

## NACSIS-CAT における最近の動向

酒井 清彦・京藤 貫

国立情報学研究所 開発・事業部コンテンツ課

### 1. NACSIS-CAT 運用状況

#### (1) 参加状況，データ作成状況，更新状況

NACSIS-CAT の参加状況（2002 年 1 月末現在）は以下の表のとおりである。今年度は雑誌データベースの全国所蔵更新作業の調査年度に当たっており，この調査を機に NACSIS-CAT(オンライン目録システム)に参加する機関が目立った。特に国立試験研究機関図書室，各省庁の附属研究所図書室及び病院図書館等の参加が顕著であった。

表 1 参加状況

機関区分	参加機関数	機関の母数	参加率	2001年度の参加機関数
国立大学	99	99	100%	0
公立大学	72	74	97.3%	2
私立大学	454	497	91.3%	42
短期大学	116	489	23.7%	18
高等専門学校	55	62	88.7%	2
その他	165	-	-	77
大学共同利用機関等	14	19	73.7%	0
合計	975	-	-	141

一方，データの作成状況（2002 年 2 月 23 日現在）は以下の表のとおりである。最近 3 ヶ月の月平均増加件数は，図書書誌が 35,685 件，図書所蔵が 523,085 件となっている。

表 2 総合目録データベースのデータ作成状況

データベース名	件数	1999 年度の増加件数	2000 年度の増加件数	2001 年度の増加件数
図書	書誌	6,106,490	447,749	448,483
	所蔵	58,623,169	6,731,165	7,186,233
雑誌	書誌	242,401	4,325	8,408
	所蔵	3,652,204	124,345	147,139
著者名典拠	1,194,066	52,520	45,682	37,972
統一書名典拠	19,560	1,167	578	971
雑誌変遷マップ	31,057	719	3,344	2,620

また、データの更新状況は以下の表のとおりである。「新規」は毎週新たに作成されるデータの総合計である。「修正」は毎週行われるデータ修正を累計したものであり、同一週内で同一データに対して修正が行われた場合は1件とカウントし、週をまたがって同一データに修正が行われた場合は2件とカウントされる。「削除」は毎週行われる削除処理の総合計である。最近の傾向として、「新規」に対する「修正」の比率が高まっていることが挙げられる。

表3 総合目録データベースのデータ更新状況

データベース名			1997年度	1998年度	1999年度	2000年度	2001年度
図書	書誌	新規	458,060	482,904	473,887	478,265	414,381
		修正	445,554	440,318	435,991	697,630	603,533
		削除	17,494	26,329	18,649	29,782	17,295
	所蔵	新規	5,489,984	6,539,390	6,957,206	7,400,358	6,507,790
		修正	909,093	1,264,515	1,364,005	3,326,640	2,600,116
		削除	78,669	104,951	127,686	214,125	336,228
雑誌	書誌	新規	6,707	9,318	8,938	8,883	7,615
		修正	67,248	39,220	43,695	46,775	52,071
		削除	99	435	4,442	475	1,732
	所蔵	新規	190,077	480,134	304,259	428,696	181,231
		修正	475,798	506,073	689,471	886,579	921,551
		削除	82,473	355,203	178,286	281,557	93,865
著者名典拠	新規	65,905	66,238	51,881	47,110	37,649	
	修正	21,206	25,136	22,482	24,223	19,265	
	削除	450	1,086	826	1,104	468	
統一書名典拠	新規	2,244	2,761	1,250	580	971	
	修正	455	386	284	102	177	
	削除	0	83	28	0	0	

(2) 参照ファイル(MARC)の収納状況

NACIS-CATで利用している参照ファイルは以下の表のとおりであり、現在同表のとおり件数となっている。

2001年度中に、米国議会図書館(Library of Congress)の書架目録用目録カードを基にして作成された目録データREMARCを参照ファイルとして導入する予定である。1890~1967年までの約420万件のデータが利用可能となる見込みである。

表4 参照ファイル構築状況

データベース名		件数	収録範囲
参照 LC (USMARC)	洋図書	6,577,725	1968年~
	非文字資料	268,147	1984年~
	洋著者名典拠	3,400,010	1977年~
	洋統一書名典拠	275,251	1977年~
	洋雑誌	872,449	1973年~
参照 JP (JAPANMARC)	和図書	2,712,427	1868年~*
	和著者名典拠	325,315	1969年~
	和雑誌	100,582	1989年~
参照 UK (UKMARC)	洋図書	1,868,636	1950年~
参照 DN (DNMARC)	洋図書	3,372,190	1945年~

