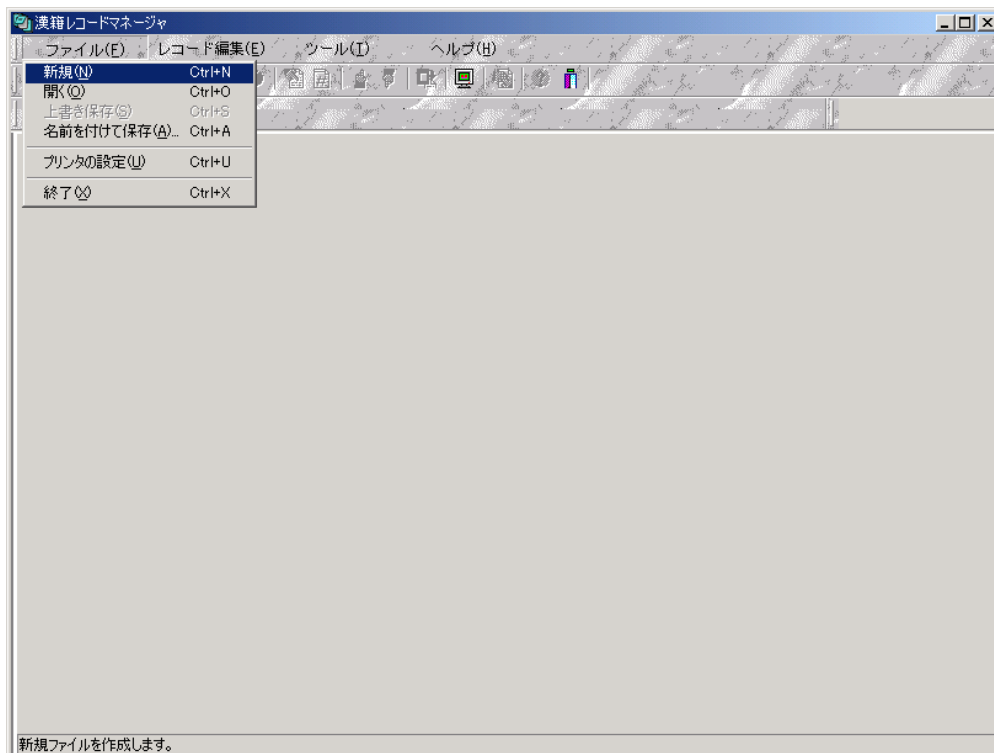


漢籍レコードエディタの簡単な使い方

安岡孝一

1 エディタの起動

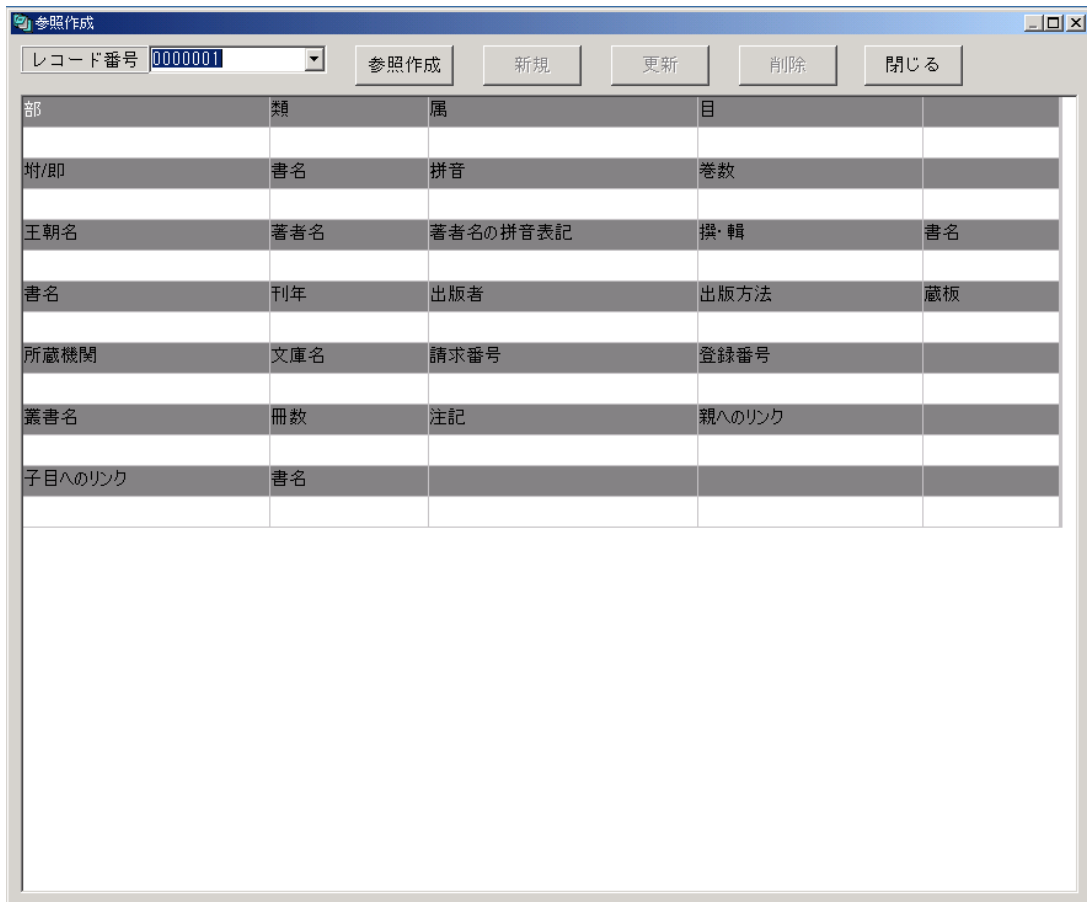
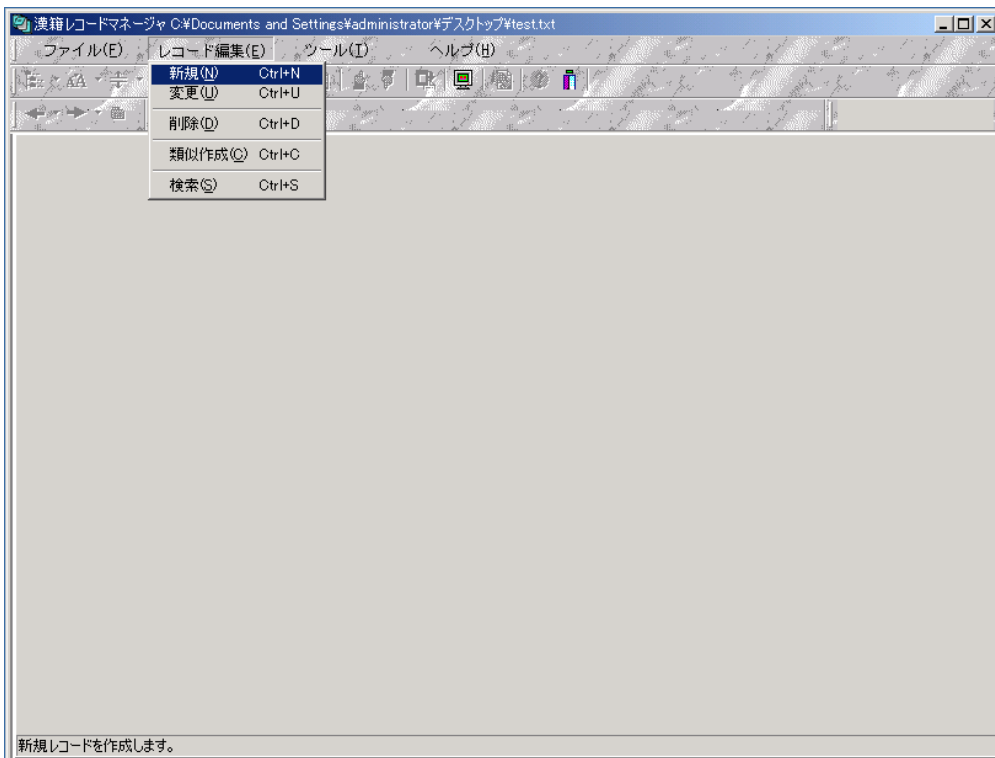
デスクトップ上の「KansekiEditor」をダブルクリックすると、漢籍レコードマネージャが起動します。ここから「ファイル」「新規」を選んで(下図)、適当な新規ファイル名を入力して下さい。



