整合服務問卷調查系統之核心引擎設計

The Design of Core Engine for Questionnaire-Based Integrated Services Survey System

曾文憲[a] 張博論[a] 馬自恆[b]

Wen-Hsien Tseng[a] **Polun Chang[a]** **Tze-Heng Ma[b]**

[a]國立陽明大學衛生資訊與決策研究所

[a]Institute of Health Informatics and Decision Making, National Yang-Ming University, Taipei, Taiwan, ROC

[b]中央研究院資訊科學研究所

[b]Institute of Information Science Academia Sinica, *Taipei, Taiwan, ROC*

g39223007@ym.edu.tw **polun@ym.edu.tw** **mada@iis.sinica.edu.tw**

摘要

問卷調查為健康醫療領域常使用的資料蒐集方法，其實施上不僅耗費人力與金錢，也常因難以避免的人為輸入錯誤而造成困擾。本研究主要利用圖型辨識技術，來研究與設計出自動化問卷處理系統(QBISSS：Questionnaire-Based Integrated Services Survey System)的核心辨識引擎(core engine)，其將可迅速且有效地辨識出問卷結果，其並可彈性地搭配問卷輔助設計系統與統計分析圖報表軟體來對整個問卷調查流程做全方位的協助。在 QBISSS 的系統架構上，依問卷處理的流程順序主要可分為三個模組，分別是電腦輔助問卷設計系統（CAQDS：computer-assisted questionnaire design system）、問卷辨識核心引擎（QR engine：questionnaire recognition engine）和問卷資料的統計與分析工具（SA tool：statistic and analysis tool）。而在研究結果中可發現，在處理問卷的速度與正確率上，自動化問卷處理系統都明顯優於傳統人工處理問卷的方式，是故其將可有效提高研究調查品質，並加速問卷調查的流程。

關鍵字：自動化、圖型辨識、問卷調查、問卷辨識

壹、 前言

在各種的田野調查中，傳統的紙本問卷調查是其最常使用的工具，然而其卻有著許多的不便與缺點，當面對要輸入大量的問卷資料到電腦中時，人為錯誤往往容易發生，如此一來不但使得研究費時費力、成本提高，其可信度也因而降低[1][12]。

而問卷調查的成本至今仍是相當高，從設計問卷、派送問卷、填寫問卷、回收問卷和輸入問卷資料至資料庫中，到之後的資料變數設定和統計分析，其中每一個過程都是需要花費相當的成本、時間和人力。平均請人去做一份問卷調查工作，包含之後將問卷結果輸入資料庫的人力成本約需花一百元左右；若選擇將回收的問卷全部交由專業統計公司來處理(包含：資料輸入、建立資料庫、建立 CODE BOOK 和基本的統計分析與描述)，平均一份問卷就已經需要約一百四十元左右的花費[3][10][11][20]。

本研究主要利用圖型辨識技術，以設計出自動化問卷處理系統的核心辨識引擎，其將可協助問卷調查流程的自動化，進而提昇問卷調查品質、降低成本和減少人為錯誤的發生[13][14][15][16][17][21]。

貳、 研究方法

一、 開發工具：

- 硬體方面：使用 NEC VERSA SXI NOTEBOOK(CPU:800MHz，Memory：256Mb RAM)；MICROTEK 5600 SCANNER。
- 軟體方面：在 Windows 2000 Professional 作業系

統平台上，使用 Microsoft Visual Studio.NET 2003 軟體開發工具與 C#語言來開發系統。

二、 系統設計架構：

如圖 1 所示，在整個自動化問卷處理系統中主要可分為三個模組，其依問卷處理的流程順序分別為：

- <u>電腦輔助問卷設計系統</u> (CAQDS：computer-assisted questionnaires design system)：其主要功能在協助使用者對問卷的設計與編排，並也包含了完成問卷資料欄位的定義(QD：questionnaires definition)與問卷辨識定位的工作。
- <u>問卷辨識核心引擎</u> (QR engine：questionnaires recognition)：其為自動化問卷處理系統的核心引擎，主要功能在自動化地準確地辨識出問卷的資料結果。而其問卷辨識率將為本研究作為結果評估的重要指標。
- <u>問卷資料的統計與分析工具</u> (SA tool：statistic and analysis tool)：其主要功能為替使用者做基本的問卷資料統計分析與圖報表的產生。

