

## 図書目録カード画像検索システムの改善 - 扱いやすく柔軟なインタフェースへの移行 -

喜田 拓也<sup>†</sup> 南 俊朗<sup>‡§</sup>

**あらまし** 多くの図書館において所蔵文献の検索には OPAC ( Online Public Access Catalog ) が用いられている。しかしながら、歴史の長い図書館では古い資料の書誌データが電子化されないままであり、そのような資料は OPAC で検索することができない。すなわち、利用者は図書目録カードが入った引出しの前まで出向き、カードを一枚一枚めぐりながら目的の資料を探さなければならない。このような状況を打開する手段として、図書目録カードのすべてを画像データ化し、Web 上でカードを検索できるシステムを我々は提案してきた。本稿では、新たに改善された本システムの概要について述べる。

**キーワード** 図書目録カード検索, イメージデータ, 電子図書館, 大量画像データベース

## Improvement of Image-based Catalog Card Search System - More Friendly and Flexible Interface -

Takuya KIDA<sup>†</sup> and Toshiro MINAMI<sup>‡§</sup>

**Abstract** Recently many libraries provide their patrons with OPAC (Online Public Access Catalog) system for retrieving their materials. However, in libraries with a long history for example, it often happens that a patron can not find very old materials by the system due to lack of data. Then one must visit the library, stand in front of the catalog card boxes, and turn the cards one by one. In order to rescue them from such troublesome jobs, we have developed a system, where all library catalog cards are scanned, saved as image database, and they can be retrieved through a web-browser. In this paper, we will describe the outline of the newly improved system.

**Keyword** Catalog card search, Image data, Digital library, Very large image database

### 1. はじめに

#### 1.1. システム構築の背景

図書館業務の電算化により、現在では多くの図書館で OPAC ( Online Public Access Catalog ) システムによる資料の検索が行えるようになった。しかし、歴史の長い図書館などでは、すべての所蔵資料が OPAC で検索できるわけではない。OPAC 導入以前の古い資料についてはその書誌情報が電子化されておらず、未だ目録カードを頼りに資料を探さなければならない。九州大学附属図書館 ( 以下、九大図書館 ) の場合、340 万件を超える蔵書のうち OPAC で検索可能な資料は約 240 万件であり、残る 100 万件は未入力である ( 2005 年 6 月 17 日現在 )。これは特別プロジェクトによる数年間

<sup>†</sup>北海道大学大学院情報科学研究科 〒060-0814 札幌市北区北 14 条西 9 丁目  
Graduate School of Information Science and Technology, Hokkaido University, Kita-14 Nishi-9, Kita-ku, Sapporo-shi,  
060-0814 Japan

<sup>‡</sup>九州情報大学経営情報学部経営情報学科 〒818-0117 福岡県太宰府市宰府 6-3-1  
Department of Management and Information, Kyushu Institute of Information Sciences, 6-3-1 Saifu, Dazaifu-shi,  
Fukuoka, 818-0117 Japan

<sup>§</sup>九州大学附属図書館研究開発室 〒812-8581 福岡市東区箱崎 6-10-1  
Research and Development Section of Kyushu University Library, 6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581  
Japan

E-mail: †kida@ist.hokudai.ac.jp, ‡minami@lib.kyushu-u.ac.jp

に渡る集中的な遡及入力の結果である。従来のペースでの遡及入力では、入力件数は蔵書件数の半分にも満たなかったであろう。

書誌情報の電子化を行うために、これまでも国立情報学研究所を中心とした全国図書館のネットワークを利用した遡及入力が進められてきた。しかし、この作業には多くの人手と費用そして年月が必要である。しかも入力には専門的知識が必要となることもあり、単純な人月計算では見積ることができない。たとえば九大図書館の例でいうと、1件当りの遡及入力コストを500円として計算した場合、残りの約100万件の資料データの入力に5億円もの費用がかかることになる。また、これらの内訳は製本雑誌・和装本・漢籍・特殊言語資料等となっており、特に製本資料以外のものは既入力分以上の手間がかかる。なぜなら、それらの入力には専門知識が必要とされ、図書館職員の更なる研修が不可欠だからである。このような状況から、現在でも入力完了時期の目処が立っていない。

