

資料

文化財科学関連文献データベースの作成と利用

西田泰民¹⁾・河西 学²⁾

作成者側から見た文化財科学関連文献データベース

西田 泰民

学会の10周年記念事業として企画担当理事であった小池裕子会員の発案により開始されたプロジェクトに協力の依頼を受けて、このデータベースに関わり、10年以上が経つ。この事業は学会設立以前に刊行された文献目録の作業の継続という意味もあったが、当初から文献目録としてではなく、コンピュータでの検索を主とするデータベースとして作製することを目的としていた。

作業を開始した時点ではNECのPC98シリーズが圧倒的シェアを持つ時代であったので、98上で走るソフトで作成することとなったのは当然のなりゆきであった。また独自の検索システムを構築することも当初は検討したが、機種別のプログラムを作製する余裕がないこともあり、ユーザーがすでにデータベースソフトを使用していることを前提に数タイプのファイル形式をサポートすることとした。当時はMS-DOS上で動くものとしてはRbase, dBase, 桐, Ninja, Databoxなどのソフトが市販されていたが、圧倒的シェアを持つものはなかった。また現在とは異なり、アプリケーションが組み込まれた状態でハードは販売されておらず、それぞれのアプリケーションもかなり高価であった(註1)。それらの中どのデータベースソフトを標準とするか選択に悩むところであったが、配布用の試用版ソフトの提供を頂けることとなったASCII社のCard 3を入力に用いることとなった。文献目録には著者、発表年、タイトル、掲載誌名、掲載頁は必ず含まれるが、論文の内容をどのように検索できるようにするかが、次の課題であった。結果として、どの時代のどの地域のどんな種類の文化財をどのような方法で分析したのかが、わかるようなコード表を作成し、それに基づいて文献を整理しデータ入力することとなった。対象、分野、方法に関するコードの作成は以前に刊行された文献目録や研究発表会での項目分けを参考とし、理事・幹事会で審議を経た後に決定した。その後新たな分野や方法の開発があれば適宜コードを増やすこととしている。

文献コードは当データベースの特徴でありしばらくはこの形で入力を行っていきたいと考えているが、当初の設計では各文献につきコードが3つまでしか割り振れないという制限があった。作業開始当時のパソコンデータベースソフトには総データ数や各データ長に制限があったり、あいまい検索や中間一致検索ができなかったためにやむを得なかったのであるが、現在ではこうした制限はほぼ無視できるよ

¹⁾ 新潟県立歴史博物館：〒940-2035 新潟県長岡市関原町1丁目2247

²⁾ 帝京大学山梨文化財研究所：〒406-0032 山梨県東八代郡石和町四日市場1566

すいいになっており、いずれはどの検索項目についてもキーワード方式で入力する方法に移行することになるであろう。

実際の入力作業が始まってからの問題は非JIS漢字の扱いとコード付加の統一性であった。前者については偏と旁に分解して表現できる場合はそのようにし、それ以外は形の近い漢字をあて、その前に\$を付加することとした。=や■を当ててしまうと、全く情報が失われてしまうからである。後者についてはおそらく図書館での本の分類でも同様の問題が起こっていると考えられる。図書館の場合は主テーマが何であるかによって分類コードが与えられて配架されるが、特に異なる分野にまたがるような内容の場合は類似の内容の本が全く別の場所の棚に置かれているのを、目にする可能性がある。データベースでは複数のコードを付与することができるので、ある程度この問題は回避できるのであるが、分析会社があらゆる分析を一手に引き受けてレポートを作成したような文献では、無制限にコードをつけるわけに行かず、扱いに困ることがあった。現在では、同一の文献ではあるが、年代測定と古環境復元など異なる分野にそれぞれ属する個別の文献として登録するようにしている。

文献情報の収集の方法もいまだ万全とはいえず模索中であるといっても良い。他の分野と大きく異なるのは掲載されるレポートの多くが発掘報告書に所収となることで専門家であればだれでも入手できるというものでもなく、一般の目に触れない場合が多い文献であることである。全国の報告書が集積される機関で文献チェックができればよいのであるが、公立の機関でこの学会の事業を受け入れるところはこれまでない。また各分野毎に散発的に様々な専門誌に投稿されるレポートがあり、カバーしなければならない学術雑誌が非常に多い。第1版の作成時には報告書所載の文献をリストアップするために各県

