

CATP - Autoを利用した遡及入力の内方

広島大学附属図書館 洋書目録情報係
岡田 英治

1. はじめに
 - 1.1. 遡及入力の必要性
 - 1.2. NACISIS - CATへのアップロードの必要性
 - 1.3. CATP - Autoの登場
2. アップロードファイルの作成
 - 2.1. 準備
 - 2.2. 検索キー、照合キー、登録データの作成
3. 自動登録結果の分析
4. 問題点
 - 4.1. ローカル側の問題
 - 4.2. CATP - Auto側の問題
5. おわりに

1. はじめに

広島大学附属図書館では、大学における教育・研究活動を支援し、各種情報関連支援サービスを迅速かつ的確に提供するため、1972年より外国雑誌の発注・契約業務を皮切りに図書館業務の機械化を開始した。1981年には洋図書の、1982年には和図書の、それぞれ新規受入図書について目録データベースへの入力が始まった。1987年以降は学術情報センターへ接続して目録登録を行い、1991年からはそれまでの情報処理センターのシステムから、図書館の専用機に乗り換えて現在に至っている。

平成9年度末現在、目録データベースへの入力を行った冊数は約97万冊であり、全蔵書冊数（約280万冊）に対する割合は34%に達する。そして平成10年度内には100万冊を突破する見込みである。

1.1. 遡及入力の必要性

前に述べたとおり、相当数の図書レコードが目録データベースに入力されているが、利用者にとってはまだ不十分なレベルにある。目録データベースに未入力の図書を検索する手段は、不便なカード目録に頼るのみで、頻繁に利用される図書館開架に並ぶ図書でさえカードをめくらないと目録や所在が確認できないことが多い。すなわち相当数の図書が検索手段が無いために有効に利用されることのないまま、置物のごとく図書館のかなりの領

域を占領している有り様である。

そこで、平成9年度から10ヶ年計画で、主に中央図書館の開架・書庫に並ぶ図書のうち、本学目録データベースに未入力 of 図書(約35万冊)の遡及入力を予定している。なお、過去2年間で、大型コレクションを中心とした約1万冊についてはNACISIS-CATを通じて遡及入力を行った。

1.2. NACISIS-CATへのアップロードの必要性

本学では1987年の学術情報センター接続以後、新規受入の図書のほとんどをNACISIS-CATに登録し、その書誌レコードを本学の目録データベースにダウンロードした後、OPACを通じて利用者に情報を提供している。学術情報センター接続以前の約6年間に目録データベースに入力した図書レコードについては、OPACを通じて検索可能ではあるものの、NACISIS-CATには登録が行われていない。その冊数は、正確には把握していないが、一年間の平均受入冊数を5万冊とすると、約30万冊にのぼる計算になる。したがってNACISIS-CATに書誌レコードがないばかりに、必然的に相互貸借や文献複写を目的とするILLの対象外となっている状況である。一単科大学の全蔵書に相当するほどの図書を眠らせておかないためにも、NACISIS-CATに本学の目録データベースのレコードを提供することが望ましいであろう。

1.3. CATP-Autoの登場

しかしながら、これまでのNACISIS-CATは、各図書館がローカルに作成し所有する遡及変換レコードのファイルをバッチ処理で一括アップロードできるシステム構造にはなっていなかった。同時に、本学図書館の専用機においても、それに対応する機能は備わっていない。そのため、こうしたローカルに作成した遡及レコードをNACISIS-CATに登録するには、1冊ごとにNACISIS-CATにおいて書誌レコードの同定をした後、所蔵レコードを作成する手順を繰り返すという地道な作業が必要であった。平成10年度の国立大学図書館協議会総会・研究集会において東京学芸大学附属図書館の村田輝氏は次のように述べている。「大学図書館の全蔵書約2億冊のうち、学術情報センターの総合目録データベースに登録されているものが1割強に過ぎないという現状の中で、目録情報の遡及入力がさらに促進されなければならない、と進言しています。しかし、従来どおりのやり方では、遡及入力の完了までには、気の遠くなるような長い時間がかかってしまいます...」。それらの期待に応えるべく、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、各図書館における目録の遡及入力を推進するための支援ツールとして開発した目録自動登録システムが、新目録システム対応自動所蔵登録クライアント(CATP-Auto)である。

CATP-Autoの特徴の一つに、アップロードファイルによる自動登録機能がある。これについての詳細は割愛するが、この機能を利用して本学の目録データベースのうちN

A C S I S - C A Tに未登録のレコードを一括してアップロードできないか、と考えたわけである。

2. アップロードファイルの作成

2.1. 準備

本学の目録データベースにおいて、書誌レコードはZ0Sと称し、9桁のID番号を持つ。それから、1987年の学術情報センター接続以前の書誌レコードは、IDの先頭2文字が和図書・洋図書でそれぞれ“19”“29”で表される。一方、学術情報センター接続以後の書誌レコードは、IDの先頭2文字が同様に“10”“20”で表される。よって、今回アップロード対象となる書誌レコードは、IDの先頭2文字が“19”“29”のものを抽出すればよい。