このような状況を改善するために、九州大学附属図書館研究開発室では、図書目録カードをイメージ化し、ウェブを通じてカードを検索できるシステムを開発・公開している。このアプローチにより、短期間かつ低コストですべての書誌情報を電子的に検索可能な状態にすることができた。実際、高速イメージスキャナを用いて、1台につき1日1万件以上を処理することができ、かかる費用も1枚約10円程度であった。

もちろん、イメージ化したデータは遡及入力されたテキストデータとは異なり、そのままでは検索することができない。しかしながら、イメージ化したことにより、以下のようなメリットが生じる。

1. ウェブ上で目録カードを検索・閲覧できるようになる。よって、利用者は図書館に足を運ぶことなく目録カードを探すことができる。
2. また、このことから、元の目録カードを倉庫に移すことができる。
3. さらに、複数の人が同時に同じ引出しを閲覧できる。このことは利用者にとっての利便性向上だけではなく、遡及入力作業自体の効率化にも役立つ。遡及入力作業のための利用者のカード検索の制限が不要であり、しかも、複数人での入力作業が可能となる。

以下、2節では旧システムの使用法および課題点について述べ、3節でそれらを改善した新システムの紹介を行う。

## 2. 旧システム

### 2.1. インタフェースおよび使用方法

本節では、以前に開発した目録カード画像検索システム[1,2]についてその概要を説明する。本システム構築における当初の目標は

1. 大量の画像を効率よく閲覧できること、
2. 学外からのアクセスでもストレス無くブラウジングできること、
3. 従来の図書カード利用者にとって理解しやすいインタフェースをもち、かつ容易に検索ができること、

であった。図1~4に旧システムの画面を示す。

図1は旧システムのトップ画面であるが、目録カードのイメージ化が済んで検索サービスの対象となっている分野の名前（「総合目録」「文学部」「教育学部」「理学部」の4つ）が並んでいる。それぞれについて、和書目録、洋書目録（文学部の場合はさらにロシア語目録）へのリンクがあり、これらを選ぶと図2に示すようにその分類の引出し一覧が表示される。

図2の引出し一覧から見たい引出しをクリックすることで、引出し内部を俯瞰した図と引出し内の実際のカードのサムネイル画像10枚分が表示される（図3）。このビューを使って、順に引出し内のカードを閲覧できる。サムネイル画像をクリックすれば、別ウィンドウに大きく表示することもできる（図4）。

このように、実際のカードを検索する動作をできるだけ忠実に再現したインタフェースを実現している。



図 1 . 旧システムのトップ画面



図 3 . 引出し一覧画面



図 2 . 引出し内部とサムネイル表示

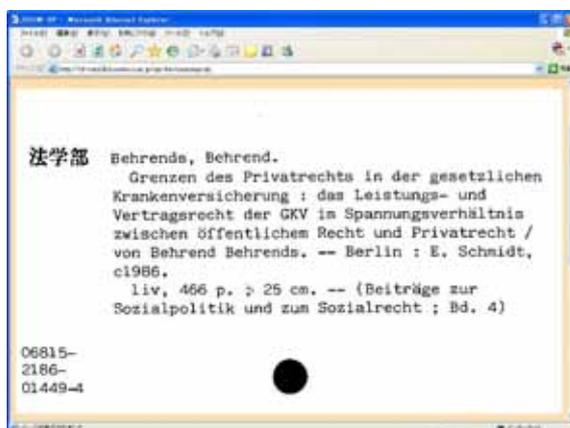


図 4 . カード画像拡大表示

## 2.2. カード画像の仕様

使用した高速イメージスキャナの制限により、オリジナルのカード画像は 2 値の TIFF 形式である。このままではウェブ上に表示できないため、同サイズの PNG 形式に変換して用いている。これら画像の解像度は、カードごとに多少のばらつきはあるが、幅約 1440 ピクセル×高さ約 850 ピクセルである。また、容量は PNG 画像 1 枚あたり約 1K~20K バイトである。字数が多いカードやノイズが多いカードほどデータ量が大きい。これとは別に幅と高さを約 3 分の 1 に縮小したサムネイル用の PNG 画像を生成して用いており、サムネイル表示画面（図 3）において画面全体のサイズが 150K バイト程度に収まるようにしている。これは ADSL 回線において実際に 1 秒～数秒以内で表示できるデータ量である。

