

みんなくりポジトリ

国立民族学博物館学術情報リポジトリ National Museum of Ethnology

属性情報を用いた標本検索

メタデータ	言語: ja 出版者: 公開日: 2015-11-19 キーワード (Ja): キーワード (En): 作成者: 黒川, 雅人 メールアドレス: 所属:
URL	http://hdl.handle.net/10502/3632

属性情報を用いた標本検索

黒川 雅人*

要旨

本稿では、国立民族学博物館の民族学研究用画像検索システムにおける文字・数値情報（属性情報）のデータベース化について、その設計概念と実際の構築手法について述べる。ここで属性情報とは、国立民族学博物館所有の情報カードに記載された収蔵標本に関する文字情報、各標本に対して付加された文化項目分類コード、標本の大きさ・重量等を示す寸法情報、及び計算機上に蓄積された標本画像の管理情報を意味する。これらの意味・内容・形式の異なる多種のデータを統合して扱うためのデータベース管理システムとして関係型モデルを用いた。

1 情報源

本共同研究で構築した民族学研究用画像検索システム *CIRES*¹⁾ で管理される属性情報源は標本資料の情報カード²⁾、HRAF の文化項目分類 (OCM) コード並びに地域・民族分類 (OWC) コード³⁾、標本画像寸法情報⁴⁾、画像管理情報の4種類であり、これらデータを一括して管理・検索の対象としようとするものである。この内、寸法情報と画像管理情報は、国立民族学博物館（民博）が所有する「標本画像自動処理装置」で得られる画像データに関するものである。本装置では、非接触で計測された標本の幅・奥行き・高さ・重量の計測値が与えられる。画像管理情報とは、画像（ここでは「カタログ画像」と呼んでいるパーソナルコンピューター PS/55 上で概覧用に表示される縮小カラー画像と、IBM5080 グラフィックス表示装置上で表示される高画質画像等）を実際に記録している媒体上（それぞれ磁気ディスクと光ディスク）でのアドレス（磁気ディスクでは画像ファイルのレコード位置、長さ等、光ディスクでは記録面、開始セクター、長さ等）をいう。

* 日本アイ・ビー・エム(株)東京基礎研究所

1) 本書「共同研究の概略——方法と成果——」参照。

2) 本書資料編A「情報カード」及び本書「標本資料の整理と活用——コンピューターの利用——」参照。

3) 本書資料編B「HRAF/文化項目分類 (OCM) コード」及びC「HRAF/地域・民族分類 (OWC) コード」参照。

4) 本書資料編D「標本画像自動処理装置」参照。

2 関係型データベース上での表現方式

上記の情報を実際に計算機上で関係型データベースを用いて管理する場合、それらをどのような関係にマッピングするかが問題となる。我々の初期のプロトタイプでは、実体-関連モデル [DATE 1983] を基礎として情報カードの持つ属性を6つの関係に分割したモデルを示した⁵⁾。これは、情報カードの各属性の持つ意味内容を検討し、共通の性質を持つと思われるものをまとめあげていった結果による。

最終的には、基本的な関係として3つの情報源すなわち情報カード、画像管理情報(寸法情報を含む)、OCMコードを3つの関係(データベース)に分割して別々に管理することとした(表1)。これは、各データの情報源としての性質から、データ入力の手続きがこれら3種類で別々に行なわれ、その結果、レコードとしての入力、更新、削除の操作がこれら3つの単位で行なわれることを考慮して、最善と判断した

表1 管理する情報 [1] サイド:光ディスクの記録面]

表名	情報カード	画像管理		OCMコード
属性 1	標本番号	標本番号		標本番号
2	標本名	カタログ	ポイント	OCMコード
3	訳名		サイズ	
4	標本の種類	カラー鳥瞰1024	サイド ¹⁾	
5	OWCコード		ポイント	
6	製法・材料		サイズ	
7	使用地	カラー鳥瞰512	サイド	
8	使用民族		ポイント	
9	使用年代		サイズ	
10	使用者	カラー鳥瞰256	サイド	
11	現地名		ポイント	
12	製作年代		サイズ	
13	製作状況	濃淡平面	サイド	
14	製作地		ポイント	
15	製作者		サイズ	
16	用途・使用法	濃淡側面	サイド	
17	変遷・分布		ポイント	
18	文献		サイズ	
19	添付資料	濃淡正面	サイド	
20	その他		ポイント	
21	原収集者1		サイズ	
22	原収集者2	幅		
23	収集年月日	奥行		
24	収集地	高さ		
25	入手状況	重量		

