

遡及入力における自動登録クライアントの有用性について

香川大学附属図書館

岩澤 尚子

1. はじめに

香川大学附属図書館では1997年5月より第2期遡及入力(資料1)を行っている。中央館書庫内社会科学系図書のうち、未登録のもの約108,000冊を非常勤職員(6時間)2名と、職員2名(日常業務と並行して)で行おうというものである。第3期計画終了予定が平成22年。15年間をかけて、確実に利用できる図書を目録データベースに登録しようという計画である。確実に利用できるものという観点から、現物を書庫よりブックトラックに乗せ運んで来て、1冊ずつ、まず自館OPACを検索。未登録のものについては、NACSIS-CAT検索・登録後、ダウンロードしたデータを自館OPACに書誌登録・所蔵登録という作業の繰り返しである。約1年半経過したところで振り返ってみると、予定冊数は順調にこなしてきているし、それに伴いILL(特に相互貸借)の受付件数(資料2)も飛躍的に伸びてきた。ただ、館内図書全てをOPACで検索できるようになるまでの長い道のりを考えると気が遠くなるようである。そんな折、平成9年度国立大学図書館協議会次期電算化システム専門委員会が、学術情報センターの新CATシステムへ各図書館等の所蔵データを自動的に登録するシステム「CATP-Auto」を開発した。現在は、ダウンロードファイルから自館目録データベースへの変換はできないが、来年2月のシステム更新を視野に入れ、CATP-Autoを使ってどのくらい遡及入力の効率化を図れるのか試みることにした。

遡及入力には様々な方法があると思うが、このレポートは1件ずつデータを入力する方法で、特に洋書について実際に書誌ルートで登録してみた結果に基づいている。なお、NCには既存の所蔵レコードがない場合を想定している。

2. アップロードファイル作成方法

CATP-Autoを実行するためにはまず、アップロードファイルを作成しなければならない。CATP-Autoには簡易入力エディタが用意されているが、入力できる項目が限られている。例えば版に関する事項での絞り込みやAKEY検索はできないなどの制限がある。また業務画面と同時に開けない場合も考えられるので、外部ツールを利用した2種類の作成方法を紹介する。

2-1. Microsoft Excel を使って

表計算ソフトを使う場合、フィールド名を入力データの前に1件ずつ入力するのは面倒なので、フィールド名の入っている列と入力データの入っている列を別に分けて作る方が便利である。ところが、タブ区切りのテキストデータでファイルを保存すると、フィールド名の後にも不要なタブ区切りが入ってしまう。

また、すべてのフィールドにデータを入力するとは限らないので、空白フィールドを残

結果

処理結果	件数	%	エラーメッセージ(資料4)
登録	158	66.4	
レコードなし	58	24.4	/*該当の書誌レコードが存在しません
特定できず	20	8.4	/*書誌レコードを特定できないため更新できません
VOLが複数存在	1	0.4	/*該当書誌はVOLが複数存在するため更新できません
処理スキップ	1	0.4	/*所蔵のVOLが特定できないため処理をスキップしました

分析

登録

登録できたもののデータチェックを行った結果、158件中8件、間違えて登録していた。

その8件の内訳は

- ・出版年など、アップロードデータそのもの間違い…………… 5件
- ・版違い(特に「international ed.」など出版年が同一の場合)… 1件
- ・出版者違い(出版年が同一の場合)…………… 1件
- ・書誌違い…………… 1件

レコードなし

58件中46件は本当に該当の書誌レコードはなかった。12件については以下の理由で検索されなかった。

- ・書誌の取り違い…………… 6件
- ・ISBNが書誌レコードになかった…………… 6件

特定できず

特定できなかった20件の理由は

- ・同一ISBNの書誌が複数存在した…………… 4件
- ・フルタイトルと出版年が同一の書誌が複数存在した…………… 16件

VOLが複数存在

これについては1件だった。その理由は該当書誌レコードにはVOLフィールドが繰り返されていた(VOL:: hbk, VOL::pbk)が、本にはその記述がなかった。

処理スキップ

原因不明。該当データはISBNによる検索であったので、Vol 処理設定画面の設定とは無関係だと思われる。

注意点

(1)「international ed.」など出版年が同一のものが複数存在する可能性のあるものは版表示を照合キーとして絞り込むなどの工夫をすること。ただし、これは書誌の記述上の問題であるが、版として記述されていない場合も多く見受けられるので、プルーフリストチェック段階での注意が必要。

注意点

(1) 書誌レコードの特定ができない点については、重複書誌の存在など、ある程度は避けられないことである。ただ、その点を考慮しても、ISBN キーが分かっている場合にはそれが一番有効なキーだと推測される。ISBN により VOL の特定ができるので、アップロードファイル中で VOL を指定しなくてもよいという利点もある。

