

# CD-ROM による国文学研究材料 データベースの配布

北 村 啓 子

## 要 旨

計算機による国文学研究支援の中でもパーソナル環境の支援として、研究材料を研究者の手元に配布することに取り組んだ。配布媒体としては CD-ROM を選び、配布材料としては、(1)目録型データ (2)フルテキストデータ (3)マルチメディアデータの3種類に分けて、順次取り組んで行くことにした。

(1)目録型データでは、高速インデックスサーチと高速表示を目的としてデータ構造を設計し、マイクロ資料目録データベースの評価版を作成した。また、試用者によるアンケートを実施し、その分析によると良好な結果が得られた。特に、ツールとしてはもちろんアイデアプロセッサ的な真の意味での研究システムとしての要素もあるという喜ばしい結果も得られた。

(2)フルテキストデータでは、高速文字列検索のためにデータ構造を設計し、日本古典文学大系の中から6作品を選んでテスト版を作成した。またこの検索システムには、日本語固有の縦書き文化に合わせた縦表示機能を開発し組み込んでいる。

(3)マルチメディアデータでは、4種類の実際の応用例について考察したのでその紹介を行う。

## 1. はじめに

国文学研究資料館では、国文学研究の計算機による支援を目的とし、各種研究材料それぞれをデータベースとして大型計算機上で構築してきた[3][4]。例えば、書誌情報データベース、所在情報データベース、目録情報データベース、フルテキストデータベース、イメージデータベースなどである。ネットワークを利用してこれらのデータベースにオンラインアクセスも可能としている。

次のステップとして、パーソナル環境の支援に取り組んでいる[1]。特に注目した要件を以下に列挙する。

- ・パソコンで使える
- ・自宅の書斎で使える（スタンドアロンが望ましい）
- ・研究材料を個人所有できる（自分だけのメモの書き込みや自分だけのデータ加工ができる研究ノートの電子版“プライベートライブラリ”）

そこで研究材料を国文学研究資料館の大型計算機上に集中型のデータベースとして置くだけでなく、個人利用のデータベースとして研究者の手元に配布することを考えた。配布媒体としては、大記憶容量、コンパクト、標準化が進んでいる、取扱いが簡単、普及性、安価、パソコン環境での使用などの点からCD-ROMを選んだ。[7]

研究材料配布の構想を研究材料の種類により（1）目録型データ（2）フルテキストデータ（3）イメージデータを含めたマルチメディアデータの3つのステップに分けて順次取り組んで行くことにした。それぞれのステップで、特別なデータ構造と検索システムを設計してきた。また、国文学者にとって使いやすいインターフェースも重要な課題である。上記3つのステップとは別に1つの大きなテーマとして取り組んでいるが、その成果は各ステップに組み込んでいる。ここでは、インターフェースについてまとめて議論はしないが、各ステッ

## CD-ROM による国文学研究材料データベースの配布（北村）

プの中で適宜触れることにする。

本論文ではこの活動のそれぞれのステップでの現在までの成果と今後の構想についてステップを追って述べる。

### 2. 目録型データ

#### 2.1. データの内容と検索システム

CD-ROM 化するデータベースとして、当館所蔵マイクロ資料目録データベースを選び、評価版を作成・プレリリーステスト利用して頂いている[2][5]。このデータベースは、オンライン公開しているデータベースの1988年度版と同じ内容である。オンライン公開の検索システムでは、検索キーは書名、著者名のみであるのに対し、刊行年（和暦、西暦）、出版地、出版者、所蔵者、請求番号も検索キーとして使用できるようにした。特に和暦は、年号だけの指定によりその年号の元年から最終年までの範囲を検索できるようにしている。

国文学固有の書名ソーラスは、オンライン公開の検索システムと同等の内容を格納している。国文学では特に重要な読みによる検索と表記（漢字）による検索の両方をオンライン公開の検索システムと同様に可能にしている。

JIS 外字の文字フォントは、オンライン公開の検索システムでは当館大型計算機直結の端末しか表示できないが、出現頻度の高いものから188/245文字（出現率としては90%）の文字フォントを格納し表示可能とした。

また、利用計算機を PC9800 シリーズに特定することにより、画面制御上の制約が少なくなるので、マンマシンインタフェース（MMI）の改良を行った。（オンライン環境では、端末として多くの種類の計算機を前提とするため計算機によって異なる画面制御は事実上難しい。）具体的には、

- ①フルスクリーンを使用し、
- ②検索初心者向けのメニュー方式（画面のガイドに従って入力）と、検索習熟者向けのコマンド方式（検索論理式を入力）の2種類の入力方式を用意し

\*\*\* 国文学研究資料館 MICRO 検索メニュー \*\*\*

書名 [ ゲンジモノガタリ ]  
 著者名 [ ]  
 請求記号 [ ]  
 西暦刊年 [ ] 年 ~ [ ]  
 出版者 [ ]  
 出版地 [ ]  
 和暦刊年 [ ]  
 所蔵者名 [ ]

次の 8 件の書名が該当します。  
 検索したい書名の番号を入力して下さい

番号	書名
1	藤原氏物語
2	源氏物語
3	源氏物語歌集
4	源氏物語類聚
5	源氏物語
6	源氏物語
7	源氏物語
8	源氏物語

番号 [ 1,6,7 ]

終了 消去 決定 前頁 次頁 検索

メニュー方式での書名シソーラスの例

\*\*\* 国文学研究資料館 MICRO 検索コマンド \*\*\*

検索式 [ A 清少納言 & B \*日記 ]

