慎逐一查看)。

(二)風險評估雛型之使用限制與建議

1. 由於目前採用中文斷詞方法，摘要內容若僅是數字、英文所組成，例如傳票編號和來源單據號碼，在中文斷詞時並無法識別為有意義的字詞；所以在資料前置處理階段視為無意義之字詞加以剔除。故如整個分錄摘要僅包含來源憑證編號卻未附加說明，將視為 REA 類別不完整傳票，可能誤判為高風險傳票。因此建議會計部門摘要記錄應更加完整，或是在分析該類傳票時，可以加入其他因素的考量，例如金額是否重大、是否為自動拋轉分錄、來源系統模組與對應的會計科目是否正確等，藉此輔助確認傳票是否為異常分錄。
2. 不同的科目所對應的字詞頻率應有明顯差異，因此建議未來可建立科目別詞彙庫或產業別詞彙庫，持續累計得到不同字詞的頻率權重。如此，不但可以提高詞彙風險估算準確度，在時間受限時，也可用基於既有歷史資料的詞頻表剔除較低風險的分錄。
3. 由於不實普通日記簿分錄或其他調整分錄，其特徵通常包括使用不相關、不尋常或甚少使用之科目。另一方面，用斷詞分析針對所有科目分錄進行斷詞與風險評估，可能資料量太大而耗費太多時間，分析結果也不易相互比較。本系統雛型建議後續可結合其他會計分錄測試方法，利用其分析結果所得之符合重大性高風險會計科目，再進行單一會計科目相關會計分錄之摘要字詞分析，以協助快速區分具潛在風險之異常會計分錄資料。

伍、結論與建議

一、研究結論與貢獻

本研究採設計科學研究法，以文本探勘技術結合 REA 模型設計了會計分錄風險評估系統雛型；並針對會審專家展示與訪談，評估該雛型之可行性及有效性。本研究透過文本探勘技術將會計分錄摘要內容進行字詞的拆解與詞性分析，透過 REA 模型進行詞性歸類與完整性分析，並依照字詞頻率及金額對傳票進行風險評估，以協助審計人員快速篩選及檢視高風險傳票內容，再實地詢問相關憑證細節。經會審專家實地使用及回饋建議，本方法應可彌補審計抽樣與現有會計分錄測試方法的不足，可增進查核的效果與效率。

由本研究設計之會計分錄風險評估系統雛型，可以有效地偵測出具有風險特徵的異常資料，未來如開發為自動化系統，將可以有效降低查核成本且同時提高審計品質。在研究程序中為評估此分析方法與系統雛型之使用效益，彙總會審專家意見及整合文獻相關研究概況，本研究具有以下三點貢獻：

1. 過去的研究中，文本探勘技術僅應用在分析財務報告，尚未有研究針對會計分錄摘要加以分析；本研究所發展的會計分錄風險評估方法，可與其他會計分錄測試方法結合使用，彌補現有會計分錄測試的不足。例如先進行各科目交易金額重大性、交易頻率、人工分錄比例等，挑選出高風險之會計科目，再針對該科目相關分錄以文本探勘進行風險評估。
2. 可針對會計分錄的摘要分析出字詞頻率較少或不尋常的特徵，迅速且有效地探測出具有風險的傳票。後續可利用特徵詞的共現性分析方法 (co-word analysis) (如：陳良駒、張正宏與陳日鑫，2010)，或是建立特定領域的語料庫 (馬偉雲、謝佑明、楊昌樺與陳克健，2001)；以利於更有效且準確地斷詞，以及建立各會計科目相關詞彙庫。日後查核活動可配合使用，透過剔除低風險字詞，快速篩選出高風險之會計傳票。
3. 運用視覺化工具設計所建立的互動式會計傳票風險矩陣圖表介面，可快速檢視高風險傳票完整分錄內容，評估是否為異常分錄，可能發掘查核人員依審計抽樣所不易發現的異常交易。

二、研究限制

1. 本研究使用中央研究院之中文斷詞系統作為斷詞工具，由於該系統僅能針對中文字詞進行拆解，對於英文、數值無法判斷其規則，因此將會影響識別摘要字詞風險的準確性。
2. 會計分錄摘要的詳細程度與完整性，依各公司會計人員的記錄習慣而異。如果摘