各画像は仕切り名ごとに連番でネーミングされており、それらが一つのディレクトリに収まっている。仕切りに対応するディレクトリが集まって一つの引出しディレクトリに収まっており、さらにそれらが集まって一つの種別・分野に対応するディレクトリに収まっている。すなわち、一つの画像へのパスは

＞ ルートディレクトリ/(分野・種別)/(引出し)/(仕切り)/(画像ファイル名[8桁連番]).PNG

となっている。例えば総合目録の和書 A の引出し「ア」の仕切り名「アジア」の中にある 5 番目の画像だと、

＞ ルートディレクトリ/sougou/wasyo/A/001a/001ajia/00000005.PNG

というパスになる。

また、検索システムが動作しているサーバとは別のサーバにこれら画像データを保持しており、http 経由で通信を行っている。

## 2.3. 課題点

この旧システムは当初の目標を達成したが、いくつかの課題も浮上した。

まず、引出しの並びおよび引出し内部のカード仕切りの並びが、現物の並び方と異なっている場合があることが挙げられる。これは、引出し名・カード仕切り名をローマ字で表記したものを UNIX のディレクトリ名として各画像ファイルを管理しているため、元の並びが辞書順と異なる並びをしている場合に不整合が起きる。

第二には、引出し名や仕切り名をキーワード検索できないことである。旧システムでは、各仕切り名の情報を info.txt というファイルで管理しており、これが各引出しのディレクトリに分散している[3]。このため、一括したキーワード検索が困難となっていた。

また、トップ画面がテンプレート化されていないため、新たな分野のカード画像が追加されるたびに、HTML ドキュメントを変更する必要があった。

その他にも、操作インタフェース部分での使い勝手など、改善すべき点が見つかった。今回、これらの課題を克服すべく新たな検索システムの構築を行った。次節で、各改善点についての詳細を述べる。

## 3. 新しいインタフェース

### 3.1. システム内部での改善点

まず、前節 2.3 で挙げた第一と第二の課題を克服するために、分散していた info.txt 内の情報を分野・種別ごとにまとめて保持するように変更を行った。あくまで情報はテキストとして保持し、特別な DB は用いていない。RDB を用いなかった主な理由は、図書館職員が SQL などの特別な知識なしにデータを保守・変更できるようにする点が第一であり、第二に文字列照合による柔軟な検索を簡単に実現するためである。すべてのテキストは Unicode (UTF-8) でエンコードされており、Windows 標準のメモ帳で編集することができる。

新たな info.ini (図 5) は、”%%”を区切りとしたまとまり(チャンク)で一つの引出しを表す。また、各行はタブ区切りのデータになっている。チャンクの先頭行は特別に”#”で始まり、「チャンクが表す引出しのディレクトリ名」、「0」、「実際に表示する引出し名の文字列」が順に並ぶ。チャンクの 2 行目からは、その引出し内の仕切り情報が並んでいる。先頭から順に「仕切りのディレクトリ名」、「ディレクトリ内のカード画像の数」、「表示される仕切り名の文字列」となっている。3 列目の仕切り名を表す文字列はいくら長くてもよい。実際には、仕切り名と共にそれを補足する情報が記入されている。例えば、

```
002aka 786 赤：紅、赭、明石、縣、吾田、茜、…
```

といったように仕切り中のより詳細な情報が書かれる。

```
%%
#001a 0      ア
001a 1432   ア (アーアサ):
002ajia 502  アジア (アーケ):
%%
#002a 0      ア
001ajia 1040 アジア (ケイーケイザイタ):
002ajia 512  アジア (ケイザイチャーコ):
%%
#003a 0      ア
001ajia 420  アジア (サート):
002ajia 664  アジア (ナーノ):
003ajia 384  アジア (ハーレ):
004a 684    ア (アシーアト):
%%
#004a 0      ア
001a 896    ア (アナーアメ):
002amerika 608   アメリカ (アーケ):
003amerika 460   アメリカ (ケーサ):
```