而為了保有系統之彈性需求，各部份皆可選擇人工作業方式或由電腦來做自動化的輔助處理。

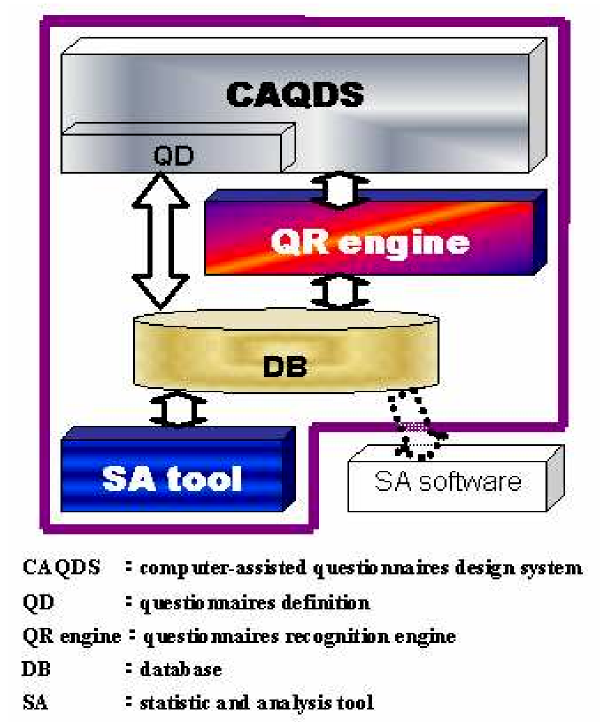


圖 1：自動化問卷處理系統架構圖

三、 問卷格式設計：

盡可能將所有形式題目改為勾選checkbox作答方式的設計，以簡化問卷的辨識困難度。而為了提高問卷辨識率與圖檔曲折校正之故，問卷背景內容須分別在左上、右上和左下等三個角落套用不同的特殊(漣漪式 45^o 角扇形)為底圖 (如圖 2 所示)，用以應付問卷掃瞄圖檔發生旋轉的問題時，可以對checkbox做三點定位的校正功能；當然若使用CAQD來輔助問卷設計的話，底圖將會自動套用而不必擔心。

圖 2：問卷背景底圖設計-(漣漪式 45^o 角扇形)

四、 掃瞄圖檔規格方面：

所有問卷將被掃瞄成 200dpi 黑白約 1700 X 2330 pixels 的 BMP 圖檔，平均每份問卷大小約為 500KB 左右。

五、 核心辨識引擎設計

- <u>定位方式</u>：問卷事先的定位方式為了彈性的需求，而不採用將定位點位置或offset內建固定於程式碼中(類似答案卡的方式)，而採用由使用者彈性利用CAQDS中的QD來自行替任一問卷原稿做事先定位工作(含checkbox和定位圖型之定位點)。
- <u>辨識校正方式</u>：問卷辨識首先的重點是要能夠先找出所有定位圖型的定位點(左上、右上和左下，共三個)，來替所有checkbox做三點定位校正。定位圖型分別位於問卷的三個空曠邊緣角落，其較身為問卷作答區主要內容之checkbox不易受問卷背景或因作答所造成的干擾而影響了辨識率，且其辨識特徵也較明顯而易取。而為了應付各種掃瞄問卷圖檔時可能發生的問題(偏移、扭曲、旋轉、部份遮蔽、模糊和大小改變)，故在校正辨識的定位點上採用offset與校正矩陣的綜合方式來做校正，而其中因校正矩陣之成功校正的機率較高，故相對給予較高的影響權重值。
- <u>offset</u>：其是指各定位點間的偏移量，利用offset可快速地替各定位點做定位校正，而由於其取值的距離短，變動風險相對減少。但較不易應付問

卷圖檔旋轉和大小改變的問題。

- <u>校正矩陣</u>：在本研究中採用的是affine mapping functions-2D affine transformation的方法。在找出各個問卷圖檔之所有定位圖型的定位點後，以新舊定位圖型的定位點共六點為係數，產生出校正矩陣，以對所有舊checkbox的定位點做校正。利用校正矩陣，可成功避免問卷圖檔偏移、旋轉和大小改變的問題。說明範例如下：