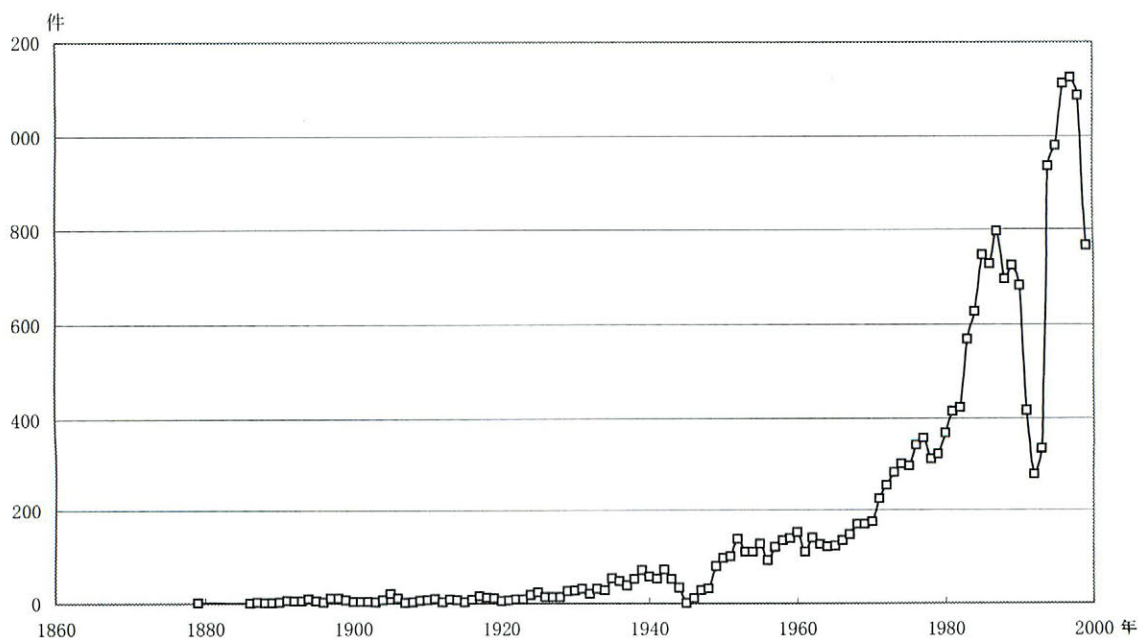


図1 明治以降文化財科学関連文献件数の推移

別に会員の中から協力者をお願いした。しかし、文献のコピー代の負担や勤務時間中の作業を認められるのかなどの問題があり、全くのボランティアとしての作業を強いることになるので、入力側としても状況は同じなのであるが、他者に依頼しにくい状態にあるため、現在はこの方式はとっていない。

なお、このデータベースの利用法としては必要とする文献の検索が第一であるが、必ずしも研究にそのまま役立てるところまではコードが細分化されていないことに限界を感じることもある。たとえば、ウマの出土例を検索しようとしても、タイトルに「ウマ」の語が用いられている文献を抽出することはできるが、コードでは哺乳類・種同定で検索をかけ、あとは実際にそれぞれの文献に当たって、ウマが含まれているかどうかを確かめなければならない。分類コードをさらに細かくすることは可能ではあるが、作業量が膨大になるので将来ともそこまではカバーできないであろう。

実はデータベースは学史の研究データとしても重要である。図は明治以降の文化財科学関連文献の件数をグラフとしたものである。さらに分野毎や研究者毎また地域毎の文献数を比較して研究動向の変化を検討することも可能である。試みに動物遺体の同定を行う研究者の数を調べてみた。このように研究の蓄積を一目で見ることができるのもデータベースの利点である。

