

『樟蔭學報』のデジタル化とデータベース試作の取り組み

— 樟蔭学園草創期資料のデータベース化とその活用 (4) —

竹内 さおり
白川 哲郎

1. はじめに

1917年(大正7)に樟蔭高等女學校として開学した本学園は、2007年に創立90周年を迎える。学園草創期の資料は、近現代の女性史及び学校教育史を考察するうえで貴重な資料であるとともに、本学大学史を講じるための重要な教材である。

平成15年度より継続的に進めてきた本学園草創期資料のデジタル化とデータベースの試作に関する研究は、作業手順をほぼ確立することができ、進捗状況は順調である。着手から3年目にあたる今年度は、保存と整理の作業そのものよりもデジタル化した資料やデータベースを教材として活用することにポイントをおいて、使い易さや楽しさをテーマに工夫をした。具体的には、1936年(昭和11)創刊の学園広報誌である『樟蔭學報』を対象として、デジタルコンテンツの利点を生かした教材の作成に取り組んだ。

本稿では、『樟蔭學報』のデジタル化とPDF (Portable Document Format)¹を用いたインタラクティブなデータベースの試作について報告する。以下、2節で研究の背景と目的を、3節でPDFの特徴と利点を述べる、4節で『樟蔭學報』の内容にふれ、5節で作業手順を説明し、具体例を示す。6節で関連研究を挙げ、最後にまとめと今後の予定を述べる。なお、PDF版『樟蔭學報』19冊分はCDに収録し、希望者に配布したいと考えている。



図1:『樟蔭學報』創刊号の表紙

2. 研究の背景と目的

本研究は、本学園の歴史的資料を授業で活用することで、学生の教育効果を期待するとともに母校への愛着を増加させたいという筆者等の考えからスタートした。平成16年度には、データベースソフト (Microsoft Access) を用いた『設立二関スル書類』のカード型データベースを試作し、紙資料のデジタル化とデータベース化を行った²。しかしながら、テスト運用の段階で以下のような指摘を受けた。

- (1) 1枚のカードで資料1ページ分の画像と各項目のテキストデータを見ることができると、

資料全体のイメージは把握しにくい。

(2) 文字や画像が小さすぎて見えにくいので、教材としては相応しくない。

(3) 画像をそのまま複製して転用される可能性があるので、データでの配布や Web で公開することは好ましくない。

これらの指摘を改善するために、データベースソフトを使用して作成することや画像データとテキストデータを並べて表示する型に固執せず、コンピュータの画面上でパラパラとページを繰るように閲覧する仕組みの実現とデータの配布や Web で公開することを目的に、新たな手法を検討した。

3. PDF の特徴と利点

PDF は、プラットフォームや使用環境に影響を受けず、利便性の高い文書フォーマットである。近年は Web 上で、データをやりとりすることが頻繁になり、PDF の利用頻度は非常に高くなっている。

文書ファイルだけでなく、画像ファイルを PDF へ変換した場合にも遜色がなく、ファイルサイズが小さくなることも注目できる。また、検索機能が利用できるのも魅力的である。ただ、ワープロソフトで作成したファイルを PDF に変換した場合は、文字をそのまま変換するので、文字列検索が可能であるが、画像ファイルの場合は、文字ではないため、既存の文字列検索機能は利用できない。

そこで、画像ファイルから生成した PDF を検索する方法として、注釈の機能を利用できないかと思いついた。つまり、撮影時に抽出したキーワードを注釈として付加し、検索に利用しようという考えである。さらに、複数の注釈を追加することが可能であるので、作成者がキーワードの付加に使うという使い方だけでなく、閲覧した利用者が新たなキーワードを追加したいと思えば、同様に注釈を作成して追加することができる。キーワードの追加に限らず、利用者がコメントを加えて、他の利用者に情報を提供することも可能である。つまり、知識共有に近い場を実現できると思う。例えば、図書館の本に書き込みをすることはできないが、デジタルコンテンツであれば書き込みは自由である。コミュニケーションや知識共有の側面から教育効果も期待できる。従来のデータの蓄積と検索という型に固執せず、動的な仕組み³の実現を望む。一方、画像の複製や転用が容易な点については、PDF のセキュリティ設定の変更により防止できることを確認している。

4. 『樟蔭学報』について

『樟蔭学報』は1939年（昭和11）に創刊された樟蔭学園広報誌であり、創刊号から第三巻第三号までの合計19冊が現存する。創刊号掲載の「樟蔭学報発刊の辭」に依れば、在学生二千名、樟蔭高等女學校本科・専攻科・高等科の卒業生同窓会員（緑蔭会員）三千数百名、樟蔭女子専門學校卒業生同窓会員（緑翠会員）一千数百名、それぞれの保護者を含めた合計一萬数千名の関係者に向けて、親交を温めるために発行されたとある。内容については、樟蔭高等女學校及び樟蔭

