

半自動モードを用いた自動登録システム 実証実験プロジェクト報告書

平成 17 年 3 月

半自動モードを用いた自動登録システム実証実験プロジェクト

1 はじめに

平成 16 年度から開始した国立情報学研究所遡及入力事業では、各大学の遡及入力事業の効率を高めることを目的として、1)多言語資料、2)人文・社会科学系資料、3)自動登録システム実証実験支援¹、の 3 項目を事業の柱として進めています。

自動登録システムは、平成 9 年度の CATP サーバの公開以来、国立大学図書館協議会(当時)の CATP-AUTO を始めとして、各メーカーから製品として販売されてきました。しかし、遡及入力の対象となる資料は、ISBN などの書誌レコードの同定に必要なユニークな情報が不足している例が多く、タイトルや著者名などの情報だけでは、機械的に処理を進めることが困難な場合もあり、遡及入力を効率よく行うツールとしては、力不足の面もありました。

平成 15 年に日本電気株式会社から「NC-AUTO Version 2」がリリースされました。これは、上記のような遡及入力における自動登録システムの問題を解決するため、半自動モードを備えたもので、第二世代の自動登録システムと言えます。

国立情報学研究所は、自動登録システム実証実験支援として、「半自動モードを用いた自動登録システム実証実験プロジェクト」を立ち上げました。このプロジェクトは、このシステムの半自動モードの遡及入力への効果の評価を目的として行ったものです。4 大学図書館から協力をいただき、中国語資料、洋書コレクション、新館建設に伴う所蔵レコードの登録、所蔵レコードの変更などの多様な資料を対象に実験を行うことができました。

その結果をここに「半自動モードを用いた自動登録システム実証実験プロジェクト報告書」としてまとめました。今後の各図書館における遡及入力の効率化に役立てば幸いです。

平成 17 年 3 月

半自動モードを用いた自動登録システム実証実験プロジェクト

¹ 平成 16 年度遡及入力事業の各種資料では、自動登録システム実証実験支援を「次世代自動登録システム実証実験支援」と表現していましたが、自動登録システムの今後の更なる発展を期待して、本報告書では、「次世代」という表現を実験名から除外することとしました。

2 プロジェクトの概要

2.1 プロジェクト名称

国立情報学研究所 平成 16 年度遡及入力事業 自動登録システム実証実験

2.2 目的

和書，洋書など種々の遡及資料に対して，半自動登録の有効性の検証を行い，成果を参加機関共同のレポートにて公開すること。

2.3 使用システム

日本電気製「NC-AUTO Version 2」(以下，NC-AUTO)

2.4 参加機関

- ・ 東京大学附属図書館
- ・ 三重大学附属図書館
- ・ 大阪大学附属図書館
- ・ 九州大学附属図書館

3 半自動モードを用いた自動登録システムとは

半自動モードとは、「アップロードファイルによる検索の結果、書誌を特定できなかった時に、検索語の追加や変更ができる」機能のことである。

具体的には、書誌レコードを検索した結果、書誌レコードが特定できない場合、対面検索画面（検索機能のみ。書誌レコードの修正機能はありません）が開き、処理が停止する。

該当する書誌レコードがない場合は、簡易 UIP 画面上で検索語を修正して書誌レコードを検索できる。

該当する書誌レコードが複数ヒットした場合、簡易 UIP 画面上で検索語を追加して更に書誌レコードを絞り込んだり、一覧表示されている書誌レコードから目的のものを選択することができる。書誌詳細確認画面で求める書誌レコードであることを「確定」すると、所蔵登録が行われ、次のデータに処理が進みます。ここで、該当する書誌レコードが無かった場合、キャンセルして次のデータに進むか、処理全体を中止することができる。

・日本電気株式会社製「NC-AUTO Ver.2」

ハードウェア： TCP/IP プロトコルでインターネットに接続された PC
OS： Windows 2000, XP
Java： Java 2 Runtime Environment Standard Edition v 1.4.2 以上
NC-AUTO Ver.2 ライセンスディスク

4 実験報告

4.1 東京大学附属図書館

(FA001787 東京大学総合図書館)

東京大学総合図書館では、目録業務への NC - AUTO 新バージョンの利用について検討するため以下の3種類の資料を対象に実証実験を行った。

- ・受入データを利用した業務簡素化の検討
柏図書館開館準備用図書（新刊図書その他、退官教官から寄贈された旧蔵書を含むため幅広い年代の資料から構成される半遡及に近い資料群）
総合図書館新刊和洋図書（新刊図書の登録状況を探る）
- ・半自動モードの有効性の検討
総合図書館洋書遡及入力資料

4.1-1 柏図書館開館準備用図書

柏図書館では平成16年5月の仮開館以来、図書の貸出業務を早急に実施する必要があり、開館準備用に用意された図書約6,000冊（新規購入及び寄贈された図書）の目録入力之急務となった。そこで受入データを利用してアップロードファイルを作成し、NC-AUTOを利用して短期間に集中して大量登録を行うことを計画した。

4.1-1.1 実験対象

対象資料	コレクション名	東京大学柏図書館開館準備用図書
	冊数	6,048冊
	和洋区分	和5,155冊・洋書893冊
	言語割合	日本語87% 英語12% 独語+仏語1%
	年代	1928~2004(1990年~2004年を中心)
	予想ヒット率	90%
	特徴	自然科学系の新刊図書が中心

実験環境	期間	平成16年7月~8月
	実日数	・請求記号付与 20日 ・アップロードファイル作成 5日 ・登録 5日 ・データ確認とエラー処理 10日
	人数	3名
	PC台数	1台
	OS	Windows XP Home edition

4.1-1.2 実験作業手順

- 1) 原ファイル準備
- 2) 請求記号の決定と入力
- 3) アップロードファイル作成
- 4) NACSIS-CAT に登録 (全自動)
NC-AUTO で NACSIS-CAT に登録。
- 5) エラー処理
全自動モードのエラーファイルを確認し, エラー処理。
 - 5-1) NACSIS-CAT に登録。(半自動)
エラーファイルから不要なデータを削除して半自動モード用のアップロードファイルを作成。半自動モードで NACSIS-CAT に登録。
 - 5-2) NACSIS-CAT に登録 (全自動)
エラーファイルを確認し, 必要な訂正を行ったうえでアップロードファイルを作成。全自動モードで NACSIS-CAT に登録。
 - 5-3) マニュアル処理
該当書誌レコードがない等 NC-AUTO で対応できないエラーについて手入力で処理。
- 6) ローカルデータベース登録
ローカルシステム取り込み用プログラムを使用してローカルデータベースに登録。
- 7) 登録データ確認
NC-AUTO のバージョンアップに伴い, ローカルシステム取り込み用プログラムも改修されたためその動作確認も兼ねて最終的なデータ確認を OPAC で行った。

4.1-1.3 アップロードファイル作成方法

- 1) 受入データ抽出
ローカルシステムの受入データからアップロードファイル作成に必要なデータを抽出。(CSV 形式でデータ取り出し) Excel におとして原ファイルを作成。
抽出項目は資料 ID, 書名, 出版者, 出版年, ISBN, 所在, 巻冊次の 7 項目。
* 基本的な検索キーは ISBN としたが, 資料の確認等に備え, 抽出項目は多めに指定した。
- 2) 請求記号の決定と入力
1) で作成された原ファイルに請求記号を入力。総合図書館, 駒場図書館と所蔵が重複する資料については原ファイルの ISBN を検索キーとしてローカルデータベースから両図書館の請求記号を原ファイルに書き出し, それを利用して付与した。
- 3) 原ファイルの加工
2) で請求記号が入力された原ファイルをアップロードファイル用に加工。(検索データと登録データに加工。) テキスト形式で保存し, アップロードファイル完成。

4.1-1.4 検索キー

対象とした資料の大半に ISBN が付与されており, 受入データにも ISBN が入力され

ていたため基本的な検索キーを ISBN とし、ISBN が付与されていない資料に限り、他の検索キーを使用した。

初回の全自動モード登録で生成されたエラーファイルを利用して半自動モード用のアップロードファイルを作成・利用したため検索キーの変更は行っていない。

検索キー	冊数（全自動）	冊数（半自動）
ISBN	5,791	240
FTITLE	28	5
FTITLE+YEAR	226	31
FTITIE+AUTHOR	2	0
ISBN+YEAR	1	0
合計	6,048	276

4.1-1.5 実験結果

4.1-1.5.1 全自動モードでの結果

全自動モード	処理件数合計		6,048
	成功		5,083
	エラー		965
	エラー 内訳	該当書誌レコードが存在しません	72
		書誌レコードを特定できないため更新できません	357
		該当書誌は親書誌のため更新できません	1
		該当書誌はVOLが複数存在するため更新できません	63
		一致するVOLがあるため処理をスキップしました	339
		該当書誌に一致するVOLが存在しないため更新できません	7
その他		126	

・登録開始当初、NC-AUTO 自体に VOL の障害が発生しており、それに気づくまでに VOL に関するエラーが大量に発生した。（障害は、現在解消されている。）

・ISBN のみを検索キーとしたため重複書誌レコード、刷違いの別書誌レコード、複数図書に付与された同じ ISBN 等が原因となってエラーが大量に発生した。

4.1-1.5.2 半自動モードでの結果

半自動モード	合計件数	276
	成功	123

	エラー		153
	エラー 内訳	該当書誌レコードが存在しません	40
		書誌レコードを特定できないため更新できません	32
		該当書誌はVOLが複数存在するため更新できません	34
		一致するVOLがあるため処理をスキップしました	42
		該当書誌に一致するVOLが存在しないため更新できません	4
		その他	1

4.1-1.5.3 エラーの再処理の結果

エラーの再処理	合計件数		545
	成功		378
	エラー		167
	エラー 内訳	該当書誌レコードが存在しません	43
		書誌レコードを特定できないため更新できません	36
		該当書誌はVOLが複数存在するため更新できません	36
		一致するVOLがあるため処理をスキップしました	42
		該当書誌に一致するVOLが存在しないため更新できません	4
その他		6	

4.1-1.6 評価

・全自動モードで書誌レコードを特定できないためにエラーとなった資料に対して、生成されたエラーファイルを再利用してアップロードファイルを作成し、半自動モードで利用できるのは有効な機能である。しかしながら多巻もの場合、検索キーによっては書誌レコードが特定できてもアップロードファイルにVOLの指定がないとエラーになるのは不便である。半自動モードでは、書誌レコードの特定とVOLの指定が同時にできることが必要な機能である。