< 検索項目 >

請求記号 → #      著者名 → A      書名 → B  
 出版者 → P      出版地 → S      和暦刊年 → J  
 西暦刊年 → Y      所蔵者名 → H

< 検索内容 >

単語間 : AND 指定 → & OR 指定 → , NOT 指定 → ^  
 項目間 : AND 指定 → & OR 指定 → , NOT 指定 → ^

終了 消去 切替 HELP 検索

コマンド方式の例

\*\*\* 国文学研究資料館 MICRO 検索メニュー \*\*\*

書名 [ ]  
 著者名 [ ]  
 請求記号 [ ]  
 西暦刊年 [ ] 年 ~ [ ]  
 出版者 [ ]  
 出版地 [ ]  
 和暦刊年 [ ]  
 所蔵者名 [ ]

次の 件 の年号が該当します。  
 検索したい年号の番号を入力して下さい

番号	年号	開始	終了
47	永長	1096	1097
48	永徳	1381	1384
49	永仁	1293	1299
50	永祿	1235	1238
51	永万	1185	1186
52	永暦	1180	1181
53	永祿	1558	1570
54	永和	1375	1379
55	延久	1068	1074
56	延慶	1308	1311

番号 [ ]

終了 消去 決定 前頁 次頁 検索

和暦による検索の例

CD-ROM による国文学研究材料データベースの配布（北村）

た。また、

- ③ルックアップ機能（サブウィンドウに候補を表示し、数字で選択できる。網羅性の向上、入力省力化）や、
- ④サンプルによるガイダンス機能（サブウィンドウに表示される通りに指定するとサンプル経験ができる）、
- ⑤ファクションキーの多用（機能が全てファクションキーに割り当ててあるの  
で、ファクションキーを押すだけで目的の機能を実行できる）を実現した。

検索結果を見る時も、オンライン環境ではデータのバッファリング（貯めておくこと）をしないため、一方向に（前へ前へ）見るだけで戻ることができなかったが、前へ入ったり戻ったり任意の場所に飛んだりできるようにしている。また、一覧表示と詳細表示の間を行ったり来たりもできるようにしている。

パーソナル環境での利用を考えると、検索結果をプリントしておいたり、他のアプリケーション（ワープロ、データベースシステム）を利用して加工したりしたくなると思われる。検索結果をプリントする機能、ならびに MS-DOS 標準ファイル形式でダウンロードする機能を持たせている。

検索キーによる高速検索と検索結果の高速表示とを目的として、CD-ROM 上のデータ構造を設計した。検索キーの索引の構造には、多層インデックス構造を採用し、各検索キーからポイントされた目録情報は検索結果の表示フォーマットで格納することにした。これらのデータ構造により、CD-ROM 自身のアクセスの遅さをカバーしている。

格納データは1976年～1988年度年度撮影分の100万件の目録で、データサイズは約100Mバイトである（CD-ROM容量の1/5程度）。データ構造と検索システムの詳細については、[2][7]を参照されたい。

## 2.2. 評価

この CD-ROM は人文系研究者、大学図書館等約60サイトに配布し試用評

価中である。広く利用者の評価を分析してシステム評価を行うためにアンケートを作成しご協力頂いた。[7]

アンケート結果の分析によると、検索結果への満足度、使い易さ、検索スピード感など定性的な評価では良好な結果が得られた。検索キー、検索機能は種類の的にはほぼ充足しており、個人単位で見ると少ない種類の組み合わせで個人の検索スタイルに合った検索方法を利用しているようである。また、単なるツールとして役立つのみならず、新しい研究スタイルの発見やアイデアなど研究の刺激となるなど、真の意味での研究支援の一面も見いだすことができた。

以下に評価方法と分析結果について述べる。

#### (1) 評価方法

大きく次の3種類に分類して質問を用意した。

- I. 想定選択肢で確定できる事実：専門、利用環境、使用した検索キーと検索機能など
- II. 定性的な評価：使った感じなどを5段階分類で選択
- III. 自由回答形式

アンケートの実物は、文献[7]の Appendix に添付したので本論文では割愛した。そちらを参照されたい。

アンケート回収状況は、1990年9月末日現在、国文学研究者11名（のうち館内5名）、その他国語学等人文科学系研究者3名、人文科学系への応用を研究している計算機研究者3名、大学図書館司書6名である。今回の分析は、1990年4月末日回収分のみを対象としたため、国文学研究者9名（のうち館内5名）、その他国語学等人文科学系研究者3名、人文学系への応用を研究している計算機研究者1名、大学図書館司書5名である。

次の観点でアンケート結果の集計を行った。今回は上記分類のI、II、のみの集計を行い、III、自由回答は傾向をつかむに留めた。なお集計とグラフ作成には Lotus 1-2-3 を使用した。以下の5項目について結果を順次紹介する。

CD-ROM による国文学研究材料データベースの配布（北村）

- ①利用者環境の把握：使用目的，使用場所，使用機器，日本語 FEP
- ②専門別と年齢別の使用した機能：入力方式，検索キー，検索機能
- ③専門別と年齢別の定性的評価：使い易さ，検索結果への満足度，スピード感
- ④どういう風に役にたったか？
- ⑤自由回答のサーベイ