女子専門学校の年中行事の紹介，経過報告，緑蔭会及び緑翠会の会記や会員の消息，生徒の詩歌や文章の発表，緑蔭会員及び緑翠会員の作品発表，保護者の声と学校側の反響，論説，文苑等となっている。創刊号の表紙（図1）はカラー印刷で，蓮池のほとりで高女生と女専生が並ぶ明るい色調の画が目を惹く。行事や校内で撮影された写真も多く掲載されていて，服装や行事の様子などもよくわかる。当時の様子がうかがえる広告などもたくさん含まれ，非常に興味深い資料であると思われる。

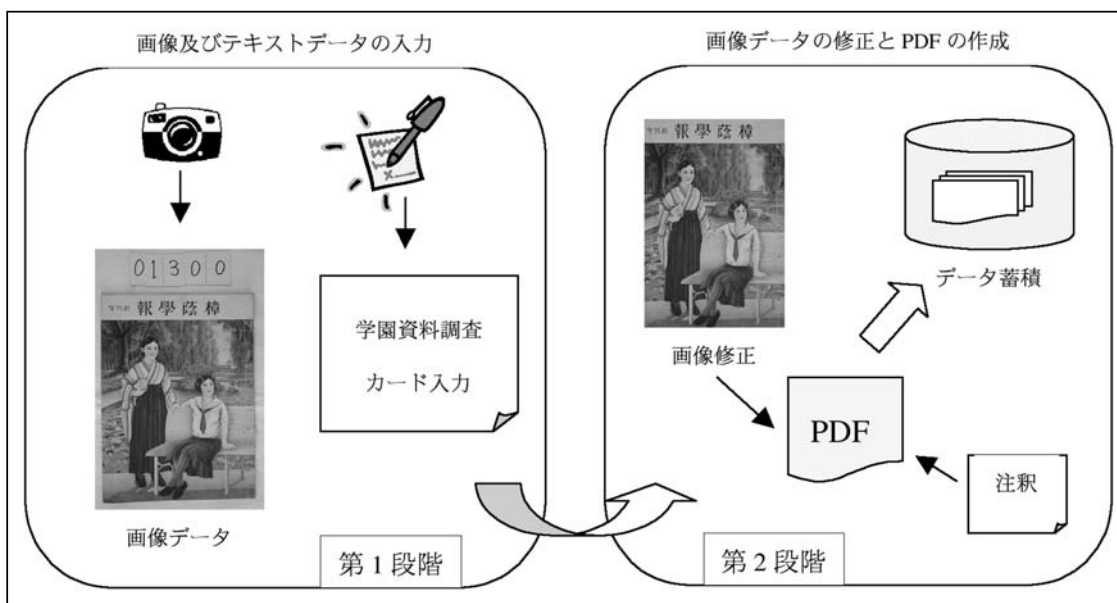


図2：作業の流れ

5. 作業の手順

図2に作業の流れを示す。

第1段階では、『樟蔭學報』の各ページをデジタルカメラで撮影しながら，学園資料調査カード（紙媒体）を記入する。デジタルデータとテキストデータを作成する部分である。第2段階では，画像データを修正し，PDFファイルに変換して検索用に注釈を付加する部分である。

第1段階は白川ゼミ卒業生の松田憲子さんに担当を依頼し，第2段階は竹内が担当した。本稿執筆の段階で，創刊号からの第一巻五号まで，5冊のPDF版『樟蔭學報』が完成し，作業を継続中である。以下では，作業内容の詳細な説明と具体例を示す。

5.1 画像の撮影と学園資料調査カードの記入

画像は，前回と同じ手順で1ページずつ順番にデジタルカメラで撮影した。デジタルカメラの誤操作によるデータの消失を避けるためにCD内蔵のデジタルカメラ⁴を使用した。学園資料調査カードの様式は，前回と全く同じA4サイズ用紙に「カードNo.，調査日，調査者，名称，資料名，形状，材質，サイズ，年記，内容，キーワード，写真，特記事項，学園整理台帳におけ

る記載の有無，記載者」の15項目で作成したものを使用した。学園資料カードの記入（手書き）とMicrosoft Excelのワークシート入力の方を行うことについて，冗長な感は否めないが，変更はしなかった。第1段階で特筆する点は，撮影後にハードディスクへ保存する方法でデータの二重化につとめていたにも関わらず，ハードディスクのクラッシュによりテキストデータの一部を消失してしまった失敗をふまえ，毎日決まった時間に別のハードディスクにバックアップを作成するタイプ⁵に更新したことである。データの消失は，保存データが増加すればするほど，大きな痛手となる。今後も，データの管理やバックアップの方法については，その都度，信頼できるハードウェアを採用していくことで解決したいと考えている。