例) 検索キーがタイトル、著者等で該当する書誌レコードが複数件ヒットした場合、対面検索画面で書誌レコードを特定し、書誌詳細表示画面で表示されるVOLを指定してもアップロードファイルでVOLを指定していないとエラーになる。(VOLに対応するISBNを検索キーに指定している場合は、アップロードファイルにVOLの指定がなくても正常登録される。)

- ・ エラーファイルに書き込まれたエラーメッセージを見るとエラー処理する側にとっては同じ内容のエラーに対して異なる言い回しのエラーメッセージがでていてエラーメッセージ自体わかりにくい。ヘルプ等をもてもエラーメッセージに対する記述がなく、初めて NC-AUTO を利用する者にとってはわかりにくい。特に今回の実証実験で東大は最初に登録を開始したためプログラムの不具合のため出たエラーなのか NC-AUTO の機能として出るべくして出たエラーなのか判断しかねる場合があり、慣れるまでに苦慮した。
- ・ 今回の実証実験では資料の性格により 200 冊～900 冊程度のグループにわけ、登録を行った。参考までに初回全自動モードによる登録にかかった所要時間を示すと以下のとおり。(PM10 時以降に予約登録で実施した結果)
 - 200 冊程度 40 分～1 時間
 - 500 冊程度 1 時間 50 分～2 時間 50 分
 - 700～900 冊程度 3 時間 20 分～4 時間 30 分
 1 冊登録するのにかかる所要時間は最短で 12 秒、最長でも 20 秒前後で処理が完了している計算になる。日中の混在した時間帯に比べ、格段に速いスピードで処理されていることがわかる。短期間に大量に登録する必要がある場合には、威力を発揮するものとする。

4.1-2 総合図書館新規購入図書

4.1-2.1 実験対象

対象資料	コレクション名	新規購入図書
	冊数	2,037 冊
	和洋区分	和 1,838 冊・洋書 199 冊
	言語割合	日本語 90% 英・仏語 10%
	年代	1990～2004 年
	予想ヒット率	95%
	特徴	新刊和洋図書

実験環境	期間	平成 16 年 9 月～2 月 15 日(現在も継続中)
	実日数	10 日
	人数	3 名
	PC 台数	3 台
	OS	Windows 2000

4.1-2.2 実験内容

* 通常の目録業務として NC-AUTO を利用した実験を行っており、特に分業体制はとらず

に 1) から 7) までの作業を同じ人間がトータルに処理。

1) 原ファイル作成

2) 請求記号の決定と入力

3) アップロードファイル作成

4) NACSIS-CAT に登録 (全自動)

NC-AUTO で NACSIS-CAT に登録。

5) エラー処理

全自動モードのエラーファイルを確認し, エラー処理。

5-1) NACSIS-CAT に登録。(半自動)

エラーファイルから不要なデータを削除して半自動モード用のアップロードファイルを作成。半自動モードで NACSIS-CAT に登録。

5-2) マニュアル処理

該当書誌レコードがない等 NC-AUTO で対応できないエラーについて手入力で処理。

6) ローカルデータベース登録

ローカルシステム取り込み用プログラムを使用してローカルデータベースに登録。

7) 登録データ確認

NC-AUTO のバージョンアップに伴いローカルシステム取り込み用プログラムも改修されたためその動作確認も兼ねて最終的なデータ確認を OPAC で行った。

4.1-2.3 アップロードファイル作成方法

1) 受入データ抽出

ローカルシステムの受入データからアップロードファイル作成に必要なデータを抽出。(CSV形式でデータ取り出し) Excel におとして原ファイルを作成。

抽出項目は資料 ID, 書名, 出版者, 出版年, ISBN, 所在, 巻冊次の 7 項目。

2) 請求記号の決定と入力

1) で作成された原ファイルに請求記号を入力。

3) 原ファイルの加工

2) で請求記号が入力された原ファイルをアップロードファイル用に加工。(検索データと登録データに加工。) テキスト形式で保存し, アップロードファイル完成。

4.1-2.4 検索キー

検索キー	冊数	冊数 (半自動で再処理)
ISBN	2,033	22
FTITLE	4	0
合計	2,037	22

4.1-2.5 実験結果

4.1-2.5.1 全自動モードでの結果

全自動モード	処理件数合計	2,037	
	成功	1,986	
	エラー	51	
	エラー内訳	該当書誌レコードが存在しません	26
		書誌レコードを特定できないため更新できません	20
		該当書誌は親書誌のため更新できません	1
一致するVOLがあるため処理をスキップしました		4	

- ・対象となった資料が市販の新刊図書のため、和書の場合はほとんど問題なく登録できた。
- ・洋書（コンピュータ関連の図書）は刊行後数年経過しているにもかかわらず書誌レコードが登録されていない場合があり、登録した冊数に比し、「該当書誌レコードなし」エラーの発生率が高くなっている。

4.1-2.5.2 半自動モードでの結果

半自動モード	合計件数	22	
	成功	16	
	エラー	6	
	エラー内訳	該当書誌レコードが存在しません	1
		該当書誌は親書誌のため更新できません	1
		一致するVOLがあるため処理をスキップしました	4

4.1-2.5.3 エラーの再処理の結果

2 回目の全自動モードの登録は必要ないため行っていない。

4.1-2.6 評価

- ・和書の新刊図書については初回の全自動モード登録でほぼ 100%に近い成功率に達しており、半自動モードを利用するまでもない状況である。自動登録に最適な状況といえよう。NC-AUTO の有効性が確認できた。特に市販の新刊図書では付与されている ISBN を検索キーとして利用することでほぼ問題なく登録することができ、エラー処理も不要なため受入データ等利用できるデータがある場合には業務の簡素化が可能であろう。

- ・ しながらアップロードファイルの作成（請求記号の決定と入力を含む）から登録後のデータ点検までにかかる時間をトータルに考えると少量の冊数に対しNC-AUTO を利用しても一冊ずつ登録するのと時間的に違いがなく，NC-AUTO 利用のメリットは少ない。NC-AUTO の利用は少なくとも 100～200 冊をひとつの単位とする大量の図書を集中して登録する際に力を発揮するのではないかと考えられる。

4.1-3 総合図書館遡及入力資料

東京大学では平成 7 年度から 10 年計画で全学遡及入力事業を展開しており，総合図書館でのこれまでの経験に基づき，ヒット率が期待できそうな資料を対象を選び，実証実験を行った。

4.1-3.1 実験対象

対象資料	コレクション名	洋書遡及入力図書
	冊数	3,098 冊
	和洋区分	洋書
	言語割合	英語 70% 独語 15% その他 15%
	年代	1910～1960 年代
	予想ヒット率	65%
	特徴	図書館学・レクラム文庫等の叢書・言語学

実験環境	期間	平成 16 年 7 月～1 月（現在も継続中）
	実日数	38 日
	人数	1 名
	PC 台数	1 台
	OS	Windows 2000

作業は当館の遡及入力担当の非常勤が行った。

4.1-3.2 実験内容

- 1) 資料の移動
書庫から資料を事務室まで移動。
 - 2) アップロードファイル作成
 - 3) NACSIS-CAT に登録（全自動）
NC-AUTO で NACSIS-CAT に登録。
 - 4) エラー処理
全自動モードのエラーファイルを確認し，エラー処理。
- 4-1) NACSIS-CAT に登録。（半自動）
エラーファイルから不要なデータを削除して半自動モード用のアップロードファイル

を作成。半自動モードで NACSIS-CAT に登録。

4-2) NACSIS-CAT に登録（全自動）

エラーファイルを確認し，必要な訂正を行ったうえでアップロードファイルを作成。

全自動モードで NACSIS-CAT に登録。

4-3) マニュアル処理

該当書誌レコードがない等 NC-AUTO で対応できないエラーについて手入力で処理。

5) ローカルデータベース登録

ローカルシステム取り込み用プログラムを使用してローカルデータベースに登録。

6) 登録データ確認

NC-AUTO のバージョンアップに伴いローカルシステム取り込み用プログラムも改修されたためその動作確認も兼ねて最終的なデータ確認を OPAC で行った。

7) 資料の返却

事務室から書庫へ資料を返却。

4.1-3.3 アップロードファイル作成方法

1) Excel のシートにアップロードファイルに必要な項目を入力。

FTITLEKEY,AUTH,YEAR 等の検索項目と LOC , RGTN , CLN 等の登録データを入力。

2) 1) のファイルをテキスト形式で保存し，アップロードファイル完成。

4.1-3.4 検索キー

ISBN 導入以前の資料のため ,CATP-AUTO 前バージョンの実証実験の報告*1 を参考に以下の検索キーを設定した。

検索キー	冊数	冊数 (半自動で再処理)
FTITLE	5	2
FTITLE+AUTHOR	27	40
FTITLE+AUTHOR+PUB	46	82
FTITLE+AUTHOR+YEAR	2,034	123
FTITLE+AUTHOR+YEAR+PUB	442	34
FTITLE+AUTHOR+YEAR+ISBN	86	0
FTITLE+AUTHOR+YEAR+NCID	93	0
FTITLE+YEAR	323	8
FTITLE+YEAR+PUB	26	3
FTITLE+YEAR+ISBN	15	0
FTITLE+ISBN	1	0
FTITLE+PUB	0	7
合計	3,098	299

*1 「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 LV p.38-53,

1999.3)

4.1-3.5 実験結果

4.1-3.5.1 全自動モードでの結果

全自動モード	処理件数合計	3,098	
	成功	1,908	
	エラー	1,190	
	エラー内訳	該当書誌レコードが存在しません	873
		書誌レコードを特定できないため更新できません	249
		該当書誌は VOL が複数存在するため更新できません	23
		一致する VOL があるため処理をスキップしました	9
		該当書誌に一致する VOL が存在しないため更新できません	14
該当書誌は VOL が空値でないため更新できません	9		
その他	13		

・英語で書かれた図書を中心にヒット率が高そうな分野を選んだが、結果としては「該当書誌レコードなし」エラーが多く発生した。

・検索キーを工夫して対応したが「書誌レコードが特定できない」エラーが発生しており、全自動モードでの検索キー選択の難しさがわかる。

4.1-3.5.2 半自動モードでの結果

半自動モード	合計件数	299	
	成功	207	
	エラー	92	
	エラー内訳	該当書誌レコードが存在しません	36
		書誌レコードを特定できないため更新できません	46
該当書誌に一致する VOL が存在しないため更新できません		8	
該当書誌は VOL が空値でないため更新できません		2	

