

審計機關創新提案表【性質：創意類】

提案範圍	<input checked="" type="checkbox"/> 審計 <input type="checkbox"/> 非審計
提案單位	<input checked="" type="checkbox"/> 單位提案（審計部新北市審計處） <input type="checkbox"/> 個人提案
提案人員	主要提案人姓名：稽察員林東慶 貢獻度：50% 參與提案人姓名：稽察盧凱偉 貢獻度：25% 稽察張士賢 貢獻度：15% 稽察兼科長杜竟源 貢獻度：10%
提案名稱	創新運用 Python 語言撰寫爬蟲程式(Web Crawler)蒐集網路資料，結合最小編輯距離法程式、Power BI 等分析工具及各類資料庫，查核火災預防相關機制運作及廣義民間參與招商案件及政府採購得標廠商履約保證金繳交情形
創新面向	<input checked="" type="checkbox"/> 產出 <input checked="" type="checkbox"/> 流程 <input type="checkbox"/> 組織 <input type="checkbox"/> 溝通 <input type="checkbox"/> 其他(敘明創新之面向)
提案緣起	<p>一、創新來源</p> <p>一般數據分析的三要素分別是「資料來源」、「技術方法」及「查核構想」(詳圖)。其中審計機關常見之「資料來源」，包括機關提供資料、OPENDATA、GBA、CBA 及 AIS 系統等，但上述資料來源仍有不足之處，舉如民國 106 年 11 月中和違建出租發生大火，造成嚴重傷亡，相關機關對民眾違建出租情事卻無法掌握，顯示如果審計機關只就機關提供或 OPENDATA 資料分析，必然產生盲點；另外在「技術方法」領域，ACL 等分析工具多擅長於結構化之數據比對分析，較難運用於網際網路之文字、影片及 GBA、CBA、AIS 系統等摘要說明欄位資料等非結構化或半結構化資料，且該等非結構化資料，由於每個人對事情描述書寫之習慣不同，雖可透過人工方式判讀，理解其內容含意，卻不利於程式或自動化方式讀取，造成運用之困難。</p> <p>因此，本案運用 Python 語言撰寫「網路爬蟲(Web Crawler)」程式蒐集網路資料，擴大審計機關資料來源(詳附件 1)，並運用 Excel VBA 巨集指令撰寫之最小編輯距離法(Minimum Edit Distance)程式(詳附件 2)，分析不同字串間之相似程度，篩選出最佳接近標的物字串，作為跨資料庫比對之橋樑及基礎，達成系統化、大規模查核之目標。茲就「網路爬蟲」程式及最小編輯距離法程式，如何運用於查核火災預防相關機制運作情形及廣義民間參與案件招商與履</p>

約管理及政府採購履約保證金繳交情形概要說明如次：

- (一)建築物公共安全檢查申報(下稱安檢申報)、消防安全設備檢修申報(下稱消檢申報)係火災預防機制最重要之設計。本案運用 Python 語言編寫「網路爬蟲」程式，進行網際網路資料探勘，蒐集「旅館」、「民宿」、「日租套房」、「護理之家」及「幼兒園」為關鍵字之商家名稱及地址，並運用最小編輯距離法程式，多方交叉比對上開爬蟲程式所蒐集之網路資料與各機關列管之安檢申報、消檢申報及各種 OPENDATA 資料，找出機關漏未列管場所，並結合文獻分析多重審計技術方法，瞭解相關機制運作情形。
- (二)各機關每年辦理之促進民間參與公共建設(下稱促參)、委託經營及標租等廣義民間參與案件眾多，惟各類案件辦理公告招商規定不一，除造成廠商或申請人無法藉由單一管道或網站取得招商資訊，不利民間參與及廠商間之競爭外，各機關是否依規定確實公開辦理招商亦待查證，爰運用 Power BI 軟體篩選 CBA 及 AIS 系統中含有「權利金」、「租金」或「標租」等關鍵字之資訊檔案，再運用最小編輯距離法程式，找出疑似未公開辦理招商公告之異常案件後，再進一步函請各機關提供相關文件。
- (三)依投標須知範本規定，廠商應於採購案決標日之次日起期限內繳交，得標廠商未於規定時間內繳妥保證金者，得處以懲罰性違約金，逾規定日之次日起 15 日內仍未繳妥者，得以解除、終止契約，爰運用最小編輯距離法程式，比對採購得標廠商資料與 CBA 及 AIS 資訊檔中繳交履約保證金情形。