以原資料庫儲存的定位圖形之舊定位點：$A(x_1,y_1)$，$B(x_2,y_2)$，$C(x_3,y_3)$，與系統辨識尋得的定位圖形之新定位點$A'(u_1,v_1)$，$B'(u_2,v_2)$，$C'(u_3,v_3)$為係數套用公式(1)，以求出反矩陣如公式(2)所示，再將所得反矩陣其中之元素重新排列，成為校正矩陣如公式(3)所示，最後再以原資料庫儲存的checkbox之舊定位點：$D(m,n)$為係數，利用校正矩陣，套入公式(4)便可求出checkbox之新定位點$D'(p,q)$。

$$\begin{bmatrix} \Sigma x_j^2 & \Sigma x_j y_j & \Sigma x_j & 0 & 0 & 0 \\ \Sigma x_j y_j & \Sigma y_j^2 & \Sigma y_j & 0 & 0 & 0 \\ \Sigma x_j & \Sigma y_j & \Sigma 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Sigma x_j^2 & \Sigma x_j y_j & \Sigma x_j \\ 0 & 0 & 0 & \Sigma x_j y_j & \Sigma y_j^2 & \Sigma y_j \\ 0 & 0 & 0 & \Sigma x_j & \Sigma y_j & \Sigma 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \begin{bmatrix} \Sigma u_j x_j \\ \Sigma u_j y_j \\ \Sigma u_j \\ \Sigma v_j x_j \\ \Sigma v_j y_j \\ \Sigma v_j \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} \quad (2) \qquad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} p \\ q \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m \\ n \\ 1 \end{bmatrix} \quad (4)$$

參、 結果

一、 系統雛型測試結果操作：

- <u>CAQDS階段</u>：使用者可依系統的問卷格式要求規定，自行編制與設計問卷之後，將問卷紙本原稿的掃瞄圖檔利用系統的問卷資料定義功能(QD：questionnaires definition) (如圖 3 所示)，先用mouse點選出checkbox的位置，系統會自動找出各個checkbox的中心點並存入資料庫(XML格式，如圖 4 所示)，完成問卷事先定位的工作，然後替問卷的資料變數做定義；或直接利用CAQDS來做問卷的設計與資料定義的工作[19][22][25][26]。

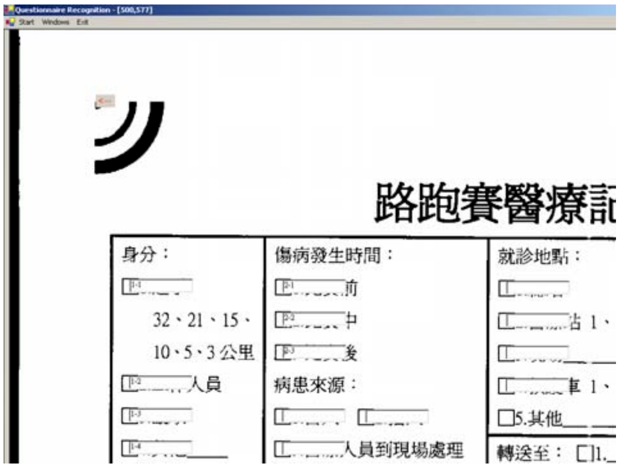
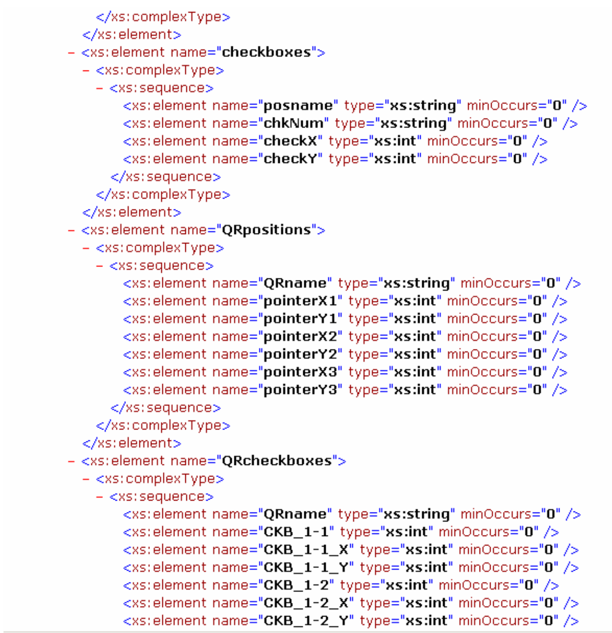