・初回全自動モードの登録で「書誌レコードが特定できない」エラーが大量に発生したが、

半自動モードで目視しながらの登録で成功率が向上した。

4.1-3.5.3 エラーの再処理の結果

エラーの再処理	合計件数	103	
	成功	60	
	エラー	43	
	エラー 内 訳	該当書誌レコードが存在しません	34
		一致するVOLがあるため処理をスキップしました	3
その他		6	

・VOLエラーに対してVOLの指定を修正し、成功率が向上した。

4.1-3.6 評価

・ISBNを検索キーとして利用できない遡及入力の自動登録では、検索キーの選択がポイントとなる。FTITLEキーとその他の検索キーの組み合わせで検索の精度を上げることは可能だが、遡及入力ではそれでも書誌レコードの特定ができない例が多く、半自動モードの存在はヒット率の向上に有効である。柏図書館の報告にも書いたように書誌レコードの特定と同時にVOLの特定も行えるようになればさらにヒット率を向上させることが可能となり、必要な機能と考えられる。(VOLの表記にコレが多く、NACSIS-CATを検索してからアップロードファイルを作成するのでは自動登録の意味がなく、本末転倒となろう。)

・登録後のデータ確認で異なる書誌レコードに登録されている例が約30件見つかった。原因は登録した図書に対応する書誌レコードがまだNACSIS-CATに登録されておらず本来ならば新規に書誌レコードを作成すべき例であった。指定した検索キーに合致している以上、自動登録では正常処理されるのは当然のことでこのあたりが自動登録の限界といえよう。ただし発生率にして1%に満たない数字であることを勘案するなら登録後のデータ確認をきちんと行うことで修正は可能であり、問題にはならないのではないかと考える。

・遡及入力では、「該当書誌レコードなし」エラーが多発しているため新刊図書を中心とした入力作業に比べ、エラー処理に時間がかかっている。

・対象資料の選択により大きくヒット率が変わってくる点やトータルな作業時間を考えるとNC-AUTOを利用しての遡及入力作業では、効率的に作業を行うための体制を整備することが必要であろう。例えば今回の作業では同一人がすべての作業を行ったが、作業をアップロードファイルの作成・登録・エラー処理・データ確認等にわけ、流れ作業的に登録作業を行うことでさらに効率が上がったのではないかと考えられる。

4.2 三重大学附属図書館

(FA002564 三重大学附属図書館)

三重大学附属図書館では、目録業務への NC - AUTO 新バージョンの利用について検討するため「土井治氏旧蔵書コレクション」を対象に実証実験を行った。

平成 16 年 11 月上旬に当コレクションを用いた展示会を予定していたため、9 月中に入力を終えることで工程表を作成した。実際には、作業の開始が諸般の理由で 1 ヶ月遅れたため、10 月中には作業を終了しなければならない、ということで日程的には日が押しつけてきついついところもあったが、もともとの工程表が余裕をみて作成してあったため、10 月末に作業を終えることができた。

余裕をみて工程表を作成していたように、今回の作業では、半自動の検証に重きをおき、作業効率の検証は特に行わなかった。

4.2.1 実験対象

対象資料	コレクション名	土井治氏旧蔵書コレクション
	冊数	4,553 冊の内から 4,275 冊
	和洋区分	和 1 : 洋 6
	言語割合	日本語 1 : 英語 6
	年代	~1850: 1%, ~1920: 11%, ~1969: 44%, 1970~: 41% 刊年不明: 3%
	予想ヒット率	76%
	特徴	・ G. Eliot, G. Gissing, G. Meredith 等の 18・19 世紀英文学研究書 約 2,700 冊 ・ 73 歳から始めた年間 1 万ページ読破の対 象となった英文書 (ホラー, 推理等) を含 む娯楽・大衆小説 約 750 冊 ・ 俳句関係書 約 165 冊

実験環境	期間	平成 16 年 8 月 17 日(火)~10 月 28 日(木)
	実日数	75 人日 (37.5 日)
	人数	2 名
	PC 台数	2 台
	OS	Windows XP

4.2.2 実験作業手順

以下の手順で行った。

- 1) アップロードファイルに書架番号の入力

- 2) アップロードファイルの整形
- 3) アップロードファイルによる自動登録
- 4) ブルーファイルによる登録結果のチェック(ブルーファイルに出力されたアップロードファイルのデータと登録先のNACSIS-CAT データをタイトルと出版年の2項目について、その異同をチェックするプログラム(perl)による)
- 5) エラーファイルから半自動用のアップロードファイル作成
- 6) NC-AUTO 半自動によるチェック及び入力
- 7) エラーファイルによる複本のローカルへの手動登録
- 8) 参照 MARC を利用した新規入力とオリジナル入力
- 9) ローカルシステムからのブルーファイルによる上記入力データのチェック

4.2.3 アップロードファイル作成方法

コレクション受入時に寄贈者に渡した「遺贈図書目録」(ISBN 記載あり)と会計上の書類, 物品請求及命令書続紙(登録番号の記載あり)が共に Excel で作成してあったので, この2ファイルを結合して一本のアップロードファイルとした。

各ファイルの項目は下記のとおり。

- ・「遺贈図書目録」原稿ファイル
 - 項目(著者名, 書名, 出版者, 出版年, シリーズ名, ISBN)
- ・物品請求及命令書続紙
 - 項目(登録番号, 書名, 著者名, 出版者, 出版年)
- ・アップロード用ファイル
 - 項目(書名, 著者名, 出版者, 出版年, ISBN, 登録番号, 図書 ID, 簡易分類番号, 書架番号)

このアップロード用ファイルに, 下記の登録項目名と検索キー名を Excel の文字列関数の機能を利用して各データの先頭に貼り付け, その後, 列の結合・分割などにより最終的なアップロードファイルを作成した。書架番号は, CLN の 2 段目と 3 段目に組み込んだ。

- ・登録項目 LOC=@, VOL=, CLN=, RGTN=, LDF=
 - ・検索キー (ISBNKEY:, FTITLEKEY:, PUBLKEY:, YEARKEY:)
- 登録項目の LDF には図書 ID(@ID:), 配置コード(@LC:c_d)等を入力

入力例:

```
LDF=@ID:504005367!?!@RGTNC:1!?!@AR:01!?!@LC:c_d!?!@SC:01!?!@ACPTP:2000!?!
@TP:01!?!@PH:01!?!@REFTP:0!?!@ST:20!?!
```

図書 ID 以外は各レコード共通の固定項目で, 図書 ID は登録番号から自動生成させた

4.2.4 検索キー

ISBN があるものは, ISBNKEY のみで検索キーとした

ISBN がないものは, 基本的に FTITLEKEY, PUBLKEY, YEARKEY の 3 個のキーを組み合わせた。

有効な検索キーとしては、既に岩澤尚子（香川大）「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」（平成 10 年度第 2 回総合目録データベース実務研修・研修レポート）等に検証されているので、その結果を利用させていただいた。ただし、_AUTH_、_TITLE_等の仮想キーは既入力データの編集に手間がかかるため使用せず、これらも含めて、今回は検索キーの検証には重点をおかなかった。

KEY として使用しなかったタイトル関連情報や出版者名の完全形は、後でチェックの時に必要になるので元のデータを「/*」でコメントアウトしておいた。

4.2.5 実験結果

4.2.5.1 全自動モードでの結果

全自動モード	処理件数合計		4,275
	成功		2,142
	エラー		2,133
	エラー内訳	一致する VOL があるため処理をスキップしました	276
		該当書誌に一致する VOL が存在しないため更新できません	17
		該当書誌は VOL が空値でないため更新できません	11
		該当書誌は VOL が複数存在するため更新できません	10
		該当書誌レコードが存在しません	1,124
		書誌レコードを特定できないため更新できません	671
システムエラー		24	

システムエラー24件の内訳は、下記の2種のエラーによるものである。
(メッセージ例)

Invalid CATP-Field (CAT130201 /*PUBLKEY) /*PUBLKEY:Little, Brown
6件

Invalid CATP-Field(CAT130201 FTITLEKEY=) "FTITLEKEY:Moby Dick ; or, the white whale "
18件

6件は、「/*」でコメントアウトしたデータで、出版者名にコンマのあるもの。18件は、Excel からテキストデータに保存する時に、引用符の削除をし忘れたものであった。

全自動を終了した段階で不適当な書誌に所蔵を付けたものは、18件であった。

4.2.5.2 半自動モードでの結果

半自動モード	合計件数	2,133
	成功	617
	エラー	1,516

	エラー内訳	一致する VOL があるため処理をスキップしました	276
		該当書誌に一致する VOL が存在しないため更新できません	17
		該当書誌は VOL が空値でないため更新できません	11
		該当書誌は VOL が複数存在するため更新できません	10
		該当書誌レコードが存在しません	1,124
		書誌レコードを特定できないため更新できません	54
		システムエラー	24

4.2.5.3 エラーの再処理の結果

エラーの再処理	合計件数		1,516
	成功		623
	エラー		893
	エラー内訳	該当書誌レコードが存在しません	893

半自動モードにより処理できなかったものは、複本（一致する VOL があるため処理をスキップしました）276 件の他、具体的な内容は以下のとおりであった。

[該当書誌に一致する VOL が存在しないため更新できません]

3 巻ものを 1 冊に合冊したもので、元の標題紙の情報からアップロードファイルでは、VOL=v. 1 のようにしていた

編者が異なるため、固有のタイトルとなるものをアップロードファイルでは、VOL=夏のようにしていた

NC 書誌では VOL=Pt. 1 であるものを、アップロードファイルでは、VOL=v. 1 のようにしていて巻号名の呼び方が異なっていた

[該当書誌は VOL が空値でないため更新できません]

すべて、NC 書誌では VOL=:: pbk. のみ、あるいは、VOL=:: hbk.のみとなっているものをアップロードファイルでは空欄であったもの

[該当書誌は VOL が複数存在するため更新できません]

Oxford University Press と Cambridge University Press 同時出版もので、VOL:: oup VOL:: cup のように一つの書誌に出版者毎の VOL が記述されていた

一つの書誌に同一 ISBN の VOL が記述されていた

[該当書誌レコードが存在しません]

893 件はメッセージどおり NACSIS-CAT に書誌がなく、新規に書誌作成を要するものであった。そして、91 件はそれら新規作成書誌へ所蔵を付けるものであった

