

『東洋学文献類目』編纂の歴史 — CHINA3

安岡孝一*

1987年11月、『東洋学文献類目』はCHINA3という名前のデータベースとして、京都大学大型計算機センターで運用が開始された。その後2000年10月、CHINA3という名称はそのままに、京都大学人文科学研究所附属漢字情報研究センターのデータベースとして、運用が継承された。本稿ではCHINA3の歴史について、ごくかいつまんで述べる。

1 CHINA1とCHINA2

CHINA3というからには先行するCHINA1とCHINA2があるはずだ、と考えるのは自然なことである。しかし、実はCHINA1とCHINA2は、『東洋学文献類目』とは何の関わりもない。CHINA1とCHINA2は、京都大学大型計算機センターのFAIRS上に構築されたデータベースの名称であり、CHINA1は『明代登科録彙編』を、CHINA2は『樊南文集』を、それぞれデータベース化したものだった[3, 5]。1980年9月、京都大学大型計算機センターでは、「FAIRS日本語処理(JEF)版」というデータベースエンジンの運用を開始していた。これを用いた漢字データベースの研究開発が企画されたが、その対象の1つとして、京都大学人文科学研究所附属東洋学文献センターとのタイアップで、『明代登科録彙編』が選ばれたのである[1]。その後、『樊南文集』と『東洋学文献類目』が、データベース化の対象としてピックアップされたことから、『明代登科録彙編』にはCHINA1、『樊南文集』にはCHINA2、『東洋学文献類目』にはCHINA3というデータベース名が、それぞれ付与されたのである。

2 CHINA3の開発

『東洋学文献類目』のデータベース開発計画が、京都大学大型計算機センターにおいて承認されたのは、1981年5月のことである[2]。7月にはフォーマットの検討が開始され、10月にはカードの書式とFAIRSのフィールド[†]が決定された(図1)。

このフォーマット中、もっとも特徴的なのは、「排列のための手がかり」というフィールドがあることである。ここでいう排列とは、冊子体の『東洋学文献類目』を作成する際の排列である。すなわち、1年分のデータをまず「分類コード」でソートし、さらに各分類内を「排列のための手がかり」でソートすれば、冊子体で印刷する順序にデータが並ぶという仕掛けになっていた。つまるところCHINA3は、共用データベースとしての運用よりは、冊子体を印刷するためのオフセット版下作成を狙ったものだった[4, 8]。

*京都大学人文科学研究所附属漢字情報研究センター

[†]MARCでのタグに相当するものだが、FAIRSではフィールド名の先頭の1文字は英字ということになっていた。

① 雑誌カード 作業年度 事務番号

010	A (C)	□□△△△△
020 (021)	\$j	\$k \$g
	\$d	
030 (031)	\$i	\$n

④ 書評論文のカード

200 (201)	○○○○○	
	\$r \$f \$z \$n	
	\$p	
	\$l B	□□△△△△
	(D)	
	\$w	

批評された単行本の
コントロール番号

⑤ 単行本カード

510	B (D)	□□△△△△
520 (521)	\$t	\$u
530	\$a \$f \$z	\$n
540 (541)	\$e \$d \$v \$p	\$h \$s
550 (551)	\$m	\$b \$c (又は\$x \$y)
560	\$q	

② 論文カード

論文番号

100 (101)	○○○○○	
	\$t \$u	
	\$a \$f \$z \$n	
	\$p \$m	
	\$b \$c (又は\$x \$y)	
	\$q	

⑥ 批評論文カード

300 (301)	○○○○○	
	\$r \$f \$z \$n	
	\$p	
	\$l A	□□△△△△○○○○○
	(C)	