図 5 . info.ini ファイル (総合和書 A)

```

%%
#ア
001a   ア
002a   ア
003a   ア
004a   ア
005a   ア

006a   ア
007a   ア
008a   ア
009a-ae アーアエ
010ao-akita   アオーアキタ

011akitsu-asashi   アキツアサシ
012asata-ashi   アサターアシ
013asu-atarashiiso   アスアタラシイソ

... (中略)...

066oki-oko   オキーオコ
067osa-owa   オサーオワ
068on   オン
%%
#カ
069ka   カ
070ka   カ
071ka   カ

```

図 6 . 引出し並び制御テキスト (総合和書 A)

```

#総合目録
和書 A   sougou_wasyo_A   sougou/wasyo/A
和書 B   sougou_wasyo_B   sougou/wasyo/B
洋書 A   sougou_yousho_A   sougou/yousho/A
洋書 B   sougou_yousho_B   sougou/yousho/B
%%
#文学部
和書     bungaku_wasyo     bungaku/wasyo
洋書     bungaku_yousho    bungaku/yousho
ロシア語  bungaku_russia    bungaku/russian
%%
#理学部
和書     rigaku_wasyo      rigaku/wasyo
洋書     rigaku_yousho     rigaku/yousho
%%
#医学部
和書     igaku_wasyo       igaku/wasyo
洋書     igaku_yousho      igaku/yousho
%%

```

図 7 . カテゴリ登録ファイル

図 2 のように、旧システムにおいては一つの分野・種別の引出しを一度に全部並べるしかなかったが、この並びの情報も別のテキストを用いて制御するように変更した (図 6)。引出し選択画面で一度に表示する引出しのまとまり (チャンク) を”%%”で区切り、横一列に並べるまとまりを空行で区切っている。チャンクの先頭行にある”#”に続く文字列はそのまとまりのタイトルを表しており、各行は引出しのディレクトリ名と表示される文字列がタブ区切りで並んでいる。このファイルによって、引出しのまとまりを自由に表現することができる。たとえば、実際の引出しの並びと同じように並べることも五十音順に並べることもできる。

各分野・種別のデータもテキストで管理できるように変更を行った (図 7)。”%%”で一つの分野のまとまりを表し、各行で種別ごとの情報 (種別名、種別 ID、対応する分野・種別ディレクトリパス) をタブ区切りで表している。種別 ID は、各 info.ini が格納されるディレクトリ名と一致している。それらは、本システムが置かれるディレクトリにある”box”ディレクトリ以下に配置される。これにより、階層の構造が異なっても同列に扱うことができる。

新システムでは、以上のようなテキストベースの設定ファイルから操作インタフェース画面を自動的に生成し表示している。またそれと同時に、検索用の索引データとしても活用している。一つの info.ini ファイルはせいぜい 200K バイト程度なので、文字列照合ベースの検索 (つまり Grep ツールと同じ方式) でも、十分高速な検索機能を実現できる。また本方式により、行を入れ替えることで辞書式順にとらわれない自由な並びでの表示が可能となる。図 8 ~ 11 に新システムの画面を示す。

以上のシステム内部の改善により、新システムでは画面の上部にある検索窓から自由なキーワードを用いて仕切り名の検索を行えるようになった。キーワードとしては、仕切り名に含まれるローマ字・カタカナ・漢字のいずれかを用いることができる。たとえば、「hana」もしくは「ハナ」を総合目録和書 A から検索すると

- ・ハツ／ハノ／ハナー／ハネ：花、鼻、華、話、咄、譚、嘶、英、離、白菖、埴、含羞、埴、羽
- がヒットする。「花」で検索した場合には、このほかに
- ・カ／花：
  - ・ケー／ケイ／華：[華]、花、化、稀、假、希有

