Web上における『法政経済社会論文総覧』の検索閲覧システムの構築

○古隅 弘樹* 周防 節雄*

The Construction of Web-version Information Retrieval System on the Database of "Housei Keizai Shakai Ronbun Souran"

OHiroki FURUZUMI* Setsuo SUOH*

"Housei Keizai Shakai Ronbun Souran"(Catalogue of Academic Papers on Law, Politics, Economics and Sociology) and its additional version edited by Keitaro Amano were published in 1927 and in 1928 respectively. They cover collection of indexes of important academic papers and articles on law, politics, economics and sociology published in academic journals and books during the period from the Meiji Restoration until Showa 2nd year, and they still remain one of main, important sources of information for surveying old documents of those days. We have been working on the construction of web-version information retrieval system on the database of the two indexes, the whole text information of which has been converted into XML style. Our system is now released on our website. In the present paper we describe how our system has been developed and how it can be used.

1. はじめに

昭和2年発行の天野敬太郎編『法政経済社会論文総覧』は明治・大正・昭和初期の法律・政治・経済・社会に関する専門雑誌や、大学紀要・論集に掲載された論文記事目録の集成であり、五十音順配列の見出し語(テーマ)によって分類されている。ここには明治初期～大正15年6月の雑誌論文・記事が収録されている。更に、昭和3年には、大正15年7月～昭和2年の論文記事目録を新たに収録した「追篇」が発行された(以下ではそれぞれを「天野正篇」、「天野追篇」と呼ぶ)。大部分の関係資料がその創刊号から収録されているため、今でも明治・大正・昭和初期の文献調査のための重要資料となっている。

著者らが主宰する書誌情報データベース研究会(代表者:周防節雄)では、「日本社会科学形成期書誌索引情報集成データベース」を構築するプロジェクトを推進中で、手始めに、天野正篇・追篇の全文入力データを作成し、この入力データを計算機処理しやすいようにXML文書化した。このXML文書を用いてデータベースを構築して、Web技術を連携させることで天野正篇・追篇に収録された論文記事目録の検索閲覧システムの開発を行った。このシステムは現在も開発を継続中であるが、機能の一部を一般公開している¹。本稿では、公開中のシステムの開発手法や検索機能について述べる。

2. データ入力とXML文書化

天野正篇・追篇の本文の情報をプログラム処理するために、データの入力規則を指定して、入力業者に委託して作成したものが初期入力ファイルである。天野正篇本文の冒頭ページを図1に、これに対応する部分の初期入力ファイルの内容を図2に示す。

*兵庫県立大学 神戸学園都市学術情報館

Academic Information Systems Centre, Kobe Gakuentoshi Campus, University of Hyogo

¹ (公開URL) <http://mighty.gk.u-hyogo.ac.jp/dabii-jass/>

図1 天野正篇本文の冒頭ページ

図2 初期入力ファイルの冒頭部分

初期入力ファイルでは、可能な限り正字（原本で使用されている漢字）で入力している。原本では書誌情報の記載レイアウトが統一されており、①タイトル、②執筆者名、③掲載雑誌・論集名（略語）、④発表年（和暦）、⑤巻、⑥号、の順に記述されている。プログラム処理で必要な情報の区切り記号（「¥」など）を挿入した他は、原本の記載どおりに入力した。ファイルの文字エンコーディングはShift_JISを使用した²が、可能な限り原本の記載と同じ漢字を使用するため、最終的なアプリケーションやデータベースではUnicodeの文字セットを使用することを想定し、Shift_JISでは入力不可能であってもUnicodeでサポートされる文字については、HTMLにおける文字実体参照の記述形式を用いて文字コードを埋め込んだ。

本文の記述構造に基づいてXML文書構造を独自に設計し、初期入力ファイルをプログラム処理してXML化を行った（図3）。Unicodeの文字セットを扱えるようXML文書におけるエンコーディングはUTF-8を使用した。

```
<?xml version="1.0" encoding="utf-8" standalone="no" ?>
- <法政経済社会文庫総覧 xmlns="http://mailto.kobeuc.ac.jp/~furuzumi/AmanoSource">
- <五十音索引 index="ア">
- <項目 name="藍" ruby="アイ" page="1" row="upper">
- <論文 id="A0100100001" page="1" row="upper">
  <タイトル>人造藍と硫化色素</タイトル>
  <執筆者>守屋源次郎</執筆者>
  <書名略語>日経</書名略語>
  <発表年 ad="1910">明四三</発表年>
  <巻 item="8">八</巻>
  <号 items="4">四</号>
</論文>
</項目>
- <項目 name="愛國心" ruby="アイコクシン" page="1" row="upper">
- <論文 id="A0100200001" page="1" row="upper">
  <タイトル>世界主義と愛國心と國際法</タイトル>
  <執筆者>中村進午</執筆者>
  <書名略語>新報</書名略語>
  <発表年 ad="1902">明三五</発表年>
  <巻 item="12">一二</巻>
  <号 items="1">一</号>
</論文>
...以下省略...
```