5.2 画像の修正とPDFの作成

デジタルカメラで撮影したjpeg形式のファイルは，コントラストと色調の微調整を行い，画質の良い状態で保存し直した後，1号（1冊）分ずつをまとめてPDFに変換する。まとめて変換することで，1枚ずつバラバラで撮影した画像をひとまとまりとして扱うことができるので，1冊の本のように閲覧することが可能となる。画像の修正にはAdobe Photo ShopをPDFの作成にはAdobe Acrobatを使用した。PDFに変換した後に，テキストデータの情報からキーワードとして抽出した語を注釈として付け加える。図3にPDFの例を示す。



図3：PDFの例

5.3 注釈を用いた検索と拡張性

図4に注釈を用いた検索の例を示す。「旅行」をキーワードに創刊号を検索したところ、3件の結果が得られた。該当する「東京地方修学旅行」のページと注釈が表示される。注釈のマッチした部分が太字になっている。ここでの検索は、注釈として付加された語だけを対象にしているため、検索文字列と全く同じもしくは一部が同じ文字列を探しているにだけである。全体を目で見て探すことを幾分補助しているにすぎない。しかしながら、検索で得られた結果が最終目標ではなく、利用者が注釈にキーワードを追加することやコメントを書き込むことによって、知識を共有するための手がかりである。

例えば、演習形式の少人数のクラスで使用する場合、ネットワーク上のPDF版『樟蔭學報』に同時に複数の学生が各自のコンピュータからアクセスして、閲覧しながら授業に参加する。授業時間内にキーワードや個々のコメントを注釈として追加することができなかったとしても、授業時間以外にも同じように利用できるため、時間的な制約も受けない。教授者はコメントに対するアドバイスを与え、議論の活発化をはかる。

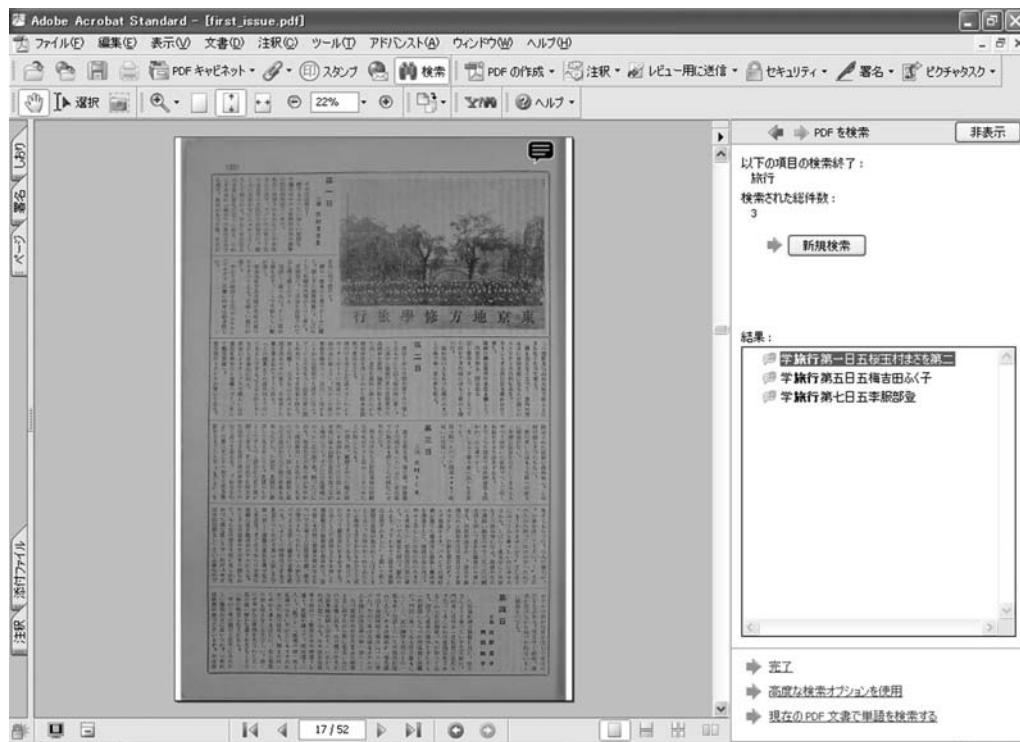


図4：注釈を用いた検索の例

5.4 作業の進捗状況

創刊号から第三巻三号までの合計19冊の撮影と学園資料カードの記入が完了し、テキストデータの入力と最近になって見つかった緑蔭会員⁶及び緑翠会員⁷向け冊子の特集記事を撮影中である。画像の修正が終了したものから順に、PDFへの変換は終了しているので、テキストデータが揃えば、注釈を加えることができる。