参照 TRC (TRCMARC)	和図書	1,009,682	1985年～
参照 GPO (GPOMARC)	洋図書	500,009	1976年～
参照 CH (CHMARC)	洋図書	474,025	1988年～
参照 BL (BLSERIAL)	洋雑誌	610,172	

\*大正期分は除く

### (3) 品質管理状況

総合目録データベースの品質管理は、3つの段階で行われている。

- A. 図書館でのレコード登録時の点検作業 (CAT入力担当者数は約3,200名)  
十分な検索を実施し、重複レコードの作成を回避するよう留意  
データ入力基準類に基づいているかどうか、誤りがないか否かの確認
- B. 図書館でのレコード利用時点での誤り発見  
データ入力基準に基づいたレコード修正 (年間約50万件)  
うち、レコード作成館との協議に基づいたレコード修正 (推計：年間約15万件)  
NIIへの重複・削除レコードの報告 (年間約2,000件)  
NIIへの修正レコードの報告 (年間約1,000件)

#### C. NIIでの品質管理

NIIでは、総合目録データベースの全レコードを点検することは物理的に不可能である。また、資料そのものを所蔵していないため、以下のような処理に限定される。

目録情報の基準、コーディングマニュアル等のデータ入力基準類の整備及び維持管理

システム操作法、基準等の目録担当者への周知、教育  
(マニュアル作成、ニュースレターの刊行、講習会の実施)

質問書/回答書データベースの運用管理

(質問受付：約60件/月、DB検索回数：約3,000件/月)

品質調査と問題点の解明、品質維持のための対策検討

品質管理のための研究開発とシステム開発

限定された範囲内での入力データの品質管理確証作業

重複レコード (書誌・典拠) の自動統合処理	約5,000件/年
削除予定レコード (書誌・典拠) の一括削除	約15,000件/年
リンク付替え、データ修正	約6,000件/年
FAXによる修正個所の連絡	約5,000件/年
電話・メールによる問合せに対する回答	約2,600件/年
雑誌レコードのリストによる目視チェック及び修正*	約33,000件/年

\* (学術雑誌総合目録編集のため、雑誌レコードについては全件チェックを行っている。)

## 2. 多言語資料への対応

### (1) 中国語資料への対応

- ・ UCS (Universal multiple-octet coded Character Set：国際符号化文字集合：日本語、中国語などの漢字圏を含めた世界的に共通なコード体系の文字セット) が制定されたことに伴い、システム面での大きな制約が取り除かれた。

- 1995 年度に設置した「中国語資料データベース化検討ワーキンググループ」において、中国語資料の取扱いについて議論を重ね、1998 年度に「中国語資料の取扱い(案)」として公表した。この案の特徴は、漢字単語単位での検索を可能とするため、中国語タイトルに日本語ヨミで分かち書きしたヨミを付与することを必須としたこと、そのための読み方や分かちの仕方を詳細に提示したことである。ピンインはヨミに併記できるようにした。
- また、著者名典拠については、原則として最初に典拠レコードを作成する時に用いた資料に表示されている字種・字体で記録するが、著名な著者や多作な著者又は清朝期までの著者については最もよく知られた形もしくは多くの参考資料に記載されている形を採用することにしている。
- 目録記述面での検討と並行して、目録システムのシステム開発を実施した。主要な開発事項は、全てのデータベース及びプログラムを UCS 対応仕様にする事、日中韓の漢字を効率よく検索できるように漢字統合インデクスを作成利用することであった。
- これらの準備を経て、2000 年 1 月から目録システムにおいて中国語資料が正式に入力することができるようになった。
- 中国語資料の入力支援のために、CHINAMARC を導入し、書誌情報の作成支援を実現している。今後は中国国内で作成される中国名著者名典拠の導入を図ることによって、典拠情報の作成支援に資する必要がある。

## (2) 韓国・朝鮮語資料への対応

- 1999 年度より、韓国・朝鮮語資料の取扱いに関して「韓国・朝鮮語資料データベース化検討ワーキンググループ」を設置(2000 年度からは「韓国・朝鮮語資料の取扱いに関する小委員会」と組織変更)し、目録記述面での検討を進め、2001 年度に取扱い方針を定めた。
- この取扱い方針の特徴は、単語単位での検索を可能にするため、単語単位での分かちが必要な場合、ハングルによるヨミを付与するようにしたことである。
- 著者名典拠については、中国語資料と同様、原則として最初に典拠レコードを作成する時に用いた資料に表示されている字体で記録するが、著名な著者等については最もよく知られた字体で記録することにしている。漢字で表記された著者名についてはハングルヨミを付与することになった。
- 2001 年度には、システム改造も併せて実施しており、これによって、2002 年度から NACSIS-CAT への入力が可能となる。
- 韓国・朝鮮語資料の入力支援のため、2002 年度から参照マークとして韓国 MARC を導入する予定である。