図3 XMLファイルの冒頭部分

² Unicodeのエンコーディングに不完全な対応のテキストエディタを使用すると、データが破損する恐れがあるため、入力データの校正作業におけるトラブルを避けるためにShift_JISを使用した。

4. XML データベース

データベースの実装にはXMLデータベース (eXist³) を使用しており、天野正篇および追篇本文のXMLファイルをそのまま格納している。初期入力ファイルは正字で入力されているため、本文のXMLファイルも正字であるが、異体字等の漢字を一つの字に統一することによって表記揺れを減らし、可能な限り新字に置換したXMLファイルを用意して二重化することで、検索時における問題点の解決を図っている。コレクション階層及びドキュメントの配置はデータベースの検索効率を考慮して、新字 (simplified) と正字 (traditional) で分けてXML文書を管理している⁴ (図4)。

図4 コレクション階層とドキュメントの配置

5. Web 検索閲覧システム

データベースの公開ページ (図5、図6) から本システムを利用できる。本システムは天野正篇・追篇の内容を格納したXMLデータベースと、その内容を検索・表示するプログラムで構成され、これらは共にJavaサーブレットとして動作するWebアプリケーションとして構築した。以下では、公開している検索機能について紹介する。

図5 データベース公開ページ

図6 「法政経済社会論文総覧」検索ページ

³ eXistは、XPathやXQueryによるクエリサービスをサポートするオープン・ソースのネイティブXMLデータベースである。現在のシステムでは2006年1月24日付けのスナップショットを使用している。詳細は<http://exist.sourceforge.net/>を参照。

⁴ eXistはXML文書をファイルシステムに似た機構で管理している。フォルダに相当するものを「コレクション」、XMLファイルを「ドキュメント」と呼び、XML:DB仕様に基いている。

5.1 条件を指定して検索

データベースの内容に対して条件を指定して検索することができる。指定できる項目は図 7 に示すとおりで、発表年以外の項目については指定文字列の部分一致で検索し、各条件は AND 結合になっている。

図 7 「条件を指定して検索」入力項目

5.2 目次から検索

天野正篇・追篇では、論文記事目録はその内容に応じて見出し項目（テーマ）毎に分類されている。見出し項目はヨミの五十音順に配列されており、見出し項目の一覧、すなわち目次を表示することで、原本の閲覧と同様の検索機能を提供している。Web ページで用意した五十音索引表（図 8）から、ヨミ毎に見出し項目の一覧（図 9）を表示することができ、選択した見出し項目から論文・記事を検索できる。また、本文中に記載されていた見出し項目の関連情報を利用して、他の関連見出しへのリンク等の情報も併せて表示している。

図 8 「目次から検索」五十音索引表

図 9 見出し一覧の表示例（正篇：「キ」部）

5.3 執筆者索引から検索

天野追篇の巻末に正篇・追篇を通じた執筆者索引が掲載されている。執筆者はヨミの五十音順に配列されており、執筆者毎に論文記事目録が記載された正篇および追篇本文のページ番号を調べることができる。ただし、漢字表記の氏名だけが五十音順に配列されており、英字や記号で始まる執筆者は「雑部」に分類されている。なお、カタカナで始まる氏名の執筆者は原本の執筆者索引には掲載されていなかったため、新たに「カタカナ部」として加えた。執筆者索引の入力データと、本文の XML データの情報を照合したところ、索引氏名の不整合や索引漏れが多数見つかったため、これらを修正、整理する作業が必要となった。この検索機能では索引表（図 10）で選択したヨミで始まる執筆者の一覧表示（図 11）から、特定の執筆者に関する論文記事を検索することができる。

