

# 小規模アーカイブズにおける電子管理システムの構築

山畑 倫志

(本論考は大学共同利用期間法人人間科文化研究機構国文学研究資料館が行った 2009 年度アーカイブズ・カレッジへ提出した終了論文をもとにしている)

## 1. はじめに

現在、小樽商科大学百年史編纂室では所蔵文書を電子的に管理することを試みている。種々検討の結果、現在は RDBMS を基盤にして Web アプリケーションとして使用することを目的としてシステムを開発している。本稿ではその結論に至った経緯を記し、小規模アーカイブズにおける所蔵資料管理の電子化について論ずる。

当編纂室ではシステム開発業者に発注する費用、およびメンテナンスに必要な恒常的な費用を捻出する見込みがしばらくはない。そのため、安価で情報技術に精通していない者でも用いることが可能なシステムが必要となる。似たような状況に置かれているアーカイブズ、文書館で資料の電子管理システムを模索している諸機関の参考となれば幸いである。

## 2. CSV と全文検索

小樽商科大学には文書館に相当する組織は存在せず、多くの大学と同様、大学史編纂室にその役割が求められている。しかし、現状では紙の目録も存在せず、一部が Excel を用いて CSV として入力されているに留まっている。しかし、Excel は本来データ管理に用いるアプリケーションではないため、入力業務と検索業務のそれぞれにおいて大きな支障をきたしていた。Excel による管理の問題点はいくつか挙げられる。まず、階層構造の表現が困難である。表計算ソフトでは原則的に一枚の表で様々な関連づけを行わなければならないため、全体として記述が冗長になり、取り扱いが難しくなる。また、重複や誤入力を避けることができず、またその発見も難しい。

Excel による管理はこのような問題があったが、まず外部利用者の検索の利便を重視して、そのまま Web 上での検索が可能となるシステムの模索を優先した。まず資料の一覧を CSV で記述する。具体的には 1 つのレコードにフォンド、サブフォンド、シリーズ、サブシリーズ、ファイル、アイテムの六つの階層それぞれの記述項目をすべて記した。そのため、一つのレコードのフィールドの数は 50 を超え、また繰り返しの情報も多いため大変冗長となってしまった。入力作業についても極端に長いレコードの入力のため、ずれの有無を確認する作業に手間がかかってしまう。たとえばあるアイテムを入力しようとした場合、その上部の階層であるファイルからフォンドまでを入力しなければならない。だがそれにより、擬似的にはあるが、資料群の各階層の構造を明確に記述することができた。

さて、資料を参照するためには検索することが必要であるが、CSV をそのまま検索するのはそのたびごとに表の初めから検索することになるため、効率がよくない。またヒットしたレコードの一覧も参照しがたい。そのままでは入力ができても参照業務が難しいため、まず無料の全文検索アプリケーションである Namazu の導入を試みた。Namazu を利用するためにはサーバー上で

CGIを動かさなければならぬが、大学のサーバーでは許可が下りなかったため、簡易なサーバーを編纂室で立ち上げ、それを利用した。Namazuはコンピュータ上のファイルを単位として検索している。Excelで入力した表では一つのファイルにすべての情報を集中させているので、そのままでは検索機能を活用できない。そこで、表を一行ごとに一枚のhtmlとして分割した。当時入力済みであった資料は5,000件ほどであったが、それをすべて一つ一つの資料情報として、カード型目録のように分割していった。日本語の解析プログラムとしてはKakasiを用い、それによって抽出された単語をインデックス登録させ、ある程度の処理速度を持つ検索システムをWeb上で公開することが可能となった。その検索例が図1である。