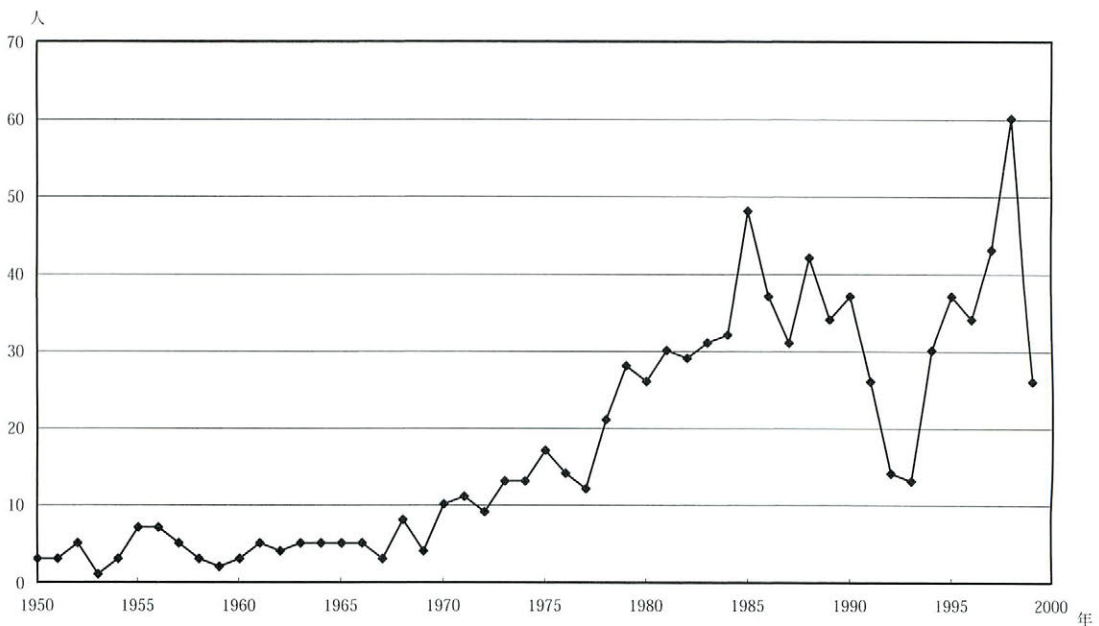


図2 動物考古学関係論文発表者数の推移

平成9年12月の第15期学術審議会学術情報資料分科会学術情報部会が行った報告（註2）では次のような問題点があげられており、いずれも当学会のデータベースについても当てはまる。

- ・学術情報データベースの作成・維持に関わる組織が、欧米に比して極めて不十分であり、研究の片手

間に行われている例が多いこと。また、研究者が作成する場合、業績として評価されにくいこと。

・学術情報データベースは、作成に相当の経費がかかるほか、恒常的に事業として自立しにくい性格を持ち、民間では供給されにくいものであるが、これまで、公的部門自ら作成することも、民間活動を支援することも、欧米に比較し極めて不十分であること。

・利用者たる研究者等は、データの追加や改良を行う生産者としての側面も有すべきであるが、その認識が低い。また、利用者側の要求との不一致があり、利用者からのフィードバックを生かす体制や利用者の要求に関する調査研究が遅れていること。

最後の問題については、作成者側としていつも気にかかっている。文献の収集の不足を補うだけでなく、入力データの間違いの指摘などをきちんと受け取ることの出来る体制を整備する必要もあるが、利用者側からの積極的な意見の御提供も是非お願いしたい。

註1 MS-DOSはマイクロソフト、RbaseはMicrorim、dBaseはボーランド、桐は管理工学研究所、Ninjaはサムシンググッドそれぞれの商標である。

註2 「学術情報データベースの整備について（報告）」(<<http://www.monbu.go.jp/singi/gaksin/00000191/>>)

利用者側から見た文化財科学関連文献データベース

河西 学

文化財科学文献データベース（以下文献データベース）は、前編版F D（1993）、後編版F D（1994）、補遺版F D（1996）、99年版C D-R（1999）が公表されており、改訂の度に改良が試みられ、利用しやすくなっている。最新のC D-R版は、20,489件の文献レコードが収録されており、C D-Rのタイトル面がカラー印刷され外見も美しい。文献データベースは、文献に関する著者、年号、タイトル、誌名、巻号、頁数の必須項目に加え、文献内容について時代・地域・対象・分野に関する分類コード項目12とこれに対応する文字表現された12項目、さらに著者名のふりがな項目の計31項目から構成される。このデータベースは、利用者の利便性を重視して作成されていて、文献内容の分類が充実している点が大きな特徴となっており、利用価値が高い。この文献データベースを通じて、文化財科学の分野における今までの研究の蓄積を多くの会員が共有し、今後の研究に活用できるということは、メリットが大きい。データベースの構築に関係された方々の多大なるご努力に対し、深く敬意を表する次第である。

