

中国書データベースの現状と課題

山田 敬三 (神戸大学)

YAMADA Keizo

文字セットとOS

現在、すでに相当数の大学や研究機関で、光ファイバー・ケーブルによる学内LAN (local area network) が形成されている。手元のパソコンを使って汎用コンピュータにアクセスしたり、ワークステーション等をサーバーとして、特定の情報を大量かつ高速にやりとりすることも可能になってきた。

中国書データベースに関しては、現時点ではまだ文部省学術情報センターのIR (一般のパソコン等で手軽に検索できるシステム) にのせることはできないが、その可能性の実現にむけて、大きく事態が動いている。それは、ISO (国際標準化機構) の多言語文字セットとしてISO-10646が確定し、その中に20902個の漢字がコード付けされたことによるものである。

この文字セットには、JIS (日本工業規格) で定められた12156個の漢字 (補助セットの5801字を含む) とともに、GB (中国国家標準) コードの基本である6763個の簡体字および繁体字、韓国の文字セットであるハンジャ、アメリカで使用されているユニコード、台湾で通用している文字セット等がほとんど網羅されている。字種としては、中国の『辞源』に収録されている約12000個の範囲に近い漢字の使用が可能になり、中国書のデータベース構築に必要な文字の大部分が、国際的に共通の基準で使用できることになる。

問題は、この文字セットを電算機上で動かすことのできるOS (基本ソフト) の開発だが、これについても、アメリカの企業を中心に一定のめどがついてきた。汎用機上での基本ソフトはまだ発表されていないものの、ウィンドウズNT版に対応するソフトのベータ版が発表された。現在のところはまだ、それに見合う文字フォントが作成されておらず、漢字はJISの範囲でしか表示されないが、これはいずれ近日中に解決されるはずである。

チャイナ・マーク

中国書に関して最も多くの蔵書を有している中国国家図書館 (北京図書館) では、1989年に日本電気 (NEC) の汎用コンピュータを導入して本格的な中国書データベースの構築に入ったが、他方では、ワークステーションを使ったLANの形成にも力を注いでいる。これまで、パソコンによって作成してきたデータベースをワークステーションに取り込む実験を行い、そのためのプログラムも今年に入って完成した。

最近の電算機事情から推定すれば、今後、中国書のデータベースは、汎用コンピュータよりは、むしろ、ワークステーションで処理する可能性がたかまっている。ISO-10646の文字セットに対応する大型計算機用のOSが、開発されるかどうかについても疑問符がついており、北京図

書館の方式は、充分考慮に値するものだと考えられる。

和書のデータベースが日本の国会図書館を中心に構築され、洋書がアメリカの議会図書館を中心としているように、中国書のデータベースに関しては、当然、北京図書館が中心になる。そこで作成されるデータ（China MARC）との互換性をも考慮にいたしたシステムを受け皿として準備する必要もある。しかも、それは日本の研究機関が現在もっているハード環境に適合したものであると同時に、国際的にも対応可能なものでなければならない。

国際規格の制定

以上のような事情を考慮して、日本中国学会に設置された中国書データベース化委員会では、国際交流基金の協力を得て、昨年六月から半年間、北京図書館自動化発展部の朱岩氏を日本に招聘、前半三か月を神戸大学、後半を東京大学でそれぞれ共同研究を実施した。その結果、これまで作成していたフォーマット（案）にも修正を加えた。

中国書のデータベース化については、漢籍目録の編集や図書館カードの作成を目的とした目録規則をそのまま使用することはできず、あくまでもコンピュータ入力を前提とした独自の中国書目録が必要となる。これについても、中国書データベース化委員会は時間をかけて検討を重ねてきたが、当面は中国の全国文献工作標準化技術委員会の作成した『著録規則』を参照することにし、その翻訳に取り組んできた。

この『著録規則』の翻訳作業はいったん完了し、その出版準備を進めている途中に、それをさらに改善した改訂版が発行されたため、現在その改訂作業を進めている。これについては、できるだけ早い機会に公開したいと考えている。

このほか、文字セットについても、ISO-10646をさらに拡大したいという要求が、中国のみならず韓国やヴェトナムからも提起され、そうした状況をふまえて、本年3月に国際会議がハノイで開催された。これには委員会の松岡委員が出席して討論に参加した。ただし、こうした拡大要求にあわせて文字セットがめまぐるしく変化するようであれば、中国書の本格的なデータベース化にも多大の支障の生ずる恐れがある。また、それに対応するソフトの開発が遅れ、実際には入力には困難になる。

幸い、冒頭にもふれたように、待ち望んでいたISO-10646を処理できる基本ソフトがウィンドウズNT版のベータ版として発表され、目下、関係者の間で検討されている。近く予想されるその本格的な完成をまって、本委員会としては、中国書データベースの構築に関する当面の方針を決定したいと考えている。NT版を開発したマイクロソフト社だけではなく、その他の企業からもこの種の基本ソフトが発表される可能性もある。

本委員会は、中国書データベース化のためのフォーマットと目録規則の制定、ならびに文字セットへの対応を中心に、すでに5年間にわたって検討を続け、その間、実験的に3万6千件のデータベースを作成してLANにも組み込んだが、むろん、それは究極の目標との間に距離のあるものであり、そのまま学情センターのIRにのせられるものでもない。ただし、ISO-10646をコンピュータ上で処理できるOSが出現したことにより、中国書データベース実現の可能性はきわめて近くなった。

しかし、このデータベースは国際規格にのっとり、世界中の漢字文献に関する情報を、相互に

インターネットを通じて交換できるようなものでなければならない。そのための規格作りには、当然、事前の国際的な合意を必要とする。今回のシンポジウムは、それに向けての橋頭堡になるものと考えている。会議の具体的な内容は次のようになっている。

C J K シンポジウム

この会議は「環太平洋 C J K (Chinese, Japanese, Korean) 書誌情報ネットワーク・システムの構築に関する国際シンポジウム」と名付けられ、国際日本文化研究センターと国際交流基金が共催する「日本研究・京都会議」の一環として開催されたが、中国国家図書館の朱岩氏が Organizer となり、日本中国学会内に設置された「中国書データベース化委員会」ならびに日本情報処理学会の CJK-JRG (Joint Research Group) その他の関係者がタイアップして開催したものである。

シンポジウムの報告者ならびに報告題目は以下の通りである。

Organizer:

朱 岩 (中国国家図書館自動化発展部) : *ISO/IEC 10646 and the Construction of the Circumpacific Documents Information Network*

Chairman:

山田敬三 (神戸大学教授) : 中国書データベースの現状と課題

Panelist:

Jack Cain (ISM Director, Asian Development): Requirements for the Internationalization of CJK Systems

和泉 新 (図書館情報大学教授) : 中国書目録のデータベース化における一つの問題
— 目録規則作成作業の中で考えたこと

宮澤 彰 (文部省学術情報センター教授) : 東アジアの書誌データベースと
I S O - 10646 U C S

杉山時之 (国会図書館情報処理課課長補佐) : 国立国会図書館の資料とデータベース化

Discussant:

岡崎由美 (早稲田大学助教授) : 新書目録規則について

佐々木勲人 (筑波大学講師) : 書誌情報検索システムについて

笹倉一広 (静岡大学助教授) : 古典目録規則について

瀬戸 宏 (摂南大学助教授) : 逐次刊行物目録規則について

松岡栄志 (東京学芸大学助教授) : キャラクターセットについて