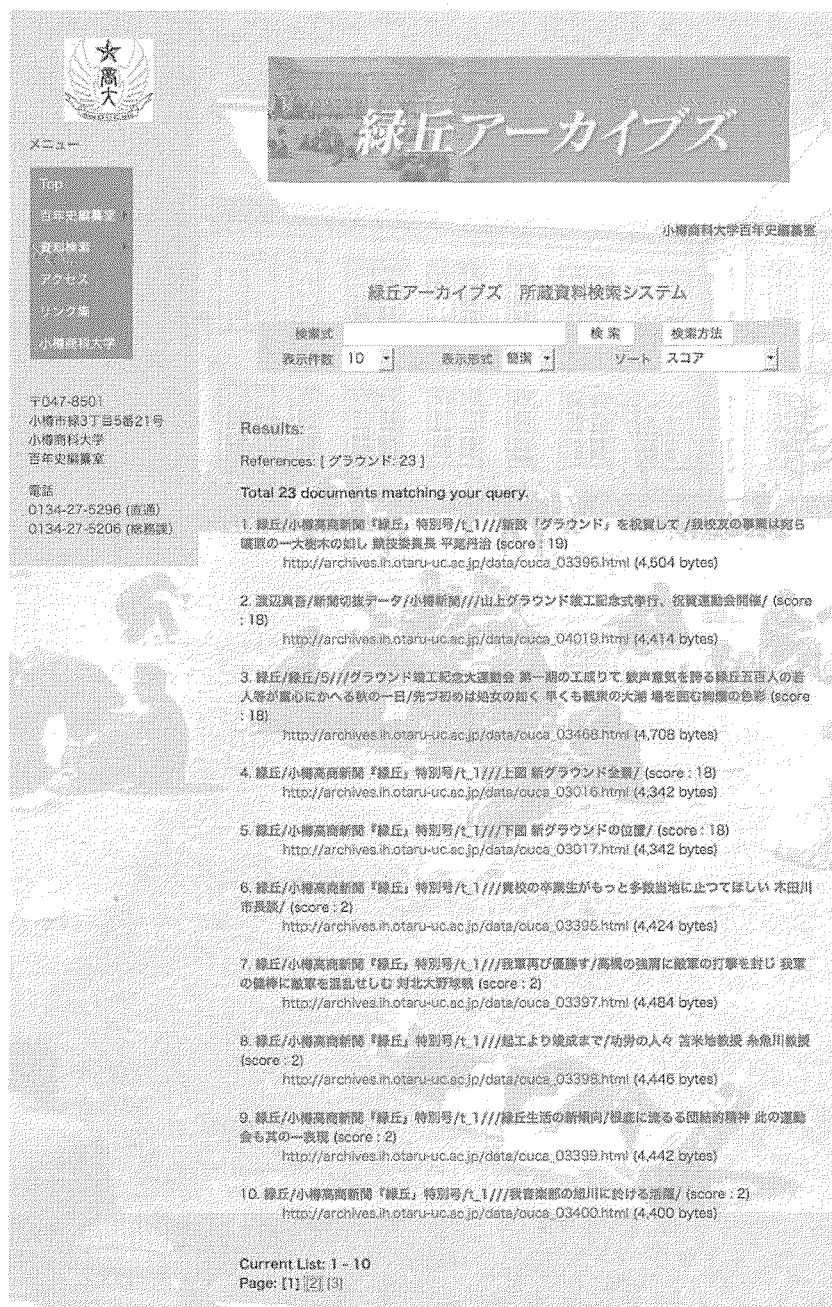


図1. Namazuによる検索システム

このシステムの構築によりExcelのみの処理に比べて目的の資料の情報を検索する手間が非常に軽減された。だが、これを将来に渡って使っていくには難しい要素が多々ある。まず、資料の更新に手間がかかり、情報の修正も容易ではない。また、検索機能を優先したため、入力業務自体はExcelで行わざるを得ず、Excelを資料情報の入力目的で用いる問題点はそのまま残ってしまっている。さらに検索結果画面から上部や下部の階層へのリンクを張りにくく、全体像が見えづらい。このような欠点、特に更新の難しさがあるため、当編纂室では電子化手段をリレーショナルデータベースの利用へと改めることとなった。

### 3. リレーショナル型データベースの利用

上に記したように、Namazu の運用のみでは入力業務に作業上のコストが多くかかり、参照業務についても柔軟性を著しく欠く。本来全文検索を目的としたアプリケーションを無理に用いた面があるため、新たなシステムの構築が必要とされた。そこで次に運用を決めたものがリレーショナル型データベースである。これは表計算ソフトのように一枚の表で管理するものではなく、複数の表を相互に関連づけて情報全体を表すシステムである。当編纂室では安価なシステムの構築を目指しているため、無料のリレーショナル型データベースの一つである MySQL を用いた。