残り 140 件は、アップロードファイルでの単純な入力ミス、和書での出版者名の揺れ（改造社、改造社出版）、洋書での出版者名の揺れ（Aubier Montaigne が書誌では Aubier とな

っていた)などの他,最も多いのは,出版年の相違で58件であった。

[書誌レコードを特定できないため更新できません]

特定不可ということで,この段階まで残ったものは,書誌の同定がなされたものである。

内訳は,重複レコード1件,出版年不明のものが12件,親書誌の有無4件,出版者の相違4件,ページ数の相違3件,大きさの相違1件,ISBNの相違4件等であった。

4.2.6 評価

今回の実験結果を一つの表にまとめると,下表のようになる。

NC-AUTO による入力					マニュアルによる入力			
全自動			半自動		NC 書誌 にヒット	最終 ヒット率	新規作成	
正常	エラー	ヒット率	ヒット	修正 ヒット率			参照	リジナル
2142	2133	50%	617	65%	623	79%	400	493

マニュアルによる入力まで含めて,最終的にNACSIS-CATにヒットしたものは,当初の予想ヒット率76%に近い結果(79%)となったが,細かく検討していくと資料の性格により予想とは異なった結果となった点 ISBNを持ちながらほとんどヒットしなかった資料群があったこととNACSIS-CATが学術図書館中心のデータベースであることを考慮すると,最終的なヒット率は予想外に高かった,と結論できるのではないだろうか。新規書誌作成についても1,100件程度になるものと予想していたが,493件と少なかったことがヒット率の高さを傍証している。

しかしながら,今回この実験を終えて感じることは,この規模・内容のコレクションで全自動登録でのヒット率が50%,半自動までのヒット率が65%であったことや,マニュアルでは623件がNACSIS-CATにヒットしたなどということは,ファクトデータとして多少の利用価値はあるものの,この数値を改善しようとするのは余り意味のないことではないかということである。全自動・半自動・マニュアルそれぞれの中でのヒット率の割合は,アップロードファイル作成の巧拙如何によって若干変わることはあっても,最大限79%の枠内で変動するだけである。マニュアルでのヒット623件の内,複本等を除いた252件がすべて半自動でヒットしたとしても,5ポイント改善して65%が70%になり,全自動・半自動・マニュアルそれぞれのヒット率構成が50%,65%,79%から50%,70%,79%へと変わるだけである。

今回の実験の目的が半自動の検証であったとはいえ,当初は,目録登録作業を省力化するための方策,つまりは可能な限り自動登録で済ませるための方策をも探ろうとしていた。しかし,2,142件中18件とはいえ,誤登録が発生するという事は,プルーフファイルによる目視でのチェックが欠かせないということである。この作業が必須であることは,自動登録のメリット感を減じさせることである。今回は,プリントアウトすれば2,000枚近くになるプルーフリストを一枚一枚肉眼で確認するのではなく,間にプログラムをかませ機械的に確認することで省力化を図ったが,プルーフリストに目を通す労力と不適當な書

誌に所蔵が付いた後の処理の煩雑さを考えると、最初から半自動で目視による確認をしながら作業を進めた方が能率的だという考え方も成り立つ。

工夫次第で半自動モードでのヒット率が改善される感触を得たので、全自動による処理と半自動での処理をうまく組み合わせることで、半自動モードを含む自動登録システムのメリットを引き出せるのではないかと考えている。

4.3 大阪大学附属図書館

(FA002848 大阪大学附属図書館)

当館では、本館に所蔵する「中国方志叢書」を実験対象資料とした。同資料は約 4,700 冊を所蔵しているコレクションであるが、その約 3 分の 2 が目録未登録となっており遡及入力の必要性が高かったものである。

なお、該当する所蔵データがないものがあり、また、存在しても書名の記述が一定でなくアップロードファイルのデータには使えないため、データ作成はすべて資料現物に基づいて行った。

4.3.1 実験対象

対象資料	コレクション名	中国方志叢書
	冊数	3,226 冊
	和洋区分	和
	言語割合	中国語 100%
	年代	1960 ~ 1970 年代
	予想ヒット率	ほぼ 100%
	特徴	全冊が一つの叢書である

実験環境	期間	平成 16 年 11 月 8 日 ~ 11 月 30 日
	実日数	16 日
	人数	2 名
	PC 台数	2 台
	OS	WindowsXP

4.3.2 実験作業手順

1) 対象資料の移動

対象資料を書庫 6 階書架から作業室に移動する。

背ラベルが読みとれないものは張り替え

資料番号ラベル未貼付の資料については資料番号ラベルを印刷・貼付

2) 登録用データ作成

Excel を使用する（「データ入力用」「整形用」の 2 枚のシートを用意）。

「データ入力用」シートにデータを入力する。

自動的に「整形」シートに各フィールド名（検索，登録とも）を追加した形式が表示される。入力内容をチェックする。

3) アップロードファイル作成

「整形用」シートをテキスト形式で保存し，全登録データに共通の項目（LOC==@ など）を追加してアップロードファイルを完成させる。

4) NC-AUTO による自動登録（全自動モード）

5) NC-AUTO による自動登録（半自動モード）

全自動モードのエラーファイルのファイル名を変更し，アップロードファイルとして半自動モードで自動登録を行う。

6) エラーの再処理

半自動モードで出たエラーは NC 書誌レコード内容を確認して原因を特定し，アップロードファイルの VOL 記述等を修正した。その後，再び半自動モードにより自動登録を行った。

7) 登録内容の確認

ヒットした書誌レコードが適切かどうかブルーファイルにより確認する（現物との照合）。

8) 対象資料の移動

処理が終わった資料を作業室から書庫 6 階書架に戻す。

9) 当館による事後作業

- ・自動登録できなかった資料の処理（レコード調整など）
- ・当館業務システムによるローカルシステムへの書誌・所蔵データの登録作業

4.3.3 アップロードファイル作成方法

1) 登録用データ作成

Excel を使用する（「データ入力用」「整形用」の 2 枚のシートを用意）。

「データ入力用」シートにデータを入力する。

入力項目：検索語，VOL，CLN，RGTN

（例）

検索語	検索語	VOL	RGTN	CLN	資料番号のチェック	ヒット確認
華北地方	第 3 号		07903065972	04-08/SEI	2	OK
華北地方	第 4 号	1	07903065980	04-08/ZOK/1	0	OK

自動的に「整形」シートに各フィールド名（検索，登録とも）を追加した形式が表示される。入力内容をチェックする。

（例）

TITLE:華北地方	TITLE:第 3 号		RGTN=07903065972	CLN=K:K2-11-04-08/SEI
TITLE:華北地方	TITLE:第 4 号	VOL=1	RGTN=07903065980	CLN=K:K2-11-04-08/ZOK/1

2) アップロードファイル作成

「整形用」シートをテキスト形式で保存し、全登録データに共通の項目（LOC==@ など）を追加してアップロードファイルを完成させる。

（例）（ はタブ， は改行を示す）

```

-----
PTBID::BN07219067  LOC==@  CPYR==@  CPYNT==@
LDF==1030150  TITLE:華北地方  TITLE:第 3 号  RGTN=07903065972
CLN=K:K2-11-04-08/SEI
TITLE:華北地方  TITLE:第 4 号  VOL=1  RGTN=07903065980
CLN=K:K2-11-04-08/ZOK/1
-----

```

4.3.4 検索キー

「中国方志叢書」シリーズは、全巻共通の親書誌レコード< BN07219067>の下に子書誌レコードが作成されており、地方名とシリーズ番号からほぼ一意に書誌レコードを特定可能である。この特性を利用して検索キーには以下のフィールドを指定、作業の効率化を図った。

- ・親書誌レコード ID (PTBID)
- ・中位の書誌のタイトル (地方名, なければ省名) (TITLEKEY)
- ・シリーズ番号 (TITLEKEY)

また、半自動モードにおいても同じ検索キーを使用し、対面検索画面で適宜変更した。

4.3.5 実験結果

4.3.5.1 全自動モードでの結果

全自動モード	処理件数合計		3,226
	成功		3,083
	エラー		143
	エラー 内 訳	該当書誌レコードが存在しません	18
		書誌レコードを特定できないため更新できません	32
		該当書誌に一致するVOLが存在しないため更新できません	2
該当書誌はVOLが空値でないため更新できません		90	
Connection timed out		1	

全自動モードによる登録率は95.7%となり、ほとんどの資料が登録できた。

4.3.5.2 半自動モードでの結果

半自動モード	合計件数	143	
	成功	64	
	エラー	79	
	エラー 内 訳	該当書誌レコードが存在しません	5
		書誌レコードを特定できないため更新できません	3
該当書誌に一致する VOL が存在しないため更新できません		2	
	該当書誌は VOL が空値でないため更新できません	69	

全自動モードでのエラーのうち、半自動モードの書誌検索により書誌レコードを特定できたものは以下のとおりである。

- ・「該当書誌レコードが存在しません」
シリーズ番号の違い(「第」「号」が抜けている)、検索語の入力ミス(1件)など
- ・「書誌レコードを特定できないため更新できません」
複数のレコードにヒット

また、「該当書誌に一致する VOL が存在しないため更新できません」および「該当書誌は VOL が空値でないため更新できません」のエラーは、書誌検索時に一致 VOL がなかった場合に半自動モードの対面検索画面が開くことなく処理がすべてスキップされたエラーである。

(例) VOL の記述の違い

- ・アップロードファイルでは「1」、書誌では「1:舊志」
- ・アップロードファイルではVOLなし、書誌では「全」

4.3.5.3 エラーの再処理の結果

エラーの再処理	合計件数	79
	成功	76
	エラー	3
	エラー 内 訳	該当書誌に一致する VOL が存在しないため更新できません
該当書誌は VOL が空値でないため更新できません		1

最終的なエラーはレコード調整が必要な書誌レコードであった。

なお、このようなエラーの再処理が行えたのは、作業者が自動登録作業に慣れており、エラー理由を指定し、アップロードファイルを修正できたためである。

4.3.6 評価

事前の予想とは異なり、作業担当者に中国語に関する知識はそれほど必要なかったが、半自動モードでは NACSIS-CAT の検索スキルが必要であった。

当館での作業からは、「半自動モード」を中心に NC-AUTO に改善の余地がかなりあると思われた。特に、「4.3.5.2 半自動モードでの結果」で述べたように、書誌検索時に一致VOLがなかった場合に半自動モードの対面検索画面が開かずに「処理スキップ」または「アップロードファイルのVOLを使用」の処理が行われる点は、改善を強く要望したい。書誌を確認し、作業者が選んだVOLに所蔵登録できる機能が必要である(それが可能であれば、「4.3.5.2 半自動モードでの結果」のエラー件数は10件以下になったと思われる)。

