

【文／編譯發展中心助理研究員 邵婉卿】

在資訊科技發達的今天，到處暴露自己的個人資料，是一件極為普遍的事，例如寫寫部落格、上上臉書或編輯一下維基百科等。據估計，2013 年全球數位資料量大約是 1,200EB，非數位資料量僅占數位資料量的 2%，資訊儲存量是經濟成長量的 4 倍，針對這麼龐大的數位資料量，牛津大學網路研究所 Mayer-Schonberger 教授與巨量資料思潮評論員 Cukier 在其合著的 *Big Data: A Revolution That Will Transform How We Live, Work, and Think* 一書中，提出「巨量資料」，即大數據（Big Data）的概念。「巨量資料分析」是一門新興科技，用來分析極大量、極凌亂的資料，作為預測潛在客戶的購物行為、全國流感是否即將爆發等資訊。

巨量資料強調樣本幾近於母體，數量比品質重要，只要從數以億筆的資料中找出相關性，就能抓住各種機會，巨量資料的重點不在於知道為何如此，只在乎知道正是如此就行，巨量資料是人類利用量化數據來理解世界的方法之一，教育界如果能夠善用，也能知道教育現況最該關注的議題是什麼。

身為智庫，本院不斷創造教育知識與內容，對教育的理念與實踐提供優質的資訊服務，編譯中心亦致力於出版專精的、權威的圖書及工具用書，在兼顧教育實務與學術價值的前提下，我們該如何面對巨量資料的發展洪流？

Amazon 的做法是，不斷蒐集購書者的紀錄並分析電子書使用者的線上使用習慣，用來知道以前作者們所不知道的事：喜好某類主題的讀者群在哪裡？大多數讀者喜歡書中的哪一段？讀者不喜歡超過幾頁內容的書？以及閱讀時的停頓與重點在哪一章？所有利用網路傳輸的線上互動教育課程所產生的使用者歷程資料，也都可以據以分析，用來改善教育資料的編輯與教學的成效。

而 Google 拼字檢查系統的做法是，藉著從每天 30 億筆搜尋字串中，找出錯的、不正確的和有問題的資料，轉為聰明的回應機制，利用「您是不是要找…」的方式，貼心地幫使用者快速找到正確的用字，成為世界上最好用的和最新型的拼字檢查功能。

傳統圖書和工具書的做法則是一種勞力密集、單調又乏味的苦工，以簡明、準確、典雅、淵博的定義聞名的《牛津英語辭典》（Oxford English Dictionary, OED）為例，歷時超過 70 年才完成首版，如此浩繁費工的 OED，是靠著收集大量英語出版物及其他記載中的用詞逐一過濾而來的，OED 強調唯有使用例證，才能充分探討一個詞在過去的發展脈絡，OED 至今仍是英語世界的每個角落中，如議會、法院、學校和研究社群中的必然引用資源。

綜上，巨量資料的確有小量資料無法達到的功能，但巨量文本也無法做到的事：因為還有許多沒有發表出來的意見、沒有被寫出來的想法，和還沒有想到的概念，這些都是巨量資料無法蒐集和處理的，巨量資料只可以用來預測，指出追尋理解的方向，目的是在通知，不在解釋，也不是最終的正確答案。而本院的各項研究成果、出版專書和工具書編輯，正如 OED 一般，是可供校正和採

信的資料來源，因為我們所產出的是可供理解的資料，目的在定義、解釋和提供權威的正確因果分析，這些都是被用來分析巨量資料的基石。

【參考文獻】

林俊宏（譯）（2013）。大數據。臺北市：天下文化。

楊傳緯（譯）（2009）。教授與瘋子。上海：上海人民。