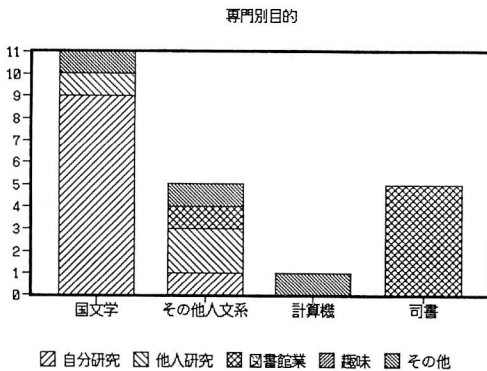
なお、②の検索キーと検索機能については、◎（いつもよく使う）を3，○（時々使う）を2，△（使ったことがある）を1として重み付けをした合計値である。人数の規格化はしていないので、グラフ上、縦軸の数字の大きさ（分類項目間の数値比較）は意味がない。分類項目内のバランスとして見ていただきたい。また、複数回答の項目（②，③）については、選択肢ごとに重複カウントしているので正確な人数ではない。それ以外は、グラフ上縦軸の数字は人数を表している。

(2) 結果

①利用者環境の把握：

a. 使用目的

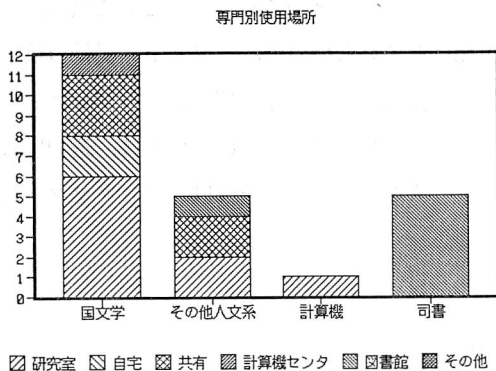
国文学研究者は自分の研究，司書は図書館業という当然の結果であるが，そ



の他人文系研究者は、他人の研究（他研究者の代行）で利用している人が多い。

b. 使用場所

国文学研究者は、20%弱の人が自宅書斎で利用している。国文学者の書斎化傾向を裏付けている。



c. 使用機器

PC9800 の機種は様々であるが、VX, VM が比較的多い。新しい機種ではRX がめだつ。

CD-ROM ドライバは、約60%が NEC (CD101, CD102 を含めて) である。他に、日立、東芝が使われている。

d. 日本語 FEP

ATOK, 松茸, NECDIC の利用者が、ほぼ同数ずついる。NECDIC 利用者には、配布状態のまま使用している人が多いと思われる。

②専門別と年齢別の使用した機能：

a. 入力方式

表記の入力：ローマ字入力→仮名漢字変換は70%弱, 仮名入力→漢字変換は20



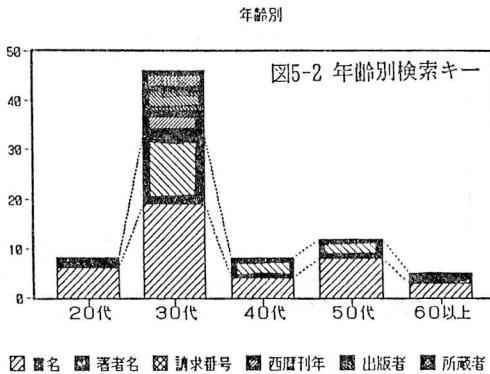
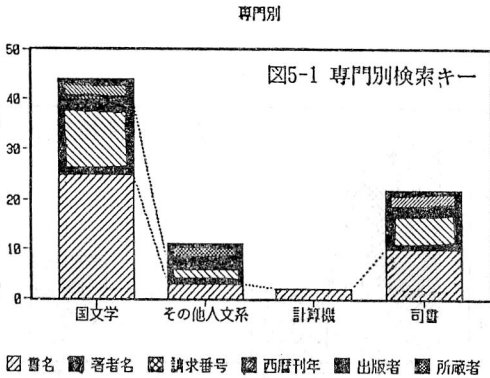
CD-ROM による国文学研究材料データベースの配布（北村）

%強である。ローマ字入力仮名漢字変換の利用者が多い。

ヨミの入力：ローマ字入力は40%，仮名入力は35%である。ヨミによる検索は行わない人が結構いるようである。

b. 検索キー

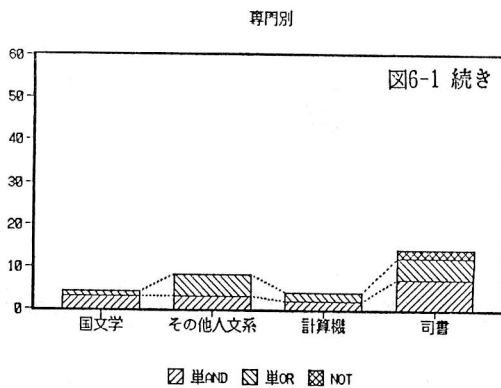
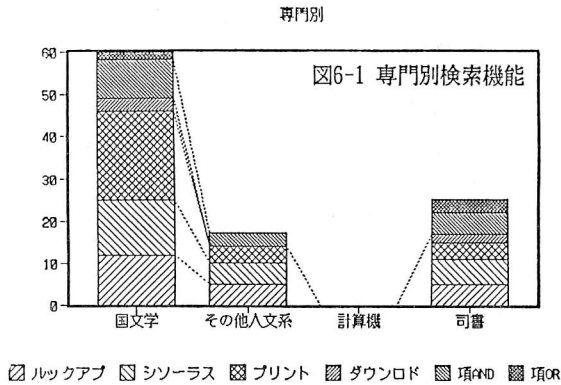
出版地，和暦刊年の利用者はゼロであった。主に書名，著者名が利用されており，その他の検索キーは特定少数の人が特定のキーのみを合わせて使用して