一方、当館での対象資料の場合、約3,200冊の現物からの遡及入力を16日間で終えることができたのは効率的だったと言える。それは、対象資料が一つの親書誌の下に子書誌レコードが概ね一定の形で登録されているシリーズものだったため、最初に決定した検索キーで書誌レコードがほぼ同定できたことによる。

4.4 九州大学附属図書館

4.4-1 (FA022084 九州大学附属図書館 筑紫分館)

九州大学では、平成16年4月に筑紫分館が新装開館し、数万件に及ぶ図書資料の移動とともに、目録データを修正する必要が生じた。そこで、移動のメンテナンスが完了したローカル所蔵レコードをもとにアップロードファイルを作成し、NACSIS-CATへの自動所蔵登録を行った。

4.4-1.1 実験対象

図書

対象資料	コレクション名	筑紫分館全蔵書(図書)
	冊数	32,306冊
	和洋区分	和 59, 洋 41%
	言語割合	jpn 58.3, eng 39.0, ger 1.6, rus 0.3, fre 0.3%
	年代	1900年代 91.9, 2000年代 7.9%
	予想ヒット率	95%以上
	特徴	理工系図書
実験環境	期間	2004年7月14日～8月9日 NIIと相談のうえ、土曜9:00～18:00も含む
	実日数	19日
	人数	1名
	PC台数	2台
	OS	Windows2000

雑誌

対象資料	コレクション名	筑紫分館全蔵書(雑誌)
	冊数	1,518タイトル
	和洋区分	和 343(22.6), 洋 1175タイトル(77.4%)

	言語割合	eng 67.7, jpn 22.3, ger 4.4, fre 1.6%
	年代	-
	予想ヒット率	95%以上
	特徴	理工系雑誌
実験環境	期間	2004年7月9日～7月12日
	実日数	2日
	人数	1名
	PC台数	1台
	OS	Windows2000

4.4-1.2 実験作業手順

ローカル所蔵レコードをもとにアップロードファイルを作成し、新規の参加組織コード (FA022084) のもとに所蔵登録を行った。

- ・ 検索データに書誌 ID を採用したため、エラーの大部分は出版物理単位特定失敗によるものになる
- ・ 出版物理単位特定失敗エラーには、所蔵登録時の書誌レコードの VOL 値が最新の書誌レコードの VOL 値と必ずしも一致しないことが原因であるものも含まれる

と予測し、半自動モードの有効性を検証しながら、エラーのパターンを整理して効率よく処理を進めるために留意すべき点をまとめることを目指した。処理の流れと件数を図 1 に示す。

4.4-1.3 アップロードファイル作成方法

ローカルデータベースから所蔵レコードを抽出し、検索フィールド名、検索指定子、区切り記号を追加することにより、アップロードファイルを作成した。

雑誌

- 1) 1 書誌レコードにつき複数存在しうるローカル所蔵レコードを、配置コード単位で 1 件に合成して (NEC 製図書館業務システム LICSU-LX の NACISIS-CAT 雑誌所蔵報告用 B フォーマットデータ作成機能を使用)、抽出
- 2) 検索フィールド名、検索指定子、区切り記号を追加 (Active perl を使用)

[例 - 雑誌所蔵ルート用] (はタブ, は改行を示す)

BID:AA11056294 LOC:@ HLV=32,34,35(1-5),36-39,40(1-2) HLYR=1997-2004 CONT=+

図書

- 1) 処理対象となる所在のローカル所蔵レコードを抽出 44,070 件
- 2) 複本を除去

Microsoft Access を使用。書誌 ID と VOL でレコードをグループ化し、所在の優先順位 (研究室よりも図書館に配架されているものを優先して NC 所蔵登録する等) に応じてソートし、グループ先頭のレコードを登録用レコードに採用した (本学では複本の場合は 1 件のみ登録している

ため)

32,306 件=NACISIS-CAT 所蔵の出版物理単位件数

書誌レコード件数=NACISIS-CAT 所蔵件数としては 28,604 件

- 3) 書誌 ID を検索データとし、配置コード「@」、ローカル所蔵レコードに VOL 値があるものは登録データにその値を設定(空値は明示せず)(Microsoft Access を使用)

[例 - 図書書誌ルート] (はタブ, は改行を示す)

ID:BN00612997 LOC=@ CLN=410.3/Ku 33 RGTN=071132000051500 CPYR=1976

ID:BN01582676 LOC=@ VOL=10 版 CLN=G/K RGTN=071132000031608 CPYR=1954

- 4) 3 ファイル (BOOK1.upl・BOOK2.upl それぞれ 10,000 件, BOOK3.upl 12,306 件) に分割し、それぞれ、中断機能を使用してファイルが終了するまで連続処理

4.4-1.4 検索キー

雑誌

検索キーは書誌 ID, 処理ルートは雑誌所蔵とした。

図書

検索キーは書誌 ID である。処理ルートを図書書誌、「オプション設定 - VOL 処理: 書誌検索時の一致 VOL なし時の処理」を「処理スキップ」とした。

1 回目の全自動モード登録後、エラー「該当書誌は VOL が空値でないため更新できません」が発生したアップロードレコードについては、「VOL=@」を追加して再度全自動モード登録を行った。

半自動モード登録時に検索キーの変更は行わなかった。

なお、アップロードファイルの作成について、当館での処理後の反省点として、以下の 3 点を挙げておく。

1. アップロードファイルの検査

NC-AUTO の現行仕様では、文法チェックのみで、フィールド名チェックは行わない。検査でエラーがなかったため夜間処理を行ったところ、フィールド名のミスによって翌朝全件エラーになっていたこともあり、注意を要する。

2. アップロードファイルのレコード順と VOL 登録順

複数の出版物理単位は、1 件の NACISIS-CAT 所蔵レコードの中ではアップロード順に登録されるので、レコードの並び順に留意したい。今回の処理では単純に文字列昇順で処理したため、例えば「v.1,v.10,v.11,v.2,v.3,...」といった順序で登録されてしまった。いずれにせよ、既登録の所蔵レコードに別出版物理単位を追加登録する場合は、所蔵レコード内の VOL 順は崩れる。

3. ISBN の扱い

書誌 ID が特定されているとはいえ、より確実に書誌レコード・出版物理単位を特定す

るために、ISBN を検索語に採用すべきであった。

4.4-1.5 実験結果

4.4-1.5.1 全自動モードでの結果

雑誌

全自動モード	処理件数合計		1,518
	成功		1,511
	エラー		7
	エラー内訳	Connection timed out	7

雑誌に関しては、検索上の問題によるエラーは発生しなかったため、1 回目の全自動モード処理で作業が完了した。以下、図書処理について記述する。

図書

全自動モード	処理件数合計		32,306
	成功		30,810
	エラー		1,496
	エラー内訳	1) 該当書誌は親書誌のため更新できません	19
		2) 該当書誌レコードが存在しません	29
		3) 該当書誌は VOL が空値でないため更新できません	77
		4) 該当書誌に一致する VOL が存在しないため更新できません	541
5) 該当書誌は VOL が複数存在するため更新できません		658	
6) Connection timed out		172	

ブルーファイル、エラーファイルを perl によって 1 行 1 レコードに変換し、Microsoft Access で分析した。

ブルーファイルをもとに、時間当たりの処理件数を表にした（表 1）。処理速度に影響する要因は複合的であるため、実行遅延タイムの最適設定値を一般化することはできないが、処理時間帯、実行遅延タイム設定値および処理速度の関係の一例として示す。概ね、1000 ミリで 90～100 件、500 ミリで 150～160 件、400 ミリで 140～160 件、300 ミリで 170～190 件程度である。

エラーの原因は、書誌レコード特定の失敗によるもの（上表 1), 2), 出版物理単位特定の失敗によるもの（同 3)～5)）に二分できる。

書誌レコード特定の失敗は、ローカルデータベースと NACSIS-CAT との不整合であるため、必要に応じて現物を確認しながら、書誌統合や所蔵付替等、ローカルデータベースお

よび NACSIS-CAT を個別にメンテナンスした。

一方，出版物理単位特定エラーのうち，3)「該当書誌は VOL が複数存在するため更新できません」について考察する。

NC-AUTO では，ISBN を含まない検索をした場合は，アップロードファイル中に所蔵登録用の VOL 指定なしであれば，「出版物理単位が単数」かつ「VOL が空値」かつ「親書誌ではない」場合に限り，所蔵レコードの登録に進む。

図書書誌ルート，書誌 VOL 値との不一致時処理スキップに設定して処理した場合，

- ・アップロードファイルに VOL 指定無しの場合
→書誌レコードの「出版物理単位が単数」かつ「VOL が空値」かつ「親書誌ではない」場合に限り，成功
- ・アップロードファイルに「VOL=@」指定の場合
→VOL が空値の出版物理単位が存在する場合，成功

となる。

1 回目の全自動モード処理では，ローカル所蔵レコード VOL 値が空値の場合はすべて VOL 指定無しとした（「VOL=@」のように VOL 空値の出版物理単位の存在を明示しなかった）。そのため，書誌レコードに複数の VOL が存在する場合，このエラーが発生した。

[例]

(アップロードファイル)(はタブ， は改行を示す)

BA06553243 @ 420.4/W 55 067212000002614 1989 2004/07/24 02:49:39 該当書誌は VOL が複数存在するため更新できません

(該当書誌レコード)

<BA06553243> CRTDT:19890307 CRTFA: RNWDT:19910914 RNWFA:FA001685

CODE:LCCN:88024532 ; YEAR1:1989 ; CNTRY:us ; TTLL:eng ; TXTL:eng

VOL: ; ISBN:0716719827

VOL:pbk. ; ISBN:0716721066

TR:The privilege of being a physicist / Victor F. Weisskopf

... (以下省略)

このエラーレコードに無条件で「VOL=@」を追加して再処理（全自動モード）すれば，VOL 値が空値のものを含めて複数の出版物理単位を持つ書誌レコードに対しては，正常に所蔵登録可能である。

再処理後に発生したエラー「該当書誌に一致する VOL が存在しないため更新できません」については，書誌レコードに VOL 値が空値の出版物理単位が存在しないということになるので，同じエラーが発生した VOL 値が空値でないレコードとともに再処理すればよい。