6. 関連研究

國學院大學日本文化研究所で進められている中村等⁸の研究は、デジタルアーカイブシステムパッケージ「でんとうなび[®]」⁹を利用した Web 版データベースの公開である。國學院大學が所蔵する考古学・民俗学・国文学・文化財学・神道学等の分野の重要な業績と膨大な研究資料の写真資料を中心に、冊子体目録・CD-ROM 版データベース・Web データベースの構築を実現した。ユーザー用メモ欄を設けるなどの工夫が参考になった。

また、大橋¹⁰の南山大学図書館における収蔵の貴重書『ローマ法大全』をデジタルアーカイブ化に関する報告では、映像を交えた資料紹介の Web ページ作成について詳細に述べられている。なかでも、エンドユーザーの立場から見て利用しやすい、かつ独創的な発信内容を満載した利用型のアーカイブを構築したい旨の将来構想には、共感するところが大きい。

多様な資料構造に対応したデジタルアーカイブシステムについて、依田等¹¹はメタデータツリーの生成という手法で、検索の柔軟性と拡張性を実現した。利用者の要求に即した情報を網羅的に収集するためには、章や節といった構成要素の単位まで検索できなければならない。検索者の意図とメタデータ間の関係を利用して、検索方法を使い分ける仕組みに注目した。

7. まとめと今後の予定

『樟蔭學報』のデジタル化と PDF の注釈機能を利用したインタラクティブなデータベースの試作について報告した。

先に報告した『設立二関スル書類』のカード型データベースにおける、いくつかの指摘をふまえたうえで、今回の仕様では、使い易さや楽しさへの工夫をテーマにデジタルコンテンツを作成した。また、データとして配布する場合や Web で公開を想定して、データのセキュリティ面にも配慮した。特徴は、PDF ファイルの注釈を利用し、キーワードによる文字列検索を可能にした点。作成者だけでなく利用者からもキーワードやコメントの追加を可能にし、知識共有をも実現しようとしている点である。確立したデータベースではなく、インタラクティブで拡張性の高いデータベースを目指し、デジタルコンテンツの特性を生かしながら、活用することで楽しく学べる教材提供を心がけているので、本学の歴史的資料にも魅力を感じてもらいたい。

今後は、残り 14 冊分の注釈を早急に作成し、利用者側の環境を整備した後、テスト運用による実験と評価を行い、実用に備えたいと考えている。

今後、検討しなければならない事項として、以下のような点を挙げておく。

1. ネットワーク上においた同一ファイルに複数の利用者が同時に注釈を付加したときに、更新が正常に行われるか検証する必要がある。
2. 1 冊ずつ個別に PDF ファイルに変換したことが検索範囲を狭めてしまい、検索結果を得にくくしているため、19 冊全てをまとめて PDF ファイルにすることを検討する。
3. コンピュータの操作に不慣れな利用者を支援するためのマニュアルづくりにも対応したい。

謝辞 本研究は、平成 17 年度特別研究助成費によるものである。ここに記して謝意を表す。

参考文献

- 1 <http://e-words.jp/w/PDF.html>
- 2 竹内さおり, 白川哲郎, 『設立二関スル書類』のデジタル化とデータベース試作の取り組み－樟蔭学園草創期資料のデータベース化とその活用 (2), 大阪樟蔭女子大学 (学芸学部) 論集第 42 号, p213-219, 2005
- 3 清水宏一, 治田嘉明, デジタルアーカイブの実践と今後への課題, 情報処理学会研究報告, 2003-CH-60, p1-8
- 4 CD マビカ (MVC-CD500), SONY 製
- 5 Link Station (HD-160LAN), Direct Station (HD-160U2), BUFFALO 製
- 6 樟蔭高等女学校本科・専攻科・高等科の卒業生同窓会員
- 7 樟蔭女子専門学校卒業生同窓会員
- 8 中村耕作, 黒崎浩行, 小川直之, 杉山林継, 学術調査資料の整理・公開システムの構築－写真資料を中心に－, 情報処理学会研究報告 2005-CH-66, p15-22
- 9 http://www.toshiba.co.jp/efort/product/digital/index_j.htm
- 10 大橋直美, 貴重書『ローマ法大全』と南山大学デジタルアーカイブ, 南山大学図書館紀要第 8 号, 2003, p109-116
- 11 依田平, 渡邊隆弘, 大月一弘, 鳩野逸生, 岩杉大輔, 多様な資料構造に対応したデジタルアーカイブシステム－神戸大学電子図書館アーカイブ検索システム－, 情報処理学会研究報告, 2003-FI-73, p45-52

※上記 URL の最終確認日は, 平成 17 年 9 月 30 日である.