圖 3：系統問卷資料定義畫面



```
</xs:complexType>
</xs:element>
- <xs:element name="checkboxes">
- <xs:complexType>
- <xs:sequence>
<xs:element name="posname" type="xs:string" minOccurs="0" />
<xs:element name="chkNum" type="xs:string" minOccurs="0" />
<xs:element name="checkX" type="xs:int" minOccurs="0" />
<xs:element name="checkY" type="xs:int" minOccurs="0" />
</xs:sequence>
</xs:complexType>
</xs:element>
- <xs:element name="QRpositions">
- <xs:complexType>
- <xs:sequence>
<xs:element name="QRname" type="xs:string" minOccurs="0" />
<xs:element name="pointerX1" type="xs:int" minOccurs="0" />
<xs:element name="pointerY1" type="xs:int" minOccurs="0" />
<xs:element name="pointerX2" type="xs:int" minOccurs="0" />
<xs:element name="pointerY2" type="xs:int" minOccurs="0" />
<xs:element name="pointerX3" type="xs:int" minOccurs="0" />
<xs:element name="pointerY3" type="xs:int" minOccurs="0" />
</xs:sequence>
</xs:complexType>
</xs:element>
- <xs:element name="QRcheckboxes">
- <xs:complexType>
- <xs:sequence>
<xs:element name="QRname" type="xs:string" minOccurs="0" />
<xs:element name="CKB_1-1" type="xs:int" minOccurs="0" />
<xs:element name="CKB_1-1_X" type="xs:int" minOccurs="0" />
<xs:element name="CKB_1-1_Y" type="xs:int" minOccurs="0" />
<xs:element name="CKB_1-2" type="xs:int" minOccurs="0" />
<xs:element name="CKB_1-2_X" type="xs:int" minOccurs="0" />
<xs:element name="CKB_1-2_Y" type="xs:int" minOccurs="0" />
```


圖 4：XML 格式資料庫檔

- <u>QR engine階段</u>：而在問卷透過各種管道回收後，接下來便是問卷調查結果資料輸入的工作了，若是電子格式的問卷資料，系統當然可自動處理，若是傳統紙本格式的問卷資料，則可透過附有自動送紙器的快速掃瞄器來將問卷轉為圖檔，再利用系統中的QR engine功能來做問卷作答結果的辨識與資料庫資料輸入的工作，系統問

卷辨識的結果檢視如圖 5 所示，系統會先自動找出問卷左上、右上和左下的輔助定位點，來修正自資料庫讀出的checkbox定位點資料，而checkbox定位點間也會彼此做修正，然後再將有作答的checkbox標示在畫面上，以供使用者檢查，並將結果存入資料庫(XML格式)中[18][23][24]。