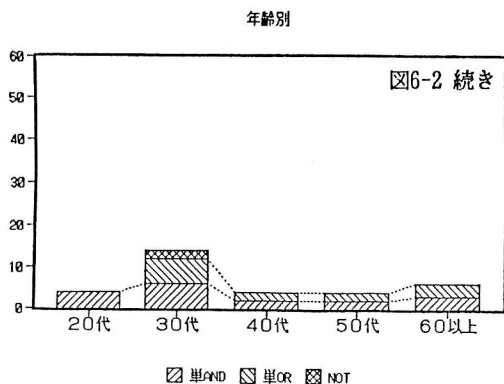
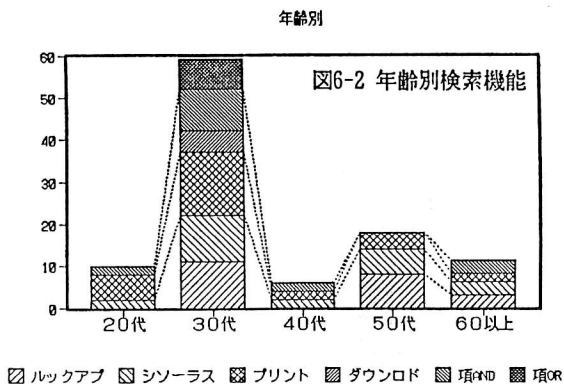
いるようである。つまり個人個人は少ない種類の検索キーしか使用していない。

c. 検索機能

どの検索機能も平均的に使用されているように思われる。プリント機能が多少多く利用されており、検索の中間結果を見ながらさらに絞り込みをするためにプルーフリスト的な使い方をしているようである。年齢別によると30代が一番多くの種類の機能に挑戦している傾向が見受けられる。



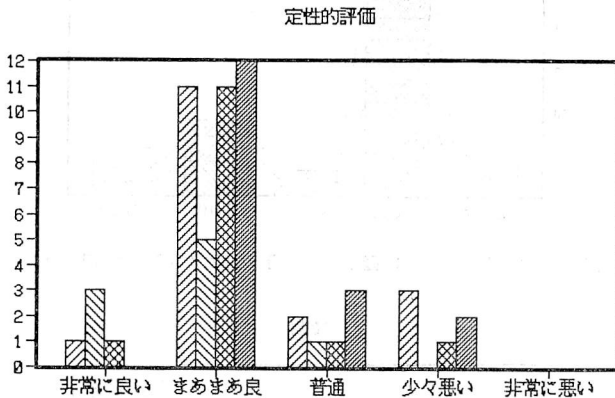
CD-ROM による国文学研材料究データベースの配布（北村）



③専門別と年齢別の定性的評価：

CD-ROM を使った感じ、オンライン検索に比べて使った感じ、検索結果の満足度、検索のスピード感の4項目をそれぞれ5段階評価で棒グラフにした。全体的に“まあまあ良い”の評価が得られている。特に検索結果の満足度について約85%の人が“まあまあ満足”以上（70%以上目的物が見つかった）と感じている点は、データベースの品質の面で高く評価できよう。スピード感はまあまあ速いの回答者が多かった。オンライン検索利用者はモデムの性能上（現

在 300bps～1200bps を使用) 遅い検索システムに慣れているのも一要因と考えられる。



CD-ROMの使い易さ
  対ウツラの使い易さ
  検索結果の満足度
  ステート感

#### ④ どういう風に役に立ったか？

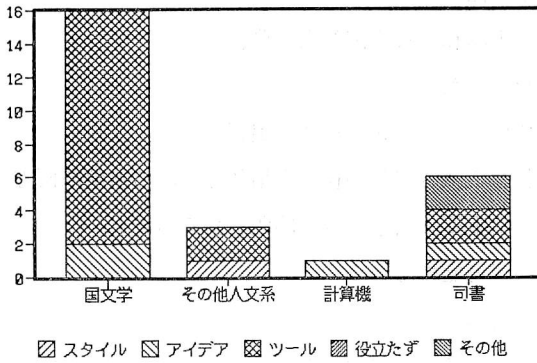
65%の人が \*ツールとして役だったと評価していることは検索システムとして合格点といえよう。単なるツールを越えて新しい研究スタイルの発見や新しいアイデアなど研究の刺激になったと答えた人が25%もいた。年齢別に見ると、若手より30代、40代、50代と研究活動への造詣が深まるにつれ、真の意味での研究の支援システムに対する考察も深まることが期待されるようである。

この検索システムが(真の意味での研究支援である)アイデアプロセッサ的な要素を多少兼ね備えていると判断してよいと思われる。

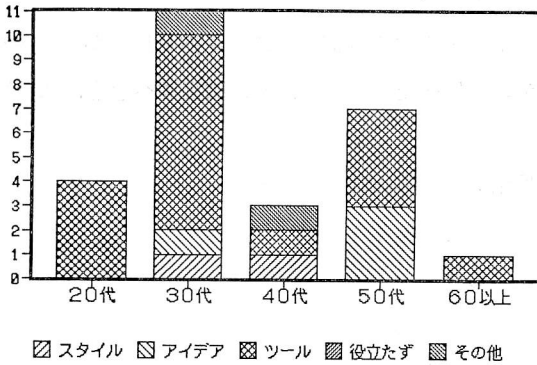
\* ツールとして役立ったと他の項目とを複数選択した人は重複カウントしているので正確な人数で言うと80%程度である。