批評された論文の
コントロール番号⑤' 多巻物の場合の全体の
タイトルのカード

510	B (D)	□□△△△△
520 (521)	\$t \$u	
530	\$a \$f \$z	

③ 子目論文カード

110 (111)	○○○○○	
	\$t \$u	
	\$a \$f \$z \$n	
	\$p \$m	

⑦ 多巻物の場合の各巻の
単行本カード

510	B (D)	□□△△△△
520 (521)	\$t \$u	
530	\$a \$f \$z	
540 (541)	⑤と同様	
550 (551)		
560		
570	\$l B (D)	□□△△△△

全体のタイトルのコントロール番号

フィールド一覧

フィールド名 内 容

BANGO	コントロール番号
A, A ₁	著者名 (漢字)
A ₂	著者名 (EBCDIC)
B, B ₁	分類コードとその漢字表記
C	排列のための手がかり (地域・時代・事項・内容)
C ₁ , C ₂	地域コードとその漢字表記
N ₃ , N ₄	時代コード (上, 下限)
C ₇	事項コード
C ₈	内容コード
D	出版年 (月)
E	出版社 // 出版地 (漢字)
E ₂ , E ₃	同上 (EBCDIC)
F	著者名 (カナ)
G	号
H	規格
I	巻, 号の注記
J	雑誌名

フィールド名 内 容

K	巻
L	リンク番号
M	ページの注記
N	雑誌・著者の注記
O	階層コード
P	ページ
Q	手がかり要語
R, R ₁	評者
S	叢書名
T, T ₁	表題 (漢字)
T ₂	表題 (EBCDIC)
U, U ₁	副書名 (漢字)
V	巻数
W	書評についての注記
X, X ₁	学者名 (漢字)
Y	学者名の排列種別
Z	著者名の排列種別

図 1: 東洋学文献類目電算化のためのフォーマット [4]

ところが、データベースとしての運用を考えた場合には、これが問題となった。冊子体を印刷する際に、これまでの『東洋学文献類目』の伝統を踏襲するならば、漢字は全て正字[‡]にしなければならない。しかし、データベースの検索者は、常用漢字で検索をおこなってしまうことが多く、その結果、正字にはマッチングが起こらない。このジレンマを解決するため、たとえば論文の表題に対するフィールドに T と T1 の 2 つを設け、T には正字による表題を、T1 には常用漢字による表題を、それぞれ格納することにしたのである [6]。

『東洋学文献類目』は 1980 年度分までは、旧来の手作業によるデータカードの分類とソートをおこない、写植組版で発行された。1981 年度分が CHINA3 の最初のデータとなり、1981 年 12 月にデータ入力を開始して、1984 年 3 月に初のオフセット版冊子体として発行された。さらに 1984 年度分までデータが入力され、1984 年度版の冊子体が発行された後、1987 年 11 月 5 日[§]に CHINA3 は、京都大学大型計算機センターの共用データベースとして公開されたのである [7, 9]。

3 WWW と CHINA3

しかし、公開後 10 年もすると、CHINA3 は非常に使いにくいものとなってしまっていた。CHINA3 は FAIRS 上のデータベースであり、漢字コードには富士通の JEF を使っていた [9]。つまり、JEF 拡張漢字を使えない端末では、表示が化けてしまうのである。

これを解決する窮余の策として、Web 版 CHINA3 が、京都大学大型計算機センターで 1998 年 2 月に公開された [10]。Web 版 CHINA3 は、京都大学大型計算機センターの CHINA3 を、WWW から検索にいくためのインターフェースであり、データベースエンジンには FAIRS をそのまま用いていた。したがって、検索には京都大学大型計算機センターの利用者番号が必要だった (図 2) し、検索結果も JEF 拡張漢字が画像で埋め込まれるものの、FAIRS の出力そのままといった趣 (図 3) のものだった。

2000 年 4 月、京都大学人文科学研究所附属東洋学文献センターが京都大学人文科学研究所附属漢字情報研究センターに改組されると同時に、CHINA3 を漢字情報研究センターに移転する作業が開始された。自前のデータベースを立ち上げることにしたのである。ただし、京都大学大型計算機センターでの 20 年分のデータを生かす必要があったことから、フィールド等は最小限の手直しとした。また、京都大学大型計算機センターでのデータ入力と冊子体版下作成は 2000 年度分を最終とし、2001 年度分からは、自前のデータベースへの直接入力かつ版下作成という形を取るようになった。