ただし、ファイル名は英数字だけにしておいた方が無難でしょう。

2 漢籍レコードの入力

漢籍レコードの入力は、基本的に「レコード編集」「新規」でおこないます(次ページ上図)。これにより、漢籍レコードの入力画面が立ち上がります(次ページ下図)。



2.1 レコード入力の基本

漢籍レコードの入力は、漢籍用図書カードの作成を、コンピュータ上でおこなうものです。たとえば

史	有象列仙全傳九卷
IX	
7	
明汪雲鵬輯	
寛政三年大阪書林柳原喜兵衛等刊本	
8冊	

という図書カードに対しては

参照作成

レコード番号 0000001

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	傳記類	雜傳之屬		
附/即	書名	拼音	卷数	
	有象列仙全傳		九卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
明	汪雲鵬		輯	
書名	刊年	出版者	出版方法	蔵板
	寛政三年	大阪書林柳原喜兵衛等	刊本	
所蔵機関	文庫名	請求番号	登録番号	
		史-IX-7		
叢書名	冊数	注記	親へのリンク	
	8			
子目へのリンク	書名			

という入力をおこなうことになります。書名と巻数、著者の王朝名、刊年と出版者および出版方法は、それぞれ別の枠に入力すること。また、冊数は数字だけを入力することに注意して下さい。

図書カード 1 枚分の入力終了したら、右上の「閉じる」をクリックして下さい。「レコードへの変更を保存しますか?」に対して「はい」を答えるのを、忘れないように。

2.2 著者が複数の場合

史	泉幣圖説六卷
XV	
6	
清吳文炳吳鸞同撰 嘉慶五年吳氏香雪山莊刊本 2 冊	

上記のような、著者が複数(吳文炳と吳鸞)の場合、1 人目の著者を入力したところでマウスの右ボタンをクリックし、「行挿入」をおこなう必要があります(下図)。

参照作成

レコード番号 0000002

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	鏡鑑泉幣雜器之属		
附/即	書名	拼音	卷数	
	泉幣圖説		六卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
清	吳文炳			
書名	刊年		出版方法	蔵板
所蔵機関	文庫名		登録番号	
叢書名	冊数	注記	親へのリンク	
子目へのリンク	書名			

行挿入 (W)
行削除 (X)
1つ上にあげる (U)
フォント (F)

これにより、著者欄を 2 行あるいはそれ以上に増やすことができます。

参照作成

レコード番号 0000002

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	鏡鑑泉幣雜器之屬		
附/即	書名	拼音	卷数	
	泉幣圖説		六卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
清	呉文炳		同撰	
清	呉鸞		同撰	
書名	刊年	出版者	出版方法	蔵板
	嘉慶五年	呉氏香雪山莊	刊本	
所蔵機関	文庫名	請求番号	登録番号	
		史-XV-6		
叢書名	冊数	注記	親へのリンク	
	2			
子目へのリンク	書名			

なお、元の図書カードでは「清呉文炳呉鸞同撰」のように、情報が一部省略されていますが、入力時にはこれらをそれぞれ「清呉文炳同撰」「清呉鸞同撰」と分ける点に注意してください。

2.3 書名が複数の場合

史	隸韻十卷碑目一卷 附碑目攷證一卷
XV	
4	

宋劉球撰 碑目攷證清秦恩復撰
嘉慶十四年江都秦氏用宋石刻本景印
12冊

上記のような、書名が複数の場合、1つ目の書名を入力したところでマウスの右ボタンをクリックし、「行挿入」をおこなう必要があります(次ページ上図)。

参照作成

レコード番号 0000003

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	鏡鑑泉幣雜器之屬		
附/即	書名	拼音	卷数	
	隸韻		十卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
書名	刊年	出版者	出版方法	蔵板
所蔵機関	文庫名	請求番号	登録番号	
叢書名	冊数	注記	親へのリンク	
子目へのリンク	書名			