CD-ROM による国文学研究材料データベースの配布（北村）

専門別 役立ちかた



年齢別 役立ちかた



⑤自由回答のサーベイ

自由回答の中から主な内容をリストアップし紹介する。

- ・手元にあるという心理的満足度が得られる
- ・計算機音痴でも使える
- ・他のデータも欲しい

例えば、目録（所蔵、国書、漢籍、論文、連歌、……）

古典作品本文（古典文学大系、和歌、俳句）

原本の画像データ

辞書（大漢和、日本国語大辞典、……）

- ・講習会やデモを行った
- ・教育に使用している（データベースの実習）
- ・今後データ更新を希望（累積の検索と増加分だけの検索の両方）
- ・ヨミによる検索は重要（新字旧字の問題があるので）
- ・オンラインと共存可能であろう
- ・中間一致検索を希望
- ・高度なマニュアルを希望（現在は平易なマニュアルしかない[5]）
- ・製品化を希望（図書館や院生も購入できるよう。CD-ROM 普及の促進や価格安定にも役立つであろう。）

なお、アンケートを実施する以前に集めた館内利用者からの声を a. オンライン検索システムと比べたメリット, b. 改良すべき点, に分類し, ここで紹介しておく。

a. オンライン検索システムと比べたメリット

- ・手軽なパソコン環境で使える
- ・自由な時間に使える
- ・家庭のパソコンでも使える
- ・中間結果のリストを見ながら絞り込みができる
- ・ダウンロードしたものに書き込みができる
- ・ヨミによる検索を行うことにより, 網羅度が向上する

例えば, リュウノスケ→竜之介

→龍之介 の両方を検索できる

- ・ルックアップ機能により, 網羅度が向上する

例えば,

リュウ\*→

竜之介 龍之介 竜之進 .....
----------------------------

の候補の中から選択するので、漏れがない

- ・漢字表記による検索を行うことにより、間違いが減少する（自分でキー入力するため）
- ・検索結果の表示でバックできる

b. 改良すべき点

- ・ダウンロードした結果をまたこのシステムで再利用したい
- ・ダウンロードした結果を市販ソフトウェアで利用したい
- ・日本語 FEP のインライン変換をして欲しい（画面の1番下でファンクションキーとの切り替えが面倒）
- ・検索集合の中で検索をしたい（検索集合に対するオペレーション）
- ・中間一致検索をしたい
- ・新データの追加をして欲しい
- ・その場合、増加分だけの検索も可能にして欲しい

3. フルテキストデータ

フルテキストデータベース（科学研究費により岩波書店刊行日本古典文学大系全100巻600作品をデータベース化中）から次の6作品を選び、テスト版を作成し内部評価を行った。〔6〕

作品：竹取物語、伊勢物語、大和物語、源氏物語5巻、枕草子、平家物語2巻

フルテキストデータベース利用上の重要な問題の一つは捜したい表現をいかに網羅的に捜せるかという点であろう。これについては、単語という概念を導

入せず、飽くまでも単なる文字列の並びとみる方針を採った。単語という概念を入れると、単語の切り方に一種の標準化が必要となり、その標準を知らないが故に網羅的な検索ができないという現在の大系化された検索キーと同様の危険に陥る可能性がある。(検索キーに比してその危険性は低いであろうが。)

実際の例として、“御前辺り”の場合、それぞれ“御前”と“辺り”それぞれを一単語とすべきか、“御前辺り”で一単語(複合語)とすべきか議論の分かれるところであろう。この場合、複合語として切っていれば、それぞれの“御前”と“辺り”のどちらで検索してもヒットしないという事態が起きる。逆に“御前”と“辺り”それぞれを一単語として切った場合も同様に“御前辺り”で検索してもヒットしない。他に次のような例が該当するであろう。

複合語の類：端近う 午前近く 御座有り 御出有り

今一度 故姫君 春三月

地名の類：近江国滋賀

官職名前の類：太政大臣鎌足 柿本朝臣人麿 法勝寺執行御坊

相模守孝義 平相模守孝義

固有名詞類：橘大夫元家九代目市村羽左衛門 東海道往來之記

単純な文字列検索の場合、単語の部分に含まれるものもヒットするため、ノイズが入る可能性はあるが漏れは少ないことが期待できる。即座に KWIC リストを見せることにより、ノイズを容易に除けるという便利さ(利用環境)でノイズが入ってしまう点をカバーしようと考えた。

機械翻訳などで形態素解析と呼ばれている、文法ルールを参照しながら大規模な辞書と比較を行うことにより単語を認識し、文の構成を解析する研究がなされている。品詞の接続関係や活用形を始めとして文法を明文化することが必要であり、また単語認識の正当率は辞書の品質(語数)に依存する。現在の研究対象は現代語のみでありそのまま利用することはできないが、古典の電子辞書を構築、文法を明文化することにより、同様の方法で単語の認識をすること



は期待できる。

実用に耐える大規模な辞書の電子化は息の長い仕事であるので、もう少し将来を目標とし、現状では文字列以上の扱いはしないことにする。どう表現されているかの可能性は、利用者の専門知識に頼ってしまう方法である。例えば、“狂言尽し”の言葉の由来、内容、思想性などを調べるため、“～づくし”の用例を集めたいという時、捜したい表現が書かれている可能性のある表記と読みすべて

尽し

盡し

づくし

つくし

の4通りをあげれば、システムは漏れなく捜すことができるという着実な方法である。

また、国文学だけでなく文学一般に言えることかもしれないが、上の例のように非独立語が重要視されているという傾向があるようである。“狂言尽し”を調べるために、独立語である“狂言”は辞書、本の索引などにより用例も比較的収集しやすい。非独立語である“～づくし”となると、用例は採集されていなかったりするようである。また、従来の索引は名詞のみ（または名詞と動詞）が一般的なもので、索引の利用はあまり期待できないであろう。非独立語に注目するという意味でも、単語の概念を入れず、単なる文字列として扱う方針を選んだ。