ところで，そもそも最初のアップロードファイルで VOL 空値の場合にすべて「VOL=@」を指定しておけば，エラー「該当書誌は VOL が複数存在するため更新できません」は発生せず，逆に出版物理単位のない書誌レコードに対して，エラー「該当書誌に一致する VOL が存在しないため更新できません」が発生する。この場合，書誌レコードに複数の出版物理単位が存在するが空値のものが無いのか，出版物理単位自体存在しない書誌レコードなのか，エラーメッセージからは区別がつかない。

したがって、書誌レコード特定のための検索データは必要充分でなければならないが、書誌レコードを1件に特定できる（検索データが書誌IDである）場合は、アップロードレコードのVOL値が空値であれば、最初は指定無しで処理し、そのエラーに対してVOL空値指定で再処理する方が、処理手順を組み立てやすい。

エラー 4)「該当書誌に一致するVOLが存在しないため更新できません」、5)「該当書誌はVOLが空値でないため更新できません」も、出版物理単位特定の失敗が原因である。4)は、アップロードファイルのVOL(値あり)が書誌レコードのVOLと一致しない場合、5)は、アップロードファイルのVOLが空値であるのに書誌レコードのVOL(1件)が有値である場合に発生するエラーである。これらについて、エラー 3)を「VOL=@」指定で再処理(全自動モード)後に生じた同メッセージのエラーと合わせて、半自動モードでの処理を試みた。

4.4-1.5.2 半自動モードでの結果

半自動モード	合計件数		991
	成功		0
	エラー		991
	エラー内訳	出版物理単位特定の失敗	991

出版物理単位特定エラーについて半自動モードでの再処理を試みたが、NC-AUTOでは、書誌レコードが1件に特定された場合、対面検索画面を展開しない仕様であるため、全件エラーになった。

4.4-1.5.3 エラーの再処理の結果

エラーの再処理	合計件数		991
	成功		988
	エラー		3
	エラー内訳	Connection timed out	3

本来、NACSIS-CATの書誌レコードのVOL値と所蔵レコードのVOL値は一致させることが原則であり、1件ずつ確認しながらメンテナンスを行うべきである。

しかし、「:us」「:pbk」などといった、出版国の相違や装丁に関するVOLの説明語句の不一致であれば、個別に書誌レコード(必要であれば現物も)を確認することは非効率にすぎること、ILL業務への影響を最小限にするために処理を急ぐことから、一旦すべてアップロードファイルのVOL値を採用して所蔵登録を行い、資料内容の相違に関わる値の場合のみに絞って個別メンテナンスを行うことにした。

1. 一旦すべてアップロードファイルのVOL値を採用して所蔵登録を行う(「オプション設定 - 出力編集」で、書誌レコード全フィールドをプルーフファイルに

出力する設定にしておく)

[結果]

エラー (Connection timed out) 3 件 個別メンテナンス済み
所蔵登録成功 988 件

2. 所蔵登録成功 988 件について、プルーフファイルを perl で加工し、書誌 ID、所蔵レコードの VOL 値、書誌レコードの VOL 値のみを抽出したファイルを作成する。VOL の値を 1 件ずつ確認し、以下の条件で、所蔵レコードの VOL 値および書誌レコードの VOL 値にレベル付けを行う。

- * 要確認レベル : 「Vol.1」「北海道編」等、内容の相違に関わる値
- * 確認不要レベル : 「: us」等、内容の相違に大きく影響しない値

[結果]

いずれも確認不要レベル : 545
いずれか、またはどちらもが要確認レベル: 443 件

3. いずれか、またはどちらもが要確認レベルと判断された 443 件について、書誌レコードと所蔵レコードの VOL 値の相違を比較し、個別メンテナンスの要否を判断する。

[結果]

メンテナンス不要 (285 件)

記述文法や表記の些細な相違・巻号等に続く説明語句の有無
(例) 「:」やスペースの有無、長音「ー」「-」の表記の相違、
「1」「1 北海道」の相違

メンテナンス対象 (158 件)

書誌 VOL 値 (空値) が所蔵 VOL 値 (有値) と不一致 78
書誌 VOL 値 (有値) が所蔵 VOL 値と不一致 80

(原因→修正処理)

- ・ローカルデータベースのレコード調整メンテナンス漏れ
→所蔵付け替え・VOL 値修正
- ・VOL 値の誤り →VOL 値の修正

最終的に、処理件数 988 件中、メンテナンス対象 158 件、ローカル所蔵 VOL 値をそのまま採用 830 件となった。

[例 - 所蔵と書誌の VOL 値比較ファイル] (はタブ, は改行を示す)

(書誌 ID) (自動登録した所蔵の VOL) (書誌の VOL[複数存在する場合は「||」区切り])

```
BA07832526        Vol. 2    [Vol. 1] ISBN:354012571X|[Vol. 1] ISBN:038712571X||Vol. 2  
ISBN:354051306X||Vol. 2 ISBN:038751306X  
BA26507544    pbk.    :pbk ISBN:0198557736  
BA24953050    1
```

なお、「オプション設定 - 出力編集」で、図書書誌の出力項目設定を「1:簡易形式」にす

ると、VOL が空値の出版物理単位が存在する場合は出力される値がないため、VOL グループの有無が不明である。VOL グループが存在しない場合と、VOL グループは存在するがVOL に値がない (ISBN, PRICE, XISBN に値がある) 場合とを区別できる出力形式であることが望ましい。

4.4-1.6 評価

本学では、検索データが書誌 ID であったため、エラーの大部分が出版物理単位特定の失敗によるものであった。このように、書誌レコードを 1 件に絞ったが出版物理単位を特定できない場合については、対面検索画面で容易に出版物理単位を特定できるものが少なくない。しかし、残念ながら現在の半自動モードでは、書誌レコードを 1 件に特定できない (書誌レコードがヒットしないか、複数件ヒットする) 場合のみ、対面検索画面が表示される。つまり、書誌レコードを 1 件に特定できる場合は、出版物理単位特定に失敗すると、即エラーになってしまう。

半自動モードで書誌レコードの VOL を選択することができれば、検索データの入力、検索ボタンの押下等、定型的作業はすべて NC-AUTO に任せられることができるので、作業の効率が上がる。半自動モードでの対面検索画面展開の判断を、書誌レコード特定のレベルではなく、出版物理単位特定のレベルで行うことで、半自動モードの利便性をさらに活かせるのではないだろうか。

4.4-2 (FA003454 九州大学附属図書館)

九州大学附属図書館中央館での除籍図書在所蔵データメンテナンスに際して、自動登録機能の検証を行った。

4.4-2.1 実験対象

対象資料	コレクション名	附属図書館中央館除籍資料
	冊数	3,826 冊
	和洋区分	和 98.7, 洋 1.3%
	言語割合	jpn 98.7, eng 1.2%
	年代	1900 年代
	予想ヒット率	95%以上
	特徴	破損等による除籍資料
実験環境	期間	2004 年 9 月 28 日 ~ 10 月 1 日
	実日数	3 日
	人数	1 名
	PC 台数	2 台
	OS	Windows2000

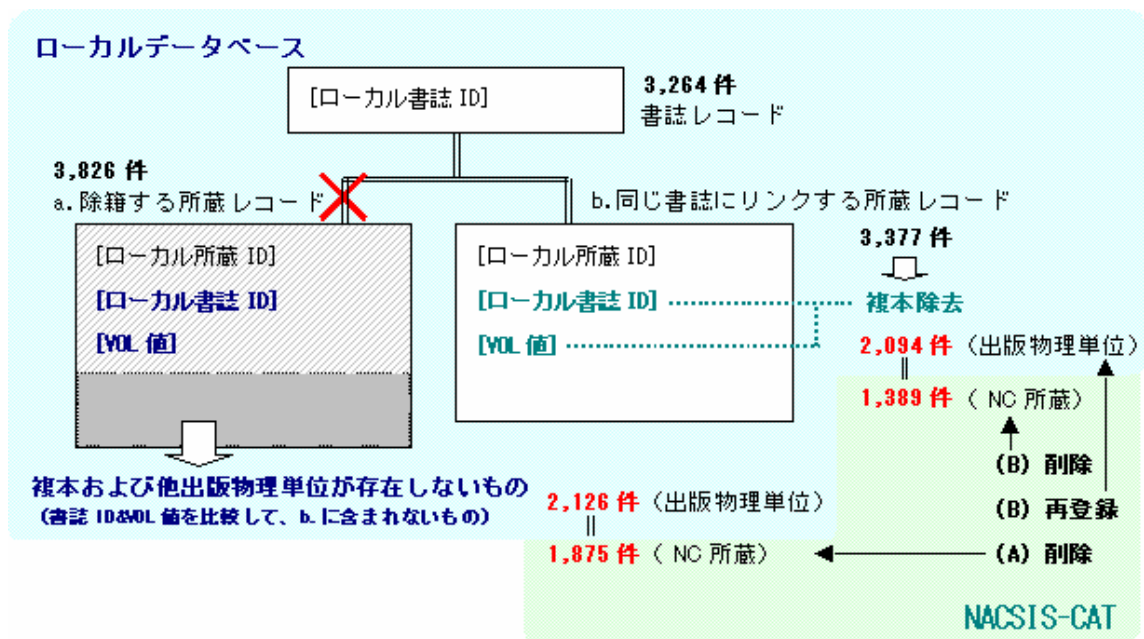
4.4-2.2 実験作業手順

NC-AUTO での図書所蔵削除処理は、NACSIS-CAT 所蔵 ID 単位であり、出版物理単位レベルではない。また、本学では、複本を所蔵する場合でも 1 冊のみを NACSIS-CAT へ登録しているのので、除籍した図書に複本が存在する場合、NACSIS-CAT に登録されているのが除籍した図書なのか、あるいはその複本なのか、ローカルデータからは判断できない。

そこで、ローカルデータベースで除籍所蔵レコードが登録されていた書誌レコードについて、NACSIS-CAT 所蔵レコード全体を一旦削除し、その後複本および別の出版物理単位がローカルデータベースに存在する場合は、それらから複本を除いたうえで NACSIS-CAT へ再度所蔵登録することにした（この処理によって、もともとの NACSIS-CAT 所蔵 ID や所蔵作成日といった情報が失われるが、特に支障は無いと判断した）。