二、創新程度

- (一)Python「網路爬蟲」程式及最小編輯距離法程式均由本單位人員自行撰寫開發，經查閱審計部歷次會報，尚未有其他審計單位提出相同功能程式。
- (二)本案由數據分析三要素中之「資料來源」及「技術方法」著手，以 Python「網路爬蟲」程式擴大審計機關「資料來源」，減輕轄審機關提供資料之負擔；以最小編輯距離法程式，提供字串之相似程度比對功能，作為跨資料庫比對之橋樑及基礎，改善「技術方法」可達成系統化、大規模查核之目標。
- (三)Python「網路爬蟲」程式可依照審計人員需要及查核主題，修改關鍵字即可蒐集網路上不同類別商家名稱、地址、電話及經緯度座標等資料；最小編輯距離法程式對於字串之比對幾無限制，對於未來推廣至各審計機關後，可配合各種「查核構想」使用彈性極大。
- (四)業將程式無償提供機關使用，協助解決問題，落實審計機關係被審核機關之協力夥伴角色，爭取機關對審計機關之認同。

	(五)火災預防機制運作情形、廣義民間參與案件招商作業及採購案件履約保證金繳交情形，查核議題新穎。
問題分析	<p>一、現況問題</p> <p>(一)審計機關請機關提供各類資料電子檔案，機關往往需要額外耗費人力時間準備或基於資訊安全考量延遲提供。</p> <p>(二)現行數據分析工具多提供字串精準比對，必須兩個字串完全相同才能挑出，常常造成比對分析之困難，舉如「台灣」與「臺灣」或「重慶路1號」與「重慶路一號」在一般精準比對功能下，便認定為不同字串，影響比對結果。</p> <p>(三)安檢申報及消檢申報係火災預防機制最重要之設計，但台灣地區違建情形普遍，民眾申報態度消極，相關機關囿於預算及人力問題，欠缺管道察覺高風險場所或民眾違建情形。</p> <p>(四)促參、委託經營及標租等廣義民間參與案件，因法令規定招商公告作法不一，造成廠商或申請人無法藉由單一管道或網站取得招商資訊，不利民間參與及廠商間之競爭，且各機關是否依規定公開辦理招商較少審計機關關注。</p> <p>(五)依投標須知範本規定，廠商應於採購案決標日之次日起期限內繳交，未於規定時間內繳妥保證金者，得處以懲罰性違約金，逾規定日之次日起15日內仍未繳妥者，得據以解除、終止契約。但各機關所辦採購案件是否依規定期限繳交履約保證金，審計機關缺乏有效工具瞭解其辦理情形。</p> <p>二、解決必要性</p> <p>(一)撰寫 Python「網路爬蟲」程式蒐集網際網路資料，可減輕被審核機關負擔，增加審計機關研析資料之多樣性。</p> <p>(二)審計機關自行撰寫最小編輯距離法程式，可強化字串比對之工具及技術，運用於多面向議題。</p> <p>(三)火災預防機制運作情形攸關民眾生活安全，協助機關善用網際網路大數據分析技術及府內各機關資訊系統，主動發掘高風險之場所積極管理，可發揮審計機關以民為本之功能。</p> <p>(四)督促機關建置單一招商公告管道或網站，可促進民間參與及廠商競爭，減少弊端，可共創民眾、機關及廠商三贏。</p> <p>(五)採購案件履約保證金攸關採購執行保證及機關權益。</p>
建議方案	<p>一、創新程度</p> <p>茲將本案運用 Python「網路爬蟲」及最小編輯距離法程式，結合 Power BI 等分析工具，進行跨資料庫研析，所獲之創新結果說明如次：</p> <p>(一) 結合各機關列管資料，查核火災預防機制運作情形部分：1. 建請在考量人力允許範圍內，妥善運用網際網路大數據分析</p>

技術及府內各機關資訊系統，主動發掘高風險之場所積極管理；2. 提供 Python「網路爬蟲」程式及最小編輯距離程式予機關，並視其需要派員協助操作及諮詢；3. 本次比對結果，發現疑似商家未申報者 517 家，疑似機關漏未列管者 308 家(詳下表)，經函請機關妥處(詳附件 3)。

單位：家

研析結果	旅館、日租套房及民宿	護理機構	幼兒園托兒所	合計
有安檢申報，但無消檢申報者	14	6	212	517
有消檢申報，但無安檢申報者	32	29	224	
安檢申報資料漏未列管	137	11	23	308
消檢申報資料漏未列管	111	6	20	