行挿入(N) 行削除(O) 1つ上にあげる(U) フォント(F)

参照作成

レコード番号 0000003

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	鏡鑑泉幣雜器之屬		
附/即	書名	拼音	卷数	
	隸韻		十卷	
	碑目		一卷	
附	碑目攷證		一卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
宋	劉球		撰	
清	秦恩復		撰	碑目攷證
書名	刊年	出版者	出版方法	蔵板
	嘉慶十四年	江都秦氏	用宋石刻本景印	
所蔵機関	文庫名	請求番号	登録番号	
		史-XV-4		
叢書名	冊数	注記	親へのリンク	
	12			
子目へのリンク	書名			

これにより、書名欄を2行あるいはそれ以上に増やすことができます。増書名や即書名については、書名欄の左端のボックスで「増」や「即」を選択して下さい。

なお、元の図書カードでは、著者の一人が「碑目攷證清秦恩復撰」となっていますが、この場合は著者に含まれる書名の部分(碑目攷證)を、著者欄の右端のボックスに入力する点に注意して下さい(前ページ下図)。

3 子目を含む場合

1	<table border="1"><tr><td>史</td></tr><tr><td>XV</td></tr><tr><td>5</td></tr></table>	史	XV	5	金石三例 清廬見曾輯 清王芑孫評 光緒四年南海馮氏讀有用書齋刊本 4冊
史					
XV					
5					
2	<table border="1"><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr></table>				金石例十卷 元潘昂霄撰 金石三例所收
3	<table border="1"><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr></table>				墓銘舉例四卷 明王行撰 金石三例所收
4 止	<table border="1"><tr><td> </td></tr><tr><td> </td></tr><tr><td> </td></tr></table>				金石要例一卷 清黃宗羲撰 金石三例所收

子目カードは、総目カードに対し、子目へのリンクをクリックすることで、入力をおこないます。上記のカードであれば、金石三例の入力が終わった段階で、子目へのリンクをダブルクリックします(下図)。

参照作成

レコード番号 0000004

参照作成 新規 更新 削除 閉じる

部	類	属	目
史部	金石類	義例之属	
附/即	書名	拼音	卷数
	金石三例		
王朝名	著者名	著者名の拼音表記	撰・輯
清	盧見曾		輯
清	王芑孫		評
書名	刊年	出版者	出版方法
	光緒四年	南海馮氏讀有用書齋	刊本
所蔵機関	文庫名	請求番号	登録番号
		史-XV-5	
叢書名	冊数	注記	親へのリンク
	4		
子目へのリンク	書名		

0000005
0000006
0000007
0000008
0000009
0000010
0000011
0000012

最初の番号を選ぶと、子目編集画面が立ち上がります(次ページ上図)。まずは、最初の子目カードである金石例を入力して(次ページ下図)、その後、子目編集画面を閉じてください。

2枚目以降の子目カードを入力するには、子目へのリンクでマウスの右ボタンをクリックし、「行挿入」をおこなって下さい(次々ページ上図)。これを繰り返していくことで、全ての子目カードを入力し、最後に総目カードを閉じる(次々ページ下図)と、子目を含むデータの完成です。

子目編集1

子目レコード番号 0000005

更新 閉じる

部	類	属	目	
附/即	書名	拼音	卷数	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
書名	刊年	出版者	出版方法	蔵板
所蔵機関	文庫名	請求番号	登録番号	
叢書名	冊数	注記	親へのリンク	
子目へのリンク	書名		0000004	

子目編集1

子目レコード番号 0000005

更新 閉じる

部	類	属	目	
史部	金石類	義例之属		
附/即	書名	拼音	卷数	
	金石例		十卷	
王朝名	著者名	著者名の拼音表記	撰・輯	書名
元	潘昂霄		撰	
書名	刊年	出版者	出版方法	蔵板
所蔵機関	文庫名	請求番号	登録番号	
叢書名	冊数	注記	親へのリンク	
金石三例			0000004	
子目へのリンク	書名			

参照作成

レコード番号 0000004

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	義例之屬		
附/即	書名	拼音	卷数	
	金石三例			
王朝名	著者名	著者名の拼音表記	撰・輯	書名
清	盧見曾		輯	
清	王芑孫		評	
書名	刊年	出版者	出版方法	蔵板
	光緒四年	南海馮氏讀有用書齋	刊本	
所蔵機関	文庫名	請求番号	登録番号	
		史-XV-5		
叢書名	冊数	注記	親へのリンク	
	4			
子目へのリンク	書名			
0000005	金石例			