もう一つの重要な問題は、いかに速く文字列を捜せるかという点である（国文学者にとっては先の網羅度に比すれば大した問題ではないかもしれないが、計算機の専門家にとっては興味あるところである。国文学者にとって必須でなくとも嬉しいであろうということここで取り上げる。）特にパソコンの計算パワーでは死活問題となるであろう。CD-ROM は他の記憶媒体（ハードディス

ク、フロッピディスク等)に比べてアクセスが遅いという欠点がある。この欠点を克服し、フルテキストデータベース一般に利用できるデータ構造の設計を行った。

我々の採ったアプローチは、テキストを一文字一文字見ながら文字列を捜すことをやめるという方法である。検索処理をある程度まであらかじめ行なっておき、その中間結果を保存しておく。実際に検索する時は、それらの情報を参照することにより、テキスト上を一文字ずつ見ることなく、捜すものに速く辿りつくという方法である。端的な例が索引である。前もって索引を沢山作って格納しておこうというアプローチである。

データ構造の設計にあたって、次の3点にポイントを置いた。

●本文はテキストをべたづめにする

CD Extension や市販アプリケーション、ユーザプログラムでの利用を可能とするため検索のための特殊な記号などを書き込まず、純粋なテキストデータのみを格納する。(オープンデータ)

●1字インデックスを格納

1文字単位で全ての文字が本文中に出現する場所のアドレスを索引にして格納しておく。捜したい文字の1字インデックスを利用することにより、本文中のどこに出現するかが即わかる。

●KWIC リストを格納

全ての文字に対する1文字単位の KWIC (Key Word In Context) を格納しておく。捜したい文字列すべての KWIC リストである。

先にもインタフェースの重要性については述べたが、国文学者にとって自然なインタフェースの研究は1つの大きな課題と認識し、取り組んでいる。マイクロ資料目録 CD-ROM では、種々の試みを検索システムに盛り込んだ。国文学を扱う場合の一般的なインタフェース上の大きな問題として、書物を縦に読み、縦に書くという日本語文化と、パソコンなどキャラクタディスプレイの

## CD-ROM による国文学研究材料データベースの配布（北村）

横書き文化のギャップがある。この問題は、特にフルテキストデータベースを扱う場合顕著に現れる。そこで、テキストを表示する際共通に使用できる縦表示機能を開発することにした。日本語文字を縦に速く表示し、左右にスムーズかつ速くスクロールする機能を持つ。パソコンはビットマップディスプレイを持たず、キャラクタディスプレイとグラフィックスディスプレイしか持っていないため、縦専用の表示関係のソフトウェアを開発することにした。この縦表示のソフトウェアはフルテキスト CD-ROM に組み込んでいる。利用者は、テキストの表示、検索結果の表示、KWIC の表示、などの機能をこの縦表示機能を通して利用できる。

このテストバージョンでは、フルテキストデータベースを CD-ROM に格納する際の我々が設計したデータ構造を評価することが主な目的であるため、文字列検索システムはその評価を行うだけの基本的な機能だけにとどめた。以下に主な機能を説明する。

### ①テキスト表示（画面1）

指定した作品のテキストを縦に表示する。日本古典文学大系本の1行は画面上の1行に入らないので、折り返して画面上2行使って表示している。横スクロール、次のページ、前のページ、先頭ページ、最終ページへの移動ができる。画面の右端に現在位置（本のページ、行数）と最後の位置（本のページ、行数）を表示している。

### ②検索文字列の指定（画面1）

キー入力ファンクションキーを押すと、検索キー入力の促進記号が表示される。MS-DOS 上の任意の日本語フロントエンドプロセッサを組み込んで利用することができる。最大23文字の文字列を指定できる。

### ③文字列の検索（画面2）

指定された文字列を捜し、①テキスト表示と同じ画面を表示し、該当する文字列をカラーの反転表示する。①テキスト表示と同様に横スクロールとページ



の移動ができる。また、ファンクションキーを押すことにより現在見ている位置の KWIC リストに移れる。

#### ④KWIC 作成 (画面 3)

指定された文字列を捜し、該当する文字列の前後を取り出し1件としてリストを作る。リストの1件を画面縦表示の1行として表示する。ファンクションキーを押すことにより現在見ている位置の③文字列の検索へ移れる。

#### ⑤ダウンロード

テキスト、KWIC リストを MS-DOS テキストファイルに格納できる。

#### ⑥プリント

テキスト、KWIC リストをプリントアウトできる。ただし、現バージョンでは、横書きで印刷される。縦書きで印刷したい場合は、PC9800 のハードコピーを利用するか、またはダウンロードした後、縦印刷のできるワードプロセッサを利用されたい。

テスト版を開発し、このデータ構造の効果を確認した。館内での紹介を行い、国文学研究者から検索機能の面（非独立語の KWIC など）、検索スピードの面、使いやすさの面（テキストの縦表示）から、好ましい評価が得られた。マルチメディア CD-ROM にもテキストを表示する場合に縦表示ソフトウェアを利用する予定である。テスト版の評価をもとに設計の見直しを行い、1991年春に改良した CD-ROM を大量作成し、人文科学系研究者に配布し試用していただく予定である。