截至 107 年 7 月 26 日止，機關初步查復 566 間商家確有營業行為，卻未申報或納入列管，其中屬商家未申報者 470 家，機關漏未列管者 96 家。

(二)結合 Power BI 程式及 CBA、AIS 等各類資料庫，查核廣義民間參與案件招商與履約管理及政府採購履約保證金繳交情形

1. 廣義民間參與案件部分：(1)廣義民間參與案件招商公告作業規範不一，致各機關學校刊登公告位置散見於各處，不利民間參與及廠商間之競爭，建請修訂相關公告規定，並要求各機關學校刊登於同一資訊網頁或進行超連結，共創業者、民眾及機關學校三贏；(2)委託經營案件契約混雜政府勞務採購契約條文，增加履約困擾，且有 2 案長期逕洽民間廠商續約迄今；(3)1 案採逕予出租案件之適用要件與規定未合，且租金標準不一；(4)各機關經管之公有財產依採購法辦理委外經營使用者，未檢討其收益不違背其事業目的或原定用途；(5)原依政府採購法第 99 條規定辦理之招商案件，卻誤引政府採購法後續擴充條款辦理續約，不符函釋規定；(6)1 案採購金額認定錯誤，逕洽廠商未公開招商等情事(詳附件 4)。
2. 政府採購履約保證金繳交部分：AIS 系統中，各級學校 106 年有 200 筆憑證登載得標廠商繳納履約保證金之日期，逾投標須知規定期限(詳附件 5)，據復：計有 14 筆廠商已繳納違約金、有 8 筆刻正追繳違約金當中、3 筆出納人員違反出納管理手冊規定逾期入帳、175 筆係因傳票日期與收款收據日期有落差所致，已於新北市各級學校總務行政資訊網刊登「轉知審計部新北市審計處有關收入傳票記載有得標廠商履約保證金之注意事項」最新資訊 1 則，通知各校爾後注意辦理(詳附件 6)；經將相同技術運用於 CBA 系統資料，又發現各機

	<p>關 106 年有 367 筆憑證登載得標廠商繳納履約保證金之日期，逾投標須知規定期限(詳附件 7)，其中 152 筆屬得標廠商延遲繳交履約保證金，147 筆屬納人員違反出納管理手冊規定逾期入帳，68 筆係因傳票日期與收款收據日期有落差所致。</p>
預期效益或影響	<p>■效益：</p> <ol style="list-style-type: none"> 1. Python「網路爬蟲」程式方面，經上網蒐集旅館、日租套房、民宿、護理之家及幼兒園等商家資訊，均能快速完成任務，且修改查詢主題方便，若能獲得審計機關參採，應能配合各查核主題，擴大審計機關蒐集資料來源，或未來結合 CBA、AIS 或機關各類資料庫，進行跨領域比對分析，均可提高審計機關之審計績效。 2. 最小編輯距離法程式方面，在火災預防機制運作情形之查核，比對出有 470 家商家未申報，機關漏未列管者 96 家；在廣義民間參與案件招商情形之查核中，比對出有 6 案未依規定公開辦理招商，其中屬長期逕洽廠商續約者 2 案，不符逕予出租案件之適用要件者 1 件，原依政府採購法第 99 條規定辦理之招商案件，卻誤引政府採購法後續擴充條款辦理續約者 3 件，屬採購金額採購金額認定錯誤，逕洽廠商未公開招商者 1 件；在採購履約保證金繳交情形之查核中，比對出機關 106 年 AIS 系統資料有 200 筆履約保證金逾規定期限繳交等異常情事，106 年 CBA 系統資料有 367 筆履約保證金繳交異常，其運算快速、成果精準，已獲豐碩成果。 <p>■影響：</p> <ol style="list-style-type: none"> 1. 本案創新撰寫 Python「網路爬蟲」程式及最小編輯距離法程式，進行網際網路文字探勘及非結構化字串相似度比對，使審計機關之資料分析範圍再一步擴大，未來可運用於各種議題，對於提高審計機關工作效率有明顯之幫助。 2. 提供 Python「網路爬蟲」程式及最小編輯距離法程式予機關使用，可增進與被審核機關之夥伴關係，建立審計機關專業形象。
相關附件	<p>附件 1：Python「網路爬蟲」程式碼及成果示範 附件 2：最小編輯距離法程式碼 附件 3：通知主動發掘建築及消防高風險之場所積極管理函 附件 4：通知委託經營及標租案應辦理事項函 附件 5：通知各學校政府採購履約保證金逾規定期限函 附件 6：機關聲復函 附件 7：通知各機關政府採購履約保證金逾規定期限函</p>
聯絡窗口	<p>姓名：稽察員林東慶 電話：02-29640947-154 Email：dclin@mail.audit.gov.tw</p>
單位評核意見	<p>本提案人員發揮審計人員創新核心價值，同時針對審計機關數據分析之「資料來源」及「技術方法」等 2 個面向提出改進方案，其中撰寫 Python「網路爬蟲」，可擴大審計機關蒐集資料來源，</p>

	<p>減輕機關提供資料負擔；以 Excel VBA 巨集指令編寫最小編輯距離法程式，可作為非結構化資料跨資料庫比對之橋樑與基礎，經運用於各項議題，均能達成預定之審計成果，足可推廣作為各審計單位查核之輔助工具，並已將該等程式提供機關使用，落實審計機關為被審核機關之協力夥伴關係，建立審計人員專業形象。綜上顯示提案人員善用創新思維及技術方法，符合「審計機關創新提案評審及獎勵要點」之「產出創新」及「流程創新」案件。</p>
--	---

附件 1 Python「網路爬蟲」程式碼及成果示範

註：以新北市重慶路112號為中心，搜尋半徑5公里內，關鍵字為「補習班」之商家，計有172家。