

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 —

南俊朗

九州大学附属図書館

〒 812-8581 福岡市東区箱崎 6-10-1

Tel&Fax: 092-642-4422, Email: minami@lib.kyushu-u.ac.jp

栗田英和

九州大学大学院システム情報科学府情報理学専攻 (現, (株) 東芝)

有川節夫

九州大学大学院システム情報科学研究院

概要

最近では、図書館所蔵の文献検索に OPAC (Online Public Access Catalog) を用いることが一般化している。しかし、歴史の長い図書館においては、OPAC 検索可能な蔵書は全体の数分の一にすぎない。ほとんどは、目録カードを用いて検索する必要がある。もちろん、これらの情報を機械可読化するための遡及入力が、全国図書館の協力の下進められて来たが、その完了までには、なお、相当の年月が必要である。本稿では、このような遡及入力のボトルネックを改善することを目的として我々が九大において研究・開発を進めている、図書目録カードのイメージを用いた検索システムに関して、その背景、意義、システムの概要、今後の課題等についての概略を示す。

キーワード

遡及入力, 目録カード検索, イメージデータ, 電子図書館

Image-based Catalog Card Search System — A Solution to Catalog Data Inputting Bottleneck —

Abstract

The OPAC(Online Public Access Catalog) system becomes very popular these days in order to search a book stored in the libraries. However most of the books are not accessible from OPAC. We have to use the old paper catalog cards most of the time. The effort for inputting these data into databases has been made under the cooperation of libraries all over in Japan. Despite of such an effort it is estimated that more than two decades would be needed to complete. In this paper, we will have an overview of a system which is has been researched and developed, with the intension to overcome such a bottleneck of inputting catalog data of books. We will briefly describe the background, importance, system organization and some problems to be solved.

Keywords

Catalog Data Inputting, Catalog Card Search, Image Data, Digital Library

1. はじめに

図書館に所蔵してある大量の蔵書類の中から探し求めている適切なものを効果的に見つけるために、目録カードは重要な役割を果たしてきた。図書館業務の電算化が進んだことにより、現在は、OPAC(Online Public Access Catalog)システムによる検索が一般的になった。しかし、全ての所蔵資料がOPACで検索可能な訳ではない。現在、全国の大学図書館には約2億件の蔵書があるともいわれている。九大附属図書館の場合、300万件を超える蔵書があるが、このうち、OPAC検索の対象となる電子化済みなのは、約70万件と、全体の1/4程度にすぎない。ほとんどの蔵書はいまだに紙製の図書カードを用いた検索に頼らざるを得ない状況にある。

蔵書情報の電子化を行うために、これまで、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、この作業は、かなりの人手と費用がかかる。このような状況を改善するために、九大では図書目録カードをイメージ入力し、ネットワークを通じて遠隔地からカード検索できるシステムを開発し、公開している。このようなアプローチを取ることにより、短期間に、しかも低コストで全ての目録カード情報を電子的に検索可能とすることができる。これは、図書館情報の電子化を図る上で極めて重要な点である。

本稿では、以下、第2節において、まず、図書目録カードをイメージ化し、短期間に検索可能とすることの意義を述べる。次に、第3節で、システムの概要と構成を示す。第4節では、今後の展開への方向性についての議論を通じて、本システムの方式の多用な可能性を探る。

2. 図書目録カードのイメージ化

前節で述べたように、現在行われている遡及入力を改善し、少ないコストで早急に電子化を進める手段として、我々は図書目録カードをイメージ化し、それに基づいた検索システムの研究を進めており、現在、そのテスト版が公開されている。本節では、その基本的なアイデアについて説明する。

書誌データの機械可読化のために、これまでも国立情報学研究所(旧学術情報センター)が中心となり、全国の図書館が協力して、データの遡及入力が進められてきた。そこでは、各図書館は、データの全体を入力するまえに、その書誌データが既に、いずれかの図書館によって入力されていないかをチェックする。もし、既に入力されているのならば、そのデータを流用することができる。もし、未入力であれば、最初から入力を行う。新たに入力されたデータは共有のデータベースに保管され、別の図書館の入力の際に参照される。