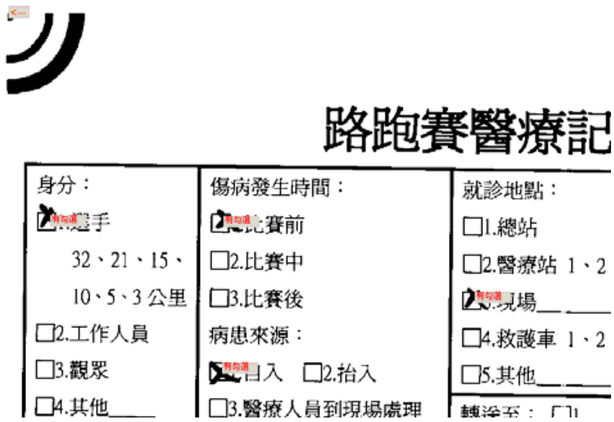


圖 5：系統問卷辨識結果檢視畫面

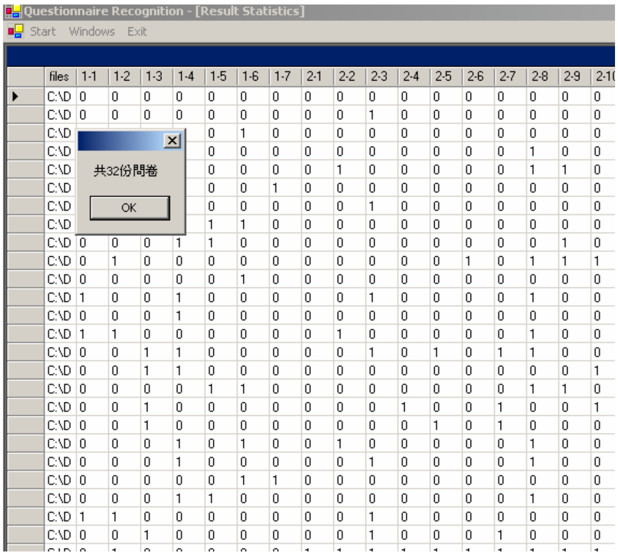


圖 6：系統之問卷統計結果畫面

- SA tool階段：系統將自動產生出問卷調結果的統計資料(如圖 6 所示，包含各欄位加總值和圖檔路徑名稱)。最後使用者可選擇使用系統內建基本的統計分析功能來做基本的統計分析報表和統計圖表的輸出，或將統計量結果輸出成純文字檔(如圖 7 所示)，而匯入到習慣使用或高階的統計軟體(如：SPSS)中做進一步的統計分析與圖報表的產生[2][4][5][6][7][8][9]。

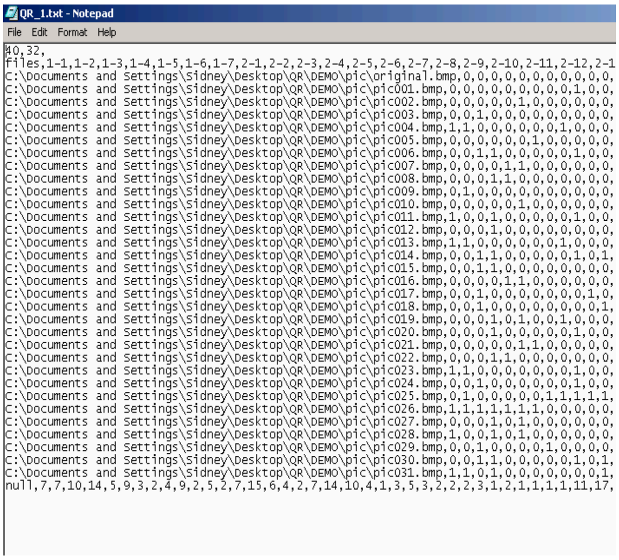


圖 7：系統之問卷統計結果-文字檔格式輸出畫面

二、 結果評估

在 CPU:800MHz，Memory：256Mb RAM 的 NOTEBOOK 上，分別共用了二份不同內容與格式的問卷做測試。

- 測試問卷一之結果：31 份各 40 題選項的紙本問卷之辨識測試結果，平均一份問卷需花 0.3 秒的辨識時間，而其圖檔的掃瞄時間約需 15 秒，問卷之辨識準確度在目前自行設計的問卷測試中約為 99.92%。
- 測試問卷二之結果：30 份各 113 題選項的紙本問卷之辨識測試結果，平均一份問卷需花 0.43 秒的辨識時間，而其圖檔的掃瞄時間約需 15 秒，問卷之辨識準確度在目前自行設計的問卷測試中約為 98.82%；。

問卷之辨識失敗主要原因是在於問卷掃瞄圖檔的旋轉、扭曲或模糊太嚴重所造成，此應歸咎於所使用之掃瞄器與自動送紙器因使用多年過於老舊，而導致送紙不正與部份區域掃瞄模糊的情況發生。

平均而言，同樣的問卷若以人工手動輸入的話(40 題題目/問卷)，一份問卷約需花 20 秒左右，且人工處理費用又較高，而人的精神狀態隨時間變化會疲憊而易造成輸入錯誤；相較之下，若能以較穩定不用休息的機器和電腦來協助研究學者處理問卷的話，必能事半功倍的，同時提昇研究的品質與量。

紙本問卷辨識的速度跟電腦硬體等級有密切的關係，CPU 的時脈高低將直接影響到辨識的速度，當換

用高一倍速度的 CPU 時，其辨識的速度也將隨之提昇一倍；而記憶體的大小則跟圖檔的大小與數量有關，當圖檔大小過大且量過多時，若其總需空間大小超過記憶體容量時，並會因需將資料暫存到硬碟的虛擬記憶上，而明顯降低辨識的速度。但基本上一份 200dpi 的黑白 BMP 問卷圖檔，每份大小才約為 500KB 左右而已，記憶體使用上並不致於不足。

