

報告2：漢籍データベース、日本／北京／台北

大木康（東大東文研・教授）

ただいまご紹介いただきました大木です。どうぞよろしく願いいたします。

今日のシンポジウムはデジタル化に関する問題がテーマですが、実は私自身、コンピューターについては特に詳しいわけではなく、ごく平凡なユーザーに過ぎません。大変お恥ずかしい限りですが、私自身がよく使っているデータベースのいくつかをご紹介させていただき、そしておそらく「おまえ、こんなもの知らないのか」というのがたくさんあるかと思しますので、そのあたりをむしろお教えいただくという甘い考えで、今日はここに立たせていただいております。

中国関係、特に私は中国の古典が専門ですが、今やこの領域でも、かなりのデータベースが構築されており、文献や語句の検索について、これらのデータベースを使えるかどうかによって、研究成果の上でもかなり決定的な違いが出てくる時代になっていることは確かだろうと思います。

今回非常に簡単なハンドアウトを作りました。発表の準備をしながら一つ思いましたことは、例えば私は中国古典文学という領域で仕事をしています。中国古典文学研究をするために必要なデータベースを探していくことになりま。そうすると、こういう便利なデータベースがあるということは見つかるわけですが、結果的にはいわゆるタコつぼなんです。つまりコンピューターをうまく使ったから、例えばフランス文学とか、あるいは国文学などのような他の近接する領域との交流が生まれるかというと、なかなかそうはいかないのが現状だろ

うと思います。これは仕方がないことなのかもしれませんが、今回の作業をやって、まず一つはそんなことを感じた次第です。

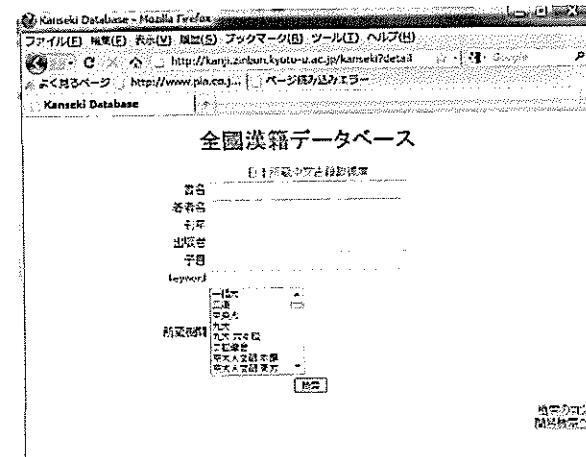
今日は、(1) 漢籍データベース全般に関する情報、(2) 書物の所在を検索するときの目録、(3) 漢籍の画像データベース、(4) 漢籍のテキストデータベースというふうに分けて、私がかたまたま知っているごくわずかな情報をお伝えしたいと思います。

(1) はじめに漢籍データベース全般についての情報ですが、これについては、京大人文研の麥谷邦夫先生のサイトである「道氣社」(道氣社については、グーグルなどで「道氣社」を引くとすぐに出てきます)に、「中国思想研究者のためのインターネット資源簡介」というコーナーがあって、ここに現在アクセスできるいろいろなデジタル情報のリストがあります。麥谷先生は道教がご専門ですが、コンピューターのことにもとてもお詳しい。

ただ、今回この発表をするにあたって、麥谷先生に直接問い合わせてみたところ、あれは古くから紹介すると言われました(2004年ぐらいが最後の改訂版)。麥谷先生に教えていただいた「やたらすナビ」というのがありますので、ごらんいただきたいと思います。こうした総合的な目録についても、充実したリストがあるぞということがあれば、ぜひお教えいただければと思う次第です。