次に書誌レコードのデータであるが、今回対象の書誌レコードはJ P - M A R CおよびL C - M A R C等の参照レコードを流用入力して作られたものであるから、N A C S I S - C A Tの目録基準とは異なる。しかし、記述の内容に差異はあっても、それぞれ共通するフィールドがあることから、今回はC A T P - A u t oに付属する簡易入力エディタにおける入力項目（I S B N、フルタイトル、著者、請求記号、登録番号など）にならって、Microsoft-Excelにてデータ作成を行った。

なお、今回は時間の都合やシステム環境の制約などによって本学の目録データベースから集められるかぎりの、中央図書館開架の和図書レコードを100冊ほど無秩序に選んだ。書誌レコードの出版年は1976年～1987年で、1985年以降が9割を占める。書誌レコードの分類はN D C 8版の類目表で言えば、社会科学...15%、自然科学...60%、芸術...10%、その他が15%である。

2.2. 検索キー、照合キー、登録データの作成

検索キー、照合キーはそれぞれsearch検索、scan検索の対象となる。したがって、まず最初に作成したアップロードファイルをC A T P - A u t oで自動登録を行い、その後検索キーや照合キーの組み合わせを替えたり、さらに検索キーのうちフルタイトルのデータを単語ごとに区切るなどして、自動登録の結果に変化があるか試みた。

登録データについては、今回は教育用モードで試行するので、任意に設定した。ちなみに、動作指示コマンドも別に設定できるが、今回は省略した。

3. 自動登録結果の分析

(詳細は別紙を参照)

教育用モードにてテストしたので、ヒットしうる書誌レコードの件数が少ないことは予想できたが、対象の100件すべてがエラーとなったことは衝撃だった。結果からいえる点をいくつか挙げると、

検索キー・照合キーを多く設定すると、自動登録の精度（絞りこみ）は向上する。
一つの検索キーのデータを分割すると、自動登録の精度（絞りこみ）は低下する。
しかし、については、サーバエラーの発生に多く見舞われたので、原因が判明するまでは明言を避けたい。

4. 問題点

4.1. ローカル側の問題

アップロードファイルのデータとして使えないデータの存在

今回対象の書誌レコードはJP-MARCおよびLC-MARC等の参照レコードを流用入力したものであることは前に述べた。しかし、これまでの蓄積データには機械的なバグ、さらに人為的なバグが数多く発生している。例えばISBNについては、情報処理センターの時代、ISBNにハイフンが含まれていても表示には問題がなかったようである。その後、図書館専用機にシステムを乗り換えた際、ISBNはハイフンを含んだまま移行された。しかし、図書館専用機ではISBNは10桁としていたため、先頭から10桁までが取り込まれ、以下は切り捨てられた。したがってISBNとしてはまったく意味のないデータが現在まで居座り続けている。ISBNは検索キーとして有効であるが、この現状ではどうにもならない。また、化け文字もいくつかあった。機種更新は必ずやってくるものなので、これまでの蓄積データが欠落することなく移行できることはもちろん、その時のチェックを細部まで行うことの大切さを痛感した。

外部のエディタを使用する時のアップロードファイル作成の不安

付属の簡易入力エディタはカード形式となっており、すでに入力すべき項目が指示してあるので、それぞれにデータを埋め込んでいけば、自動的にアップロードファイルが作成される。しかし、入力すべきデータが多い場合や、簡易入力エディタを使用できない環境下では、市販のワープロソフトや表計算・データベースソフトなどのエディタを使用した方が圧倒的に能率がよい。問題なのは、そういった外部のエディタを使用した時に、アップロードファイルのデータがうまく反映できない場合が見られることである。今回、エラーとなった原因を探っていくうちに、アップロードファイル作成時には入力した覚えがない文字がなぜか含まれていた。外部のエディタでアップロードファイルを作成するには、データを“タブ区切り”にする必要があるが、それを読み込むときに影響があるのかもしれない。ともかく原因は不明である。

外部エディタを使用する場合は、CATP-Autoとの相性をあらかじめ確かめておかないと、エラーが出るたびにその対処に追われるという余計な仕事を抱え込んでしまうおそれがある。

4.2. CATP-Auto側の問題

入力エディタの推敲

今回アップロードファイルの作成には Microsoft-Excel を使用したが、作成までの道のりは決して楽ではなかった。本学の目録データベースから必要な書誌レコードを掃き出したが、それらは一つのフィールドが一行ごとに出力される。Microsoft-Excel ではそれらのフィールドの中から検索および登録レコードに必要なものを抜き出して一行にまとめなければならない。対象レコードが100件程度なら、付属の簡易入力エディタで済ませることができるだろうが、それが膨大な量であれば、どうしたら効率よく入力できるか。ローカル側がデータを掃き出す機能を有するか、どういう形式で出力されるか、さらにそれを編集する入力エディタは何を選べばよいか、もしも外部のエディタを選んだ場合にそのエディタを使いこなすことができるか等、さまざまな点を考慮する必要がある。