文献データベース単独での利用

Windows版のC D-Rは、タブ区切りのテキスト形式、DBPro形式、MicrosoftExcel形式、およびMicrosoftAccess形式とそれぞれの利用者により便利なファイル形式に保存されている。データベースを利用する場合、日常親しんでいるデータベースソフトで利用した方が自分なりのデータの加工や画面およ

び出力レポートのレイアウトなどのカスタマイズがしやすい。このように多様な形式にデータが保存されていれば、他のデータベースソフトで利用しようとする場合、データの変換が容易で便利である。

検索は、著者名、タイトル、分類項目などで検索するが多い。検索結果は、テーブルとして表現されればとりあえず活用できるが、さらに印刷用のレポートを設計しておくと同回もの検索結果の印刷書式が統一されてより便利である。他の分野の文献データベースにおいて、Microsoft Access 上で検索画面や印刷フォームをのせたアプリケーションが整備されているものも存在するが、利用者側でのカスタマイズができないセキュリティ設定になっていると、使い勝手が悪くかえって利用しにくい場合もある。文献データベースのようにカスタマイズ可能な Access データとして配布していただく方が利用者としてはありがたい。

以前のデータベースは、著者名の一部の姓と名とが半角スペースで区切られているために、検索時に戸惑うことがあったが、CD-R 版では姓と名が区切り無しで連続しているために検索が容易になっている。

検索結果の一部を引用文献などのためにテキストデータとしてワープロソフト上で利用する場合には、たとえば検索結果の各項目値（文字型）を用いて「著者（年号）タイトル. 誌名, 巻号, 頁数.」となるような新たな項目をもつテーブルを利用者が使用しているデータベースソフトで作成してからテキストファイル化する。利用してみると連名の著者の区切り文字で「;」と「/」とが混在している点に気づく。しかしこのような小さな問題は、文章作成段階で容易に修正されてしまうものであることからほとんど気にならない。

個々のレコードにおいては、補充や修正が必要な部分が見つかる場合もある。しかしこのような膨大なデータを限られた条件下で完全な形で構築するのはきわめて困難である。多少の不備があっても最新情報が多く含まれているデータをできるだけ早い時期に公開していただく方が利用者としては有用である。データの補充や修正に関しては、統一した入力フォームを配布し、該当個所についての情報を広く会員や関係機関から得るようにしたらいかがであろうか。

異なるデータベースへの文献データベースの利用

リレーショナルデータベース（RDB）では、同じ項目を持ったテーブルを結合することが可能である。RDBではテーブル中のレコードを特定できる項目（キー）が必要となる。キーとしてよく用いられるのがオートナンバーあるいはカウンタと呼ばれる整数値である。

今回のCD-R 版 Access のファイルには項目「ID」にオートナンバーが用いられていて、「ID」は主キー項目として設定されている。ただ Excel やテキストファイルの文献データには項目「ID」は存在していないことから、Access における項目「ID」は、おそらく Excel あるいはテキストファイルから Access にデータをインポートした際、ソフト上の操作によって自動的に付加された主キー項目である可能性が考えられる。このようなキー項目「ID」を利用して他のテーブルと文献データベースとをリンクさせることは容易である。他のテーブルにID値を入力できる項目を作成し、各レコードに対応し

た文献データのID値を入力すればよい。しかし文献データベースをリンクさせた他のデータベースを長期間利用する場合には、今後も更新されるであろう文献データベースの「ID」が従来の「ID」を踏襲していくことが条件である。「ID」を統一的に管理することは、データベース作成者に新たな負担が増大する可能性があることから、筆者としてはあまり望まない。

利用者としては、文献データベースを外部データとして、検索によって得られた文献データをレコード単位で自分で構築したデータベースに取り込んで活用する方法が良いのではではないかと考えている。

おわりに

今後とも利用価値の高い文化財科学文献データベースが継続的に構築されていくことを願います。本稿を草するにあたり、西田泰民氏には情報を提供していただいた。記して感謝いたします。