参照作成

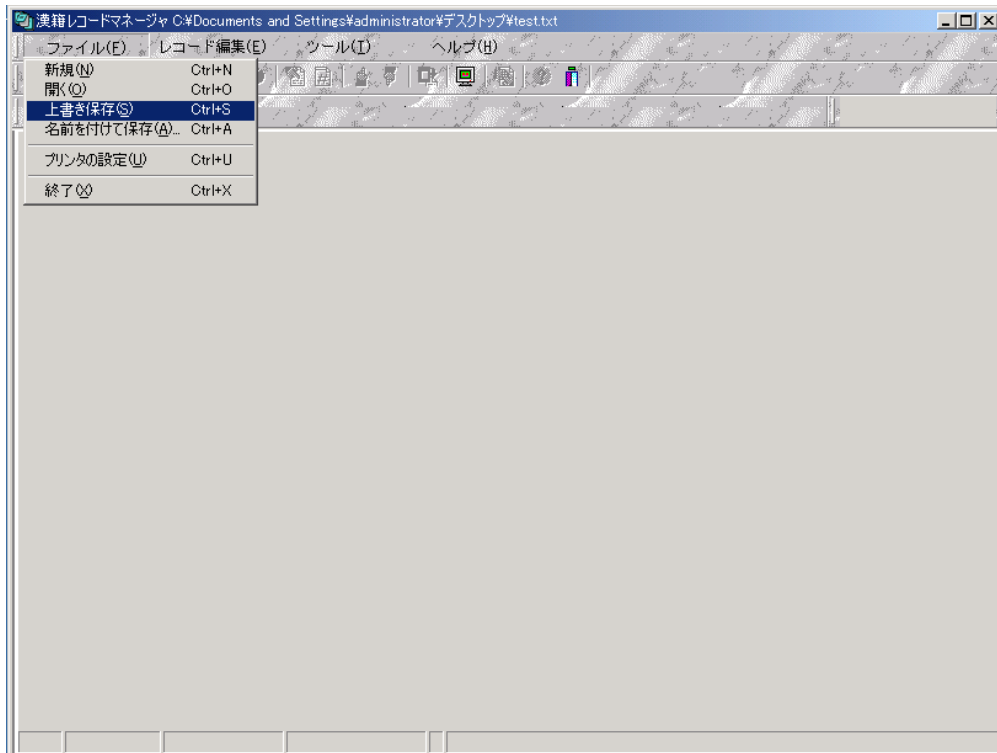
レコード番号 0000004

参照作成 新規 更新 削除 閉じる

部	類	属	目	
史部	金石類	義例之屬		
附/即	書名	拼音	卷数	
	金石三例			
王朝名	著者名	著者名の拼音表記	撰・輯	書名
清	盧見曾		輯	
清	王芑孫		評	
書名	刊年	出版者	出版方法	蔵板
	光緒四年	南海馮氏讀有用書齋	刊本	
所蔵機関	文庫名	請求番号	登録番号	
		史-XV-5		
叢書名	冊数	注記	親へのリンク	
	4			
子目へのリンク	書名			
0000005	金石例			
0000006	墓銘舉例			
0000007	金石要例			

4 エディタの終了

レコードの入力が終了したら、「ファイル」 「上書き保存」(下図)でファイルに保存して下さい。



さらに「ファイル」 「終了」とすれば、漢籍レコードマネージャは終了します。なお、ここまでの4例を順に入力した場合、保存されたファイルは次ページのようになっているはずですが、確かめてみて下さい。また、保存されたファイルの、各タグの意味は次々ページの通りです。

nu:0000001
fi:史部
sf:傳記類
tg:雜傳之屬
ti:有象列仙傳,九卷
au:明,汪雲鵬,輯,
yr:寬政三年
pb:大阪書林柳原喜兵衛等
ed:刊本
vi:8
si:史-IX-7

nu:0000002
fi:史部
sf:金石類
tg:鏡鑑泉幣雜器之屬
ti:泉幣圖說,六卷
au:清,吳文炳,同撰,
au2:清,吳鸞,同撰,
yr:嘉慶五年
pb:吳氏香雪山莊
ed:刊本
vi:2
si:史-XV-6

nu:0000003
fi:史部
sf:金石類
tg:鏡鑑泉幣雜器之屬
ti:隸韻,十卷
ti2:碑目,一卷
st:碑目攷證,一卷
au:宋,劉球,撰,
au2:清,秦恩復,撰,碑目攷證
yr:嘉慶十四年
pb:江都秦氏
ed:用宋石刻本景印
vi:12
si:史-XV-4

nu:0000004
fi:史部
sf:金石類
tg:義例之屬
ti:金石三例,
au:清,廬見曾,輯
au2:清,王芑孫,評
yr:光緒四年
pb:南海馮氏讀有用書齋
ed:刊本
ko:0000005
ko2:0000006
ko3:0000007
vi:4
si:史-XV-5

nu:0000005
fi:史部
sf:金石類
tg:義例之屬
ti:金石例,十卷
au:元,潘昂霄,撰
oy:0000004
co:金石三例

nu:0000006
fi:史部
sf:金石類
tg:義例之屬
ti:墓銘舉例,四卷
au:明,王行,撰
oy:0000004
co:金石三例

nu:0000007
fi:史部
sf:金石類
tg:義例之屬
ti:金石要例,一卷
au:清,黃宗羲,撰
oy:0000004
co:金石三例

1. レコード番号に関するタグ
 - nu:レコード番号
 - ko:子目へのリンク
 - oy:子目からの逆リンク
2. 四部分類に関するタグ
 - fi:部
 - sf:類
 - tg:属
 - ki:目
3. 書名に関するタグ
 - ti:書名
 - TI:書名の拼音表記
 - st:附録(坵)書名
 - ST:附録(坵)書名の拼音表記
 - pt:別名(即)
 - PT:別名(即)の拼音表記
4. 著者名に関するタグ
 - au:著者名
 - AU:著者名の拼音表記
5. 出版事項に関するタグ
 - tp:書名(複数の出版事項がある時に用いる)
 - yr:刊年
 - pb:出版者
 - ed:出版方法
 - sd:蔵板
6. 所蔵に関するタグ
 - or:所蔵機関
 - se:文庫名
 - si:請求番号
 - rn:登録番号
7. その他のタグ
 - co:所収書名
 - vi:冊数
 - no:注記

全国漢籍データベースの設計とWWWでの運用

安岡孝一*

1 はじめに

全国漢籍データベースは、国公立の研究機関や図書館等が所蔵する漢籍に対し、その所在情報の網羅を最終目標としたデータベースである。漢籍が他の図書にはない特殊性を有することから、NACSIS-CAT等とは別個のデータベースとして構築する、ということが、2001年3月9日の全国漢籍データベース協議会において決定されたものである。では、漢籍は他の図書に較べ、一体どのような特殊性を有し、それはデータベースの設計や運用にどのように影響してくるのか。本稿では、全国漢籍データベースの構築という側面から、これらについて述べる。

2 全国漢籍データベースの設計

全国漢籍データベースを設計するにあたっては、従来、漢籍がどのように分類・整理されてきたかを考慮する必要がある。漢籍も図書の一種であるから、図書館等においては、図書カードという形で整理されている。ただし、漢籍においては、図書カードの書き方が特殊であり、この特殊性がデータベース化を阻んできた一因でもあった。この章では、まず、漢籍における図書カードについて概要を述べ、次に、図書カードをどのようにレコード化したかについて述べる。

2.1 漢籍における図書カード

例として、呉文炳と呉鸞が共同執筆した「泉幣圖説」という漢籍の図書カードを見てみよう。漢籍は伝統的に四部分類という分類法にしたがうため、この図書カードの左上にも「史-XV-6」すなわち、史部の金石類の鏡鑑泉幣雜器之屬に分類

