

## 遡及入力における CATP-Auto の利用について

静岡大学附属図書館

溜淵 文子

### 1. 概要

本学蔵書 105 万冊の内、目録データベース化されているのは 32 万冊。その中には、十分な目録になっておらず、日頃の検索にも不十分な書誌が含まれている。電算導入時に目録カードからパンチ外注して入力した開架図書 6 万冊について、これらのオリジナル簡略書誌を NC 書誌にしたいとその必要性を感じてきた。

今回の研修のテーマとして、パンチャー入力した信頼性に乏しいデータをもとに自動登録が可能か検証したい。実際の登録では、自動登録出来なかったデータについては、再度検索キーを変えて自動登録をしたり、手作業で入力をしたり、の処理になると思われる。ローカルデータへの反映・登録方法等、種々の問題はあがるが、今回は、効率的な検索キーを探ることと、実際に自動登録された結果が信頼できるものかを検証し、CATP-Auto での自動登録が可能かを探る。

### 2. 自動登録のデータ作成

現在のローカルデータは Microsoft Access で構築されている。これから登録に必要なデータを抽出・編集し、CATP-Auto 用のアップロードファイルを作成した。本来開架図書は学生用図書であり、オリジナル書誌 6 万冊の内、現在もまだ開架にある和図書（洋書・視聴覚資料・雑誌特集号を除く）は 48,834 冊。出版年代としては、1970 - 1987 年のものがほとんどで、1976-1985 年で 79% を占める。この中からサンプルデータを抽出した。

#### 2.1. Microsoft Access からの抽出

クエリーを使って、開架図書であり、オリジナル書誌である和書を抽出した。この内教育関係図書 3,603 件をサンプルデータとした。

#### 2.2. Microsoft Excel にエクスポート。

タイトル中の一部を巻号に移したり、シリーズ名をタイトルに、タイトルを巻号に移すもの等、個々にデータ修正を行なう書誌が多数あり、修正には約 5 時間を要した。ここで、検索フィールド名や区切り記号を追加したが、この手順は「マクロの記録」機能で、マクロ化しておくと思われる。

#### 2.3. アップロードファイルを作成

タブ区切りのテキストファイルに落とし、アップロードファイルを作成した。

この前後で、検索キーとしてふさわしくないものは、気の付いた時点で修正した。

#### 2.4. 検索キー作成時の注意点

- ・タイトルについては、TITLEKEY で検索することとした。スペース、「『 』」「 - 」 「・」「：」などの特殊記号で検索キーを切るため、テキストファイルにしてから

「置換」機能を使用してもよい。ただし、「.」「/」「=」「:」「;」等、請求記号などで使用している記号は、Excel のときに分割しておく。

- ・シリーズ名は scan 検索でのみ可能。
- ・静大データでは資料番号が 10 桁なのに対して、フィールドは 16 桁。Access で変更した。
- ・著者については、静大データは漢字フィールドに「著」「編」などの役職名がついていたこともあり、ヨミを検索キーとした。
- ・出版社は、今回の静大データには入力されていないため、検証からはずした。
- ・年については、データが「刷年」のため、最初からの検索キーからはずした。2 回目の検索で使用する可能性があるので、「/\*」を入れて、残しておいた。
- ・テキストファイルにした後、未入力の項目の所はタブが二重になってしまうので、エディタを使い、タブ 2 回を 1 回に「置換」する方法で修正した。
- ・検索キーを作成する時は、使用する可能性がある場合は事前に作成しておいた方がよいと思われる。

## 2.5.その他

アップロードファイルを作成するまでの所要時間中には、ローカルデータの内容によりかなり差が出ることが予想されるが、慣れるに従って、時間は短縮できるものと思われる。今回「置換」機能は便利に多用したが、簡単な Perl のようなプログラムを使用すれば、効率的にできると思われる。

## 3. CATP-Auto での自動登録

### 3.1. 登録方法

【第 1 回検索】検索キー = TITLEKEY:+AUTHKEY:+ED:+PTBTR;

登録失敗のデータについて検索キーを変えて再自動登録。

【第 2 回検索】検索キー = 上記 + YEARKEY:

エラーの内、「書誌が特定できない」に対しては、シリーズ名があればこれを追加し、ないものは年(刷年)を追加した。「該当書誌なし」については、簡単に分かるものは適宜修正して、再検索した。