エラー発生時等の必要な場合を除いて、原則として現物確認は行わず、データ操作のみ行った。

4.4-2.3 アップロードファイル作成方法



▲ 除籍処理用アップロードファイルの作成法

アップロードファイルは、ローカル所蔵レコードを抽出して作成した（Microsoft Access を使用）。

除籍済所蔵レコードは 3,826 件、これらがリンクしていた書誌レコードは 3,264 件である。この書誌レコード 3,264 件は、

- (A) 除籍済所蔵レコードのみがリンクしている書誌レコード
- (B) 除籍済所蔵レコード以外に複本や別の出版物理単位がリンクしている書誌レコード

に二分することができる。

- ・ この書誌レコード 3,264 件にリンクする在籍所蔵レコードは 3,377 件である。
- ・ ここから複本を除去する（複本・別出版物理単位の所蔵レコードについて、書誌 ID - VOL でグループ化し、その中からそれぞれ登録日が最新のレコード 1 件ずつのみを抜き出し、登録用所蔵レコードとして採用する）と、2,094 件である。
- ・ この所蔵レコードにリンクする書誌レコード、つまり、除籍済所蔵レコード以外にリンクする所蔵がある書誌レコード(B)は 1,389 件である。
- ・ したがって、除籍済所蔵レコードのみがリンクする書誌レコード(A)は最初の 3,264 件から(B)の 1,389 件を除いた差分 1,875 件である。

アップロードファイルは以下の 3 ファイルである。

(1) 削除処理（図書所蔵ルート）

(A) 1,875 件

(B) 1,389 件

(2) 登録処理（図書書誌ルート、書誌所蔵 VOL 不一致は処理スキップ）*

(B) 2,094 件

* 登録処理の手順は「4.4-1 新装開館 - 大量図書資料の移動と所蔵データメンテナンス（FA022084 九州大学附属図書館 筑紫分館）」の場合と同様である。

[例 - 図書所蔵削除：図書所蔵ルート用]（ はタブ、 は改行を示す）

delete BID:BA00027011 LOC:図

4.4-2.4 検索キー

削除処理については、あらかじめ、動作指示コマンド *getdelete* によって（書誌 ID と配置コードで所蔵レコードを検索）、削除用アップロードファイル（所蔵 ID が検索キー）を作成しておいた。

登録用アップロードファイルについては、書誌 ID を検索キーに採用した。

4.4-2.5 実験結果

4.4-2.5.1 全自動モードでの結果

		(1) 削除	(2) 登録
全自動 モード	処理件数合計	3,264	2,094
	成功	3,192	1,970
	エラー	72	124
	エラー内訳	該当所蔵が存在しません	72 *下表参照

*登録処理 エラー内訳	1) 該当書誌レコードが存在しません	2
	2) 該当書誌は VOL が空値でないため更新できません	2
	3) 該当書誌に一致する VOL が存在しないため更新できません	96
	4) 該当書誌は VOL が複数存在するため更新できません	14
	5) 一致する VOL があるため処理をスキップしました	4
	6) Connection timed out	6

4.4-2.5.2 半自動モードでの結果

半自動モード処理対象となるレコードは発生しなかった。

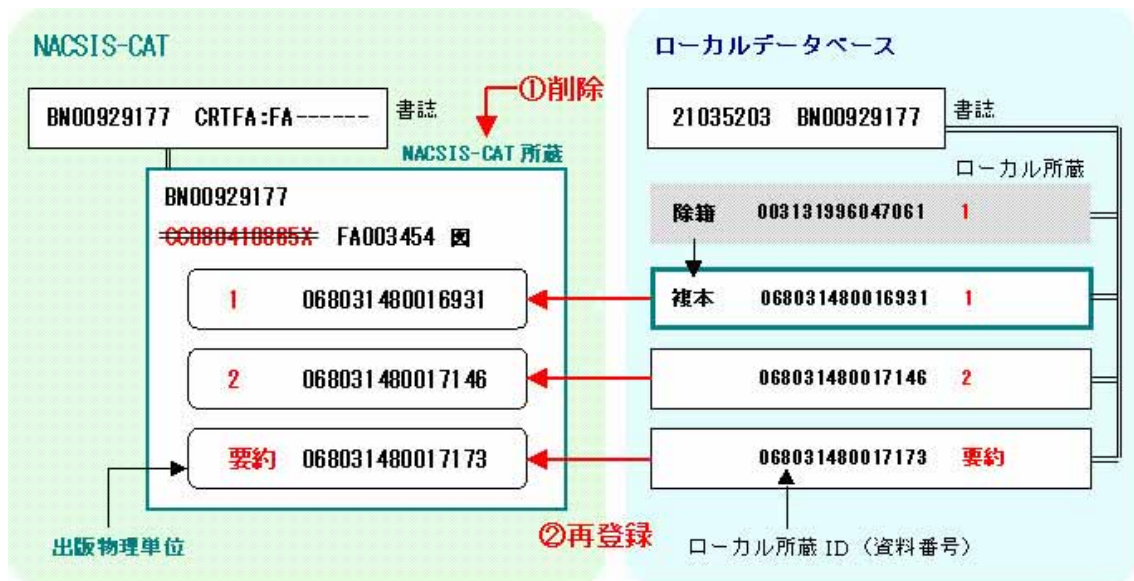
4.4-2.5.3 エラーの再処理の結果

エラーの再処理	合計件数	110
	成功	110
	エラー	0
	エラー内訳	0

エラー 3) および 4) について全自動モードで再処理し(アップロードファイルの VOL 値を採用), VOL 値の確認を行った。

4.4-2.6 評価

4.4-2.6.1 削除処理の単位



▲ 図書所蔵：削除の処理単位

NC-AUTO での図書所蔵削除処理は、NACSIS-CAT 所蔵 ID 単位である。実際には、今回の処理のように削除の後再登録することによって、出版物理単位レベルでの削除と同等の処理が可能であるが、一処理で出版物理単位レベルでの削除が可能であれば、効率がよい。

4.4-2.6.2 書誌レコードの削除予定レコード化

自動削除処理では、自館所蔵削除によって書誌レコードが書誌・所蔵リンク無しになっていないかどうか、確認できない。しかし、全件個別に確認するのであれば自動処理の意味がない。

そこで、書誌作成参加組織コードが自館の FA003454 であればリンク無し書誌レコードになっている可能性が高いと判断し、それらについてのみ個別に確認することにした。

1. 削除処理後に生成されたダウンロードファイル(ブルーファイルでも可)(A, 1835 件)を perl で加工して、書誌作成参加組織コードが FA003454 であるものの書誌 ID リストを作成 → 25 件を個別に確認
2. 個別確認の結果、書誌・所蔵リンク無しである 5 件を削除予定レコード化

4.4-2.6.3 アップロードファイルの検索指定子と処理速度

アップロードファイルで同一検索データが続く場合、1 件目に検索指定子「::」を指定すると、以降はその検索データを省略することができる。同一書誌レコードの異なる出版物理単位について、検索指定子「::」を書誌 ID に一括指定したアップロードファイルとそうでないものを作成し、処理速度を比較してみると、一定の高速化を期待できる。

- (2) 登録処理：アップロードファイルを(B-1)、(B-2)に分割し、PC2 台を同時接続して処理
- (B-1) 1,000 件(「ID::」で検索指定)
- (B-2) 1,094 件(「ID:」で検索指定)

[例 - (B-1) 図書書誌ルート用] (はタブ, は改行を示す)

```
ID::BA34695778 VOL=上 LOC=図 CLN=428.4/Ki 79/2c-1B RGTN=003111999025496
CPYR=1999
VOL=下 LOC=図 CLN=428.4/Ki 79/2c-2A RGTN=003111998017714 CPYR=1998
```

検索指定子	所要時間	処理速度
B1「::」指定(1000件中447件にVOLあり)	3.75h	266件/h
B2「:」指定(1094件中558件にVOLあり)	5.00h	220件/h

しかし、一括指定により検索データを省略したアップロードレコードについては、エラーファイルに検索データが出力されないため、エラー処理時に検索データを確認する手間がかかる。NC-AUTO の仕様改善が望まれる点である。