## (3) WebUIP

- 多言語資料の入力を行うためには、UCS に対応したクライアントシステムを導入する必要がある。しかしながら、通常図書館のシステムは 4~6 年程度のサイクルで更新されるため、図書館によってはローカルシステムがすぐに UCS に対応できない場合もありうる。
- UCS 対応クライアントがない場合、UCS で入力された文字(中国語の簡体字など)は全てコード情報として表示されることになる。例えば、「中国古代佚名哲学名著 U8BC4 述」のように、で囲まれた UCS コード番号が表示されるが、これを見ただけではどの文字が該当するのかが判断できない。

- ・ そのような場合でも NACSIS-CAT への入力及びデータ表示だけは可能にするため、国立情報学研究所では UCS 対応のクライアントを開発し、暫定的利用という形式で 2001 年 1 月から提供している。

#### (4) Webcat

- ・ 2000 年 1 月からいくつかの図書館を中心に、中国語資料のデータ入力が積極的に実施されている。2002 年 2 月時点で、多言語資料対応クライアントを利用して入力された中国語資料は約 8 万件になっている。
- ・ これらのデータも他のデータと同様、インターネットで総合目録データベースを検索できるシステム Webcat を通して見ることができる。
- ・ しかしながら、入力システムと同様、Webcat も多言語対応していないと UCS で入力された文字は全てコード情報として表示されることになる。
- ・ そこで、Webcat についても多言語対応のためのシステム開発を行い、2001 年 1 月から簡体字についても利用者のブラウザ上で表示ができるようにした。利用者は、UCS に対応したブラウザ（Internet Explorer もしくは Netscape の最新版など）を用意し、Webcat の英語版を使って検索表示させることができる。
- ・ Webcat にも、漢字統合インデクス機能があるため、簡体字を意識することなく検索することが可能となっている。

### 3. その他の動向

#### (1) Z39.50 ゲートウェイサーバの公開（クライアントの開発）

- ・ CULCON（日米文化教育交流会議）等からの要請を受け、日本の書誌データベースへのアクセス改善を図るため、国際標準である情報検索プロトコル Z39.50 で総合目録データベースを検索できるよう、システム開発を行い、総合目録データベース検索専用サーバのプロトタイプを開発し、2001 年 6 月から試行運用を開始した。
- ・ これにより、海外書誌ユーティリティや各図書館システムなどで Z39.50 クライアント機能を有しているシステムから総合目録データベースを検索利用することが可能となっている。
- ・ プロトタイプは当面、教育用データベース（図書雑誌の書誌所蔵データ）をサービス対象とし、提供形式 3 形式（MARC21 形式、NACSIS-CATP 形式及び SUTRS）のうちから選択することができる。
- ・ 一方、海外書誌ユーティリティが保有する書誌データベースのうち、Z39.50 サーバで公開されているものについて、NACSIS-CAT において参照ファイル的に利用するために、Z39.50 クライアント機能を開発中である。
- ・ 現在、各国の国立図書館等から MARC データをテープ形式（最近では ftp での提供も多くなったが）で購入し、バッチ処理によって NACSIS-CAT 形式へフォーマット変換して利用に供している。クライアント機能が完成した暁にはこのようなバックグラウンド業務は不要となることが期待される。

#### (2) SPCAT

- ・ SPCAT は、NACSIS-CAT に登録した書誌所蔵データを参加図書館単位で抽出して検索ソフトウェアとともに提供するサービスである。同様の機能を持つサー

ビスとして、従来は個別版 CD-ROM 提供サービスがあったが、今回の SPCAT はこのサービスに代わるインターネットに対応したサービスである。

- ・ システム構成はクライアント/サーバ型であり、検索サーバは WWW の仕組みを利用して検索機能を提供するので、学内 LAN 等を利用したネットワーク経由での提供ができる。また、クライアントの検索システムは WWW ブラウザが動作する環境で利用できる。
- ・ データは、従来の独自形式から MS-Access2000 互換形式に変更したので、データをハードディスクにコピーして利用することも、MS-Access2000 で利用することも可能となっている。
- ・ NACSIS-CAT で新規に登録したデータについては、個別版検索ソフトウェアを用いて検索結果を作り、随時ネットワークを通して入手し、既存データに反映させることができるようになっている。これによって、ローカルシステムで迅速な OPAC 提供が実現する。
- ・ 現在、2002 年 3 月までをめでに試験提供を行っており、4 月以降本格サービスを開始する予定である。