### 3.2. 第 2 回目の検索キーの作成方法

エラーファイルから、特定のメッセージを持つレコードをそれぞれ抽出した(jperl を使用)。エラーファイルを最初から見直すには、かなりの時間と労力を要する。

```
while(<>){
    if (/書誌レコードを特定/) {
        print ;
    }
}
```

「書誌レコードを特定できない」:「YEARKEY:」を有効にし、検索キーになっていなかったシリーズ名(タイトルの一部になっていたデータ等があった)を追加し

たりした。

「該当の書誌レコードが存在しません」：各データを見て適宜修正した。明らかに入力ミスと思われるもの等は目録カードを見直し、区切りの違いの可能性があるものは変更し、容易に修正可能なものに止めて修正した。

「該当書誌に一致する VOL が存在しない」：修正しても、自動登録は難しいと考え、登録可能と思われるもののみとした。

### 3.3. 登録の結果

検索は 3 回に分けて行ない、第 1 回検索は夜間(22:00-8:00)、最後は昼間に実施した。

検索キー	処理件数	ヒット件数	ヒット率
第 1 回検索 タイトル+著者ヨミ+版+シリーズ	3,603 件	2,201 件	61.1 %
第 2 回検索(再検索) タイトル+著者ヨミ+版+シリーズ+年	(778)	433	
全 体	3,603	2,496	69.3

実施時間： 第 1 回 1,666 件 (10 時間 50 分) + 1,476 件 (8 時間 40 分) + 461 件 (1 時間 50 分)

第 2 回 543 件 (3 時間 30 分) + 235 件 (2 時間)

### 3.4. エラーメッセージ内訳

メッセージ	第 1 回検索 (1,402 件)	第 2 回検索 (345 件)
	件	件
書誌レコードを特定できない	500	114
該当の書誌レコードが存在しません	674	154
該当書誌は VOL が空値でない	7	1
該当書誌は VOL が複数存在する	30	6
該当書誌に一致する VOL が存在しない	178	45
所蔵の VOL が特定できない	12	7
サーバエラーを検出しました	1	18

## 4 . 登録結果の分析

### 4.1. 登録済の検証

所蔵登録されたもの 800 件について、正しく登録されたか目録カードとダウンロードデータを照合した。

誤った書誌に登録した件数 2 件

この内、1件はシリーズ名が入力されていれば防げたものであった。

残り1件は、同タイトル・著者・年で、出版社が異なるものであり、所蔵の書誌は学情データにないもので、出版社を検索キーとしていない今回は避けることはできなかったものであった。

#### 4.2. 登録されなかったものの分析

登録されなかったもの314件について、webcatで再検索し、検証した。

##### 同定失敗の原因

原因	件数
a. データミス(入力ミス：ヨミ、漢字)	33 件
b. ケアレスミス(検索キー作成ミス)	20
c. タイトルの取り方の違い	18
d. 著者等のヨミ・翻字の表記違い	14
e. 版表示が異なる、該当するVOLがない	3
f. 学情に該当書誌がない	9
g. 同タイトルで別版がある	53
h. 同タイトルで異本がある	116
i. 書誌レコード作成単位の違い	24
j. 重複書誌	16
k. VOLが複数あり、特定できない	8

#### 4.3. 検索キーの検討

検索キーの作成には、ローカルデータの入力形式により大きく左右される。実際には作業を繰り返しながら、現状に合わせていくことになると思われる。

静大固有の問題かもしれないが、カード目録を見ずに検索キーを作成する時の欠点等を参考までに例を挙げる。

- ・タイプミスと思われるデータミスがやはり多い。  
例)「こどものころ」「子どものころ」
- ・検索キーが少ないため、書誌レコードを特定できないものが多い。
- ・タイトルの区切りが、学情書誌と異なる。「・」やスペースで、正式書名が連続しているのに分割して検索したり、その逆があった。
- ・シリーズ名を持つもので、固有タイトルがあるものと、ないものの区別がつけにくい。
- ・多巻物で、巻号表記が数字以外も含まれている時、VOLを同定できない。  
例)「双書・『にんげん』実践の研究 .1 集団主義編」
- ・ヨミについて、表記法の違いや、翻字、区切り位置の違い。  
例)「ヅ」と「ズ」、「トウヤマ」と「トオヤマ」