図1: 新装開館-大量図書資料の移動と所蔵データメンテナンス(FA022084 九州大学附属図書館 筑紫分館)処理フロー

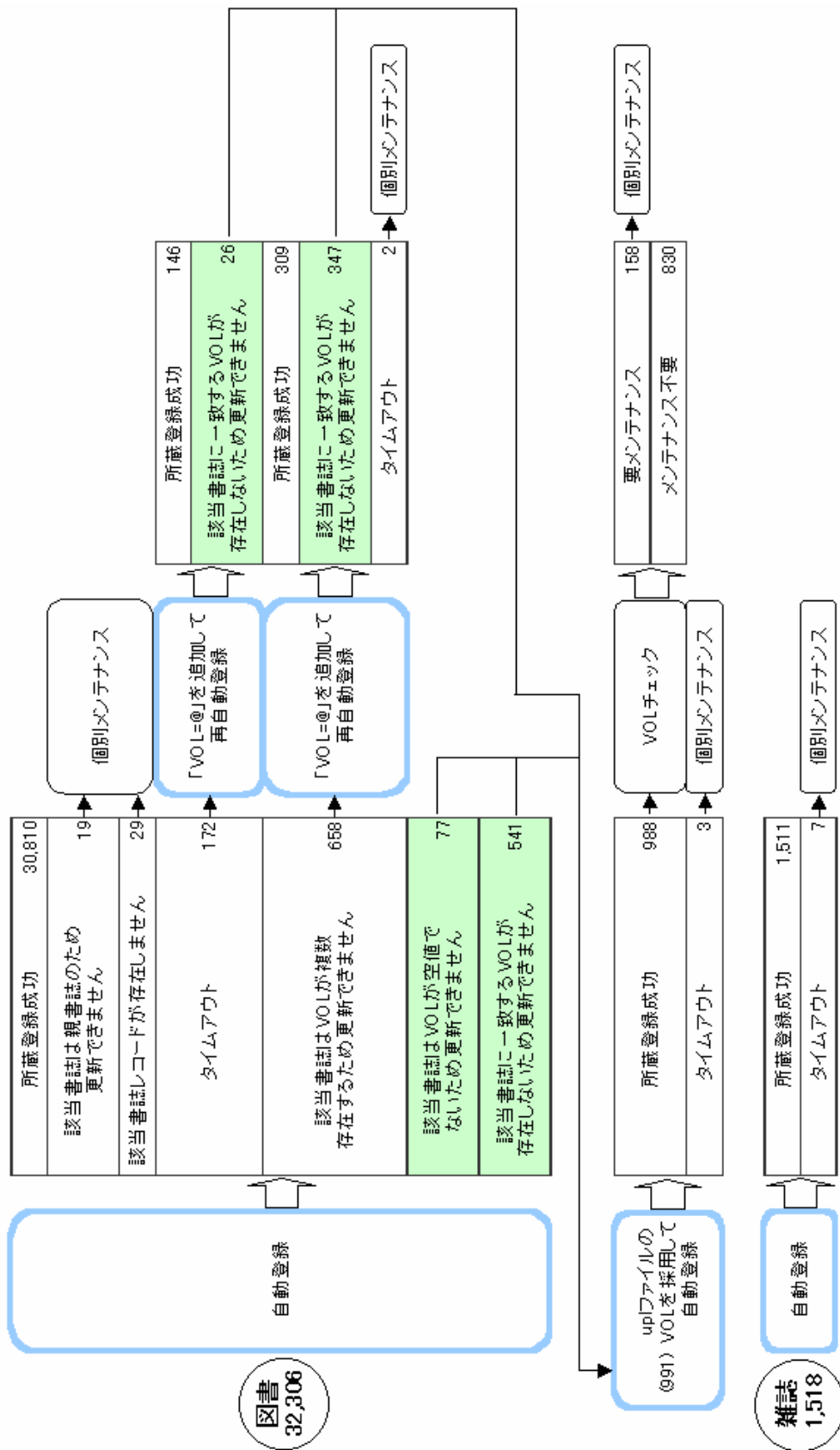


表 1: 処理時間帯，実行遅延タイム設定値および処理件数記録

ファイル	日付	曜	遅延タイム	時	正常処理 件数	ファイル	日付	曜	遅延タイム	時	正常処理 件数		
BOOK1.upl	2004/7/14	水	1000	12	13	BOOK2.upl	2004/7/22	木	500	0	145		
			1000	17	97				500	1	149		
			1000	18	100				500	2	150		
			1000	19	97				500	3	150		
			1000	22	101				500	4	153		
			1000	23	99				500	5	158		
	2004/7/17	土	1000	9	96				500	6	151		
			1000	10	94				500	7	144		
			1000	11	94				500	12	117		
			1000	12	90				500	17	117		
			1000	13	92				2004/7/24	土	500	0	149
			1000	14	102						500	1	147
			1000	15	94						500	2	151
			1000	16	95						500	3	155
	1000	17	84	500	4						152		
	500	5	151	500	5						151		
	2004/7/22	木	400	1	157				500	6	154		
			400	2	159				500	7	140		
			400	3	162	BOOK3.upl	2004/7/28	水	300	0	180		
			400	4	161				300	1	173		
			400	5	162				300	2	184		
			400	6	157				300	3	182		
			400	7	143				300	4	194		
			400	12	118				300	5	179		
			400	17	121				300	6	191		
			2004/7/24	土	400				9	146	300	7	179
	400	10			152				300	9	180		
	400	11			139				300	10	178		
	400	12			147	300	11	179					
	400	13			149	300	12	173					
	400	14			148	300	13	171					
	400	15			148	300	14	174					
	400	16			142	300	15	175					
400	17	131			300	16	169						

5 効果的な作業方法の提案

今後自動登録処理を試みる際の参考となるように、留意すべき点や効果的な方法を、全般として、また資料種別に以下に述べる。

5.1. 全般

5.1.1 自動登録処理を行うかどうかの判断

自動登録処理を行うためには、登録処理そのもの以外に、アップロードファイルを作成し、登録データを確認するといった作業が発生する。

それらに必要な作業量と処理件数とを考慮したうえで、総体的に高い効果が得られると判断できる場合に、自動登録処理を選択する。

5.1.2 アップロードファイルの作成

自動登録処理を行うにあたって、検索キーの選択は、誤登録を防いだり書誌レコードをなるべく一意にヒットさせたりするために、最も重要な点である。

検索キーは、類似書誌を確実に排除できるものを設定しなければならない。特に複数書誌がヒットするエラーが生じた場合は、登録に成功したレコードについて、検索キーに曖昧さがなかったか、再確認すべきである。

5.1.3 自動登録を行う時間帯

予約登録で夜間に行うと処理速度が速い。

5.1.4 半自動モード処理

半自動モードを活用すれば、ヒット率を高めることができる。検索キーに採用しなかったデータも削除せずアップロードファイル内にコメントアウトしておけば、半自動モードの対面検索画面で検索語を修正する場合に役立つ。

5.1.5 登録成功レコードの確認

書誌ルートで自動登録を行った場合、所蔵登録した書誌レコードが適切かどうかの確認が非常に重要である。登録データやプルーフファイルに対して現物と照合する方法、または、アップロードファイル等に検索・登録データ以外のデータを持っておき、それと照合する方法などがある。

一例として、登録成功レコードについて生成されるプルーフファイルには、アップロードファイルとヒットした書誌レコードの内容が出力されるので、テキストデータ加工プログラム等を利用すれば、登録内容の確認を効率よく行うことができる。

5.2 資料種別の効果的作業方法案

5.2.1 新刊書以外

既に所蔵している資料の遡及入力に関する自動登録の利用には、多くの図書館が関心を持っていると思われる。各館の実験報告によれば、自動登録の対象資料の選択と検索キーの選択が非常に重要であると言える。

新刊書ではない資料の作業方法としては以下の2通りが考えられる。

- 1) 受入データやローカル所蔵データからアップロードファイルを作成する。
- 2) 資料現物からデータ入力を行い、アップロードファイルを作成する。

1)は、基本的に資料現物の移動を伴わずデータ上の処理が中心となる。

2)は、図書館内の所蔵図書を遡及入力する場合に採る方法である。この場合、現物を手元においてアップロードファイル作成からプルーフファイルのチェックまでを行うのであれば、ブックトラック1台～数台の単位で作業するのが効率的と考えられる。

また、アップロードファイルには、検索キーに採用しなかったデータをコメントアウトしておくが良い。半自動モードの対面検索画面で検索語を修正する場合に役立つ。

ただし、新刊書でない資料はヒット率が低い場合もあるため、少量の資料なら自動登録を使わずに1冊ずつ登録するほうが早い可能性があり、自動登録を使うかどうかの判断が必要である。

5.2.2 新刊書

東京大学総合図書館の新規購入図書の実験報告より、ヒット率が高く自動登録に適していることがわかる。受入データからISBNを検索キーとするアップロードファイルを作成するのが効率的である。

5.2.3 大量データ移行

大量データ移行のケースとしては、

- ・新たな分館等の開館に伴い、新しい参加組織番号または配置コードで所蔵登録する
 - ・分館・部局図書室の統合等に伴い、異なる参加組織番号または配置コードの所蔵を再構成する
 - ・除籍に伴う所蔵データのメンテナンス
- などが考えられる。

これらの場合、ローカルデータベースからアップロードファイルを作成することができること、そのためNC書誌IDなど書誌レコードを特定する情報を確実に持っていること、またILL業務を円滑に進めるために処理を急ぐ必要があることから、自動登録処理を行う効果が非常に高い。

6 結論

自動登録ソフトの利用を考える際には、「5 効果的な作業方法の提案」で述べたように、各館それぞれで異なる事情を考慮する必要がある。

- ・ アップロードファイルの作成方法
 - 受入データやローカル所蔵データを加工して簡単に作成できるか
 - 現物をもとにデータ入力をする必要があるか
(現物の移動距離も勘案して検討する)
- ・ 登録成功データのチェック方法
 - 現物に当たる必要があるか，データ確認のみでよいか
- ・ 処理件数
 - 上記の準備，処理後の確認作業，予想ヒット率を考え合わせても，作業の効率化が期待できる件数であるか

自動登録の有効性としては，各館の実験報告を参考にすると以下の図式が成り立つと考えられる。

書誌 ID または所蔵 ID 判明時の所蔵データの登録・削除 新刊和書 (ISBN あり)
新刊洋書 (ISBN あり) > ISBN なし図書

大量の資料またはデータを扱う際には自動登録を利用するメリットがあると思われる。特に今回の実証実験では，半自動モードが加わったことにより，全自動モードでエラーになった資料の検索キーや登録データを再入力することなく書誌レコードを確定・所蔵登録できるという一定の効果があった。全自動モードのみの場合は，エラーになった資料を通常の目録処理と同じく業務システムにより検索・登録するため，(アップロードファイルを手入力で作成したときは) 検索キーや登録データを 2 回手入力することになる。その点が半自動モードにより改善されたことは評価したい。

一方，各館の実験報告に述べられているように，実際の利用を考えると半自動モードの改良が不可欠である。主な改善点の要望は「別表：機能改善の提案」に示したが，特に，各館が指摘している VOL 不一致時の処理についてはメーカーに対処を望みたい。また，メーカーに報告されたその他の改善要望についても対処されれば，自動登録ソフトの使い勝手が向上し，各図書館が前向きに利用を検討できるものとなるであろう。

別表：機能改善の提案

機能	現行仕様	改善の提案
半自動モード対 面検索画面への 展開条件	書誌が1件に特定できない場合 のみ、半自動モードで対面検索 画面へ展開する。	図書所蔵登録について、書誌は1件に 特定したが出版物理単位を特定でき ない場合に、詳細書誌情報が表示さ れ、アップロードファイルを参照しな がら書誌に相当と判断できる出版物 理単位が存在すればそれを選択して 確定し、書誌と同じ記述内容のVOL で所蔵登録することができる(半自動 モードでの対面検索画面展開を、出版 物理単位が特定できなかった場合に も行う)。
図書所蔵出版物 理単位での削除	-	図書所蔵の出版物理単位での削除処 理(所蔵単位では更新処理)にも対応 する。
所蔵削除後のリ ンク無し書誌チ ェック	-	削除処理によって子書誌リンク・所蔵 リンクが無くなった書誌IDのリスト を出力する。
アップロードフ ァイルの検査	文法チェックを行う。	文法チェックだけでなく、フィールド 名チェックも行う。
プルーフファイ ルの図書書誌 VOL出力	オプション設定・出力編集で、図 書書誌の出力項目設定を「1:簡 易形式」にすると、VOLが空値 の出版物理単位が存在する場合 は項目名称も表示しない。	VOLグループが存在しない場合と、 VOLグループは存在するがVOLに 値がない(ISBN, PRICE, XISBNに 値がある)場合とを区別できる出力形 式であることが望ましい。