### (3) 電子ジャーナルの記述

- ・ 近年、インターネット上のコンテンツとして利用できる情報資源が急速に増加している。特に逐次刊行物はいわゆる電子ジャーナルとして、商業出版社から学協会に至るまで電子化が進められている。
- ・ これらの電子ジャーナルを総合目録データベースに登録できるようにするため、総合目録小委員会で審議検討した結果、「暫定的な方法」で登録することとした。今後多くの事例を踏まえて再度検討・改訂することを含んだ扱いとしてある。
- ・ 今回コーディングマニュアルにおいて定められた記録方法は概略次のとおりである。
- ・ 適用範囲としては、インターネット上で提供されている、逐次刊行物の定義に合致するもののうち、図書館でアクセスを保證できるもの、とした。具体的には、出版者から購入しているオンラインジャーナルや図書館等で作成・提供しているオンライン研究紀要類などが該当する。ILL で利用可能かどうかは問わない。
- ・ 従来の記述を行う上で基準となっていた初号は、単独の情報源として存在するとは限らず、情報源自体が最新のものに変更されることから、確認できる最新の情報源を記述の基準とする。
- ・ さらに、電子ジャーナル以外のネットワーク上の情報資源をどのように取扱うかについても、現在検討組織を設置して継続検討中である。こちらの範疇のものは、メタデータデータベースとして総合目録データベースとは別の取扱いをすることが想定されている。

### (4) ポータルサイト（引用文献リンク、メタデータデータベースシステム）

- ・ 国立情報学研究所では、従来から提供してきている各種情報サービスを統合し、より高度化したサービスとして提供することを計画している。
- ・ それは、これまで個々のサービスが独立した形で情報提供してきたものを、一つの入り口から最適なサービスへとナビゲーションするポータルサイトとして想定されている。このサービスの中核をなすものが、引用文献リンクシステムであり、メタデータデータベースである。

引用文献リンクシステム

- ・ 国立情報学研究所で構築しているさまざまなデータベースから引用文献情報を抽出し、引用リンクデータベースを構築し、併せて引用文献情報の編集・統合を行う。
- ・ 電子図書館サービス、目録所在情報サービス及び情報検索サービスなどの国立情報学研究所内でのサービスデータベースの他に、Cross Ref や JST 等の外部で運用されているリンクサーバとも連携して検索リクエストのやりとりを行う。
- ・ 論文名、雑誌名、著者名等が主たるデータ要素となる。これらの情報を適切に同定識別することがこのシステムが効果的に動作するかどうかの分岐点となる。
- ・ 特に著者名については、図書の著者（主として目録所在情報サービスでの著者名典拠としてエントリーされる）と論文の著者（主として研究者ディレクトリにエントリーされる）の統合的管理が必要である。  
メタデータデータベースシステムプロトタイプ開発
- ・ ネットワーク上の情報資源の取扱いに関する検討が必要であることが、2000年度総合目録委員会において指摘され、2001年度に検討会議を2回開催し検討を行った。その結果、ネットワーク上の情報資源の組織化を行う上で必要なメタデータを、共同でデータベース構築していくことを中心としたネットワーク上の情報資源の取扱い（案）を決定した。
- ・ 現在、実務者レベルのワーキンググループにおいて、記述要素、選定基準、システム機能の検討を行っており、これらを基にメタデータデータベースシステムプロトタイプの開発を行う計画である。

#### 4. まとめ

以上、NACSIS-CATに関連したいくつかの動向を紹介した。これらを著者名典拠の観点から見ると、

- (1) 著者名典拠の作成に当って、各国作成の典拠情報を導入すること
- (2) 図書の著者以外の論文著者、特にネットワーク上の情報資源における著者名等についてのコントロールが必要

といった課題があるだろう。

国立情報学研究所が構築しようとしているメタデータデータベースシステムでは、著者名に対するコントロールを NACSIS-CAT の著者名典拠レコードとリンクという形では当面行わない、とした。これは NACSIS-CAT 典拠レコードにこれまで登録されていない著者名が頻出する可能性が高く、新たに著者名典拠レコードを作成する作業負担が大きいと想定されたためである。

NACSIS-CAT の典拠レコードとのリンクという形は採らないにしても、何らかのコントロールは必要である。それをどのような形で行うか、統合的な情報提供サービスを提供していく中でさらに検討しておく必要がある。