このような努力によっても、書誌データの入力には多大な期間とコストがかかる。遡及入力すべき目録カードの数は、全国の大学附属図書館全体で、3500万件余りと見積もられている。九大図書館の場合で、約161万件に上り、1年間の入力件数は実績6～7万件程度であることを考えると、入力の完了までには、なお、20年以上かかるペースである。これは、図書館サービスの電子化を推進するための大きな障害となる。

それに対して、目録カードのイメージ入力は極めて高速に行うことができる。高速のイメージスキャナを用いることで、1台につき、1日で、1万件以上を処理することができる。九大図書館において、文学部のデータ約38万件を、約2週間程度で処理することができた。

入力が高速に行われることの一番のメリットは、入力されたデータを用いることにより、早急に完全なデータベースを提供できることにある。次節で説明を行う、カードイメージ検索システムによって、ネットワークを利用した、遠隔地からの検索が可能である。また、同時に同じデータに対する検索を行うことが可能であることも大きな利点である。そのため、これまでは図書館に向いて紙のカードを検索する必要のあった蔵書検索の利便性が格段によくなる。

一方、必要な経費に関しても、イメージ入力は低コストである。1枚につき約10円程度で行うことができる。通常の遡及入力のためには、1枚につき約500～700円程度かかることを考えると極めて低コストであることが分かる。

もちろん、イメージ化したデータそのままでは、従来の遡及入力されたデータとは異なるものである。そのままでは、OPACシステムによるキーワード検索のためのデータとして用いることはできない。しかし、第4節で述べるように、イメージデータを利用することにより、OPACシステムとの補完的な検索が可能となり、また、遡及入力作業自体の効率化を図ることもできる。そのため、総合的な意味でもイメージ化のメリットは極めて大きい。

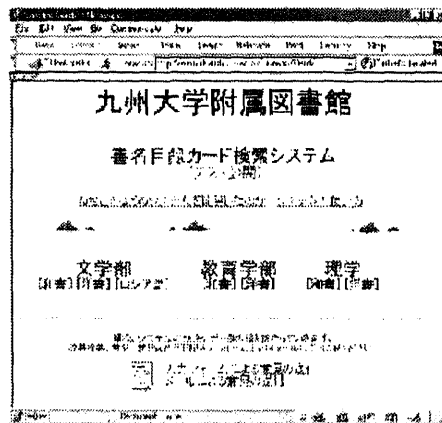


図1. システム最初の画面

3. カードイメージ検索システム

本節において、目録カードのイメージデータ検索システム [1][2] について、その概要を説明する。まず、第3.1節において、画面を示しながら、その操作法を示す。現実の図書カードを検索する際の操作法や目録カードの配置をネットワーク上に、できるだけ忠実に実現することにより、パソコン操作に十分なじんでいない利用者でも容易に検索が可能となるように配慮している。次に、第3.2節において、このようなシステムを実現するためのシステムの構成を説明する。カードのイメージデータを保存する場所と、イメージデータに関する様々な属性を分離して扱う方式を採用することにより、保守の安全性を高め、また、それらを別サーバー上で運用することを容易にするなどの工夫を行っている。

3.1 検索システムの操作法

図1に本システムのホームページの画面を示す。画面中央部分には、イメージデータ入力が済んでおり、従って、検索サービスの対象となっている学部名である「文学部」、「教育学部」、「理学」の名前が表示されている。また、その上には、このようなシステムを開発した背景や使い方の説明へのリンクがあり、下には、利用者からの意見を集めるための意見送付画面へのリンクが用意されている。残念ながら、他の多くのシステムの場合と同様に、利用者からの反応は十分ではなく、何か気づいた場合に、ホームページ以外の画面からも、いつでも簡単に、要望などが出せるようにするなど、今後なんらかの工夫が必要であると認識している。

それぞれの学部名に対して、和書目録、洋書目録、そして文学部の場合は、ロシア語目録へのリンクが用意されている。この分類項目は、実際の目録カードの分類に合わせて決められている。この例に見られるように、部局に属する図書館でごとに、独自の分類を行ったり、規則を決めていたりする場合は多いものと予想され、本稿のようなシステムを開発する際には注意が必要である。

