Workshop on Authority Control among Chinese, Korean and Japanese Languages (CJK Authority 3)
March 14 – 18, 2002, Karuizawa, Tokyo, Kyoto

Current Situation of NACSIS-CAT as of FY 2001

SAKAI Kiyohiko & KYOTO Toru
Contents Division, Development and Operations Department,
National Institute of Informatics (NII)

Contents

1 Multilingual Processing

(1) Chinese Materials

In January 2000, UCS (Universal multiple-octet coded Character Set) was incorporated into NACSIS-CAT which extended the possibilities.

A Working Group on Database Construction of Chinese Materials was established in 1995, and the WG reached to a draft of "Guideline of Cataloging of Chinese Materials" in 1998. Particular feature of the draft is: Japanese reading with word division is assigned to Chinese title for retrieving by words. The Guideline includes rules of reading and word division. Parallel transcription by Pinyin to Japanese reading is also devised.

Fonts applied are based on the materials. However, famous authors, authors with many works, authors until the Qing dynasty are transcribed in well known fonts or fonts that appear in most-used reference books.

Systems development had been carried out together with the development of catalog description. It intended to implement UCS capability to all the databases and application software, to create and implement an *Integrated Index of Kanji* for efficient retrieval of Chinese Characters in China, Japan and Korea.

With these development and preparation, inputting became effective in January 2000 for Chinese materials.

CHINAMARC was introduced at the same time (January 2000) for assisting cataloging. Further possibility is being discussed for implementing "Chinese Name Authority File"

(2) Korean Materials

A Working Group on Database Construction of Korean Materials was established in 1999. The WG was extended to a "Subcommittee on Handling of Korean Materials" and descriptive rules were determined. A handling principle is defined in FY 2001 for applying it from April 2002.

Particular feature of the draft is: when word-division is necessary, reading in Hangul is accepted. Japanese reading with word division is assigned to Chinese title for

retrieving by words.

Scripts applied are based on the materials. However, as is done in Chinese materials, famous authors, authors with many works are transcribed in well known fonts. Author names in Chinese characters are assigned readings in Hangul.

Systems revision has been carried out during the fiscal year 2001, thus it becomes effective in April 2002 for inputting to NACSIS-CAT.

KOR MARC is to be implemented as a reference MARC for cataloging Korean materials in FY 2002

(3) WebUIP

In order to input multilingual data, it is necessary to install a client system that operates under UCS. However, there would be libraries that are unable to introduce UCS client systems, because procurement regulations force libraries to keep the current systems for more than 4 to 6 years.

When there is no UCS client, all the UCS characters, such as Simplified Chinese characters, is displayed in code strings. For Example, a character is put between ◆ (black diamond symbol) like「中国古代佚名哲学名著◆U8BC4◆述」, and UCS code number is appeared between the black diamonds "◆". It is not easy to detect what character it is.

A UCS client system, that should be developed at each local systems, is developed at NII for meeting such demands as assisting cataloging and displaying. It was released in January 2001 for provisional use.

(4) Webcat

Data input of Chinese materials has been actively carried out since January 2000 by a few libraries. As of February 2002, some 80,000 records have been registered through multilingual client systems.

These multilingual records are, as any other NACSIS-CAT records, uploaded to the Webcat, which make the union catalog database available to the public through the Internet.

Multilingual provision was installed to Webcat in January 2001, for displaying simplified Chinese characters at user's browsers. Users with UCS browsers (the latest versions of Internet Explorer or Netscape) are able to access the simplified Chinese characters through the English version of Webcat.

Webcat is facilitated the Integrated Index of Kanji, users are able to access Webcat regardless of character type.

2. Other Developments

(1) Z39.50 Gateway Server (Development of a client system)

Upon the request by the CULCON (The United States Japan Conference on Cultural and Educational Interchange) for improving accessing Japanese bibliographic databases, Z39.50 was implemented for the union catalog database. A prototype of retrieval server systems for the union catalog database was developed and its trial operation was started in June 2001.

Now our database is available by systems with Z39.50 client functions from anywhere of the world.

The prototype retrieval server covers only the training database (bibliography and

holding data of monographs and serials) for the time being, three types of record formats are available (MARC 21, NACSIS CATP and SUTRS).