史
XV
6

泉幣圖説六卷

清呉文炳呉鸞同撰
嘉慶五年呉氏香雪山莊刊本 2冊

京大人文研 東方

されることが示されている。最初の行が、書名と巻数である。その次の行がいわゆる著者にあたるが、漢籍では著者名だけでなく、どの時代の人かが示されている。呉文炳と呉鸞は清の時代の人なので「清」の一字が付いており、末尾には共同執筆を示す「同撰」の文字がある。その次の行が出版事項で、嘉慶五年(1800年)に呉

*京都大学人文科学研究所附属漢字情報研究センター

氏が香雪山莊から出版した木版本2冊、であることが示されている。すなわち、この図書カードにおいては、分類だけがやや特殊で、残りの部分は一般の図書カードと大差ない内容が書かれている。漢籍の図書カードがこのようなものばかりであれば、データベース化には特に問題がない。しかし、このような図書カードは、

漢籍ではむしろ稀である。

例えば、清の盧見曾が編集した「金石三例」は、元の潘昂霄が書いた「金石例」と、明の王行が書いた「墓銘舉例」と、清の黄宗羲が書いた「金石要例」の三つを、一つにまとめて出版しなおしたものである。したがって、「金石三例」の図書カードは4枚にわたっている。一枚目には「金石三例」全体に関する事項が、二枚目には「金石例」に関する事項が、三枚目には「墓銘舉例」に関する事項が、四枚目には「金石要例」に関する事項が、それぞれ記されており、二～四枚目は、他の本に含まれている（これを漢籍では「子目」という）ことを示すために、「金石三例所收」と記すと同時に、分類が欄外に示されている。このような、本の中に本を含むような出版形態は、漢籍では珍しいことではない。例えば、潘昂霄の「金石例」は、清の章壽康が編集した「式訓堂叢書」にも収録されている。あるいは「金石三例」全体が、民国56年(1966年)に王雲五が編集した「國學基本叢書四百種」で復刊されていたりするのである。なお、「金石三例」に収録

1

史
XV
5

金石三例

清盧見曾輯 清王芑孫評
光緒四年南海馮氏讀有用書齋刊本 4冊

京大人文研 東方

2

金石例十卷

元潘昂霄撰
金石三例所收

史-XV-5

3

墓銘舉例四卷

明王行撰
金石三例所收

史-XV-5

4 止

金石要例一卷

清黄宗羲撰
金石三例所收

史-XV-5

されている三つは、いずれもたまたま「史-XV-5」すなわち史部の金石類の義例之屬に分類されているが、分類の異なる本が合本されて出版されることも、珍しくはない。

ただし、複数の本が合本されているからと言って、図書カードが必ずしも複数になるとは限らない。例えば、清の唐與崑が書いた「制錢通考」は、本人がその2年前に刊行した「泉幣彙考」と合本されており、しかも「泉幣彙考」の続編といえるものである。このような場合には、図書カードは複数に分けず、一枚の図書カードに複数の書名を並記することになる。このような並記は、それぞれの本の著者が異なる場合でも、起こりうる。例えば、宋の劉球が書いた「隸韻」は、色々な

史
XV
6

**泉幣彙考十六卷
制錢通考四卷**

清唐與崑撰
彙考咸豐元年通考三年
山陰唐氏紅葯山房刊本 8冊

京大人文研 東方

石碑に記された漢字を集めてきたものであり、「碑目」という石碑の一覧表が付いている。それを嘉慶十四年(1809年)に秦恩復が復刻した際、「碑目」に関する考証を附録として付けたのである。このような場合、「碑目攷證」はあくまで「碑目」に

史
XV
4

**隸韻十卷碑目一卷
附碑目攷證一卷**

宋劉球撰 碑目攷證清秦恩復撰
嘉慶十四年江都秦氏用宋石刻本景印
12冊

京大人文研 東方

対する附録なので、これら三つの本を、一枚の図書カードにまとめて記すことになる。すなわち、複数の本が合本されている場合、それらが著作上たがいに関係がある場合には一枚の図書カードに記し、関係がない場合には複数の図書カードに記す、というのが、大雑把なルールである。

2.2 漢籍レコードの設計

漢籍の図書カードに書かれた情報を、データベース上に載せることを考えてみよう。前提として、図書カードを実際に採録している現場のことを考慮するならば、図書カードの形式や枚数は変えるべきではない。すなわち、前節で述べた図書カードの形式はできるだけ変えずに、それをコンピュータで扱えるフォーマットに直す必要がある。しかも、漢籍といえど図書データベースなのだから、書名、著者名、刊年、出版者名くらいの検索キーは必要だろう。

データベースへの適用性を考慮した場合、このような要求に応えうるのは、やはりSGMLである。しかし、テキストエディタでの入力や編集を考えた場合、オープンタグとクローズタグの両方が必要で、しかもタグが入れ子になりうるSGMLは、あまりに繁雑と言わざるをえない。そこで、図書カードを入力・編集するための漢籍レコードには、タグ付きフィールドを用いることにした。タグ付きフィールドであれば、各行が意味を持つ図書カードの構造に近いと考えられるし、データベースで用いるSGMLへの変換も容易だからである。

では、前節で示した図書カードに対する漢籍レコードを、それぞれ見ていくことにしよう。まずは「泉幣圖説」である。

nu:00000001
fi:史部
sf:金石類
tg:鏡鑑泉幣雜器之屬
ti:泉幣圖説, 六卷
au:清, 吳文炳, 同撰
au:清, 吳鸞, 同撰
yr:嘉慶五年
pb:吳氏香雪山莊
ed:刊本
vi:2
or:京大人文研 東方
si:史-XV-6

nu は、この漢籍レコードに対する通し番号である。この例では 8 桁の数字となっているが、全レコードで一意でさえあれば、桁数は自由である。fi、sf、tg は、「泉幣圖説」に対する四部分類を示したものである。ti は書名であり、巻数との間をコンマで区切っている。これは書名の部分だけを検索できるようにするための仕掛けである。同様の仕掛けが著者名 au にもあり、著者名の前後をコンマで区切っている。yr、pb、ed、vi はそれぞれ刊年、出版者名、出版方法、冊数である。or と si はそれぞれ所蔵機関と請求番号である。漢籍レコードは、図書カードのテキストをほぼそのまま並べ、必要な情報を補完したものになっているのがわかる。