リレーショナル型データベースの利点はいくつか挙げることができるが、表計算ソフトと比較すると次のようなものがある。まず、複数の表に分割することができるため、不要な重複を避けることができる。また、SQL 文を用いた柔軟な検索が可能である。さらに入力業務のためのアプリケーション製作が可能となる。Excel でもマクロを用いてアプリケーションを作ることはできるが、複数人で入力業務を行うことを想定すると、個々のコンピュータの環境を問わないシステムが必須であるため、サーバー側で処理される MySQL が望ましい。

このシステムの開発は 2008 年 6 月から始まり、2009 年 3 月に一定の機能の実装に目途がついたため、Web 上での一般公開に至った。データベースの設計は次のように行っている。まず fonds, subfonds, series, subseries, file, item の六つのテーブルを作り、それぞれに必要なフィールドを設定する。さらにそれぞれのテーブルのレコードに主キーとして連番を振り id フィールドとする、それを外部キーとして下位の階層と上位の階層を一对多の関係で関連づける。これにより、擬似的にはあるが木構造が表現されるものと考えた。

問題点はいくつかある。各テーブルの参照先を直近の階層とのみ関連づけているため、本来 series や file など中間の階層が必要ない場合でも仮の階層を設定しなければならない。現在は仮の階層の値をすべて 0 とし、id のみあるレコードとなっている。特に series, subseries, file といったテーブルにはこの仮の階層が多く、将来的にエラーのもとになることが想定される。これを解消するためにはフォンドごとにテーブルを作成し、またその下位分類としてサブフォンドのテーブルをそれぞれ作成し、といったことを item まで続ける必要が出てくるが、テーブルの数が膨大となり、管理も難しくなるため現実的ではない。いずれなんらかの形で解決が期待されるが、現状ではそのまま製作中である。

MySQL はコマンドラインで操作することが基本であるが、コンピュータ操作に習熟していない者にとっては困難であるため、入力、検索の利便に益するために扱いやすいアプリケーションを開発する必要がある。当編纂室ではデータの Web 公開も同時に行う予定のため、ブラウザ上で扱えるものが望ましい。そのため開発言語として CFML (ColdFusion Markup Language) を選択した。これは Adobe 社の製品であり、有償である。安価なシステム開発を求めている中で、あえて有償のものを使うには理由がある。同等の言語として PHP や Perl、JavaScript など多くの種類の言語が無料で利用可能である。しかし現在の編纂室のスタッフは情報技術のリテラシーがなく、その習得コストが過大であると考えた。そこで比較的習得しやすいとされる CFML を選択した。また当編纂室ではパンフレットや冊子の製作も頻繁に行うため Adobe 社の製品をまとめて購入している。その中のホームページ作成ソフトである DreamWeaver との連携がスムーズなのも理由の一つである。

