職場から

書誌レコードの遡及入力についての報告

別府大学附属図書館
財　前　聡　担

＜遡及入力について＞

図書館では、今年度後半から遡及入力を開始した。作業内容は、本学図書館が現在所蔵している蔵書の目録データを国立情報学研究所（以下､ＮＩＩ）が所蔵する全国標準の高品位なデータに更新し、同時に本学の蔵書でＮＩＩに所蔵登録されていないデータを登録するというものである。

現在本学のデータベースには約27万件の目録データが入力されているが、そのうち20万件は良質とは言えないデータである。例えば、著者名のヨミで「大宅（オオヤ)」が「ダイタク」になっていたり「土方（ヒジカタ)」が「ドカタ」になっていたりと、蔵書検索の際に何かと支障をきたす状況である。また、古いデータについては書名や刊行年等がＮＩＩのデータと異なっている場合が多い。これらは当時の基準で作成されており、致し方ない面もあるが、現在のようにコンピューターが普及しＯＰＡＣやＩＬＬ等全国規模のネットワーク環境が整備され、本学もそれらに参加しているという状況を考えれば、標準化を進めていかないと相互協力において支障が生じてくる。

本学に限らず、現在多くの大学や研究機関でこの作業が進められている。ＮＩＩも平成16年度よりこの遡及入力に対する支援事業を行っている。これは事業計画書による公募方式で、平成18年度は108件の応募があり、37機関57件が採択された。対象となる事業は、多言語資料の遡及入力、人文社会科学関係資料の遡及入力、自動登録支援である。本学が応募したのは自動登録支援で、これに採択されたのは上記の57件のうち、本学を含めて７件であった。

今回の遡及入力は、このＮＩＩの支援とＮＣ－Ａｕｔｏという遡及入力専用ソフトによる全自動、半自動入力の二本立てで並行的に実施した。ＮＩＩの具体的な支援内容は以下のとおりである。

まずＮＩＩよりＮＣ－Ａｕｔｏの提供を受け、それを使って本学のデータをＮＩＩのサーバーにアップロードする。検索キーが一致しヒットしたデータは自動的にＮＩＩへ所蔵登録がなされ、ＮＩＩ所蔵の良質なデータが本学のコンピューターにダウンロードされるが、ヒットしなかったデータについての入力作業（半自動入力と新規入力）をＴＲＣ（株式会社図書館流通センター）が請負い、その費用をＮＩＩが補助するというものである。

以下、作業手順や結果等について述べるが、初めての試みであり、またＮＣ－Ａｕｔｏについての簡単な操作マニュアルしかなかったため、ほとんど手探りといってよい状況であった。途中何度か方針を変更したりしているので、データが統一していない部分もあり、そのために記述が煩雑になっている部分もあるかと思うが、ご容赦願いたい。

＜ＮＣ－Ａｕｔｏについて＞

ＮＣ－ＡｕｔｏとはNECが作成した書誌レコード遡及入力専用ソフトの名称である。これはISBN、タイトル、著者名等を検索キーとしてＮＩＩのデータと照合し、前述のとおり所蔵登録やダウンロードを自動的に行う機能を有するものである。

ダウンロードされたデータは、ローカルシステムへのアップロード用のファイルとプルーフファイル、エラーファイルの３つに分かれて保存される。

下にプルーフファイルにダウンロードされたＮＩＩ書誌レコードの１つを例として掲げる。

〈ＮＩＩ書誌レコードの例〉
FTITLEKEY: 近代日本の児童文化　AUTH: 滑川道夫　YEAR:1972
LOC=2F 閲覧室　CLN=909:N2　RGTN=0235861

DBNAME:BOOK
ID:BN01848740
TRD: 近代日本の児童文化 / 滑川道夫 , 菅忠道編著
PUBL: 新評論 PUBDT:1972.4
DBNAME:BHOLD
ID:CC0950524301
FANO:FA009097
LIBABL: 別府大
LOC:2F 閲覧室