子目に対しては、各レコード間のリンク情報を何らかの形で挿入しなければならない。これは図書カードには明示されていない情報だが、データベース上は必要となる。「金石三例」の漢籍レコードを見てみよう。

nu:00000002
fi:史部
sf:金石類
tg:義例之屬
ti:金石三例
au:清, 廬見曾, 輯
au:清, 王芑孫, 評
yr:光緒四年
pb:南海馮氏讀有用書齋
ed:刊本
vi:4
or:京大人文研 東方
si:史-XV-5
ko:00000003
ko:00000004
ko:00000005

nu:00000003
fi:史部
sf:金石類
tg:義例之屬
ti:金石例, 十卷
au:元, 潘昂霄, 撰
co:金石三例
oy:00000002
nu:00000004
fi:史部
sf:金石類
tg:義例之屬
ti:墓銘舉例, 四卷
au:明, 王行, 撰
co:金石三例
oy:00000002

nu:00000005
fi:史部
sf:金石類
tg:義例之屬
ti:金石要例, 一卷
au:清, 黃宗羲, 撰
co:金石三例
oy:00000002

子目のレコードへのリンクには ko を用い、その逆リンクには oy を用いて、双方向リンクとしている。リンクの内容として書かれているのは、それぞれのレコードの nu である。これに加え子目のレコードには、所収書名を co に入れているが、これは図書カードの形を踏襲するためである。もし、本 A の中に本 B があって、本 B の中にさらに本 C があるような親子孫の構造を持つ漢籍であれば、リンクは本 A と本 B の間、および本 B と本 C の間にそれぞれ張ることとし、本 A と本 C の間は直接のリンクは張らない。

では、複数の題目を持つ図書カードに対する漢籍レコードはようになるだろう。「泉幣彙考」と「隸韻」の例を見てみよう。

nu:00000006	nu:00000007
fi:史部	fi:史部
sf:金石類	sf:金石類
tg:鏡鑑泉幣雜器之屬	tg:彙考文字之屬
ti:泉幣彙考,十六卷	ti:隸韻,十卷
ti:制錢通考,四卷	ti:碑目,一卷
au:清,唐與崑,撰	st:碑目攷證,一卷
tp:彙考	au:宋,劉球,撰
yr:咸豐元年	au:清,秦恩復,撰,碑目攷證
tp:通考	yr:嘉慶十四年
yr:咸豐三年	pb:江都秦氏
pb:山陰唐氏紅葯山房	ed:用宋石刻本景印
ed:刊本	vi:12
vi:8	or:京大人文研 東方
or:京大人文研 東方	si:史-XV-4
si:史-XV-6	

「泉幣彙考」では二つの ti が存在する。図書カードでは、出版事項が「彙考咸豐元年通考三年山陰唐氏紅葯山房刊本」となっていたところを、検索のために「三年」の前に「咸豐」を補い、それぞれの出版事項がどの ti に関するものなのかを表す部分を、tp によって示している。「隸韻」では、附録を st で示していると同時に、秦恩復が附録の著者であることを示すために、au に四番目の項目を追加している。なお、tp も、au の四番目の項目も、いずれも表示のためのものであるから、ti と完全に一致する必要はない。

2.3 漢籍レコード中のタグ一覧

前節のようなやり方で作成した漢籍レコードから、必要となるタグを帰納的に導き出した。以下に分類して示す。

1. レコード番号に関するタグ

- nu:レコード番号
- ko:子目へのリンク

- oy: 子目からの逆リンク

2. 四部分類に関するタグ

- fi: 部
- sf: 類
- tg: 属
- ki: 目

3. 書名に関するタグ

- ti: 書名
- TI: 書名の拼音表記
- st: 附録 (坿) 書名
- ST: 附録 (坿) 書名の拼音表記
- pt: 別名 (即)
- PT: 別名 (即) の拼音表記

4. 著者名に関するタグ

- au: 著者名
- AU: 著者名の拼音表記

5. 出版事項に関するタグ

- tp: 書名 (複数の出版事項がある時に用いる)
- yr: 刊年
- pb: 出版者
- ed: 出版方法
- sd: 蔵板

6. 所蔵に関するタグ

- or: 所蔵機関
- se: 文庫名
- si: 請求番号
- rn: 登録番号

7. その他のタグ

- co: 所収書名
- vi: 冊数
- no: 注記

3 全国漢籍データベースの運用

前章で述べた漢籍レコードの集合すなわち全国漢籍データベースを、一般に公開運用したい。インターネットを視野に入れるならば、WWWから全国漢籍デー

データベースを検索できることが、運用の最低条件であろう。となると、運用側としては、この「WWWから」と「検索できる」を、どのように実現するかが問題となる。本章では、全国漢籍データベースの運用を、「検索できる」「WWWから」の順に述べる。

3.1 検索エンジンの運用

全国漢籍データベースの検索エンジンには、OpenText を用いることにした。文字列の検索という点では、圧倒的な性能を誇るからである。ただし OpenText は、タグ付きフィールドよりは SGML の検索向きにチューニングされていることから、漢籍レコードを SGML ファイルに変換して、検索に用いることにした。例えば

nu:00000001	<nu>00000001</nu>
fi:史部	<fi>史部</fi>
sf:金石類	<sf>金石類</sf>
tg:鏡鑑泉幣雜器之屬	<tg>鏡鑑泉幣雜器之屬</tg>
ti:泉幣圖說, 六卷	<ti><key>泉幣圖說</key>六卷</ti>
TI:quan2 bi4 tu2 shuol	<pinyin><ti><key>quan2 bi4 tu2 shuol</key></ti></pinyin>
au:清, 吳文炳, 同撰	<au>清<key>吳文炳</key>同撰</au>
AU:wu2 wen2 bing3	<pinyin><au><key>wu2 wen2 bing3</key></au></pinyin>
au:清, 吳鸞, 同撰	<au>清<key>吳鸞</key>同撰</au>
AU:wu2 luan2	<pinyin><au><key>wu2 luan2</key></au></pinyin>
yr:嘉慶五年	<yr>嘉慶五年</yr>
pb:吳氏香雪山莊	<pb>吳氏香雪山莊</pb>
ed:刊本	<ed>刊本</ed>
vi:2	<vi>2</vi>
or:京大人文研 東方	<or>京大人文研 東方</or>
si:史-XV-6	<si>史-XV-6</si>