小樽,大学が検索キーワードです。

以上2個のキーワードで検索しました。

フォンドは2件でした。

サブフォンドは2件でした。

シリーズは5件でした。

サブシリーズは13件でした。

ファイルは3件でした。

アイテムは170件でした。

合計で195件ヒットしました。

小樽商科大学

小樽商科大学ゼミナール協議会

小樽商科大学 > 小樽商科大学短期大学部

小樽商科大学ゼミナール協議会 > 小樽商科大学ゼミナール協議会

伝承文学研究会 > 伝承文学研究会 > 小樽商科大学

小樽商科大学ゼミナール協議会 > 小樽商科大学ゼミナール協議会 > 小樽商科大学ゼミナール協議会

小樽商科大学 > 緑丘アーカイブズ > 小樽商科大学史料展示室

小樽商科大学 > 小樽商科大学短期大学部 > 小樽商科大学短期大学部

小樽商科大学 > 総務課 > 小樽商科大学資金募集に関する綴

小樽商科大学 > 総務課 > 学校一覧 > 小樽商科大学一覧

小樽商科大学 > 総務課 > 学校一覧 > 小樽商科大学一覧

小樽商科大学 > 総務課 > 学校一覧 > 小樽商科大学一覧

小樽商科大学 > 総務課 > 学校一覧 > 小樽商科大学一覧

小樽商科大学 > 総務課 > 小樽商科大学資金募集に関する綴 > 小樽商科大学資金募集に関する綴

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学設置申請書

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学設置申請書

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学短期大学部設置申請書

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学夜間短期大学部設置期成会役員

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学設置申請書

小樽商科大学 > 学務課 > 学生寮 > 小樽商科大学学生寮(智明寮)取り壊しに関する経緯

小樽商科大学 > 総務課 > 大学設置 > 小樽商科大学短期大学部設置申請

小樽商科大学 > 総務課 > 皇室 > 昭和二十九年八月十九日 小樽商科大学短期大学部

小樽商科大学、総務課、学術団体との連絡、学術団体との連絡、122小樽商科大学公

図2. Coldfusion と MySQL による検索システム (検索結果画面)

よって当編纂室の資料管理システムはサーバーの OS として Linux (Debian)、http サーバーとして Apache、そしてデータベースとして MySQL、開発言語として CFML といった構成をとっている。現段階のシステムにおける検索結果画面は次の図 2 のようになっている。

fonds から item までのそれぞれのテーブルのタイトルフィールドを検索しマッチしたレコードをフォンドから順に並べて表示させている。また当該レコードの上位階層も表示し、整理の状況が判別できるようにしてある。

ただ、リレーショナル型データベースは本来、階層構造を示すことを想定されていない。また、柔軟な検索や入力といった点でも困難が認められる。その困難を越えるためには XML によって記述されたデータベースがもっとも望ましいとされており、当室でもそう考えている。しかし、XML データベースは無料のものが少なく、運用のノウハウも手軽に手に入るとは言い難い。また各種開発言語との連携についても現段階の技術では手に負えないため、導入を断念した。将来的に安価でマニュアル化されたモデルが登場することが望まれる。

id	169
タイトル	雑件綴
日付	
内容	
階層	小樽商科大学 > 総務課

雑件綴

雑件等綴自昭和39年3月23日 至昭和40年6月1日

雑件綴昭和40年度以降

雑件綴自昭和48年度 至昭和50年度

雑件綴昭和50年度

雑件綴1967

雑件1979 1981

図 3. Coldfusion と MySQL による検索システム (結果詳細画面)

#### 4. 画像閲覧アプリケーションの構築

さて、本年度 (2008 年度)、当編纂室では科研費で収蔵文書の電子画像化を行っている。これらは Web 上で外部の人間も利用できるように公開するためのものである。これも基本的には MySQL に画像のテーブルを用意し、それと目録情報のテーブルとを関連づけていく予定であるが、表示の点で困難が生じている。画像は見開きをカメラで写したものが一つの TIFF データとして納入される。しかし、資料は書籍や新聞の形態をとるものが多く、それを単に jpg に変換して表示するのでは、ページ間の移動が難しく、利用しやすいとは言い難い。pdf として結合して配布するのも方法の一つではあるが、利用者のほとんどはざっと確認することが目的と予想されるため、これも利便性に欠ける。CFML だけではその問題を解決することは難しいため、画像表示のために別のアプリケーションを作成することが必要となってくる。

ページの移動や指定が容易に行え、データベースとの連携も生かせるものとなると、ノウハウが比較的手に入りやすいものとして、JavaScript を用いた Ajax の技術がある。これによる画像閲覧のシステムはいくつか前例があり、実現可能であろうと考えている。現在は暫定的に図 4 のような形で Web 上での閲覧を可能にしている。

