

民博コスチュームデータベース (MCD) の過去・現在・未来
—集中型システム (nihuONE) の試用を通して—

中川 隆
国立民族学博物館

高橋晴子
大阪樟蔭女子大学

及川昭文
総合研究大学院大学

国立民族学博物館では、1984年から、身装—身体と装い—に関連する情報を集めた民博コスチュームデータベース (略称 MCD) を構築してきている。この MCD を、人間文化研究機構の研究資源共有化事業において構築された集中型システム (略称 nihuONE) へ移行する試みを行った。

ここでは、MCD の概要、MCD を検索するために用いてきたシステムの変遷、nihuONE の概要、MCD の移行作業、nihuONE の試用結果、及び、その移行過程で行った、検索システムの機能とデータ項目との関連についての検討、MCD の構成の見直し、データベース作成のあり方についての考察結果を報告する。

The Past, Present, and Future of the Minpaku Costume Database
- Through a trial use of the nihuONE -

Takashi Nakagawa
National Museum of
Ethnology

Haruko Takahashi
Osaka-Shoin Women's
University

Akihumi Oikawa
The Graduate University
for Advanced Studies

In 1984 the National Museum of Ethnology started constructing the Minpaku Costume Database (MCD), which stores the information of clothing culture in the world. We made attempt to transfer this MCD to the nihuONE that is a centralized system of the National Institutes for the Humanities. This paper presents the outline of MCD, transition of the system used to search MCD, the outline of nihuONE, and trial results of nihuONE. It also makes reference to the operation results of this transition process, such as review of MCD format, as well as the relationship between the function of search system and the data item.

1. はじめに

国立民族学博物館 (以下、民博) では大阪樟蔭女子大学衣料情報室 (2001年4月からは、高橋研究室) と協力して、1984年から身装関連分野の情報をデータベース化するための研究を行ってきており、その成果は、民博コスチュームデータベース (MCD) として Web 上で公開している。[1]

一方、人間文化研究機構では、所属する5機関が持つ多様な研究資源を共有化し、先進的な研究・教育を進めるため、研究資源共有化事業を展開しており、その一つとして集中型システム (nihuONE) が構築されつつある。[2]

今回、研究資源共有化事業の一環として、MCD の nihuONE への移行を試行した。その主な目的は「nihuONE の評価」ということであったが、その過程で「MCD の再検討」や「データベース作成のあり方」などについても考察を行った。

表1 MCDを構成するデータベース
(2007年10月現在)

DB-id	データベース名	データ件数	画像件数
MCD1	服装関連日本語雑誌記事:カレント	73,723件	
MCD2	服装関連外国語民族誌	1,651件	
MCD3	服装関連日本語図書	26,720件	
MCD4	服装関連日本語雑誌記事:戦前編	8,896件	
MCD5	衣履標本データベース	10,776件	39,679件
MCD6	アクセサリー・身装文化デジタルアーカイブ	5,293件	22,865件
MCD7	服装関連外国語雑誌記事	3,193件	

2. MCDの概要

MCDは、表1「MCDを構成するデータベース」に示すように、7本のデータベースからなっており、総データ件数は189,603件に及んでいる。このうち、MCD1, MCD2, MCD3, MCD4とMCD7の5本は文献データベースであり、MCD5と6は画像データベースである。

これらの検索項目は、例えばMCD1については、表2「MCD1の検索項目」の通りであり、MCD5や6は、地域・民族名、現地名の他、形態、素材、構造技術などからなっている。

表2 MCD1の検索項目一覧

MCD1の検索項目			
文献番号		時代コード	
著者名	発行年月日	時代通称名	時代通称名ヨミ
論文名	論文名ヨミ	キーワード	キーワードヨミ
掲載誌名	掲載誌名ヨミ	服装専門分類コード	服装専門分類用語
OWC	地域・民族名	身装概念コード	身装概念用語
特定地域名	特定地域名ヨミ	抄録	抄録ヨミ