＜対象とするデータの選択＞

支援事業の対象となるデータは１万件程度である。本学システム所蔵のデータをEXCELにダウンロードして詳しく見てみると、現在のシステムが導入された2000年以降は比較的良質なデータが入っているようである。また検索キーとして最も有効と思われるISBNが入力されているデータが全体で10万件程あった。（表１参照）

そこで、今回はISBNの無いデータのうち2000年以前の古いデータの中から１万件を選ぶことにした。ちょうど昨年２Ｆ閲覧室の蔵書点検を行っていたので、所在が２Ｆ閲覧室になっているデータを抽出してみたところ13,424件だった。ちょうど良い数字だが、アップロードファイル作成作業と通常業務との兼合い等の事情で、結果的にはこの中から12,000件を対象とした。

【表１】

データの種類	件数	左の件数のうち	
		1999/12/11以前のデータ件数（日付なしを含む）	2000/01/01以降のデータ件数
ISBN 有	102,776	64,895	37,881
ISBN 無	164,466	131,175	33,291
2F 閲覧室（内数）	14575	13424	1151
合　計	267,242	196,070	71,172

＜アップロードファイルの作成＞

データをアップロードするには、まずアップロードファイルを作成しなければならない。

本学システムからダウンロードした目録データはタイトルや著者名、出版社等々全ての書誌情報が区切りもなく連なって入っているので、その中から検索キーとなるタイトルや著者名等を抽出しなければならないのである。

この作業をEXCELで行うことにしたが、12,000件の書誌情報は膨大なので、これを資料番号順に2,000件ずつに分け、それぞれを職員に１人ずつ割り当て計６人で作業を行うことにした。

各々通常業務もあるのでこの作業に専念できない状況だったが、所要時間は３日～５日といったところであろう。

＜アップロード＞

上記の作業が終わったものから順にＮＩＩへのアップロードを行った。EXCELのファイルではアップロードできないのでテキストファイルに変換して行う。検索キーとして選んだのはFTITLE(フルタイトル）＋AUTH（編著者名）＋YEAR（刊行年）である。

アップロードには下記のとおり時間が決められている。

月、火、水、金　12：00～13：00、17：00～20：00、22：00～翌朝８：00
木、土　　12：00～13：00、17：00～18：00

2,000件のデータのアップロードはかなり時間がかかることが予想されたので、主に22：00以降の時間に行った。

＜結果＞

まず最初の2,000件をアップロードしたところヒット件数は50件と予想外に低い数字であった。アップロードファイル作成時からAUTHキーをどういう形で入力するか職員のあいだで問題になっていたので原因はそこにあるだろうと考え、同じファイルからヒットしたデータを除いた1,950件のデータからAUTHキーのデータを消去して、FTITLE+YEARで再度アップロードを実行したところ299件ヒットした。これ以上検索キーを減らすことはできないので結局最初の2,000件のヒット件数は349件であった。

次に同じくFTITLE+AUTH+YEAR の検索キーで4,000件を実行したところヒット件数は110件。ノーヒットデータからAUTHキーを削除して再度実行したところ473件で計583件であった。
ヒット率は検索キーがFTITLE+AUTH+YEARでは最初の2,000件が２．５％、次の4,000件が２.75％。AUTHキーを削除してFTITLE+YEARでヒットしたものを加えれば、最初が17．１％、次が14．６％でほぼ同じような割合である。

AUTHキーをどういう形で入力するかが問題になったというのは、編著者が複数の場合最初に記述された１名でよいのか、姓と名の間を１文字分空ける必要があるのか、また原著者が西洋人の場合カタカナ表記でよいのか等々である。重要な点であるが、ＮＣ－Ａｕｔｏの操作マニュアルやインターネットなどで調べてみても必要な情報はどこにもない。そこで姓のみの前方検索でやってみたり、「編」や「著」などの役割表示を入れてやってみたりしたが、大した効果はなかった。結局FTITLE+AUTH+YEARとFTITLE+YEARの検索キーで10,000件のデータをアップロードしたのだが、【表２】の通り惨憺たる結果であった。