上図左の漢籍レコードであれば、上図右の形に変換し、00000001.dat という SGML ファイルに格納する。SGML ファイルの文字コードには、UTF-8 を用いることにした。UTF-8 を採用した理由は、康熙字典を全て含む 72911 字の漢字を現時点で収録している点、ASCII 互換であり OpenText でのタグ処理に適している点、現在の WWW ブラウザの多くで言語環境を問わず直接表示可能な点、の 3 点である。さらに、key、ti、st、pt、au、yr、pb、oy、or の各オープンタグとクローズタグに挟まれる領域に限定した検索をおこなえるよう、OpenText を設定した。次節で述べる簡易検索および詳細検索を、それぞれ可能とするためである。

検索エンジンの運用において、最も問題となるのは、異体字処理である。例えば、「泉幣図説」を含む<ti>~</ti>を OpenText で検索する場合は

```
>> "泉幣図説"  
>> region ti including %
```

を実行すればよい[†]。ところがこの検索では、上記の 00000001.dat はマッチングしない。00000001.dat の<ti>~</ti>に含まれているのは「泉幣圖説」であって

[†]1 行目は「泉幣圖説」という文字列の全文検索を、2 行目は直前の結果を含む<ti>~</ti>の検索を、それぞれ意味する。

「泉幣図説」ではないからである。「泉幣図説」に対して「泉幣圖説」をマッチングさせるためには、「図」で「圖」を、「説」で「說」を検索できるような仕組みを、検索エンジン中に持つ必要があるのである。

この問題に対し筆者は、「漢字袋」[1]で制作した異体字シソーラスを、OpenTextのフロントエンドとして組み込んだ。このフロントエンドは、文字列分解と異体字置換の2つの部分から成る。文字列分解の部分では、文字列検索を、各文字の検索のシーケンスに分解する。例えば、前ページの検索例に対しては

```
>> "説"  
>> "図" fby %  
>> "幣" fby %  
>> "泉" fby %  
>> region ti including %
```

という風に、「泉幣図説」という文字列検索を、「説」 その直前の「図」 その直前の「幣」 その直前の「泉」を順次検索する形に分解する。シーケンスが逆順になっているのは、最終検索点が文字列の先頭に来るようエミュレートするためである。異体字置換の部分では、各文字の検索を、その文字の異体字検索に置き換える。上の例であれば

```
>> ( "説" + "說" + "说" + "詔" )  
>> ( "図" + "图" + "圖" + "圖" + "囧" + "囧" + "圖" + "囧" + "盟" ) fby %  
>> ( "幣" + "币" + "幣" + "幣" ) fby %  
>> ( "泉" + "淥" + "嶮" + "泉" + "泉" ) fby %  
>> region ti including %
```

という形に置き換える。この結果、「泉幣図説」に対して $5 \times 4 \times 9 \times 4 = 720$ 種類の異体字列が同時に検索されることになり、SGMLファイルの中身が「泉幣図説」であろうと「泉幣圖説」であろうと「泉幣圖説」であろうと、マッチングが起こることになる。

以上述べてきたとおり、全国漢籍データベースでは、UTF-8という国際的的文字コードを使用すると同時に、異体字シソーラスを検索エンジンのフロントエンドとして組み込むことにより、利用者が異体字を気にする必要のない環境を実現した。日本の常用漢字だろうと中国の簡化字だろうと、「漢字」の使える環境でありさえすれば、世界中のどこからでも検索可能なデータベースとするためである。

3.2 WWW インターフェースの実現

全国漢籍データベースの検索用画面として、apache上のCGIで実現したインターフェースは、簡易検索画面、詳細検索画面、レコード表示画面、の三つである。いずれの画面も、文字コードはUTF-8であり、検索エンジンで用いた文字コードと一致させている。以下、それぞれの画面について、詳細を述べる。

3.2.1 簡易検索画面

全國漢籍データベース

日本所藏中文古籍數據庫

[詳細検索へ](#)
[全國漢籍データベース協議會へ](#)

kanseki@kanji.zinbun.kyoto-u.ac.jp

簡易検索画面の CGI は、入力された検索文字列に対して、SGML ファイル中の <key> ~ </key> の部分、すなわち書名および著者名を検索し、該当するファイルの書名、著者名、出版事項、所収書名、所蔵機関を表示する。「金石例」を検索した結果を、例として以下に示す。

全國漢籍データベース

日本所藏中文古籍數據庫

金石例

21 レコード見つかりました

1. [金石例十卷](#) 元 潘昂霄 撰 金石三例 京大人文研 東方
2. [金石例十卷](#) 元 潘昂霄 撰 式訓堂叢書三集 京大人文研 東方
3. [金石例十卷](#) 元 潘昂霄 撰 國學基本叢書四百種金石三例 京大人文研 東方
4. [金石例十卷](#) 元 潘昂霄 撰 今改用金石三例本 百部叢書集成第七十三式訓堂叢書 京大人文研 東方
5. [金石例十卷](#) 元 潘昂霄 撰 金石全例金石三例 東大東文研
6. [金石例補二卷](#) 清 郭麐 撰 光緒三年刊 孫谿朱氏金石叢書 京大人文研 東方
7. [金石例補二卷](#) 清 郭麐 撰 光緒四年刊 式訓堂叢書 京大人文研 東方

⋮

アンダーラインが引かれた書名部分は、レコード表示画面への HTML リンクとなっており、例えば、「1. 金石例十卷」をクリックすると、次々ページ下の画面へジャンプする。なお、複数の検索文字列を与えた場合には、基本的には AND 検索