5) 本書「共同研究の概略——方法と成果——」図2, 表1参照。

めである。これらの3つの関係は、標本番号を主キーとして関連付けられており、各標本を一意に指し示すようになっている。

実際のシステムでは、上記の3種類のデータはホスト計算機上の関係型データベース (SQL/DS) として与えられる。SQL/DS の特徴として、実際に扱うとすることができるデータ型は次の4種類である [IBM 1984] :

1. 文字型 (最大200バイトの文字列)
2. 文章型 (最大4000バイトの文字列で、検索には使えない)
3. 数値型 (4バイトの符号付き整数)
4. 画像型 (最大1024×1024画素のカラーと濃淡画像)

これらのデータの入力、更新、削除等は利用者との会話的入力にはよらず、基本的にはデータベース管理者がホスト計算機上の入力プログラムを用いて一括して行なう。

3 検索系の実現方式

3.1 検索系の構成

検索系は、マイクロメインフレーム結合の方式で構成されている。利用者は、ホスト計算機に接続された自分のパーソナルコンピュータから検索要求を出し、ホスト計算機で検索が行なわれ、検索結果がパーソナルコンピュータに送られて表示される。この方法を用いることによって、パーソナルコンピュータの優れた対話性を生かし、使いやすいユーザーインターフェースを提供できる。特に利用者が計算機をあまり知らない場合、これは大きな利点となる。また一方では、ホスト計算機の大容量性と高速性を生かし大量の属性情報と画像データを一元的に管理できる。

実際の検索処理では、特に利用者が計算機の専門家でないことを前提としていることから、次の2点に留意した:

1. SQL/DS 上での検索文である SELECT 文を意識しなくとも検索をかけられること。
2. 利用者が実際にデータベースを操作していることを自覚しながら操作ができるように直接的操作ができること。

システムはホスト計算機として汎用の大型計算機である IBM3081 (民博では IBM3090)、パーソナルコンピュータとして IBM5560、マイクロメインフレーム結

合用のソフトウェアとして日本語 PC-VM Bond を用いている。

3.2 検索時の動作

実際のシステムにおける検索は、カードタイプのインターフェースによって行なわれる⁶⁾。利用者は「検索カード」と呼ぶカードメニュー上に並ぶ各検索項目に対し個別に値を指定する。入力条件としては各検索項目に対して、文字属性のものには検索したい語及びそれらの論理演算、数値属性のものには比較演算子及びそれらの論理演算が記入できる。利用者は、記入した検索カードをマウスでドラッグし、それを画面中の資料室アイコン上に持っていくことで検索処理を起動する。入力されたこれらの条件に対してパーソナルコンピュータ上で SQL/DS の SELECT 文への変換が行なわれ、文字属性に関しては基本的に部分文字列一致の LIKE 述語節に、数値属性に関しては対応する演算子に変換される。また各検索項目間の条件はすべて AND で結ばれる。作成された SELECT 文は、PC-VM Bond を介してホスト計算機にファイル転送され、ホスト計算機で SQL/DS の検索が行われる。検索結果としては、最初に、条件に合致したレコードの数が返される。これが利用者の希望する値と大きく異なる場合、利用者は検索カードになんらかの修正を加えることによって検索を繰り返して行なう。検索結果の数値が利用者の満足できる値になった場合、該当する標本の画像データと属性データをパーソナルコンピュータ上にファイル転送し、以後のユーザーインターフェース処理に渡され、表示される⁷⁾。この際、検索条件は検索の記録として検索カード上に残され、その後確認や再利用することができる。

4 おわりに

民族学研究用画像検索システムの属性データ管理について述べた。本システムでは、マイクロメインフレーム結合を用いることによって、ホスト計算機の持つ大容量性及び高速性を生かした大規模データ管理を行ない、またパーソナルコンピュータの持つ優れた対話性を生かしたユーザーインターフェースを提供している。この両者により、国立民族学博物館の標本資料に関わるデータ群について、それらを論理的に一元管理している点が本システムの特徴といえる。今後の課題としては、特に検索処理速度の向上が挙げられ、より高速なマイクロメインフレーム通信方式の適用、及び

6) 本書「民族学研究用画像検索システムの視覚化ユーザーインターフェース」写真9参照。

7) 本書「民族学研究用画像検索システムの視覚化ユーザーインターフェース」参照。

黒川 属性情報を用いた標本検索

SQL/DS の検索速度向上のための索引テーブルの作成が望まれる。

文 献

DATE, C.J.

1983 *An Introduction to Database Systems*. Addison-Wesley.

IBM Corp.

1984 *SQL/DS System Application Programming for VM/System Product*. SH24-5068.