### 4. マルチメディアデータ

次に上記2種類のタイプのデータにイメージデータも含めたマルチメディアの試行を考えている。マルチメディアのデータベースでは、利用者がデータタイプの違いを意識しなくても良いという点が重要である。データの種類を越えたデータ間のリンク付けが最重要課題である。システムのユーザインタフェー

スとしては、マルチウィンドウシステム上でビットマップディスプレイを使用し、イメージデータの表示、テキストデータの縦表示、ウィンドウからウィンドウへの移動や情報を受け渡しなどの機能が必要であろう。

現在応用例として、科学研究成果[8]である次の4種類を考えている。科研では、大型計算機上にそれぞれ異なるメディアのデータベースが存在するという前提での研究であった。それらのデータベースのサブセットをCD-ROMに格納する方法、また逆に数枚のCD-ROMに分割して全データベースを格納してしまいう方法も考えられる。いづれにせよ、科研の成果はCD-ROMでも実現することが可能である。以下に4種類の応用例をそれぞれ簡単に紹介する。

### (1) 原文献資料探し

国文学者が日常的に原本を捜(し紙焼きを入手)すときに行っている活動を計算機支援する形態である。書誌情報や所在情報を別ウィンドウで参照しながら古典籍目録データベースやマイクロ目録データベースを検索し、範囲を絞ってにおいて、その原本のイメージデータをブラウジングすることにより、最終的に必要な本を決定する。

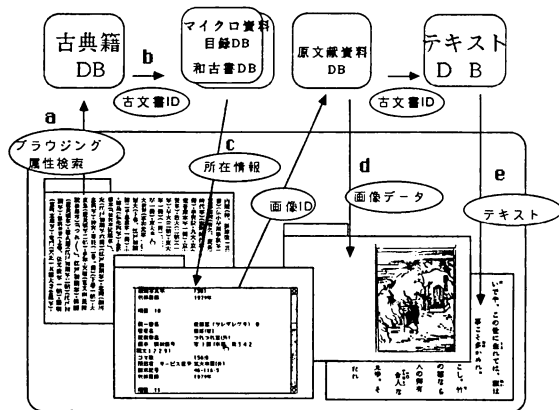


図1 原文献資料探しへの応用例

(2) 挿絵の研究

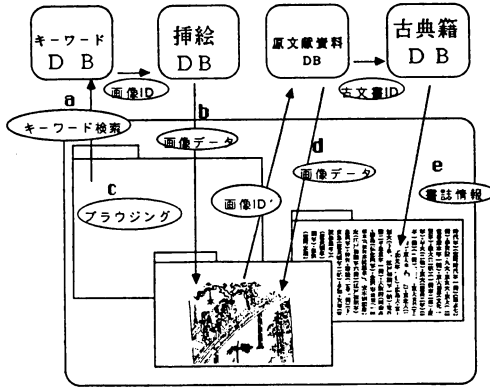


図2 挿絵の研究への応用例

これは科研“挿絵データベースシステム”の成果[10]を応用した例である。挿絵に描かれた構成物をピックアップし、それをその挿し絵のキーワードとしてデータベース化する。キーワードを検索することにより、該当する挿絵を捜すという研究である。例えば、刀を持った侍が出て来る絵を捜すといった具合である。刀や侍、また刀のさし方などが時代とともにどう変遷してきたかなどを調べるのに有効であろう。キーワード検索→イメージデータブラウジング→挿絵の前後の本文を読む（挿絵を持つ原本のイメージを見る）→挿絵を持つ本の書誌情報を見て、傾向、特徴などを分析するといった典型的な研究の流れが考えられる。

(3) 翻刻作業

翻刻作業は、1つの底本を多くの異本と照らし合わせながらの最適な解釈が要求される。同時に多くの異本を覗けることが必要である。

マルチウィンドウを利用して、底本（イメージ）を中心に、複数の異本（イメージ）を見ながら翻刻したテストを別のウィンドウにワードプロセッサを

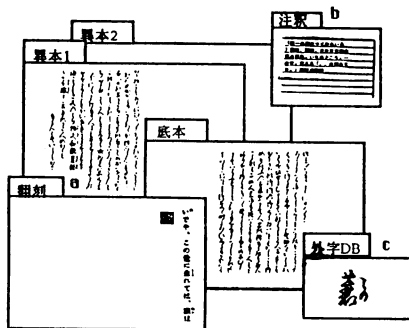


図3 翻刻作業への応用例

使って入力していく。注釈を入れたいところでサブウィンドウを開き、テキスト入力する。

同様のツールを利用して、複数の人で定本（スタンダード）を作成するのも可能であろう。1つの翻刻テキストに対して、各人の注釈・意見をサブウィンドウに書き込んで行き、全員がそれらの意見を読み合いながら1つの新しい著作物を作り上げていく。Hypertext 技術の典型的応用例である。

#### (4) CAI (教育用)

ツールとしては、(3)翻刻作業とほぼ同じで実現できるのだが、目的を原文献資料を読むという教育の支援とする。具体的には、原文献資料（イメージ）を1つのウィンドウで読んでいく。つまった時には翻刻されたテキスト（定本）の該当場所を参照しながら読む。また、不明な言葉はサブウィンドウで注釈を見たり、電子辞書を引いたりしながら読む。それでも不明な点、気付いた点は原文献資料に書き込みをしておく。また、(3)翻刻作業と同様に外字の参照も役立つであろう。