・サターサマ／サテーサマ：却説、悟、里、柳巷、酈郷、花街、青樓、諭、早苗、猿投、眞田、讃岐、寛、鯖、後貝加爾、裁、鏜、寂、淋、様、彷徨

という仕切り名もヒットする。



図 8 . 新システムのトップ画面



図 10 . 新システムの引出し内部表示画面



図 9 . 新システムの引出し一覧画面



図 11 . kuliv による拡大画像表示

### 3.2. ブラウジング面での改善点

より快適なブラウジングを実現するために、いくつか細かなレイアウトの変更や技術の導入を行った。

一つは、図 10 における引出しの中身のナビゲーションにおいて、現在引出しのどの部分をサムネイル表示しているのかを赤いバーで示すよう改善したことである。ボタンによる移動だけでなく、このバーを移動させることでもサムネイルの表示位置を変更することができる。

加えて、旧システムでは引出し内部の仕切り名とカードの並び（図 3 の左側）が実際の引出しとは逆の方向、すなわち上から下へ並んでいたものを、引出しの見栄えと同様に下から上へ並ぶように変更した。

また、過去に開いた引出しの名前を画面左下にある「ボックス履歴」下に表示する機能を追加した。これによって、どの引出しをこれまで開いたのか、確認することができる。

カード画像の拡大表示には、九大図書館研究開発室が開発した高精細画像ビューア (kuliv) のスク립トの一部を流用している。この kuliv は、本図書館の「古医書画像データベース (<http://herakles.lib.kyushu-u.ac.jp/icomb/>)」や「竹田文庫 (<http://herakles.lib.kyushu-u.ac.jp/t-bunko/>)」における高精細画像の表示に用いられている。

#### 4. 今後の課題

今回の改善によってより柔軟で使いやすくなり、我々が本来目指していたシステムに近づくことができた。しかしながら、[2]で述べた以下のような目標が未だ実現できていない。

1. システムの管理を支援する機能：  
たとえば、蓄えられているイメージデータの中から不良なものを検出する機能や、ウェブ上でのデータ更新作業を行う機能の組み込み。
2. システムのパーソナル化：  
利用者ごとに、任意の引出しあるいは画像を集めて自由に分類したり、画像ごとにコメントを入力したりする機能の追加。
3. OPAC との連携：  
既に遡及入力済みのカードについては、OPAC へのリンクをつけるなど。

また、本システムを元に、様々な構造を持った画像データベースに対応可能な、より柔軟な画像検索・閲覧システムを開発することが考えられる。検索機能に関しても、含まれる文字列検索だけではなく、目的の仕切り名や目録カードの所在位置を推定し、より高速に、かつ容易に目的の文献情報が得られるような知的検索法の研究・実装も興味深い課題である。

システム適用事例に関しても次のような課題がある。現在旧システムを福井大学附属図書館に導入しているが、新システムの開発を機に同図書館システムのバージョンアップを図りたい。さらに、本システムを他の図書館に適用し利用者サービスの向上などに役立ててもらおうこと、また、そのような経験を踏まえ、本システムの適用範囲を広げ、完成度を上げていくことも今後の大きな課題である。

#### 5. さいごに

本稿では、古い資料のための書誌情報をイメージ化することで遡及入力することなく電子化する方法について述べ、それら大量の目録カード画像データをすばやく検索・閲覧するための新システムを紹介した。なお本システムは、<http://herakles.lib.kyushu-u.ac.jp/cciss/cgi-bin/frame.cgi> にて試験運用中である。

#### 謝辞

本システムを実装し、運用する上で九州大学附属図書館の皆様には多くのご支援をいただいています。特に、研究開発室、コンテンツ整備課、利用支援課の関係者の方々に感謝の意を表します。

#### 文 献

- [1] Toshiro Minami, Hidekazu Kurita and Setsuo Arikawa: Putting Old Data into New System: Web-based Catalog Card Image Searching, Proc. 2000 Kyoto International Conference on Digital Libraries (ICDL2000), pp.296-303, 2000.
- [2] 南俊朗, 栗田英和, 有川節夫: イメージによる図書目録カード検索システム - 遡及入力問題の一解決法 -, デジタル図書館 (ISSN1340-7287), No.18, pp.27-35, Sep. 2000.
- [3] 松川伸一, 南俊朗: 図書目録カードイメージ入力のボトルネック 大量データの正当性を検証する, デジタル図書館 (ISSN1340-7287), No.19, pp.5-18, Nov. 2000.