2000 年 10 月に京都大学人文科学研究所附属漢字情報研究センターで公開された CHINA3 for WWW は、データベースエンジンに xemacs を使用したものだ。検索に利用者番号などは必要なく (図 4)、検索結果の表示もわかりやすいものとなった (図 5)。CHINA3 は、真の意味での共用データベースになったのである。

[‡]康熙字典でいうところの正字。

[§]翌週には漢籍担当者職員講習会 (漢籍電算処理) がひかれており、そこでは CHINA3 の実習が予定されていた。

File Edit View Go Bookmarks Options Directory Window
Help

Back Home Reload Open Print Find

Location: http://www.kudpc.kyoto-u.ac.jp/Services/DataBaseServices/china3.ht

What's New? What's Cool? Destinations Net Search People Software

CHINA3 for WWW

東洋学文献目録(京都大学人文科学研究所附属東洋学文献センター刊行)のタイトルを WWW から検索します。

- 検索は京都大学大型計算機センターのMSPシステム上で行なわれますので、利用者番号とMSPシステムのパスワードが必要です。
- 検索に要したCPU時間は課金されます。検索を途中で中止した場合、課金が非常に大きくなる可能性があります。検索は中止しないで下さい。
- 検索のために、MSP上に@CHINA3.DATAというデータセットが電脳時に作られます。
- 表示の都合上、できれば日本語版Netscapeをお使い下さい。sakuraではnetscape-jaというコマンド名で運用しています。
- さらにきめ細かい検索を行ないたい方は、MSPシステムのIRS CHINA3コマンドを直接お使い下さい。
- このページに関するご質問は [ここ](#) へどうぞ。
- 利用者番号の取得については [ここ](#) へどうぞ。

利用者番号

パスワード

ジャンル

^歴史	✓宗教	✓金石・古文書学
✓地理	✓学術思想 附教育	✓民族学
✓社会	✓科学	✓言語文字学
✓経済	✓文学	✓書誌学
✓政治	✓芸術	✓雑纂 附革命文物
✓法制	✓考古学	✓学会消息

検索文字列

補助漢字
☒表示できる
☐表示できない

検索開始

図 2: Web 版 CHINA3 表紙画面 [11]