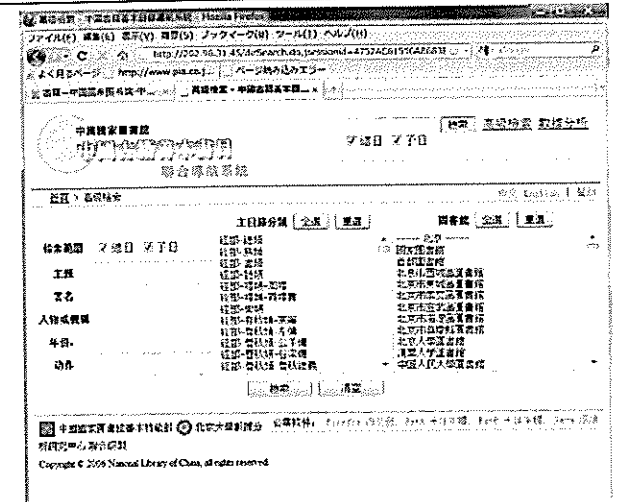
(2) 次に漢籍目録に関してですが、日本国内に所蔵されているいわゆる漢籍に関しては、京都大学の人文科学研究所が中心になって作っております「全国漢籍データベース」がありま

す。まず京大の人文研の所蔵漢籍から始まって、そのあと東大の東文研が入ってというふうにふえてきたわけですが、最近では東洋文庫とか内閣文庫とか、参加する図書館がどんどんふえてまいりました。この成長過程を見ながら、ああ、データベースというのはこういうものなんだ、ということがよくわかったわけです。一つ二つの図書館の蔵書のデータベースでは、冊子体の目録を引くのとは違って変わらなかったんですが、参加する図書館がふえることによって、どんどんパワーアップしていき、どんどん便利になりました。

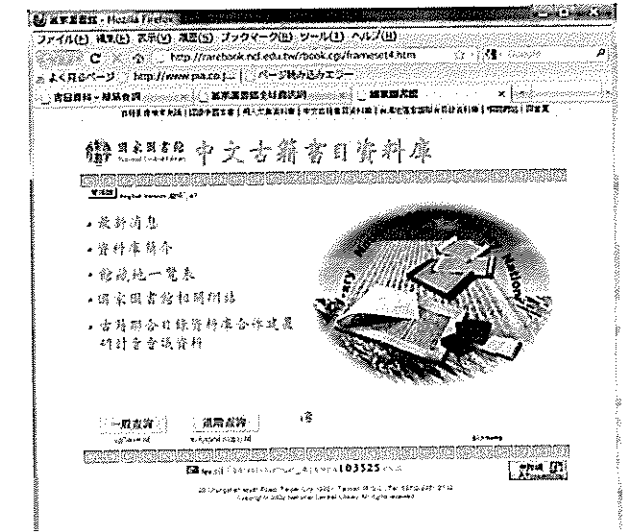


http://kanji.zinbun.kyoto-u.ac.jp/kanseki?detail

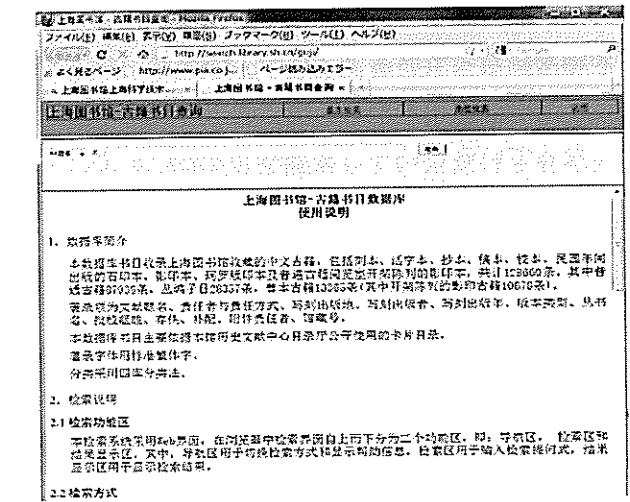
それと同じように、中国では、国家図書館(北京図書館)が中心になって、「中国古籍善本書目」、台湾の台北にある国家図書館のほうでは「中文古籍書目資料庫」というものをつくっております。台湾の国家図書館のトップページに古籍文献という項目があって、そこに「中文古籍書目資料庫」があります。これは台湾の国家図書館だけではなく、中国の国家図書館、アメリカのプリンストン大学などなど、ほかの図書館の漢籍も一緒に引けるようになっています。東文研の漢籍目録も、これに加入しております。



http://202.96.31.45/dirSearch.do?sessionId=4752AC61596A8683E0EB47DA2B240351?method=goToGaoJiQuery



http://rarebook.ncl.edu.tw/rbook.cgi/frameset4.htm



http://search.library.sh.cn/guji/

ほかにも個別には、例えば上海図書館の本も、上海図書館のサイトから入っていくと、今蔵書目録が全部データベースになって検索できるようになっており、非常に便利に使えるようになっています。