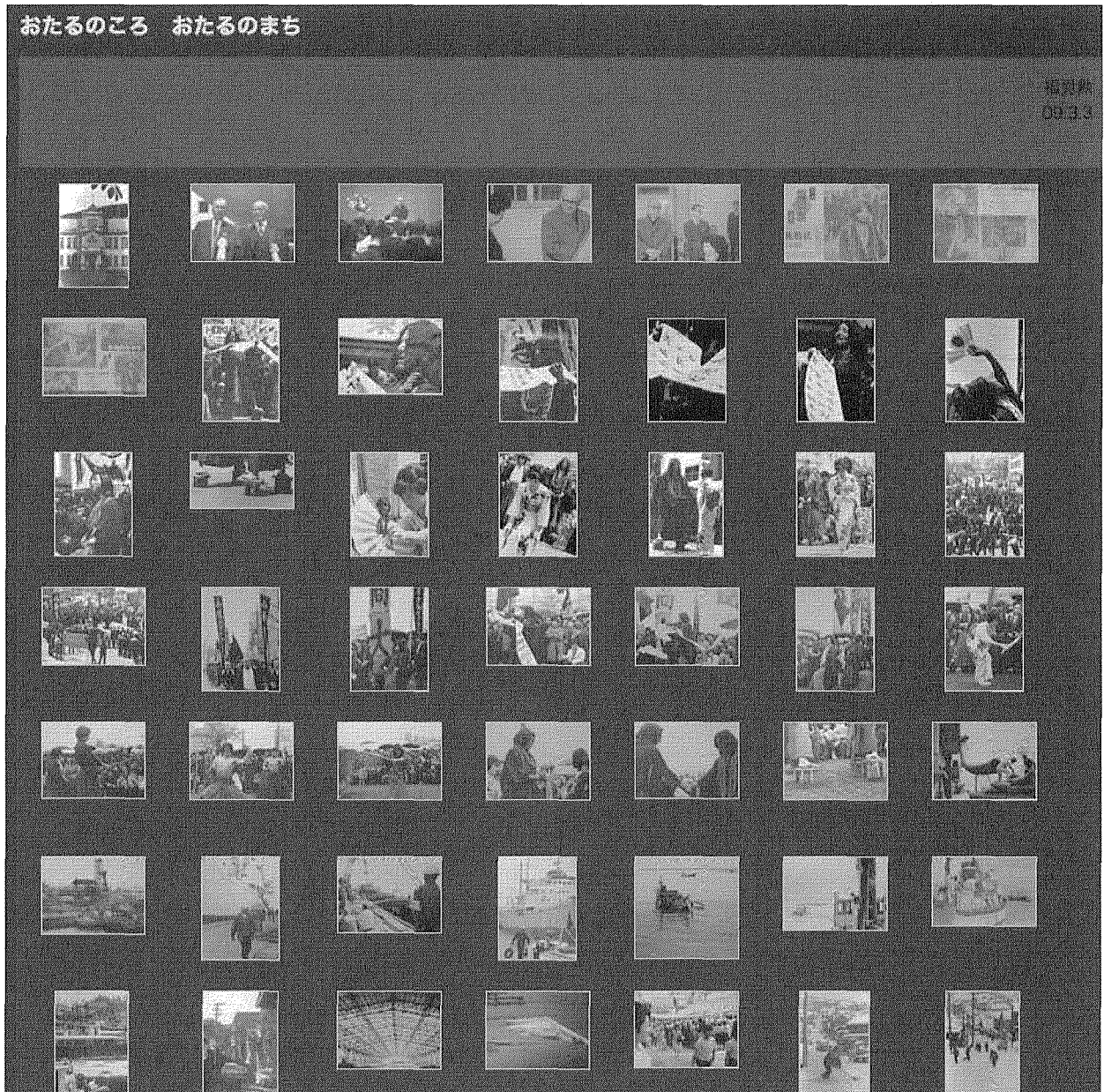


図4. 画像閲覧システム

#### 4. まとめ ——低コストでの管理システムの可能性

以上、当編纂室における資料管理システム構築の事例を述べてきた。作業開始時は2008年6月であるが、スタッフ二人が半年強の時間で模索してきた結果、今年度中には公開の運びとなる予定である。小規模なアーカイブズでは時間、予算、人員の点で複雑なシステムを構築することは難しいが、当編纂室が作成した規模のものであれば、情報技術専門のスタッフが常駐せずとも、構築可能であるとの感触を得た。

今後、このようなノウハウを各アーカイブズ間で共有し、規模の小さい部署でも、低コストで管理システムを有することができるようになれば資料情報の共有への敷居もいくらかは下がるものと思われる。