自動登録しやすいような検索キー、照合キーの作成

苦勞して作成したアップロードファイルをいざ自動登録機能にかけて、その結果がエラーだらけでは、せっかくの苦勞も水の泡である。もちろん、結果がエラーであっても、検索キーや登録データはエラーファイルとして保存されるので、後のレコード作成に役立つのは確かであるが。今回は教育用モードのため、正常に登録した件数はごくわずかであったが、1回目はエラーだった検索キーや照合キーをいろいろ組み替えると、2回目には書誌レコードにヒットすることがあった。つまり、ただ単に検索キーや照合キーを並べても書誌レコードにヒットする確率は低い。したがって、なんとかヒットする率を向上させる工夫が必要となる。例えば、検索キーは単独ではなく複数設定する、フルタイトルでなく単語に区切る、不要な検索キーや照合キーを削除する、など。そうした場面では、検索キーや照合キーの作成は目録検索に疎い者よりも、むしろ経験豊かな者が担当するのが適している。

5. おわりに

CATP - Autoの流れは、分業体制の発想からきていると思う。従来は私たちがNACISIS - CATに目録登録する際、その書誌レコードが新規入力なのか、参照MARCを基にした流用入力なのか、所蔵更新だけでよいのか、全く予見のないままレコード検索を行っていた。その結果、おのずと経験豊かな者とそうでない者とは目録登録のスピードやレコードの出来に差が生じてきた。それは、検索の手法や書誌レコードの同定などの点で両者にはっきりとした違いがあるにもかかわらず、一貫した同じ仕事をさせているからである。

そこで、一連の流れを分割したらどうなるか。例えば 検索キーの作成、 書誌レコードの同定、 所蔵登録、 書誌レコード新規(または流用)入力、の各段階に分ける。そして はパンチャー業者、 と はCATP - Autoで処理し、その結果をブルーファイルで図書館職員が確認し、 のみ経験豊かな者が、それぞれ担当することにすれば作業の能率は非常に向上するのではないだろうか。さらに、レコード1件あたりの単価は、すべて業者まかせ、または図書館がアルバイトを雇って教育しながらこなす場合よりも確

実に下がるであろう。

今後、各大学図書館では遡及入力に取り組む機会が多くなると予想される。その時、限られた予算、時間、人材の中で、いかにして多くの件数をこなし、かつレコードの品質を保っていくか。そうした要望に対して、CATP - Autoを取り入れた分業体制が果たす役割は大きいと考える。その反面、CATP - Autoがすべてとは一概にいけない。特殊なコレクションや古書など、NACISIS - CATに書誌レコードがありそうにないものは、従来どおり1件ずつ処理したほうがいい場合もあるので、入力対象全体を見通した計画を立てることが大切であろう。

別紙

自動登録の結果

【表の見方】

《検索キーおよび照合キーの組み替え》

- ・パターン1 ... <検索キー> ISBN: NBN: FTITLEKEY: AUTHKEY: PUB: YEAR: <照合キー> ED;
- ・パターン2 ... <検索キー> FTITLEKEY: AUTHKEY: PUB: YEAR: <照合キー> ED;
- ・パターン3 ... <検索キー> FTITLEKEY: AUTHKEY: PUB: YEAR:
- ・パターン4 ... <検索キー> FTITLEKEY (副題以下はカット): AUTHKEY: PUB: YEAR:
- ・パターン5 ... <検索キー> _TITLE_: _AUTH_: PUB: YEAR:
- ・パターン6 ... <検索キー> _TITLE_ (2語): _AUTH_ (2語): PUB: YEAR:
- ・パターン7 ... <検索キー> _TITLE_ (他の2語): _AUTH_ (2語): PUB: YEAR:

《エラーメッセージ一覧》

(下欄は考えられる原因)

無印.../* 該当の書誌レコードが存在しません

(原因) 検索キー・照合キーが適切でない。適切であったが書誌レコードが見つからなかった

.../* サーバエラーを検出しました。[Invalid Object-Body(LOC)@CAT306201]

(原因) アップロードファイルに異常が発生した、あるいはサーバに異常が発生した

.../* 書誌レコードを特定できないため更新できません

(原因) 該当の書誌レコードが複数見つかった

.../* 該当書誌はVOLが複数存在するため更新できません

(原因) 該当の書誌は見つかったが、VOLを検索キー・照合キーに指定していない

結果の内訳	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6	パターン7
対象件数	100	100	100	100	100	100	100
うち							
	1	2	2	2	3	4	3
	1	1	1	1	6	7	3
無印	97	83	83	76	53	46	80
	1	14	14	21	38	43	14

レコードno. 1 ~ 5 0

レコード No.	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6	パターン7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							
48							
49							
50							

レコードno. 51 ~ 100

レコード No.	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6	パターン7
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
74							
75							
76							
77							
78							
79							
80							
81							
82							
83							
84							
85							
86							
87							
88							
89							
90							
91							
92							
93							
94							
95							
96							
97							
98							
99							
100							