肆、 討論

而在資訊科技的輔佐之下，問卷調查未來勢必將可因而跳脫舊制人工問卷調查的範圍限制，如圖 8 所示，單一問卷可被轉為各種不同媒體型態的傳輸格式（text：paper、web、e-mail；voice）來派送到不同的環境的訪問對像，並可以透過整合各種不同型式的傳播工具與設備（PC、tablet computer、notebook、PDA、mobile phone、telephone），來做全方位多面向性質的問卷調查研究。

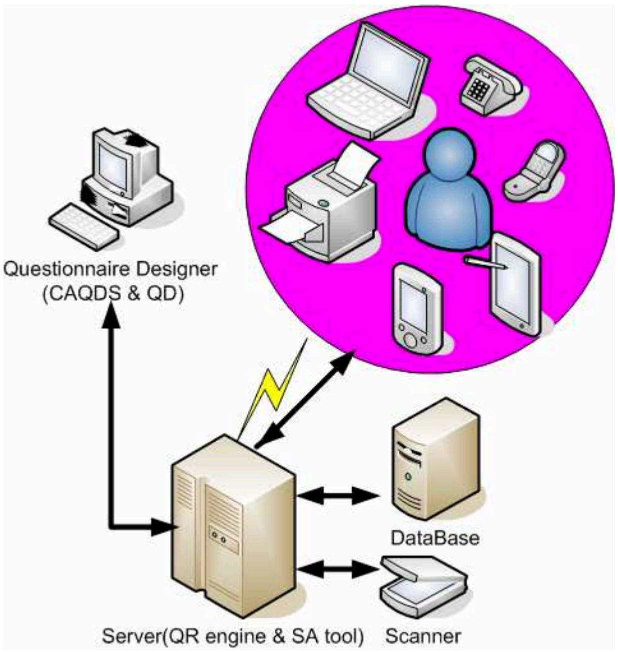


圖 8：系統架構與關聯圖

在自動化紙本問卷處理中，最耗時的是掃瞄的時間，不過若能如圖 8 所示，在 server 處由中央做統一處理的系統規劃下，各個使用者在透過網路傳送問卷原稿的設計和資料定義給中央 server 後，便可由中央所提供高效能的計算伺服器、快速列表機、自動送紙器和掃瞄器來完成工作，此法不但可集中資源的運用，也使得使用者能夠只付出低廉的成本，就可共同來租用高等級的設備和運算能力，以提昇問卷處理速度，更可降低因使用不同規格設備，而造成的問卷辨識準確率下降等的問題發生。

伍、 結論

由本研究結果可以發現，自動化問卷辨識在成本、時間與資料輸入正確率上，都明顯優於傳統人工紙本問卷調查作業的方式，其擁有以下幾點的好處：

- <u>提昇問卷資料輸入的效率</u>：利用圖型辨識之相關技術來完成問卷的快速辨識，以有效改善傳統人工閱卷模式下速度緩慢的問題。
- <u>減少人為錯誤的發生</u>：人工閱卷的模式時常會發生眼花或按錯鍵等人為錯誤，其錯誤發生機率不但高的驚人，也較不易被發覺，許多錯誤的資料往往無形間被研究學者引用成理論的解釋依據，其問題發生點因可歸類於人工閱卷的變動影響因素太大，隨著人的心情、生理和精神狀態的好壞變化，樣樣都會影響到問卷資料輸入的正確性；是故以前許多無聊的運輸帶式傳統手工作業，現在也同樣都改為機械自動化來處理了，機械不容易可以像人類般做出精緻的手工品，但卻較人類適合做簡單而重覆的工作。
- <u>降低成本與人力的浪費</u>：在整個問卷調查的流程中可以發現，其主要的成本花費都是在人力的費用支出，因為傳統的問卷調查及處理是大量依靠人力的工作，在問卷處理的自動化後，人力需求減輕了，當然成本也因而下降了。
- <u>提昇研究的品質與可信度</u>：一個好的學術研究，問卷調查的優劣往往直接或間接地影響其研究的品質與可信度，當問卷調查成本降低後，自然有多餘的經費來增加樣本量，再加上資料輸入的正確率提高，使得樣本資料對母體解釋和描述的可信度也因而大大的提昇。
- <u>提昇問卷調查流程的流暢度與方便性</u>：問卷處理多元化與自動化，促進整個問卷調查流程的流暢度，流程中各個節點也因而可以有更緊密的關聯性，其方便度也無形的增加不少。
- <u>提供單一問卷多元化的傳播媒體型態</u>：為了要讓問卷調查更容易以不同的角度深入訪談到各個不同的訪談對像，不同型態的問卷媒體傳輸格式是必要的，而若能將同一份問卷自動轉為各種不同媒體型態的格式，並自動化回收整合在同一份