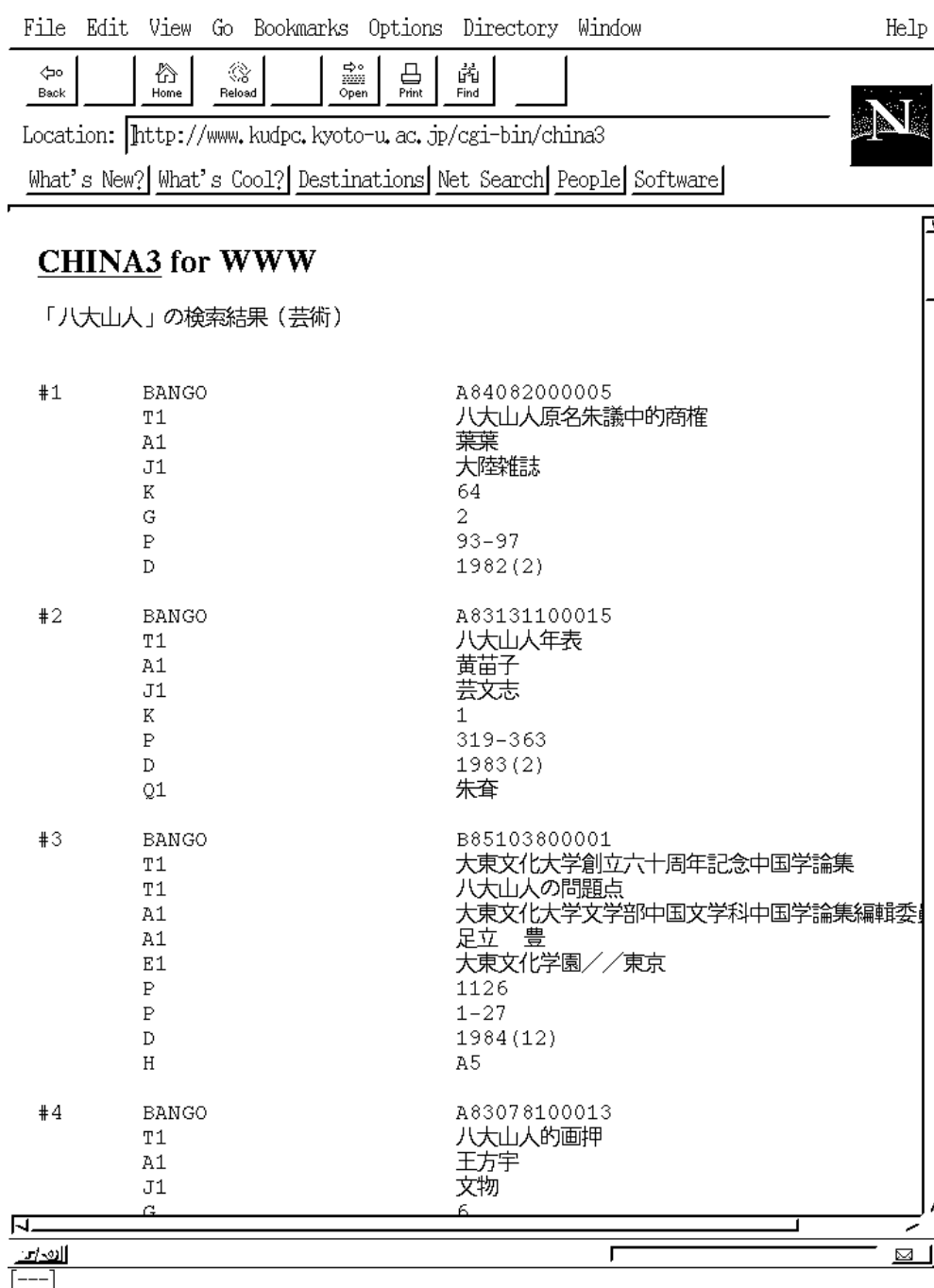


図 3: Web 版 CHINA3 検索結果 [11]



図 4: CHINA3 for WWW 表紙画面 [12]