これらのデータベースには、書誌的な事項に関する項目の他に、地域・民族名(OWC)、時代コード、フリーキーワード、服装専門分類コード、身装概念コード(統制検索語)など、文献の内容を分析したデータも収録している。

分析データのひとつである「OWC」とは、民族学で使われている地域・民族分類で、エール大学で開発されたHuman Relations Area Filesにおいて作成されたものである。MCDではこれを簡略化して用いており、例えば、朝鮮半島は「AA」、日本は「AB」というように一部の地域を除いてアルファベット2文字で表している。

「服装専門分類コード」は、ファセット構造を持つ独自の分類で、横軸に生産から消費への流れを、縦軸に物としての個々の衣服あるいは関連品を挙げた形の表として定義され、英文字2桁と数字1桁の組合せで表されている。

「身装概念コード」は、近い内容を持った言葉をグルーピングし、それに服装専門分類コードの英文字2桁と、数字2桁から5桁を組み合わせたものを付したものである。

「時代コード」はMCD独自の表記法で、例えば、20世紀前半は「A20F」、後半は「A20S」というように表している。

現状では、雑誌記事のカレント(MCD1)と戦前編(MCD4)のように、かなり似通った項目を持つデータ群であっても、それぞれを独立させたデータベースとして扱っている。その理由は、それぞれのデータ群の性格を重視したことのほかに、検索システムの制約によることが大きかった。これについては次項で述べる。

3. 検索システムの変遷

MCDを利用するために用いてきた検索システムを古いものから順に挙げると、次のとおりである。

1) STAIRS (STorage and Information Retrieval System) 日本アイ・ビー・エム社の製品

1979年に導入されたもので、汎用計算機上で稼動するオンライン会話型の情報検索システムである。

当時のシステムでは、一般的には、漢字データを扱うことができなかったためカスタマイズを行い、漢字データをJISコードで格納し、数台の特殊な端末に表示できるようにしていた。検索は、漢字データからは行えず、英語のように分かち書きされた1バイト文字列のテキストから単語単位で行うというのが基本であった。従って検索の対象となる項目にはすべて、半角カタカナで分かち書きしたデータを付加しておく必要があった。汎用計算機の端末で漢字が使用できるものが導入されたのは、1980年代になってからである。

2) JAIRS (JAPANESE INFORMATION RETRIEVAL SYSTEM) 日本アイ・ビー・エム社の製品

1991年に導入したもので、1)と同様に汎用計算機上で稼動する製品であるが、漢字データでの検索が可能になった。

1993年10月には、ゲートウェイを追加し、インターネット経由、又は電話回線に接続した300~9600bpsのモデムを経由して、パソコンのVT100 端末エミュレータを使い、民博外からもMCDが利用できるようになった。

3) MMIR (Minpaku Multimedia Information Retrieval)

1997年に民博が策定した仕様に基づいて開発したシステムで、UNIX系のサーバ上で稼動するものである。全文検索エンジンとRDBMSを組み合わせ、C++でプログラミングしたCGIアプリケーションである。MMIR内での横断検索機能や、検索結果の一覧や詳細表示時に表示する項目を指定したり、表示順序・長さ等もユーザが指定できる多機能なシステムである。一方、民博の写真資料と標本資料のデータを強く意識して設計されていて、各レコードに付けられる画像は1枚だけ又は4の倍数枚でなければならないという制約がある。

1999年には、画像用のサーバを導入し、大量の画像を公開できるようになった。

4) MCD 検索システム

1997年末から民博内で独自に開発を始め、次に述べるMARSに移行するまで使用したUNIX上で稼動するCGIアプリケーションである。3)のMMIRでは制約があった任意枚数の画像表示や、検索条件を入力する画面のカスタマイズなどの要望に応えて開発されたものであるが、一方、画像の管理方式をMCDに特化しすぎたためにメンテナンスが困難になってきた。

5) MARS (Multi Archives Retrieval System)

