

中華書局の古籍デジタル化と『中華經典古籍庫』

李晨光

中華書局が開発した大型データベース製品『中華經典古籍庫』は2014年6月に正式に発表された。

『中華經典古籍庫』は中華書局版点校本古籍の初めてのデジタル化であり、「二十四史」、『清史稿』、『資治通鑑』、「新編諸子集成」、「清人十三經注疏」、「史料筆記叢刊」、「學術筆記叢刊」などの古典的シリーズを含めた代表的整理本が網羅され、第一期の図書目録は約300種類、2億字にのぼる。

『中華經典古籍庫』は豊富な機能を持っており、特に検索機能に特色があり、書名検索と全文検索がある。点校本古籍の整理の成果を保つことにより、この製品は本文、校注、書名、固有名詞、見出しなどの小範囲内での全文検索を実現しており、異なる研究の需要を満足させることができる。システムはいかなる検索ワードに対しても簡体字と繁体字の関連づけを自動的におこない、しかも合計で47,000以上の異体字字典の内容をも収録しており、これにより異体字の関連づけも実現している。人物の情報をできるだけ全面的に検索するため、製品は繁、簡、異体字の関連検索の基礎の上に、人名別称並列検索を開発し、現在すでに約15万もの人名別称を網羅している。例えば、人名の「曹操」を検索する時、曹操の31ある別称のうち、「孟徳」、「魏公」などを選択することができ、システムは同時にマッチングするすべての別称の結果を示すことができる。ユーザーの閲読と研究をサポートするため、この製品はまたオンライン字典と紀年換算表の2つのツールを加えている。オンライン字典には現在『中華大字典』全ての内容が収録されている。『中華大字典』は中華書局が1915年に整理出版した繁体字典で、4.8万字余りを収録し、すべての字は本義、派生義、付会義の順序で解釈をおこない、注音は『集韻』に準拠しており、特に古文の閲読に適している。オンライン字典も繁、簡、異体字の関連をサポートしており、たとえば「歴」を検索すると、同時に「曆」などの関連字の結果が示され、間違えやすい字義の見分け方に使用できる。紀年の換算は西暦紀元、干支、帝王、年号など重層的紀年検索方法が使える、紀年情報は『中国歴史紀年表』をもとにして、複数の紀年表を参考にして、何度も修正をおこなっており、内容は完備されている。その他に、この製品もウインドウの拡大、縮小が可能で、閲読履歴、検索履歴の保存およびしおり機能などの基本メニューを備え、ユーザーはテキストの内容をいつでもコピーして引用することができる。

『中華經典古籍庫』は、開発の過程で古籍のデジタル化における普遍的な問題の大部分を解決した。まずデータ整理には統一規格が存在しないということ。整理の成果を保存するために、中華書局はそれぞれの書籍について作業をおこなう際に、例えば本文、注釈文、見出しなど異なる内容に対しては異なる書式で注記している。またすべての書籍の体裁は一樣で

はなく、大部分は具体的な分析を必要としており、その特定の構造によってデータベースにどのように見出されるかを研究し、見出しのレベルを調整したり、必要な見出し情報を追加したりした。たとえば、正史の人物伝記は往々にして巻頭に小文字の体裁で多数の人名を列挙しているが、中華書局は整理の過程でこれに対応する人物伝の前に並び替え、検索に便利だけでなく、データの階層構造も一目瞭然となった。不断の修正と総括の結果、中華書局は古籍データ整理の上で比較的通用する標準規格を制定して、今後のデータ整理のための参考に資することができる。次に**古籍デジタル化の使用文字の問題**について。古籍の中の多くの漢字はすでにある文字コード体系が表示できる範囲を大きく超えており、よく見られる方法は単独の漢字を画像形式で行間に挿入するものであるが、しかしこれは検索の分断化を招き、しかも美観に影響する。中華書局は前期プロジェクトの蓄積を基礎として、漢字規格と国際コード規格に厳格に準拠したフォントの整理をおこない、これによって作字の処理と相似するフォントの判断を研究した。一期データの作字量は 9000 余りにのぼり、すべて標準性と検索性をそなえ、後期のデータ加工を大きくサポートすることとなった。次に、**ソフトウェア機能の調整**について。あらゆるプログラムの開発には絶え間ないテストと修正の繰り返しが必要であり、また古籍庫はその内容の特殊性のためにテストの難易度が高く、中華書局はプログラムの設計と機能の上で改善を重ね、絶えず更新をおこない、**特にプログラムの中での作字、固有名詞に引く傍線などの表示の問題を解決し、検索機能の最適化を完成した**。洪濤古籍資源部主任はプログラム開発の主旨を説明してこう語る。「私達は製品の目線で開発を進めるように絶えず自覚し、ユーザー体験を念頭に置いて、まず自分を満足させることができこそ、広範なユーザーに気に入ってもらうことができる」。