問卷規格的資料庫中，其勢必會替研究學者帶來更豐富的研究資源，也許也會因而幫助與促成不少研究上的新突破與發現。

致謝: 國立陽明大學醫學系范佩貞教授提供相關資料

參考文獻

[1] 文崇一，問卷設計， 收錄於楊國樞、文崇一、吳聰賢與李亦園編（1988），社會及行為科學研究法，東華書局，頁 405-438。

[2] 吳明隆 (2000) 編著，SPSS 統計應用實務 (二版) ，松崗出版社。

[3] 呂以榮譯(2002)，A.N. Oppenheim 原著，問卷設計訪談及態度測量，六合出版社。

[4] 吳明隆（2003），SPSS 統計應用學習實務－問卷分析與應用統計，知城數位科技股份有限公司。

[5] 林清山 (1970)，多變項分析統計法，台北，東華書局。

[6] 林惠玲、陳正倉 (2004)，應用統計學，台北，雙葉書廊有限公司，頁 18-19。

[7] 邱皓政（2003），量化研究與統計分析，SPSS 中文視窗版資料分析範例與解析，五南圖書出版公司。

[8] 張紹勳、張紹評、林秀娟 (2002)，SPSS for Windows 統計分析—初等統計與高等統計(下冊)，松崗電腦。

[9] 陳順宇(2004)，多變量分析，華泰書局。

[10] 陳德禹(1992)，論文寫作研究-問卷設計的探討，增訂新版，三民書局，頁 214。

[11] 謝邦昌（2002），問卷設計，資商訊息股份有限公司。

[12] Anderson, J. F. (1990), Questionnaire design and use revisited: Recent developments and issues in survey research. (ERIC NO. ED271501).

[13] Baecker RM, et al. (editors) (1995), Readings in human-computer interaction: toward the year 2000, 2nd ed., San Francisco, CA: Morgan Kaufmann Publishers, Inc.

[14] Bates DW. (2000), Using information technology to reduce rates of medication errors in hospitals. [Comment]., BMJ 320(7237): 788-91.

[15] Birk-Jenson, Natalie (1986), Problems with questionnaire design in citizen preference surveys, University of Nevada, Reno.

[16] Block G, Hartman AM, Dresser CM, Carroll MD, Cannon J and Gardner L (1986)， A data-based approach to diet questionnaire design and testing. Am J Epid 124: 453-469.

[17] Bradburn, N. M. and Sudman, S. (1979), Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.

[18] G. Carpenter and S. Grossberg (1986), “A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine,” Computer Vision, Graphics, and Image Processing, Vol. 37, pp. 54-115.

[19] Gillham, B., (2000), Developing a Questionnaire, (pp. 49-84), London, Wellington House.

[20] Lu Ann Aday. ”Designing and conducting health surveys. A comprehensive Guide.” Second Edition. Lossey-Bass Publishers. San Francisco, USA.

[21] Mary Carmen Cupito (1998)：Wireless LAN：Emerging to maturing technology. Health Management Technology; 19(3); 15.

[22] Oppenheim, A. N. (1992). Questionnaire design, interviewing and attitude measurement. New York: St. Martins’s Press.

[23] P.A. Devijver and J. Kittler (1982), Pattern Recognition: A Statistical Approach.

[24] R.O. Duda, P.E. Hart, and D.G. Stork (2001), Pattern Classification, John Wiley.

[25] Sudman, Seymour; Bradburn, Norman M (1982), Asking Questions-A Practical Guide to Questionnaire Design; 1st ed. San Francisco, Jossey-Bass Publishers.

[26] Oppenheim, A. N., (1996), Questionnaire Design, Interviewing and Attitude Measurement, (pp. 112-115), London: Wellington House.