3)のMMIRの後継となるもので、2003年からレンタルを始めたシステム向けに、民博の定めた仕様に基づいて作られた汎用の検索システムである。JSPとjavaによって記述したもので、RDBMSやXML関連の標準的な技術を用いたシンプルなシステムとして構成されていた。しかし、その後、様々な拡張の要求が出され、データベースごとに検索条件の入力画面を用意できるよ

うな機能(例えば、条件をプルダウンメニューで指定できるようにすること)や結果一覧をソートする機能などを追加していき、次第に複雑なものになっていった。(図1,2)

また、Z39.50のサーバ機能及びゲートウェイ機能を持っており、データベースの横断検索も可能なシステムである。

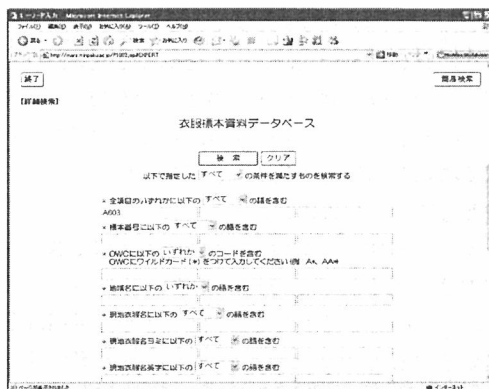


図1 MARSの検索条件入力画面の例



図2 MARSでの詳細表示画面の例

4)の検索システムを除いて、汎用的に使用できるシステムであり、MCD以外の民博の他のデータベースにも使用してきている。なお、これら以

外に、プロトタイプとして作成したものがいくつかあるが、ここでは省略する。[3]

4. システムの仕様とデータ作成との関連

前述のように 1)の STAIRS では検索が行えるのは、1 バイト文字で分かち書きされた項目のみという制約があり、検索項目には、例えば、抄録にもすべて、半角カタカナで分かち書きしたデータを用意しておかなければならなかった。従って、データの作成には多くの手間と人手を要した。また、このころでは、データの入力にパンチカードが使われていた時代であり、専門のパンチャーに依頼してデータ入力を行っていた。

また、1)や 2)の時代の汎用機の端末では、カタカナと英小文字とが同一画面上に表示できないという制約があったため、同じような項目から構成される日本語文献と外国語文献のデータベースであっても、それぞれ別のデータベースとして作成せざるを得なかった。

次に、画面のレイアウトや検索時の振る舞いについてふれると、まず、データベースごとに設定できる事項は、1)や 2)のシステムでは、項目名や表示順序をデータベース設計時に指定できる程度であった。また、検索の操作は、専用のコマンドまたはコマンドが割り当てられたファンクションキーによって行う方式であった。システムには USER EXIT 等と呼ばれる拡張のための入り口がいくつか用意されており、多少の拡張は可能であったが、アセンブラ言語等の知識が必要で、カスタマイズは容易ではなく、必要最小限のものを導入時に業者が行うという程度であった。

カスタマイズの制約から、利用記録のログ情報も不十分であり、利用者の需要動向を知るためには、データベースを分割しなければならなかったケースもあった。例えば、利用者がどの年代のデータをよく参照したかを知るためには、同じ雑誌記事のデータであっても、カレント(MCD1)と戦前編(MCD4)のように分割しておく必要があった。

3)のシステム以降では Web ブラウザをクライアントとして使用するものとなり、画面上に機能

を表す名称を付けたボタン等を表示しておき、それをクリックする形で進めていくようになった。一方、3)や 4)を導入する頃には、適当な機能・性能を持った製品としての検索システムが見当たらず、仕様を策定し特注するという形になり、仕様の策定に手間取るようになっていった。また、この頃から、3)では C 言語、5)では JSP についてのある程度の知識があれば、導入後に民博内でもカスタマイズが行えるようになった。

5. nihuONE の概要

nihuONE は、Web ブラウザ上での簡単な操作により、データベースの定義から項目の設定、データのアップロード、Web 公開までが行えるシステムで、コンピュータの専門家でなくてもデータベースを作ることができるように配慮されたものである。