現在、中国古籍データベース市場の製品は雨後の筍のごとく現れ、膨大なデータ量を誇る総合データベース、それぞれの特徴を持った専門データベースなどがあり、それぞれ各高等教育機関、図書館などの機関で良好なシェアを確保している。『中華經典古籍庫』は他の古籍データベースと比べて以下のような特徴と利点を持っている。「**中華書局版点校本古籍はこれまでに学会から広く承認されており、これらの古籍文献を現代の情報技術で処理することにより、さらなる普及と利用の便宜を図ることができる**。古籍整理の専門出版社として、私達のデータのバージョンの出所は明確で、データの加工もよりスタンダードで、精密であり、データの正確性も高く、信頼に足る。原文の版式とデータテキストの対照機能はテキストの信頼性を増すだけでなく、比較研究にも便利である」。中華書局点校本古籍は文学と史学界の専門家の学術成果をまとめ、特に「二十四史」と『清史稿』は 20 年の歳月をかけて完成した古籍整理プロジェクトであり、整理に参加した現代の史学者たちはすべて当代の碩学であり、新中国における史学界の各時代史研究の最高レベルを代表している。点校本「二十四史」と『清史稿』が出版されると、国内外の学術界で最も権威ある、最も通用するバージョンとなり、「国史」の基準本との評価を得た。「**前期のデータ収集作業それ自体が不断の最適化の過程であり、後期の人的作業は二次編集の意味合いが強く、整理の完成後にはさらにデータ変換ツールを使ってエラーの排除をおこなった**。こうした三段構えの措置により、データの品質を紙本と同レベルに維持し、さらには紙本を超えることができた。」

現在の成果はすべて前期の蓄積の上に成し遂げられている。すでに 100 周年を迎えた中華書局はデジタル出版に関する探求をとどめることはなかった。中華書局は 2003 年に古籍資源部を開設し、「中華古籍コーパス」プロジェクトを立ち上げた。当時はまだ全く「デジタル出版」という概念はなく、古籍資源部は主に古籍のデジタル化編集作業に従事し、6 年という時間をかけて 3 億字の整理本古籍のデジタル化処理をおこない、プラットフォーム構築と資料蓄積作業を完成した。この過程で特殊な難題に遭遇した。それはコンピュータ用字の大量欠如である。しかも文字コード以外の漢字処理には参照できる基準がなかった。古籍の高品質なデジタル化のために、中華書局はコンピュータ用字の規格化プロジェクトの建設に大きく力を入れ、例えば「中華字庫」、「国家デジタル図書館漢字規格処理プロジェクト」、「新聞出版用大字符集」などのプロジェクトに参加した。「これらのプロジェクト建設を通して、私達はコンピュータの古籍用字を規格化し、大量の漢字属性データを蓄積して、古籍のデジタル化作業の基礎作りをサポートした。」これらの基礎によって、『中華經典古籍庫』の作字作業は順調におこなわれたのである。

『中華經典古籍庫』はデータ更新を続け、計画では毎年ひとつのデータパックの更新をおこない、新しく出版される優秀な整理本古籍を持続的に収録し、品質保証の前提のもとに順番にデータ量の拡充をおこない、同時にデータの改訂と完備をおこなっていく。それ以外に、人名異称関連表とオンライン字典も持続して拡充し、ユーザーに更に豊富なサービスを提供していく。中華書局も製品ラインを絶えず拡充し、より多くの専門データベースや小型データベースの開発を通じて異なるユーザーの需要を満足させる。商周銅器銘文知識データベースは前期リソース整理の基礎にたち製品化を実現していく。中華書局の 2 つ目の大型データベース製品『中華基本史籍知識庫』もすでに起動しており、この製品は古籍庫の基礎のうえに、学術著作と辞典などの書籍を収録し、「史籍分析システム」プロジェクトの建設経験を参考にして、人物、時間、場所などの史籍知識元間の関連づけをおこない、その語義関係を可視的に示して、研究者により専門的知識サービスを提供する。「私達の最終目標は整理本古籍のデジタル化を系統的に完成して、体系的な知識ネットワークを構築し、知識の間にリンクを設け、専門家と読者の間の知識と情報の壁を打ち破り、本当に価値のある古籍デジタル化製品を社会に提供することである」。中華書局は一步一步着実に歩みを進め、盲目的な追従はおこなわず、最良のデジタルサービスの提供に心を尽くしていく。

李晨光 (Li Chenguang)

1988 年、東北師範大学歴史系古典文献学専業卒業、中華書局に入社し『文史』の編集を担当。後に漢学編集室主任、歴史編集室主任を歴任。『西域水道記』などの古籍整理図書を編集し、『近三百年人物年譜知見録』『伊朗瑣羅亞斯德教村落 (イランのゾロアスター教村落)』などの編集責任者を務める。漢学編集室の任にあたっていた時に「日本中国学文萃」叢書を編集出版し、国家重点出版工程「大中華文庫」の一部の書籍の編集出版の仕事に携わる。現在は中華書局デジタル出版センター主任。デジタル出版専門プロジェクトの管理を担当している。