注意を要するのは、SET の ISBN を指定してしまった場合、いくらアップロードファイルで VOL の指定をしても、書誌レコードの ISBN に対応する VOL(すなわち「: SET」)が登録されてしまうこと。もうひとつは、特殊な例かもしれないが、書誌レコードで ISBN が繰り返されているにもかかわらず VOL に記述がない場合、いくら一致する VOL が1個であっても、「/*該当書誌は VOL が複数存在するため更新できません」というメッセージにより登録されないことである。

(2) 「Vol処理設定画面」の「書誌検索時の一致VOLなし時の処理」のパラメータ設定がわかりにくかったので、まとめておく。設定により結果が異なるのは、「ISBNによらない検索」をして「所蔵登録用のVOLを指定した時」だけである。ISBN以外のフィールドではVOLの特定が困難なため、アップロードファイルの中に指定した所蔵登録用のVOLと、書誌レコードのVOLを比べて一致しなかった場合、登録するかどうかの判断をする。

- ・「処理スキップ」にした場合

確実ではあるが、例えばアップロードファイルで「VOL=: pbk」と指定したが、書誌レコードのVOLに記述がなかった場合は登録できない。

- ・「アップロードファイルのVOLを使用」にした場合

例えばアップロードファイルで「VOL=v . 2」と指定したが、書誌レコードのVOLには「VOL=v . 1」の記述しかなくても、所蔵レコードに「VOL : v . 2」が登録されてしまう。プルーフリストのチェックにより書誌レコードにVOLを追加する必要がある。

3-3 . 有効な検索キー

3-1 . 3-2 . の結果から推測できる有効な検索フィールド

ISBNKEY

なるべくユニークな_TITLE_(数個) + _AUTH_(1個) + YEARKEY

FTITLEKEY + YEARKEY

- ・これらのキーの組み合わせに、場合によって、ED で絞り込んだり、PUBLKEY を追加すればより有効性は高まる。

- ・ISBNKEY 以外のキーで検索する場合には、版を特定するために YEARKEY を組み合わせることが望ましい。

- ・フルタイトルキーはタイトルが短い場合や一般的な場合には書誌レコードの特定が難しいこと、逆に長い場合にはアップロードデータ作成が面倒かつミスを生みやすいことから、特に洋書の場合には の方法が向いているのではないかと。

[参考] 検索キーを作る上で注意すべき点の詳細：

図書書誌における更新定義

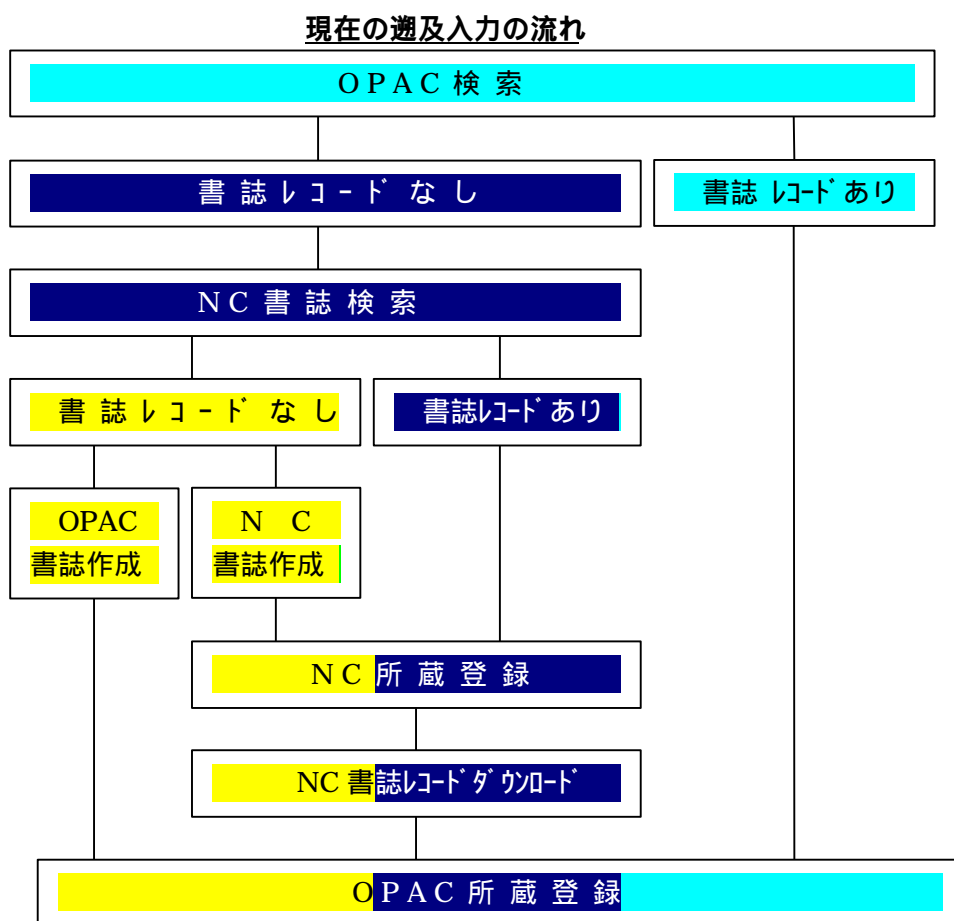
http://www.cat.op.nacsis.ac.jp/INFO/newcat/jissou_siyo/bbib.update.html