図3 nihuONE の項目定義画面

データベースの作成を支援するための機能として、データのチェック機能やテーブル変換機能がある。テーブル変換機能とは、予めコードとその値の入ったテーブルを用意しておけば、アップロード時にコードに対応する値に展開し、検索や検索結果の表示の対象にできるというものである。また、コードにミスがないかチェックしてくれる機能 (RDBMS での参照整合性) も持っており、別途にチェック用のプログラムを作る必要がない。

セルのVBAを用いて数ステップの簡単なプログラムを作り、エクセル上でhtmlに変換しておくことにした。

アップロードについては、今回は、エクセルから一度に出力するCSVファイルのレコード数を、1万件以下になるようにし、数回に分けて行った。アップロードの開始からデータのチェックと検索エンジンへのロードが完了するまでの処理時間は、1万件当たり16分程である。アップロードが完了すれば直ちに利用できる。

7. データベース分析

nihuONEの特長の一つである統計機能を使用して見た結果を紹介する。

1) 頻度統計の例1

例えば、「現在の研究者が、近代日本における身装分野のうち、さらには、こういった主題に関心があるのか」を知ろうとする場合、服装関連日本語雑誌記事:カレント(MCD1)を、地域:民族が日本(OVC=AB)という条件で検索すると57693件ヒットする。その内訳として時代コードの頻度を見ると、20世紀後半に関する研究が58%を占めているものの、近代にあたる19世紀後半および20世紀前半に関する研究は、頻度統計の6位、7位に位置していることがわかる。そして、この文献件数は、1562件にのぼる。さらに、この1562件について、統制語にあたる身装概念で頻度統計をとると、「紡績業、製糸業」、「服装史」、「西欧化」や、「風俗」の主題についての関心が1位~5位までを占めることが把握できる。

No.	身装概念	記事数	記事件数	累積件数
1	紡績業、製糸業	142	3.68	3.68
2	服装史	131	3.40	7.08
3	風俗	120	3.27	10.35
4	西洋服装	90	2.33	12.78
5	洋装	89	2.31	15.09
6	社会服装	78	2.05	17.14
7	服装材料	68	1.87	19.01
8	洋装、洋装雑誌、洋装雑誌、ファッション誌	47	1.22	20.23
9	洋装、洋装雑誌、ファッション雑誌	42	1.09	21.32
10	洋装雑誌	39	1.01	22.33
11	婦人服装	33	0.86	23.19
12	ファッション雑誌	34	0.89	24.08
13	洋装雑誌	32	0.83	24.91
14	ファッション雑誌	32	0.83	25.74
15	洋装雑誌	31	0.81	26.55
16	洋装雑誌	31	0.81	27.36
17	洋装雑誌	31	0.81	28.17
18	洋装雑誌	31	0.81	28.98
19	洋装雑誌	31	0.81	29.79
20	洋装雑誌	31	0.81	30.60

図6 nihuONEによる頻度統計の例

2) 頻度統計の例2

エクセル等の表計算ソフトと頻度統計の機能とを組み合わせれば、もう少し複雑な集計も可能である。簡単な例として、例えば上で使ったMCD1に20世紀後半という条件を付けてヒットしたものの頻度統計を取って、それをダウンロードしたものを、OpenOfficeのCalcにあるデータパイロット(エクセルのピボットテーブルでも同様)を使って、クロス集計してみると表3「クロス集計の例」のような結果が得られる。

表3 クロス集計の例

身装概念コードの2桁目	身装概念コードの1桁目						合計結果
	A	B	C	D	E	F	
P	611	5221	10265	13161	4359	9504	43211
Q	3198		51	141	9145	1	12536
R				4	2	34	40
S				21	12	31	64
T	321				6	39	366
U	3617						3617
V	561	46		8	615	175	1405
W	2639		4	4	49	185	2881
X				1328		153	1481
Y		306	8	31	1978		2323
合計結果	10947	5573	10328	14698	16166	10212	67924