学部名および、和書・洋書等の分類の中から希望するものを選びクリックすることで、図2に示されるような引き出し一覧が別ウインドウに表示される。



図2. カード引き出し一覧表示

図2の例は「理学」の「和書」に関する画面を示している。これら引き出しの並びも、図書館にある実際の目録カード引き出しの配列に準じたものとなっており、カード検索に慣れた利用者にとって、違和感が少ないよう配慮されている。ただし、使用上の違和感がそれほどはなかろうとの判断に基づき、縦方向は連続して表示されており、現実のカード配置と完全に同じ形態をとっている訳ではない。

引き出しに付随したラベルを参考に検索したい引き出しを選択する。なお、図2に示した、「理学・和書」の場合は、ラベルはローマ字表記されており、従って、分類もローマ字表記のアルファベット順により、なされている。この辺りも、学部ごとに決められている。例えば、教育学部の場合は、かな表記を採用し、あいうえお順に配置されているため、それぞれの事情に合わせた取り扱いが可能なようにシステム設計上配慮している。

図2に示されたウインドウにおいて、引き出し Kai-Kar を選択した場合に表示される画面例を、図3に示す。

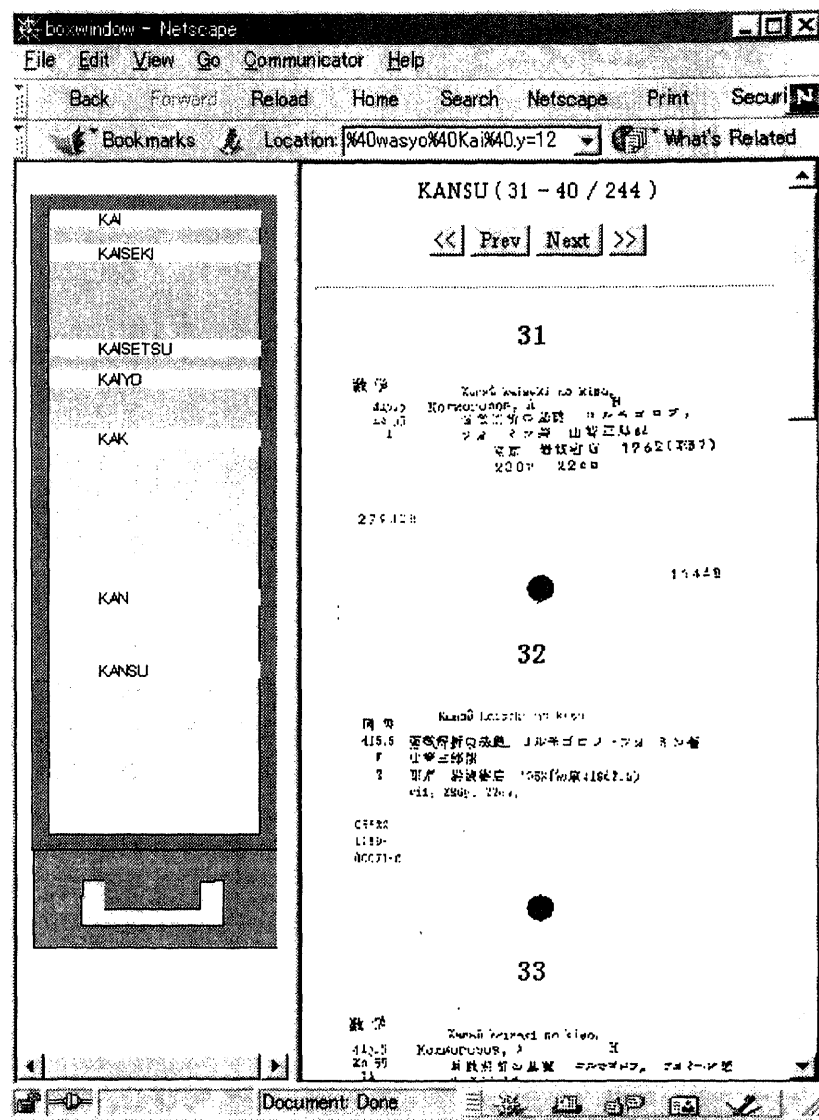


図3. 引き出し内部とカード表示

画面は、左右のフレームに分かれている。左のフレームには、引き出しを鳥瞰した図が、そして、右には引き出し内の実際のカードが10枚表示される。引き出し内の見出しカードが表示されているため、目的のカードの場所を絞ることができる。目的のカードのありそうな部分をクリックすると、右フレームは、その部分のカード10枚の表示に置き換わる。全体の中での場所の特定が容易となるよう、見出しカードの位置は、その見出しに属するカード枚数の割合に応じて決められている。実際に利用したところ目的のカードの位置を見つけるのは予想以上に容易であった。右フレーム内のボタンを押すことで前後のカードの表示に移ることもできる。また、カード部分をクリックすることにより、そのカードを単独で、別ウインドウに大きく表示することができ、小さな画面を用いている場合や、小さな文字を読み取りたい場合、便利である。このように、本システムは現実の目録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ーフェースを実現している。