CO-ROM による国文学研究材料データベースの配布（北村）

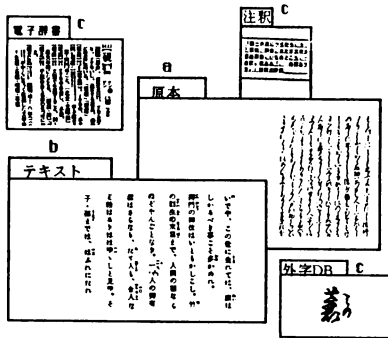


図4 CAI（教育用）への応用例

ここであげた(1)(2)の応用例については、先出の科研報告書[8]に付録として採録してある通り、Machintosh Hypercard 上でプロトタイプを作成しユーザインタフェースについて現在評価中である。

また、CD-ROM 上のデータ構造についても現在検討中である。画像データのサイズが大きいため、大容量 CD-ROM でも一般的な圧縮方法を採用して概算で 3000~4000 枚分の画像データしか格納できない。より効率のよい圧縮方式の検討、複数 CD-ROM 化の検討など、技術的な検討課題が多く存在する。またパソコン環境では、イメージデータを表示するのにディスプレイ精度が不十分である。より高精度のディスプレイ上でのシステム化考えるか、またはディスプレイ上でアウトライン表示やズーム機能などマルチビューの環境を整備するか、ディスプレイと紙（印刷物）のように機能を分担し組み合わせて利用する方式を検討するかなどについても検討中である。

## 5. ま と め

CD-ROM による研究材料配布の構想と現状を述べた。目録型データとフルテキストデータの CD-ROM 上のデータ構造と検索システムの設計を行った。目録型 CD-ROM は、評価版をプレリリースし、利用者アンケートの分析により、良好な結果が得られた。フルテキスト CD-ROM は、テスト版の開発までこぎつけ、データ構造の有効性を確認した。現在 CD-ROM というメディアで配布するということに関しては、利用者の関心が高く、国文学の研究スタイルにも変化を与えているようである。国文学固有のインターフェースとて開発し、CD-ROM システムに組み込んだ縦表示機能も利用者から好評を得た。

アンケートの回収状況は 23 名で分析の母集団としては小さかったのが残念である。この分析は引き続き行いたい。また自由回答の分析手法も開発されてきているので、それらを利用した分析にも取り組みたい。

次なる目標のマルチメディアデータについては、データタイプの違いを知らずに使えることが重要である。インターフェースとしてハイパメディアの考え方を応用したシステムを考えている。マルチメディアのデータを扱う場合、利用者は沢山の異なるタイプのデータを参照するであろうから、大きなディスプレイが必要になると思われる。そろそろパソコンから WS への過渡期かもしれない。

研究材料を配布する媒体として、CD-ROM が適していることを確認した。これまで、パソコンでの国文学者向けの使いやすいインターフェースや高速検索の実現を行ってきた。このアプローチはこれからも続けるが、同時に計算機パワーやユーザインタフェースの面でより可能性の高い WS 上での CD-ROM システムにも取り組みたい。

[謝辞]

本研究の機会を与えてくださった、国文学研究資料館長小山教授、同研究情

## CO-ROM による国文学研究材料データベースの配布 (北村)

報部情報処理室長安永教授に感謝いたします。また、日頃から国文学研究者の立場から計算機への要求や貴重なアドバイスを頂いている、同研究情報部長新井教授、武井助教授を始め同部教官の方々、ならびに同文献資料部教官の方々に感謝いたします。最後に、科研“挿絵データベースシステム”の研究に参加させていただいた同文献資料部岡教授に感謝いたします。

### (参考文献)

- (1) 北村啓子, “国文学とパーソナルデータベース”, 京大型計算機センター第28回研究セミナー, ISSN 0910-3201, pp.15-16, (1990)
- (2) 北村啓子, 安永尚志, “マイクロ資料目録 CD-ROM と検索システムの開発—パーソナル環境での国文学研究支援—”, 情処全大第40回, 6R-1, pp. 1578, (1990)
- (3) 安永尚志, “国文学研究支援のためのコンピュータ利用”, 人文科学とコンピュータ研究会, 89-CH-2, 2-6, (1989)
- (4) 安永尚志, “国文学データベースの形成, 管理, 利用”, 国文学研究資料館紀要, vol.16 pp.124, (1990)
- (5) 国文学研究資料館, “マイクロ資料目録 CD-ROM 検索システムユーザ利用マニュアル”, (1989)
- (6) 北村啓子, 安永尚志, “古典作品テキスト CD-ROM と文字列検索システムの開発—パーソナル環境での国文学研究支援—”, 情処全第41回, 1F-7, vol. 4, pp.142-143, (1990)
- (7) 北村啓子, 安永尚志, “目録型 CD-ROM 検索システムの評価”, 情処人文科学とコンピュータ研究会, 90-CH-5, 5-6, (1990)
- (8) 北村啓子, “Hypertext 技術を応用した横断の利用技術の提案—既存 DB を活用して—”, 文部省科研費一般研究(B) 63450059 “国文学異次データベース群間の横断の利用方式に関する研究” 研究成果報告書, pp. 27-50, 1990年3月
- (9) 北村啓子, “テキスト DB に対するユーザ要求のモデル化とプログラム合成”, AI学会知的ソフトウェア開発研究部知的ソフトウェア開発第1回ワークショップ論文集, 1989年8月
- (10) 岡雅彦(代表者), “挿絵データベースシステム”, 文部省科研費一般研究研(A) 62410010 究成果報告書, pp. 22-30, 1989年3月
- (11) J. Conklin, “Hypertext: An Introduction and Survey” Computer, Vol. 20, No. 9, pp.17-41, (1987)