さらに、これを発行年の項目で絞り込んで、発行年代ごとに同様の処理を行うことによって(発行年代ごとの集計表は割愛するが)、20世紀後半は、全体的には、(Q-素材)の(E-製作・製造業)に関する研究、並びに、すべてのアイテムを含む(P-総記)の(D-デザイン)に関する研究が多い。しかし、各年代のクロス集計をみると、80年代と90年代には、(P-総記)の(C-ビジネス・物流)および(F-小売)に関する研究が増加し、とくに90年代には、(F-小売)に関する記事が、80年代の約3倍となっている。この結果から、研究の対象が、川上の問題から川中・川下の問題に移行していることが理解できる。衣服の生産地が国外に移り、国内ではマーチャンダイジング等の流通に力を注いでいる日本の繊維・ファッション産業のあり方が、そのまま研究に反映されている結果となっていることがわかる。

3) KWIC機能の使用例

次にKWIC機能について触れておくと、身装関連分野では、第二次世界大戦後、「服装」「服飾」「衣服」「被服」の概念が混乱している状態が続

イズにも設定ファイルで指定できる範囲（設定ファイルにも単純な設定しかできないものから、テンプレートや XSLT のようにプログラミングの要素を持ったものまでいろいろであろうか）のものから、ソースコードを修正するものまでいろいろであるが、いずれにしろ、レイアウトや画面遷移に凝ろうとするとかなりのコスト（時間と経費）がかかることになる。

また、MCD については、データ項目の見直しや、データ内容についても見直しを検討する時期に来ていると考えられる。

まず、項目についてはデータ作成の省力化の点からも見直す必要がある。検索システムの更新時には、移行作業に追われ、とりあえず移行することのみ関心が向く傾向にあり、項目の見直しは十分に行えているとはいえない。例えば、旧システムでは必要だった分かち書きや読み仮名データが、現在のシステムにも引き継がれている、といったことである。この点はデータ作成の手間と利便性との兼ね合いから再検討する必要がある。

また、nihuONE の横断検索機能を有効に使うには、検索項目名をそろえておくことも必要となる。

さらに、MCD を単なる検索の対象から分析の対象へと発展させるためには、コードで入力している項目の内容も見直す必要がある。例えば、「発行年」の項目は、年と月が入力されている形が基本であるが、他に月が複数入っていたり、年と月の組が複数入っていたりするものもある。このため、頻度統計の機能を使って、発行年と身装概念コードのように 2 項目を掛け合わせた集計を取ろうとすると、組合せの数が多くなりすぎて処理できない、とのメッセージが出る。これは「年」だけを独立させた項目を設ければ防げる。即ち、検索のためのコードから、分析のためのコードに変えていく必要がある。

データベース全体の構成面では、データ群の分割は、検索時のレスポンスや、操作の手間といった面から見直したい。高速な検索システムでもいくつものデータベースを対象に処理すれば、遅く

なるのは明らかであるから、まとめられるものはまとめた方がよい。利用者の需要動向の調査を取りたいために分割しているデータベースについても、検索条件の入力画面で、どの範囲から検索するのかが指定でき、利用記録にその情報が残るような機能があれば、分割する必要はなくなる。

民博では、2007 年度末に計算機システムの更新が行われ検索システムも新しくなる。MCD にとっては、当面は、新システムへの移行や現在開発中の身装年表データベースとの連携などが課題となるが、nihuONE についても並行して稼働させていき、利用記録等を参考に今後の方針を検討していくことになろう。

参考文献

- [1] <http://www.minpaku.ac.jp/>
なお、現在、MCD の研究およびデータ作成は、MCD プロジェクトが行っている。
- [2] <http://nihuone.nihu.jp/>
- [3] 八村広三郎、高橋晴子他 “民博コスチュームデータベース” 情報処理学会: 人文科学とコンピュータ研究会、1993.11.26
- [4] 高橋晴子 “民博コスチュームデータベース [MCD] の現状および今後の展望”、情報管理、35-8、pp.665-674 (1992)
- [5] 木暮 仁 “システム開発におけるユーザーニーズは絶対か？”等、システム部門 Q&A <http://www.atmarkit.co.jp> (2003-)