各検索キーの作成仕様（例えばフルタイトルキー）

http://www.cat.op.nacsis.ac.jp/INFO/newcat/jissou_siyo/idx_ftitle.html

4 .CATP-Auto を利用しての遡及入力業務の今後

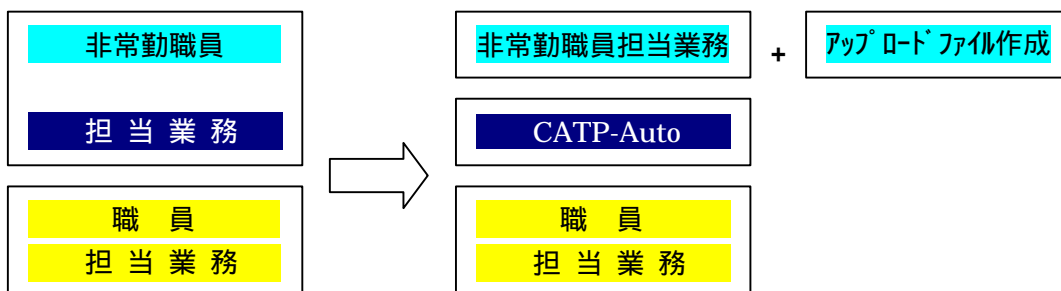
4-1 . 現在の遡及入力業務がどう変わるか



自動登録クライアント導入後は と ~ の一部が自動化される。

現在

自動登録クライアント導入後



図を見ても分かるように、CATP-Auto などの自動登録クライアントを使用することによって、今まで非常勤職員が担当していた業務のかなりの部分が自動化されることになる。アップロードファイルを作成する業務が増えるが、自館の OPAC 検索時に同時に検索キーを作成するなど工夫すれば、それほどの負担にはならないであろう。

登録レコードのチェックについては、現在は「 OPAC 所蔵登録」が終わった時点でリストを出力し行っている。自動登録クライアントを使用して登録したものについても、どのような検索を行ったとしても入念なチェックが必要である。そこで登録ミスが発生した場合は「 NC 書誌検索」に戻って、人手による作業を繰り返さなければならない。しかし、すべての書誌についてコンピュータ画面を見つめて書誌同定作業をすることに比べると、身体的な負担がかなり軽減されるのではないかと思った。

4-2. 今後の遡及入力業務

現在は、ダウンロードファイルを自館目録データベースへ変換できないために、チェック済みのプルーフリストを見ながら再度 1 件ずつダウンロードし、OPAC へ手作業で所蔵登録を行っているので、全体としてどのくらい省力化されたかは不明である。しかし、先にも述べたが、コンピュータ画面を見つめる時間が減ったため、少なくとも身体的負担が軽減されたことは確かである。限られた少ない人手で、継続的に行わなければならない遡及入力業務において、このことは充分評価できると思う。

また、CATP-Auto は現在、オンライン利用の少ない時間帯 (12:00 ~ 13:00 と 17:00 ~ 20:00) に利用するように限定されている。しかし今後、業務用サーバの夜間運転を行うようになれば、昼間作成したアップロードファイルを夜間実行させて、翌日プルーフリストとエラーリストのチェック。その結果に応じて検索・登録ミスの修正やオリジナルレコードの作成などを行うことになるであろう。

具体的には CATP-Auto 導入前、洋書については月平均約 1,000 冊の遡及入力を行っていた。まだ 2 週間の試みなのではっきりした数値ではないが、試行錯誤しながらも 2 週間で約 600 冊の処理 (しかも自館目録データベースへは手作業でダウンロード、所蔵登録) を行った点を考えれば、CATP-Auto を遡及入力業務に導入することはかなりの省力化につながるのではないかと。ダウンロードデータの自館目録データベースへの変換が可能となり、業務用サーバの夜間運転が行われれば、さらにかなり処理件数が増えることが予想される。もちろん、遡及の進んでいる分野とそうでない分野では結果は異なるが、省力化された分、書誌のオリジナル登録に時間が費やせるのではないかと思う。