Z39.50 client function is being developed for NACSIS-CAT for accessing bibliographic databases held by Z39.50 servers to use as a reference MARC.

MARC data are being imported from national libraries abroad in magnetic tape format, (currently the main stream is of ftp), then they are processed and converted into NACSIS-CAT format in batch mode. It is expected that these background work shall be left out when the client function is fully developed.

(2) SPCAT

SPCAT is a retrieval software for participating library(ies) to request registered data of bibliography and holdings through the Internet, and to update the local database. The similar service in the past was the local catalog data in CD-ROM, that is now replaced by SPCAT. SPCAT systems configuration is of client/server type. Its retrieval server offers Web interface. Therefore, the participating library can extend the service to the end users through the campus LAN. Data format is now of MS-Access 2000 compatible that was changed form the unique format of the past. User libraries are able to apply easily on MS-Access 2000.

SPCAT realizes to retrieve, acquire the registered data, and to reflect them to the existing database through the network. It allows a fresh OPAC service at the local systems.

Trial provision is being made until the end of March 2002, and a formal service shall be started in April 2002.

(3) Description of Electronic Journals

The internet resources are rapidly increased. Serials, especially, are being made into digital format as electronic journals by commercial publishers, by academic societies. A tentative measure, being discussed at the Union Catalog Subcommittee, was introduced for registering electronic journals into the union catalog database. It is decided to be tentative for future determination by examining further cases. Significant characteristics of the measures defined in the coding manual are as follows:

Coverage is of those serials on the Internet which library is able to assure access. In practice, it is those of online journals subscribed from publishers, and/or those online research bulletins created and maintained by libraries and other institutions, irrespective of ILL availability.

The first issue, that was the setting unit of traditional description, may not exist as a unique source of information, and it may be changed to be the latest, the starting point of description is to be the latest source of information available.

Description of information resources on the networks, other than electronic journals, is also discussed by a special WG. These are categorized as metadata databases and assumed to be treated separately from the union catalog database.

(4) Portal Site (Citation database links, Metadata database system)

NII plans to offer a new service which integrates and enhances the existing services.

The new service provides a portal site of the existing various services, that enable users to navigate from a single entrance to the most suitable service. The core of this portal service is the Citation database links and the Metadata database.

i) Citation database link system

NII plans to implement a system of citation link database by extracting citation

information from the existing databases at NII, together with compilation of citation information at NII.

Retrieval requests shall be processed to NII databases such as NACSIS-ELS, NACSIS-CAT and NACSIS-IR as well as to external link servers such Cross Ref or JST.

Article titles, serials titles, and author names become the major data elements. Identification of these information is the critical factor of the effectiveness of the system.

Particularly of author names, an integrated control of monographs authors and journal article authors shall be necessary. Monographs authors come from NACSIS-CAT, and the journal article authors come from the Researcher Directory.

ii) Prototype system of metadata database system

It was pointed out at the Union Catalog Committee in FY 2000 that the treatment of network resources should be discussed, and two discussion meetings were held in FY 2001. As a result, a draft guideline was proposed for network information resources that a metadata database should be constructed through a shared database construction.

In FY 2001, a WG composed of young practitioners has been analyzing descriptive elements, criteria of choice, systems functionality, and a prototype metadata database system shall be developed based on the analysis.

3. In Summary

Current situation of NACSIS-CAT is reported. From the point of the name authority control there are urgent tasks such as:

(1) import of national authority data for maintaining our name authority
(2) control of article author names, other than monographs authors, especially of the network information resources

In the planning of the metadata database system at NII, it was decided not to link with the NACSIS-CAT authority record, because s high probability is expected that new names are not yet registered to the NACSIS-CAT authority database, and high work load and work cost are expected to create new authority records.

Some kind of control is necessary for author names in the Metadata database other than having linking with the NACSIS-CAT authority database. It is planned to continue discussion and analysis for a new form of data control by operating the integrated information provision services.

Appendix Current Status of NACSIS-CAT Operation (March 2002)

(1) Participation, Data Construction, and Updating

975 institutions are participating to NACSIS-CAT as of January 2002 (See Table 1). An increase of participation was observed during FY 2001 because it was a year of national holding renewal of the Union Catalog of Serials, particularly by libraries of national experiment and research institutions, ministerial research laboratories, and hospital libraries.