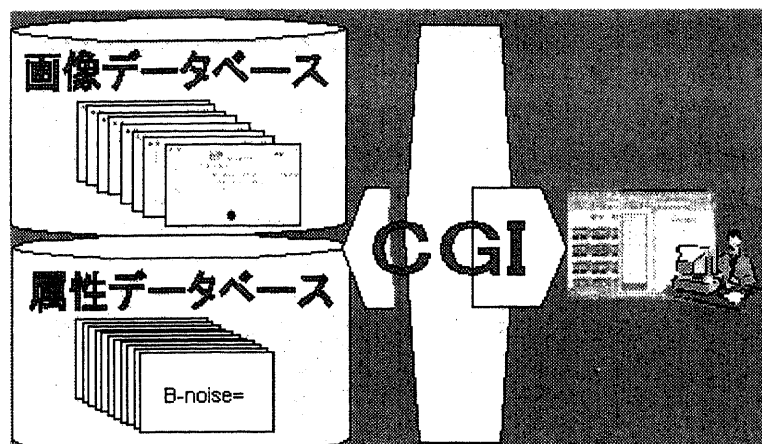


図4. イメージ検索システム構成

3.2 検索システムの構成

図4に、イメージ検索システムの構成を示す。右端にいる利用者はブラウザを用いて本システムにアクセスする。ホームページを含む少数の静的な文書はHTML文書として用意されているが、カード表示分を含むほとんどのページは、Perl言語を用いたCGIプログラムにより動的に生成される。CGIプログラムはユーザからのリクエストに応じて、必要な画像データを画像データベースから取得し、また、表示に必要なその他の情報を属性データベースから集める。システムは得られた結果をHTML形式で表現し、それをユーザの使用しているブラウザに返す。画像データベースと属性データベースを分けることにより、画像データベースを変更することなく、属性の追加や変更などを行うことができる。これは、属性更新等の処理の際、何らかの障害により画像データを破壊する危険性を少なくできる利点がある。実際、現在の運用においては、文学部の画像データは他のデータと異なるサーバーにより提供されている。

画像データならびに属性データは、ともに、〈学部名〉／〈分類名〉／〈引き出し名〉／〈仕切り名〉／〈個々のデータ〉という階層構造による木構造となっている。システムは利用者の選択に応じて、この木構造をたどり、目的の画像データおよび属性データを取得する。属性データは、属性名と属性値の組による一般的な構造を採用している。そのため、第4節で述べる今後の機能拡張によって、取り扱う属性の種類が増加しても、現在の属性名に新たな属性名を追加して記述し、また、プログラム側で、新たな属性に関する処理を追加することにより、対処可能である。

なお、現在システムで取り扱われているデータは、理学と教育学部がともに7万件弱、文学部が約3.8万件的計約5.5万件ほどである。

4. 今後の展開

本節では、まず、図5を用いて、本システムの今後の展開に関する見通しを述べる。我々の目指すシステム開発全体は大きく3つのステップから構成される。

(1) 基本的検索を実現するステップ

カードのイメージ化を行い、画像データベースをアクセスする検索インターフェースを通じて、利用者にサービスを提供するステップである。これまで説明してきたように、こは、すでに完成している。