要文字過於簡略句型結構不完整，中文斷詞將更不容易或造成詞性混淆。

3. 本研究因考量資料保護，針對公司名稱及人員進行去識別化處理與適當明細資料保密，故無法展示高風險傳票的完整內容。

三、研究建議

建議未來會計師事務所如大量持續使用此類系統時，可透過專家經驗知識或以機器學習演算法進行詞彙與 REA 類別之自動歸類，以及建置各產業別或各不同會計科目別的相關詞彙庫，使系統可以更快速且自動化地進行會計分錄風險評估，以提升查核活動之效果及效率。另外建議也可以結合目前日漸普及之機器流程自動化（Robotic Process Automation, RPA）將上述資料分析程序與中文斷詞系統、視覺化分析等不同軟體平臺工具加以整合自動化，使得分析程序更為快速有效率。

參考文獻

- 王明昌、鄭揚耀與柯建全，2015，企業財務危機前之媒體管理，會計評論，第 61 期：77-119。
- 周濟群，2016，審計專業如何使用大數據？，會計研究月刊，第 362 期：82-89。
- 周濟群與連子杰，2011，運用文字探勘與 XBRL 技術提升企業資訊擷取與整合效益之研究，當代會計，第 12 卷第 1 期：85-114。
- 馬偉雲、謝佑明、楊昌樺與陳克健，2001，中文語料庫構建及管理系統設計，第十四屆自然語言與語音處理研討會，國立成功大學，臺南。
- 孫嘉明，2018，雲端運算環境下審計數據分析之發展趨勢與挑戰，月旦會計實務研究，第 7 期：54-61。
- 財團法人中華民國會計研究發展基金會，2006 審計準則公報第四十三號—查核財務報表對舞弊之考量，臺北：財團法人中華民國會計研究發展基金會。
- 財團法人中華民國會計研究發展基金會，2020，審計準則公報第七十四號—查核財務報表對舞弊之責任，臺北：財團法人中華民國會計研究發展基金會。
- 張立程，2015，以資料探勘技術建置持續性監督機制，國立雲林科技大學會計系未出版碩士論文。
- 陳世榮，2015，社會科學研究中的文字探勘應用：以文意為基礎的文件分類及其問題，人文及社會科學集刊，第 27 卷第 4 期：683-718。
- 陳良駒、張正宏與陳日鑫，2010，以特徵詞共現特性探討知識管理研究議題相關性—使用共詞與關聯法則分析，資訊管理學報，第 17 卷第 4 期：31-60。
- 陳裕菘、謝邦昌、李勝輝與陳郁婷，2014，運用文字探勘與資料採礦技術建立匯率預測模型—以人民幣兌新台幣為例，數據分析，第 9 卷第 1 期：133-146。
- 曾元顯，2004，應用於資訊檢索的中文 OCR 錯誤詞彙自動更正，中國圖書館學會會報，第 72 期：23-31。
- 廖紫吟與孫嘉明，2014，以本體論為基礎之語意稽核系統架構設計，電腦稽核，第 29 期：62-77。
- 譚家蘭，2012，導入國際會計準則之資訊揭露—應用資料探勘與文字探勘技術，會計研究月刊，第 318 期：38-43。
- 顏素英，2018，應用本體論方法建構審計資料分析查核程序—以日記簿分錄測試為例，國立雲林科技大學會計系未出版碩士論文。
- Alles, M., and G. L. Gray. 2016. Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors. *International Journal of Accounting Information Systems* 22: 44-59.
- Ata, H. A., and I. H. Seyrek. 2009. The use of data mining techniques in detecting