## 5 . ローカルデータへの反映と注意点

ローカルシステムへの反映は、システムにより異なるため、方法等は各々検討する必要がある。現在静大では、CATP-Auto のダウンロードデータからローカルシステムにデータを反映する方法を検討中であり、100 件のみ、従来の方法で再ダウンロードし、NC データになるか検証した。

静大 (Looks21/日立) では、書誌は「1 : 学情置換可」、所蔵は「2 : 学情置換不可」のデフォルト値になっているため、このままダウンロードすると、書誌は学情書誌ができるが、既に存在している資料 ID は何もせずオリジナル書誌についてままである。このため、所蔵も「1 : 学情置換可」にして、学情書誌にリンク変更されるか、「貸出中」でも支障がないことを検証した。

手順 :

- ・ダウンロードしたファイルから、資料 ID を抽出し (Perl の使用例)

```
while(<){  
    if (/RGTN=(\d\d\d\d\d\d\d\d\d\d)/) {  
        print "$1\n";  
    }  
}
```

- ・バックアップを兼ねてテーブルに抽出、この分のみ「1」に変更。この時、書誌番号が変更になると支障のある「予約情報」をチェックし、これを持つものは対象外とする。

- ・データ変換すると、学情書誌にリンク変更されるが、配架コード等がデフォルト値になってしまうため、必要に応じて で作成しておいたデータに戻す。

- ・不要になったオリジナル書誌は、オンラインに残されたままになるので、削除する必要がある。

## 6 . まとめと改善点

今回、検索キーを増やしていく方法を取ったが、検証の結果、たとえ刷年であっても検索キーとして使用した方が、ヒットしない件数よりも一つの書誌に絞ってヒットできた件数の方が多かった。検索キーを減らして再検索をしていく方が良いと思われる。

一般の実験結果 (参考文献 2)) に比べて、多少ヒット率は下がるが、大きな違いは無いように思われる。ヒット率が悪かった原因として、教育学分野という理由ではなく、入力ミスが多かったこと、出版社のデータがないものを使用したこと、目録カードから直接の入力ではないため、効果的な検索キーがデータに入らなかったこと、等の理由と考えられる。

結果として、多少不十分なデータであっても、アップロードファイルを作成し、CATP-Auto を使用しての遡及は、十分可能と思われる。ヒット率は多少落ちるにしても、その後のヒットしなかったデータについて考慮すれば良い事を考えれば、労力・コストからかなり有効と思われる。一方、登録できなかったデータについては、

データの修正を考えると、他の半自動登録システムとの併用を考慮した方が効率的と思われる。

CATP-Auto について、半自動登録システムか、数件ヒットした場合はヒットした書誌の一覧が見られるか、いずれかの機能があれば効率よく書誌の同定ができると感じた。

## 謝辞

最後になりましたが、研修期間中、いろいろとサポートしていただいたり、レポートを作成するにあたって親切にご助言いただいた学術情報センターの皆様、特に米澤氏、熊淵氏、研修に参加されている皆様、ありがとうございました。

## 参考文献

- 1) 「CATP-Auto 運用手順書 (平成 10 年 3 月)」  
国立大学図書館協議会図書館情報システム特別委員会時期電算化システム専門委員会
  - 2) 増田元ほか “ 遡及入力における目録自動登録システムの有効性 ” 『大学図書館研究』 55, 1999.3, pp.38-53
  - 3) 星野雅英 “ 遡及変換と総合目録データベース ” 『情報の科学と技術』 42(3), pp.219-228(1992)
- <http://www.cat.op.nacsis.ac.jp/PUB/nl/nl-66-02.html>
  - [http://www.cat.op.nacsis.ac.jp/INFO/newcat/jissou\\_siyo/bbib.search.html](http://www.cat.op.nacsis.ac.jp/INFO/newcat/jissou_siyo/bbib.search.html)
  - [http://www.cat.op.nacsis.ac.jp/INFO/newcat/catpauto\\_mura/009.html](http://www.cat.op.nacsis.ac.jp/INFO/newcat/catpauto_mura/009.html)
  - <http://wwwsoc.nacsis.ac.jp/anul/material/faq.html>