

くても役に立っていた。またマルチウインドウのような工夫によって、いくつかの情報を同時に表示することができた。しかし画像を扱うようになると大型のスクリーンが必要になることがすぐに分る。国立民族学博物館では2000×2000ドットのディスプレイを導入して、標本資料の画像を複数個同時に表示するシステムを構築している。しかし、例えばランドサットやスポットなどのデータは1シーン当り縦横6000から9000ドットあり、大型のディスプレイでも一度には表示できない。

今回導入した装置は8層のフレームを持ち、各層は64メガドットの容量があるのでスポットのデータもスッポリ入ってしまう。そしてスムーズな高速スクロールによって任意の領域を表示させることができる。これは全体が一覧できるというわけではないが、それに匹敵する表示能力を有していると考えられる。とくにこの装置では、拡大、縮小が非常にスムーズに行なわれるので全体から局部まで自由に見ることができる。これも専用ハードウェアとコンピュータとの連動による新しい機能の実現である。

4. イメージデータに対する高速パターン認識の実現へ

文字の自動読取りや図形の認識は、コンピュータの歴史と共に研究の対象となってきた問題であるが、まだ解決されていない部分が多い。しかし文字列の高速検索や画像の高速スクロールや拡大縮小が行なえるようになると、文字パターンや図柄の検索も不可能でないように思われる。文字や図形の特徴を分析するのではなく、ドットパターンの一致の度合を高速にチェックする機能を活用すれば、実現できそうである。電子ファイルにイメージとして入力されているドキュメントに対して、その中に記述されている文字を認識することが可能かもしれない。そうすればフルテキストデータベースの作成も非常に容易になる。既に出版されている書籍などは電子ファイルに蓄積しておけば、やがてそれをコードに変換することができるであろう。それを期待している。

国文学研究とパーソナルデータベース

国文学研究資料館 北村 啓子・安永 尚志

国文学者の中でも身近なパソコンを研究の道具として使う人が増えてきており、パーソナル環境での研究支援の要望が多い。また、パソコン上で様々なメディアを比較的容易に扱えるようになってきており、メディアを越えて相互にリファレンスしながらの研究環境が期待されている。

そこで、従来の“大型計算機環境”と新しい“パーソナル環境”それぞれでの研究支援方法、および両者の有機的な利用形態の構想を試みる。ここではそれぞれの環境に加え、両者をコミュニケーションする環境として、オンラインで結ぶ“ネットワーク環境”、ニューメディアを介して(オフラインで)結ぶための“プロバイダ”(環境というよりシステム)を考え、これら4者が担うべき機能の分担を行い、それぞれを有機的に利用する研究支援環境の提案を行った。またこの環境において、利用者がパーソナルデータベースをいかにして獲得、利用していけるかを、利用者の立場から見た具体的イメージで説明した。

パーソナル環境での研究支援として、最初に着手したのが目録型データベースをCD-ROMで提供することである。今回開発したマイクロ資料目録CD-ROMについて、データの内容、データ構造上の特徴、ならびに検索システムの特徴と主な機能を紹介した。特に検索システム

については、既にオンラインによる公開サービスを行っている大型計算機上の検索システムと比較しながら説明した。最後に、現在までに館内で出てきた改良すべきいくつかの点にも触れた。

自動品詞分解を通しての古語のデータベース化

大阪樟蔭女子大学 学芸学部 西 端 幸 雄

わが国の古典文学作品の語彙（古語）データベースを構築しようとする場合の作業の流れを振り返ると、

単位語句句切 → 語句の諸情報付加 → 手書きでカード化 → キーボードで入力

というように、すべての作業過程で人間が介在する。それに、その作業に携わる人間は、対象とする作品の大きさにもよるが、一般的には複数人間が必要となる。このように、複数人間によって作業を進めようとする、いくらしっかりした作業マニュアルを作っておいたとしても、結果的には、不統一な面が残る可能性が高くなってしまふ。

こうした人間が介在することによって生じる不統一を、パソコンを用いることにより極力少なくし、さらに作業の流れを自動化するというのが今回の試みである。なお、今回の作業では平安時代の和歌集を資料とした。

ところで、この作業のねらいとするところが、データベース作成に関わる作業過程での人間の介在を少なくする点にあるので、事前に単位語句句切と品詞分解を行なう際の基準となる『古語辞典』（約4万語）を作成した。この『古語辞典』には自立語辞書・助詞辞書・助動詞辞書の3種類がある。

まず、入力された本文データから、14文字を最長として、漸次1文字ずつ文字長を減らし、『古語辞典』との一致を検索させ、一致すれば、その一致語句の末尾に句切符号を付けるようにした。現在のところ、この単位語句句切の正答率は約90%で、誤った句切についてはキーボードより修正できるようにしている。

次に、すべての本文を単位語句に句切ることができた後、その句切符号付きの本文データに対して、品詞分解を行なわせる。この場合も、『古語辞典』を用いて、本文データの後部から前部に向けて順次取り出した句切符号までの語句との一致を検索し、一致すれば、『古語辞典』内に登録してある文法情報や接続情報を基に、当該語句の文法的性格を決定させるようにした。この場合の正答率も約90%である。

以上の作業が完了すれば、次のようなフォーマットで、データベースとして登録するのである。

A	B	C	D	E	F	G	H	I	J	K
A	歌番号			4	バイト	(半角4桁)				
B	行数			3	バイト	(半角3桁)				
C	本文語句			60	バイト	(全角30文字)				
D	見出し用語句			60	バイト	(全角30文字)				
E	漢字			60	バイト	(全角30文字)				
F	参照語句			60	バイト	(全角30文字)				