- fraudulent financial statements: An application on manufacturing firms. *Suleyman Demirel University Journal of Faculty of Economics and Administrative Sciences* 14 (2): 157-170.
- Auditing Standards Board of the American Institute of Certified Public Accountants (AICPA). 2002. *Statement on Auditing Standards No. 99: Consideration of Fraud in a Financial Statement Audit*. New York, NY: AICPA.
- Brown-Liburd, H., and M. A. Vasarhelyi. 2015. Big Data and audit evidence. *Journal of Emerging Technologies in Accounting* 12 (1): 1-16.
- Byrnes, P., T. Criste, T. Stewart, and M. Vasarhelyi. 2014. *Reimagining auditing in a wired world*. AICPA White Paper. New York, NY: AICPA.
- Center for Audit Quality (CAQ). 2008. *Practice Aid for Testing Journal Entries and Other Adjustments Pursuant to AU Section 316*. Washington, D.C.: CAQ.
- Dalal, M. K., and M. A. Zaveri. 2011. *Heuristics based automatic text summarization of unstructured text*. Proceedings of the International Conference & Workshop on Emerging Trends in Technology (ICWET), Mumbai, India.
- Debreceny, R. S., and G. L. Gray. 2010. Data mining journal entries for fraud detection: An exploratory study. *International Journal of Accounting Information Systems* 11 (3): 157-181.
- Dunn, C., G. J. Gerard, and S. V. Grabski. 2016. Resources-events-agents design theory: A revolutionary approach to enterprise system design. *Communications of the Association for Information Systems* 38: 554-595.
- Fay, R., and E. M. Negangard. 2017. Manual journal entry testing: Data analytics and the risk of fraud. *Journal of Accounting Education* 38: 37-49.
- Financial Reporting Council (FRC). 2017. *The Use of Data Analytics in the Audit of Financial Statements*. Available at: https://www.frc.org.uk/getattachment/4fd19a18-1beb-4959-8737-ae2dca80af67/AQTR_Audit-Data-Analytics-Jan-2017.pdf (last access May 1, 2018).
- Fisher, I. E., M. R. Garnsey, and M. E. Hughes. 2016. Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management* 23 (3): 157-214.
- Frawley, W. J., G. Piatetsky-Shapiro, and C. J. Matheus. 1992. Knowledge discovery in databases: An overview. *AI Magazine* 13 (3): 57-70.
- Fukukawa, H., and T. J. Mock. 2012. Auditors' evidence evaluation and aggregation using beliefs and probabilities. *International Journal of Approximate Reasoning* 53 (2): 190-199.

- Geerts, G. L. 2008. Introduction to the REA 25th anniversary special section. *Journal of Information Systems* 22 (2): 215-217.
- Guan, J., A. S. Levitan, and J. R. Kuhn Jr. 2013. How AIS can progress along with ontology research in IS. *International Journal of Accounting Information Systems* 14 (1): 21-38.
- Hand, D. J., H. Mannila, and P. Smyth. 2001. *Principles of Data Mining (Adaptive Computation and Machine Learning)*. 1th edition. Cambridge, MA: MIT Press.
- Kloptchenko, A., T. Eklund, J. Karlsson, B. Back, H. Vanharanta, and A. Visa. 2004. Combining data and text mining techniques for analysing financial reports. *Intelligent Systems in Accounting, Finance and Management* 12 (1): 29-41.
- Liu, Y., and K. C. Moffitt. 2016. Text mining to uncover the intensity of SEC comment letters and its association with the probability of 10-K restatement. *Journal of Emerging Technologies in Accounting* 13 (1): 85-94.
- Loughran, T., and B. McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54 (4): 1187-1230.
- Lungu, I., and T. Vătuuiu. 2007. Computer assisted audit techniques. *Annals of the University of Petrosani Economics* 7: 217-224.
- McCarthy, W. E. 1982. The REA accounting model: A generalized framework for accounting systems in a shared data environment. *The Accounting Review* 57 (3): 554-578.
- Ngai, E. W. T., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems* 50 (3): 559-569.
- Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. 2007. A design science research methodology for information systems research. *Journal of Management Information Systems* 24 (3): 45-77.
- Public Company Accounting Oversight Board (PCAOB). 2002. *AU Section 316: Consideration of Fraud in a Financial Statement Audit*. Washington, D.C.: PCAOB.
- Shirata, C. Y., and M. Sakagami. 2008. An analysis of the “going concern assumption”: Text mining from Japanese financial reports. *Journal of Emerging Technologies in Accounting* 5 (1): 1-16.
- Sullivan, D. 2001. *Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing, and Sales*. Hoboken, NJ: John Wiley and Sons.
- Sun, T., and M. A. Vasarhelyi. 2018. Embracing textual data analytics in auditing with deep learning. *The International Journal of Digital Accounting Research* 18: 49-67.

- Tan, P. N., M. Steinbach, and V. Kumar. 2006. *Introduction to Data Mining*. 1st edition. Boston, MA: Pearson Addison Wesley.
- Tsai, M. F., and C. J. Wang. 2017. On the risk prediction and analysis of soft information in finance reports. *European Journal of Operational Research* 257 (1): 243-250.