(2) システムの管理を支援し、また、検索の高度化を図るステップ

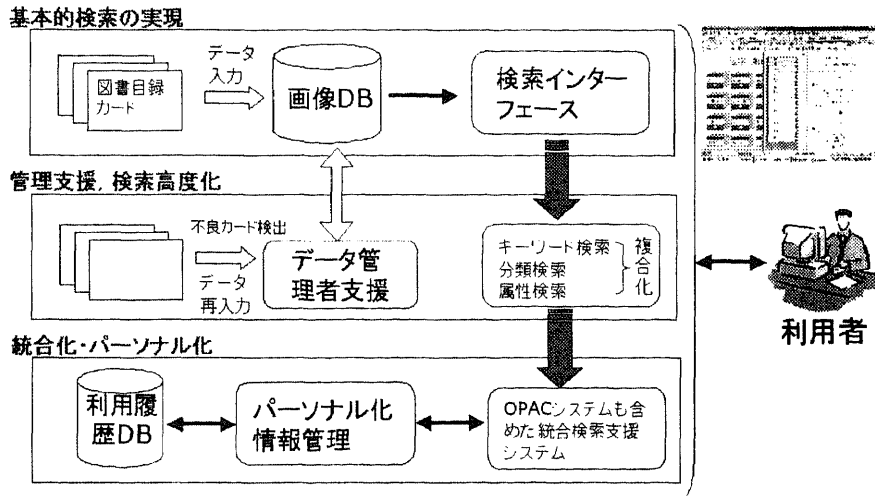


図5. 今後の展開の見通し

現在、このステップの研究を進めている。その1つは、イメージデータの管理支援機能を提供することである。たとえば、入力されたイメージデータの中から不良なものを検出する機能である。必要に応じて、データの再入力のための情報を提示する機能も求められる。また、前ステップに実現された検索は、分類木を1つ1つたどって目的の書誌情報にたどり着くという、いわば、分類検索と呼ばれるものである。次の段階として、カードの配置情報や、図形的特徴などの属性情報を利用した検索、更には、カードの内容を、部分的であるにしる、文字情報化し、それをを用いたキーワード検索といった、様々な検索スタイルを複合的に用いた検索の実現を目指している。

(3) サービスのパーソナル化を進め、それを含めた統合的システムへと発展させるステップ

更に、その先のステップとして、利用者個人の状況にあった検索の実現を目指している。たとえば、建築学の分野では「ラーメン」という用語は建物の構造を指すために用いられており、食べ物の「ラーメン」とは別物である。もし、利用者が建築学の分野で仕事をしていることが分かれば、「ラーメン」という用語に対して、建築関係の文書を検索結果リストの初めの方に提示するなり、強調して表示するなりすることによって、利用者の利便性を高めることができる。このようなパーソナル化機能を実現し、また、OPACシステムとの連携機能や強化されたサービス管理機能等を統合化したシステムの実現がこのステップにおける目標となる。

我々の研究の特徴的な部分は、イメージデータによる分類検索を出発点に特徴検索やキーワード検索をシームレスに統合化した検索を目指しているところにある。文字情報としてコード化された情報による検索においては、手がかりとなるものは文字情報だけである。そのため、検索方法として、自然とキーワード検索に限定されがちである。一方、我々人間の検索の場合、単なる文字情報のみではなく、形や色、配置といった画像等に関するアナログ的もしくは直感的な「情報」をも合わせた検索を行っている。そのような性格を持った我々の感性に近い検索を実現するためには、イメージ化された情報の利用は欠かせないものと考えられる。そのような認識に基づき、我々は現在の分類検索からキーワード検索までをシームレスに統合化することを目指しているわけである。

目録カードをイメージデータ化することにより、遡及入力そのものを効率化することも可能である。ネットワークを通じて、カードのデータにアクセスできることより、現実の目録カードを1枚1枚と手に取る

ことなく遡及入力が可能となる。機械可読化されたデータ間の対応を管理するのは容易であるため、たとえば、カードのイメージとその文字コード化された書誌情報を並べて表示し、任意の部分、および任意の順序で情報を入力することができる。この方法を利用すると、キーワード検索に有用なタイトルと著者名の部分を先に遡及入力し、配架情報や分類情報などの通常の遡及入力においては必要な情報も含めたそれ以外の情報の入力は後に回すことができる。そうすることにより、キーワード検索に必要な情報を従来の遡及入力の場合の数分の1から数10分1程度の期間で入力することが可能となる。検索結果はイメージとして表示されるため、蔵書へのアクセスも容易である。その後、その他の情報を入力し、最終的に遡及入力として必要な情報を全て入力することにすれば、早急なキーワード検索の実現と完全な書誌情報データベースの完成の両者を負担の増加なく実現可能である。