図 10 「執筆者索引から検索」索引表

図 11 執筆者索引の表示例（ア部）

5.4 検索結果の表示

表示に使用する漢字(正字または新字)、および検索結果の一覧表示の一頁あたりの件数(10件または20件)を各検索ページで指定することができる。検索結果については、ヒットした件数と論文記事の一覧を指定件数ずつ表示する(図12)。

各レコードの「詳細表示」ボタンで更に詳細な情報を閲覧できる(図13)。論文記事によっては、同一タイトルが複数の雑誌巻号で掲載(あるいは連載)されたことを表現する複雑な構造を持つ情報もある(図14)。

図 12 検索結果の一覧表示例

図 13 詳細表示例（1）

図 14 詳細表示例（2）

6. 今後の予定と課題

XMLの文書構造については、天野正篇・追篇における本文の記述内容や構造を可能な限り原本どおりに表現することを主眼に置いたため、独自に定義したものを使用した。検索効率の向上や新たな検索機能の追加、他のデータベースやアプリケーションと

の連携など、今後のシステム開発に向けた記述形式の見直しや変換規則の定義について再検討したいと考えている。

システムの検索機能については、現在公開中の機能に加えて、目録情報における掲載雑誌・論集名から検索する機能を構築するためにデータを整理中である。具体的には、目録情報にある発表年と雑誌の巻号の情報は対応するため、これらを整理すれば、雑誌毎にある程度の記事索引を作成することができる。この他にも、各検索機能間の連携や検索結果の絞り込み機能など、追加したい機能は多い。

データベースで収録している目録情報には、執筆者の氏名が索引とは異なることや、掲載雑誌の巻号と発表年の対応が一致しないなど、明らかに記述内容の修正が必要なものがある。また、天野正篇・追篇では読者の検索効率の観点から同一書誌が複数の見出しに掲載されているので、検索した結果、同一書誌が重複して出てくることがあり、冗長な情報表示を排除できるように対処する必要がある。

現在、神戸高等商業学校商業研究所編『経済法律文献目録』（寶文館発行）のデータベース化作業も平行して進めているが、目録情報の収録時期が天野正篇・追篇と重なるので、収録内容の照合作業に利用でき、重複書誌をコントロールすることによって統合データベースを構築することも構想している。

本システムは未だ開発途上であり、新たに追加する検索機能やデータベースの収録内容の更新分などを逐次公開していく予定である。是非現行システムをご利用頂き、システムに関する要望やデータに関する不具合などをご連絡いただければ幸いである。

謝辞

本研究は、(独)日本学術振興会の平成16年度科学研究費補助金（研究成果公開促進費：課題番号168131：作成代表者[周防節雄]）により行われた。書誌情報データベース研究会のメンバーである松田芳郎青森公立大学教授（一橋大学名誉教授、東京国際大学名誉教授）から底本の選定やデータベース構築の基本方針に関する貴重な助言を頂いた。同メンバーの松井幸子筑波大学名誉教授からは、原本目録情報の記載形式の調査などで多くのご助力を頂いた。また、天野正篇・追篇の復刻版出版元である皓星社の藤巻修一氏からも助言を頂いた。ここに記して謝意に替えたい。

参考文献

- [1] 天野敬太郎編『法政経済社会論文総覧』、刀江書院、1927年
- [2] 天野敬太郎編『法政経済社会論文総覧 追篇』、刀江書院、1928年
- [3] 石山洋他編『明治・大正・昭和前期 雑誌記事索引集成』、社会科学編 第7～9巻、皓星社、1994～1997年（天野正篇・追篇の復刻版）
- [4] 周防節雄・古隅弘樹、『「統計資料解題」のデジタル化に関する技法とシステム構築』、神戸商科大学研究叢書 LXIX、神戸商科大学経済研究所、2003年。
- [5] 周防節雄・古隅弘樹、『旧植民地に関する統計書誌情報データベースシステムの構築』、兵庫県立大学経済経営研究叢書 LXVII、兵庫県立大学経済経営研究所、2006年。
- [6] クン・イー・ファン著、原隆文訳、『XSLT Web 開発ガイド』、ピアソン・エデュケーション 2001年。
- [7] 屋内恭輔・安陪隆明、『XML スキーマ書法』、毎日コミュニケーションズ、2003年
- [8] Dan Chang・Dan Harkey 著、高際理監修、石井史子訳、『Java & XML データアクセスガイド』、翔泳社、2000年。