

[論文名稱] 漢籍電子文獻資料庫建置的回顧與展望

[作者]: 劉錚雲 中央研究院 歷史語言研究所

[摘要] 「漢籍電子文獻資料庫」(以下簡稱「漢籍」)的建置肇始於1984年,目標是收錄對中國傳統人文研究具有重要價值的文獻,並建立全文電子資料庫,以作為學術研究的輔助工具。資料庫內容包括經、史、子、集四部,其中以史部為主,經、子、集部為輔。若以類別相屬,又可略分為宗教文獻、醫藥文獻、文學與文集、政書、類書與史料彙編等,內容幾乎涵括了所有重要的典籍。經過近三十多年的建置與發展,「漢籍」目前收錄了歷代典籍934種,525,960,000萬字;如果經費額度維持不變,「漢籍」可以每年新增二千萬字的規模繼續成長。

本文主要說明「漢籍」這三十年來的發展與未來的計畫。「漢籍」原為院裡謝清俊教授主持的計算中心於1984年開始建置,2008年委外建立新系統,以後逐年擴充新功能,服務全球學者。作為一個學術機構,本院建置「漢籍」,在態度與方法上自然與坊間的商業公司不同。「漢籍」不僅在校對輸入文字上儘量做到正確無誤,而且對古籍上的訛誤,無論是出於當年的出版者,或是由於今日的出版商所造成的,也儘可能的補正。為因應數位化學術潮流發展,漢籍不能僅止於檢索功能的提供,必須同時加強其分析功能。有鑑於此,「漢籍」今年除了新增了一項內文自動比對的功能,同時也著手建立標誌管理系統,嘗試在若干清代典籍上標誌人名、地名與職官,為全文資料庫建立結構,俾研究者得以就檢得的資料進行統計分析。