これまで述べてきたように、イメージ化による電子図書館化を更に発展させることにより、ややもすると忘れられがちな、過去から受け継いできた文化遺産である蔵書類を見捨てることなく、新しい時代に合った情報として有効に利用できる手段が開発されるものと考えている。

5. まとめ

本稿では、はじめに九大図書館をはじめ多くの図書館において、大量の蔵書目録情報が依然紙の形態のみ存在していること、そして、それらの情報を機械可読化するための遡及入力の努力が行われているにも関わらず、今後なお多くの労力と費用が必要なことを指摘した。

その上で、このような状況を打破するための1つの方法として、目録カードをイメージデータ化し、それを用いたカード検索を実現したことを述べた。本方法によると、普通の遡及入力と比較して高速かつ、低コストで機械的検索が実現できることを示した。

現在実現されているシステムによって、利用者になじみのある図書目録カードの検索と同様な方法での蔵書検索を、研究室等の遠隔地から可能である。また、現実の目録カードを模した配列となっているため、その検索に慣れた利用者にとって、場所に関する直感が働きやすい。

さらに、イメージデータを用いることで、遡及入力作業そのものを効率化し、またOPAC等によるキーワード検索とシームレスにつなげることが可能となるなど様々なメリットがあることを述べた。

我々のシステムと同様に、目録カードをイメージデータ化し、利用している図書館がいくつか存在する。Virginia図書館[3]においては、図書館収集の資料に関して、その目録カードをイメージ化している。一部の資料については、手書き原稿等をそのままイメージ化したカードの元資料へリンクされている。このように、カードイメージの利用法としては示唆に富むものの、本システムのイメージカードは1枚単位のみであり、近傍のカードをざっとみるには適していない。Princeton大学図書館[4]においては、1980年以前のデータに関してカードのイメージ化を行っている。それ以降のものはOPACで検索するようにと切り分けておらず、遡及入力なしで、新旧のカード検索を実現する方針を採用したものと見受けられる。これもカードイメージの利用法として、注目に値するシステムである。慶応義塾大学図書館[5]においては、中国・朝鮮・アラビア・ロシア語資料に対する検索システムとして、読みのアルファベット表記による検索システムと連携し、検索された内容をそのカードイメージにより詳細に確認できる。この方式も、OPAC等のシステムから見捨てられがちな言語に関する検索システムとして興味深い。これらいずれの図書館も、我々が目指している総合的な書誌情報検索システムを実現するよりも、それぞれの問題意識に基づき、部分的にイメージによる取り扱いに適した用途に限ってイメージによる書誌情報検索システムを実現している。この点が、我々のシステムとの大きな相違点であり、また、我々の研究を進める上で参考になる部分である。

本稿で指摘した多くの利点を考慮すると、本システムのアプローチは、今後、図書館の電子化、デジタル化を進める上で考慮すべき重要な選択肢になるものと期待される。

なお、本検索システムは現在（2000年8月）、<http://www.i.kyushu-u.ac.jp/~minami/Card> において、仮運用を行っている。また、九州大学附属図書館のホームページ <http://www.lib.kyushu-u.ac.jp/> からリンクが張られている。

参考文献

- [1] 栗田英和：イメージデータ化された図書目録カードの検索システム，九州大学大学院システム情報科学研究科情報理学専攻修士論文，九州大学，2000。
- [2] Toshiro Minami, Hidekazu Kurita and Setsuo Arikawa: Putting Old Data into New System: Web-based Catalog Card Image Searching, 2000 Kyoto International Conference on Digital Libraries: Research and Practice, 2000. (投稿中)
- [3] The Library of Virginia: <http://image.vtls.com/collections/>
- [4] Princeton University Library: <http://imagecat1.princeton.edu/ECC/>
- [5] 慶応義塾大学図書館： <http://catalog.lib.keio.ac.jp/ckabooks/>