5. おわりに

大学図書館における電子図書館的機能の必要性が叫ばれる中、その基本となるべき目録情報の遡及入力が重要であることは言うまでもない。各館が少しでも効率よく遡及入力を進めていこうと努力している現状は、過去の総合目録データベース実務研修のレポートからも伺える。私は、学術情報センターの新 CAT システムで作成が簡単になった自動所蔵登録クライアント (CATP-Auto) を有効に利用することによって、単に遡及入力件数を上げるだけでなく、それで生じるであろう人力をオリジナル書誌登録に費やしたいと思う。

予算や人手が限られる中、目録の目的である利用者が求める資料の所在情報をより多く

正しく提供するための手段として、自動所蔵登録クライアントシステムの活用を積極的に考えるべきではないだろうか。遡及入力にはいくつかの方法があると思うが、アップロードファイルや検索キーの作成方法を各館の事情に合わせて少し工夫すれば、自動所蔵登録クライアントシステムは充分使えるものである。その際のひとつの事例として少しでも参考になれば幸いである。

最後になってしまったが、この研修にあたって様々な面からサポートしていただいた学術情報センターの方々、Perl について丁寧に教えてくださった筑波大学図書館部情報システム課洋書データベース系の近藤務氏をはじめ研修仲間の皆様、多忙な中快く研修に出させてくれた香川大学附属図書館の皆様、ありがとうございました。

[参考文献]

- ・「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次期電算化システム専門委員会
- ・村田輝「新目録システム対応自動所蔵登録クライアント(CATP-Auto)」
http://www.cat.op.nacsis.ac.jp/INFO/newcat/catpauto_mura.html

[注意]

CATP-Auto を業務用として利用する場合、下記を学術情報センターに申請すること

- ・「新 CAT/ILL システムの利用申請」
- ・「自動登録ソフト使用届出書」

<資料 1 >

香川大学附属図書館遡及入力年次計画

	第 1 期	第 2 期						第 3 期
年 度	S.63~H.1	H.9	H.10	H.11	H.12	H.13	H.14	H.15~H.22
冊 数	60,000	12,000	19,200	19,200	19,200	19,200	19,200	154,000
冊数計	60,000	108,000						154,000

対象資料

第 1 期：開架図書

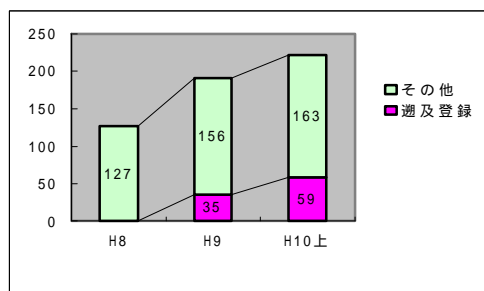
第 2 期：中央館書庫内社会科学系図書未登録のもの

第 3 期：中央館書庫内社会科学系図書以外の図書(分館書庫内図書を含む)

注：製本雑誌は対象外。研究室備付図書のうち未入力図書の遡及入力は返却された都度行う。

<資料 2 >

遡及登録件数と I L L 貸借(受付) 件数の推移



<資料 3 >

Perl プログラム例

```
#!/usr/local/bin/perl
while(<>){
    chop();
    $work = $_;
    while($work){
        if ($work =~ /^s*([a-z]+[:;=])\t([^\t]*)/i){
            &print_field($1, $2);
            $work = $';
        }
    }
    print "\n";
    $tab_flag = 0;
}
```

```

sub print_field{
    local($field, $value) = @_;
    if ($value eq ""){
        return;
    }
    print "¥t" if $tab_flag;
    print "$field$value";
    $tab_flag = 1;
}

```

< 資料 4 >

エラーメッセージ一覧

/*該当の書誌レコードが存在しません

検索結果が 0 件

/*書誌レコードを特定できないため更新できません

検索結果が2件以上

/*該当書誌は親書誌のため更新できません

検索結果が親書誌だった

/*該当書誌はVOLが複数存在するため更新できません

- ・ ISBN を含まない検索 登録 VOL 指定あり 書誌の VOL と不一致
(注：書誌検索時の一致 VOL なし時の処理=処理スキップの場合)
- ・ ISBN を含まない検索 登録 VOL 指定なし 書誌の VOL 繰り返しあり

/*所蔵の VOL が特定できないため処理をスキップしました

/*サーバエラーを検出しました

アップロードファイルに不備があった

「アップロードファイルのエラーチェック」ではチェックされない

- ・ タグの記述ミス 例えば「YEARKEY:」を「YERAKEY:」
- ・ 所蔵データの必須データ「配置コード」の指定忘れ
- ・ 空白フィールドが存在する時

< 資料 5 >

OPAC における CATP-Auto 使用時の書誌の作られ方の割合