残りの2,000件については、AUTHキーの代りにPUB（出版社名）を検索キーとして採用した。それは、ＮＩＩのホームページに筑波大学附属図書館外３機関が実施した自動登録の実証実験報告が掲載されているのを見つけ、それによるとどの機関もAUTHを検索キーとして採用しているところはなく、むしろPUBを採用していたからである。

結果は格段に違った。【表２】を見ればわかるように、ヒット率は全体で30％近くもアップしている。全体で、というのは、最初にFTITLE+PUB+YEARの検索キーでアップロードを行い、次にノーヒットのデータからYEARを削除してFTITLE+PUBの検索キーで行ったからである。YEARを削除したのは、古いデータの中には当時の目録規則に則って出版年を刷年で記述しているものがあるが、現行の目録規則では初版の出版年を記述することになっておりＮＩＩのデータもこれに則っているからである。内訳は最初がヒット件数2,000件中609件で30．５％、次が残りの1,391件中322件で23．１％、計931件、46．６％である。上記の自動登録実証実験報告によると、ISBNの無いデータで全自動入力の場合多い所でせいぜい５割程度のヒット率で、46．６％はほぼそれに近い数字である。おそらく現状ではISBNの無いデータで全自動入力をした場合、これ以上の数字は見込めないのかもしれない。

【表２】

検索キー	ヒット件数	ノーヒット件数	計	ヒット率
FTITLE ＋（AUTH）＋ YEAR	1,779	8,221	10,000	17.8％
FTITLE ＋ PUB ＋（YEAR）	931	1,069	2,000	46.6％
計	2,710	9,290	12,000	22.6％

＜ISBN が入力されているデータのアップロード＞

試行錯誤の結果、多くのことを学んだが、12,000件中2,710件のヒット件数はあまりにも少ない。そこで ISBN のあるデータもアップロードしてみることにした。この場合検索キーは ISBN のみなのでアップロードファイルの作成は容易である。所在が２Ｆ閲覧室になっているデータのうち ISBN が入力されているものは23,233件、その中から１回目は6,000件、２回目は3,000件、計9,000件をアップロードした。結果は【表３】を参照。

１回目のヒット率が55. ３％と低いので、新たに3,000件をアップロードしてみたのだが、これも最初は1,882件のヒットで、ヒット率62. ７％だった。検索キーが ISBN ならばヒット率９割は堅いと思っていたのでショックであった。結局原因は検索キーではなく登録キーの CLN（請求記号）にあった。アップロードファイルを EXCEL からテキストファイルに変換した際、文字列を含むデータの中で、‘CLN ＝913. 6 ：IKE’ のようにシングルクオーテーションで囲われた形で変換されているものがある。どうしてそうなるのか原因は不明だが、こういうデータがＮＩＩのサーバーから無効なデータとしてはじかれていたようである。ＮＩＩへ登録するデータとしては、最低限 RGTN（資料番号）さえあれば十分なので、CLN を削除して残りの1,118件を再度アップロードしてみたところ新たに826件ヒットした。これで3,000件中計2,708件のヒット、ヒット率90. ３％と見込み通りの９割を達成した。

これで一応ヒット件数の総数は8,736件となり、現在 TRC に半自動入力及び新規入力を委託している1,400件余のデータと合わせて10,000件を超えることができた。

【表３】

	ヒット件数	ノーヒット件数	計	ヒット率
ISBN 有１回目	3,318	2,682	6,000	55.3％
ISBN 有２回目	2,708	292	3,000	90.3％
計	6,026	2,974	9,000	67.0％

（ざいぜん　としひろ）