Table 1: Participating Institutions

Type	No. of Participating Institutions	Population	Participation Ratio	No. of New Participation during FY
National university	99	99	100%	0
Municipal university	72	74	97.3%	2
Private university	454	497	91.3%	42
College	116	489	23.7%	18
Technical college	55	62	88.7%	2
Other	165	—	—	77
National Cooperative Research Institution	14	19	73.7%	0
TOTAL	975	—	—	141

Data construction exceeded 6 million unique titles as of 23rd February 2002. Monthly increase numbers in the last three months up to February 2002 are: bibliography records of monographs is 35,685, holding is 523,085.

Table 2: Records of NACSIS-CAT Database

Type		No. Record	Increase in FY 1999	Increase in FY 2000	Increase in FY 2001
Monographs	Bibliography	6,106,490	447,749	448,483	397,086
	Holdings	58,623,169	6,731,165	7,186,233	6,171,565
Serials	Bibliography	242,401	4,325	8,408	5,883
	Holdings	3,652,204	124,345	147,139	87,366
Author Name Authority		1,194,066	52,520	45,682	37,972
Uniform Title Authority		19,560	1,167	578	971
Serials Title Change Map		31,057	719	3,344	2,620

Data updating has been active during FY 2001. New is a weekly total of new addition. Updates is a total of data correction in one week. When corrections are made within the same week to the same record, it is counted as one. If the actions happen over more than a week, it is counted as 2. Delete is the weekly total of deletion. recently a trend is observed that the ration of correction to addition is increasing.

Table 3: Updating of the NSCSIS-CAT Union Catalog Database

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

(2) Reference Files (MARC)

A variety of reference files or Source MARCs are incorporated in NACSIS-CAT. During FY 2001, "REMARC", based on the US Library of Congress is to be uploaded, and 4.2 million records for 1890-1967 are going to be available.

Table 4: Reference Files (MARC)

Database		No. of Record	Period
LC (USMARC)	Monographs	6,577,725	1968 -
	Non book materials	268,147	1984 -
	Author Name	3,400,010	1977 -
	Uniform title	275,251	1977 -
	Serials	872,449	1973 -
JP (JAPANMARC)	Monographs	2,712,427	1868 -*
	Author Name	325,315	1969 -
	Serials	100,582	1989 -
UK (UKMARC)	Monographs	1,868,636	1950 -
DN (DNMARC)	Monographs	3,372,190	1945 -
TRC (TRCMARC)	Japanese Monographs	1,009,682	1985 -
GPO (GPOMARC)	Monographs	500,009	1976 -
CH (CHMARC)	Monographs	474,025	1988 -
BL (BLSERIAL)	Serials	610,172	

* excludes the Taisho period.

(3) Quality Control

Quality control operation is carried out in three steps (ABC).

A. Input checking at a local library, when registering (currently the total number of catalogers inputting is 3,200).

① duplicate input shall be avoided by thorough search.

② confirm to the coding manuals.

B. Finding errors at a local library, when using records.

① data correction according to the coding manuals (some 500,000 corrections annually),

② among the total correction, 150,000 records are corrected based on the communication of the original inputting library.

③ data correction based on reporting of duplicates or date to be deleted to NII (2,000 records annually).

④ report to NII of correction (1,000 records annually).

C. Quality control at NII

It is impossible to check all the records of the Union Catalog Database at NII, which also does not hold materials. The following measures are introduced and operated.

① devising and maintaining the Cataloging Information Standards, the Coding Manuals, etc.

② publicity to and training of cataloging operators on systems manipulation, manuals. Publishing manuals, newsletters, and holding training courses.

③ Questions and Answers Database (60 questions/month, 3,000 access/month)

④ analysis on quality check and determination on solutions, new measures of quality control.

⑤ research and development as well as systems development for quality control.

⑥ Quality verification operation on the data registered within a limitation.

automatic unification processing of duplicate records (bib and holdings)	5,000 / year
collective record deletion	15,000 / year
change linking, linking data correction	6,000 / year
reporting on correction by telefax	5,000 / year
answering the questions through phone and e-mail	2,600 / year
eye-check by print-list of serials records and correction*	33,000 / year

* (For serials record, entire records are to be checked for compilation of the Union Catalog of Serials)