となるが、文字列の先頭に「+」を付加した場合はOR検索を、「-」を付加した場合はNOT検索をおこなうよう、CGIを実装している。

3.2.2 詳細検索画面

詳細検索画面のCGIは、各フィールドに入力された検索文字列に対して、SGML中の該当するタグの部分を検索し、SGMLファイルを簡易検索画面と同じフォーマットで表示する。すなわち、書名フィールドであれば<ti>・<st>・<pt>中の<key>を、著者名フィールドであれば<au>中の<key>を、刊年フィールドであれば<yr>を、出版者フィールドであれば<pb>を、keywordフィールドであれば全文を、所蔵機関フィールドであれば<or>を、それぞれ検索する。子目フィールドに入力された検索文字列に対しては、まず<ti>・<st>・<pt>中の<key>部分を検索し、該当するファイルの<oy>を順に遡っていき、<oy>のないSGMLファイル、すなわち最も先祖にあたるレコードを出力する。「金石例」と「墓銘挙例」の両方を子目として検索した例を、以下に示す。

全國漢籍データベース

日本所藏中文古籍數據庫

書名	<input type="text"/>
著者名	<input type="text"/>
刊年	<input type="text"/>
出版者	<input type="text"/>
子目	<input type="text" value="金石例 墓銘挙例"/>
keyword	<input type="text"/>
所蔵機関	<input type="text"/>

5レコード見つかりました

1. [金石三例](#) 清 盧見曾 輯 清 王世孫 評 光緒四年 南海馮氏讀有用書齋 刊本 京大人文研 東方
2. [金石全例](#) 清 朱記榮 輯 光緒十八年 潛縣朱氏槐廬 校刊本 東大東文研
3. [式訓堂叢書](#) 清 章壽康 輯 光緒中 會稽章氏 刊本 京大人文研 東方
4. [國學基本叢書四百種](#) 王雲五 輯 民國五十六年 臺北臺灣商務印書館 景排印本 京大人文研 東方
5. [百部叢書集成](#) 藝文印書館 輯 民國五十三年至五十九年 臺北藝文印書館 景印本 京大人文研 東方

アンダーラインが引かれた書名部分は、レコード表示画面へのHTMLリンクとなっており、例えば、「1. 金石三例」をクリックすると、次ページ上の画面へジャンプする。なお、刊年フィールドに関しては、将来的には西暦での入力、それも「1796年から1820年まで」のような検索をおこないたいと考えているが、現時点では文字列のマッチングのみとなっている。

3.2.3 レコード表示画面

レコード表示画面は、SGMLファイルの情報を表示する画面である。ただし、子目に関するリンクは、それぞれを辿った先の情報まで表示するようにしている。例えば、「金石三例」のレコード表示画面は

史部 金石類 義例之屬

金石三例

清 盧見曾 輯 清 王芑孫 評 光緒四年 南海馮氏讀有用書齋 刊本 4冊

京大人文研 東方 史-XV-5

金石例十卷 元 潘昂霄 撰

墓銘學例四卷 明 王行 撰

金石要例一卷 清 黃宗羲 撰

[検索画面に戻る](#)

となる。アンダーラインが引かれた書名部分は、子目のレコード表示画面へのHTMLリンクとなっており、例えば、「金石例十卷」をクリックすると

金石三例 清 盧見曾 輯 清 王芑孫 評 光緒四年 南海馮氏讀有用書齋 刊本

史部 金石類 義例之屬

金石例十卷

元 潘昂霄 撰

[検索画面に戻る](#)

となる。この画面においても、アンダーラインが引かれた書名部分は、レコード

表示画面への HTML リンクとなっており、子目の<oy>を遡れる構造になっている。ちなみに「國學基本叢書四百種」に含まれている「金石三例」のレコードは

國學基本叢書四百種 王雲五 輯 民國五十六年 臺北臺灣商務印書館 景排印本

史部 金石類 義例之屬

金石三例

清 盧見曾 輯

金石例十卷 元 潘昂霄 撰

墓銘舉例四卷 明 王行 撰

金石要例一卷 清 黃宗羲 撰

[檢索畫面に戻る](#)

という形で、「國學基本叢書四百種」中の「金石三例」に含まれている「金石例」のレコードは

國學基本叢書四百種 王雲五 輯 民國五十六年 臺北臺灣商務印書館 景排印本

金石三例 清 盧見曾 輯

史部 金石類 義例之屬

金石例十卷

元 潘昂霄 撰

[檢索畫面に戻る](#)

という形で表示する。「國學基本叢書四百種」が「金石三例」を含んでいて、その「金石三例」が「金石例」を含んでいるという親子孫の構造が、一目でわかるようにするためである。ただし、「國學基本叢書四百種」のレコードは

叢書部 雜叢類 民國之屬

國學基本叢書四百種

王雲五 輯 民國五十六年 臺北臺灣商務印書館 景排印本 2380 冊

京大人文研 東方 叢-I-6

崇文總目五卷 補遺一卷 附録一卷 宋 王堯臣 等敕編次 清 錢東垣 輯釋 清
錢繹 輯釋 清 錢侗 輯釋 清 金錫鬯 輯釋 清 秦鑑 輯釋 清 錢侗 輯補遺附録

⋮

語石十卷 清 葉昌熾 撰

金石三例 清 盧見曾 輯

樂律全書 明 朱載堉 撰 景印本

⋮

という形で表示し、孫 (あるいはその子孫) は表示しない。全ての先祖と直下の子目だけを表示する形式にしたのは、もし子目をツリー状に全て展開すると、表示に時間がかかり過ぎる場合がある、というのが最大の理由である。

4 おわりに

漢籍の図書カードをレコード化し、それを WWW から検索できるようにする、という視点から、全国漢籍データベースの設計とその運用について述べた。全国漢籍データベースは、現在 <http://kanji.zinbun.kyoto-u.ac.jp/kanseki> で運用中である。なお、現時点のデータは、まだまだ国内の網羅にはほど遠い。今後、各所蔵機関とのコラボレーションにより、データベースを充実させていきたい。

参考文献

- [1] 安岡孝一、安岡素子: WWW 異体字典「漢字袋」の制作, 京都大学大型計算機センター第 57 回研究セミナー報告 (1997 年 3 月), pp.27-44.