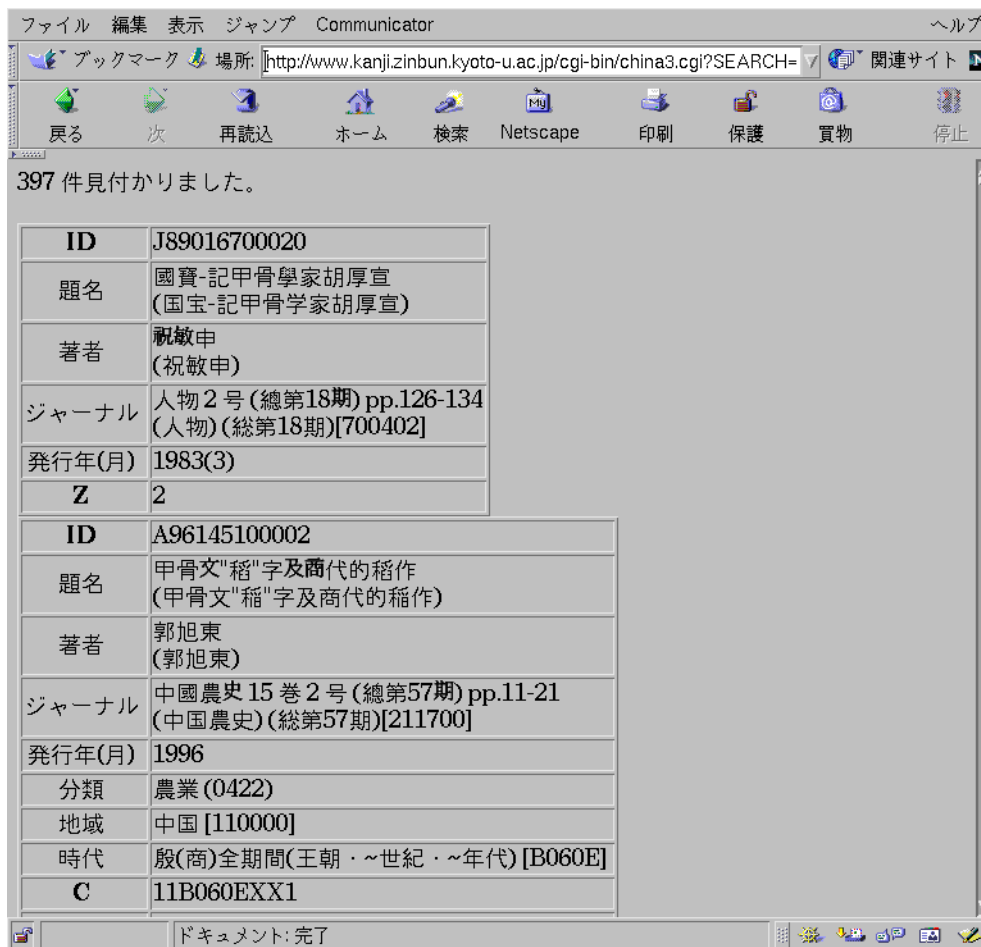


図 5: CHINA3 for WWW 検索結果 [12]

現在の CHINA3 は、データベースの文字コードを UTF-8 に変更すると同時に、データベースエンジンに PostgreSQL を採用しており、より高速で汎用性の高いデータベースとなっている [13]。また、1980 年度以前のデータへの遡及入力も開始しており、『東洋学文献類目』発刊以来の全データが揃う日も、そう遠くない。

CHINA3 の今後の発展に期待されたい。

参考文献

- [1] 星野聡, 今城一夫, 河野典, 村尾義和, 勝村哲也: 漢字処理に関する研究開発, 全国共同利用大型計算機センター研究開発論文集, No.2 (1980 年 11 月), pp.119-128.
- [2] 昭和 56 年度開発計画 (第 1 期) について, 京都大学大型計算機センター広報, Vol.14, No.3 (1981 年 6 月), pp.118-119.
- [3] 勝村哲也, 星野聡: データベース CHINA1 について, 京都大学大型計算機センター広報, Vol.14, No.3 (1981 年 6 月), pp.128-132.
- [4] 都築澄子, 志水喜久子: 「東洋学文献類目」編集作業の電算化について, 静脩, Vol.19, No.3 (1982 年 12 月), pp.3-7.
- [5] 星野聡, 三重まゆみ, 勝村哲也: データベース CHINA2 について, 京都大学大型計算機センター広報, Vol.16, No.3 (1983 年 6 月), pp.122-127.
- [6] 星野聡, 勝村哲也: 東洋学文献類目データベースの研究と開発, 情報処理学会論文誌, Vol.25, No.2 (1984 年 3 月), pp.187-193.
- [7] 電算機利用による「東洋学文献類目」のフォーマット, 京都大学人文科学研究所附属東洋学文献センター昭和 62 年度漢籍担当職員講習会 (漢籍電算処理) 資料 2 (1987 年 11 月).
- [8] 河野典: 東洋学文献類目のデータベース化と冊子体作成について, 京都大学人文科学研究所附属東洋学文献センター昭和 62 年度漢籍担当職員講習会 (漢籍電算処理) 資料 3 (1987 年 11 月).
- [9] 都築澄子, 河野典, 村尾義和, 桶谷猪久夫: データベース CHINA3 について, 京都大学大型計算機センター広報, Vol.20, No.6 (1987 年 12 月), pp.308-317.
- [10] データベース CHINA3 のデータ更新について, 京都大学大型計算機センター広報, Vol.31, No.2 (1998 年 4 月), p.54.
- [11] 安岡孝一: Web 版東洋学文献類目 (CHINA3) の実現, 京都大学大型計算機センター第 62 回研究セミナー報告 (1999 年 3 月), pp.24-36.
- [12] 守岡知彦: CHINA3 for WWW, 京都大学人文科学研究所附属漢字情報研究センター平成 13 年度漢籍担当職員講習会 (漢籍電算処理) 資料 8 (2001 年 10 月).
- [13] 守岡知彦: 東洋学文献類目データベース・システムについて, 京都大学人文科学研究所附属漢字情報研究センター平成 14 年度漢籍担当職員講習会 (中級) 資料 4-1 (2002 年 11 月).