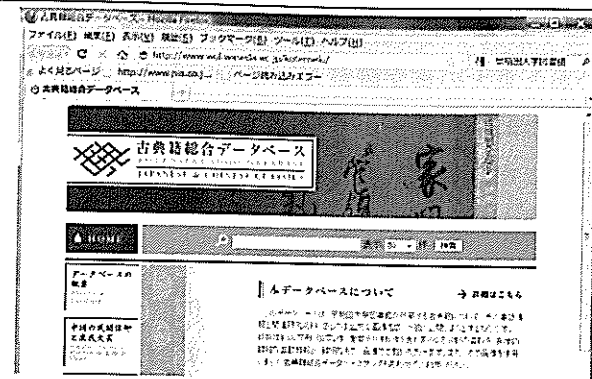
(3) 次に漢籍の画像に関するデータベースですが、これについては私も東洋文化研究所で「漢籍善本全文画像資料庫」というのをやっています。これは、東洋文化研究所蔵の善本の全体が、画像で読めるように作ってあります。



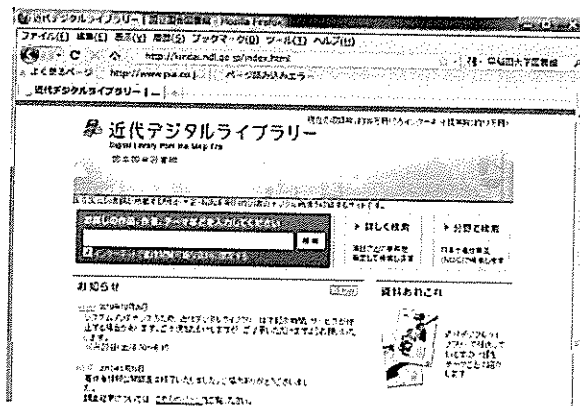
<http://shanben.ioc.u-tokyo.ac.jp/>

東文研の全文画像データベースは、基本的に全世界で無料で使えるようになっております。これは日本の科研費などを使って制作したわけですが、データベース科研は、公開が原則になっているわけです。データベースについて、海外では、お金を払わなければ見られないというのが趨勢になっているようです。ですが、日本では公開が原則ですから、逆に制限をかけられなくなっているということだろうと思います。

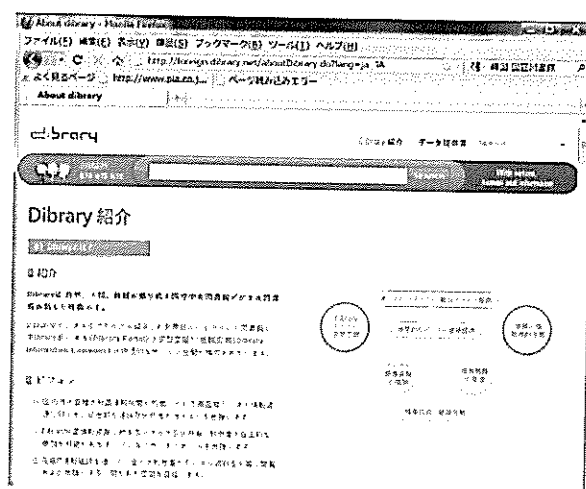
早稲田大学の図書館でも「古典籍総合データベース」を作っており、私も時々使いますが、これはただで見ることができます。国会図書館でも日本の明治以後の書物をデジタル化しております（「近代デジタルライブラリー」）。純粋な漢籍ではありませんが、明治以後の漢学者が書いた漢文の文献は、ここでよく引っかけ、見ることができます。



<http://www.wul.waseda.ac.jp/kotenseki/>



<http://kindai.ndl.go.jp/index.html>

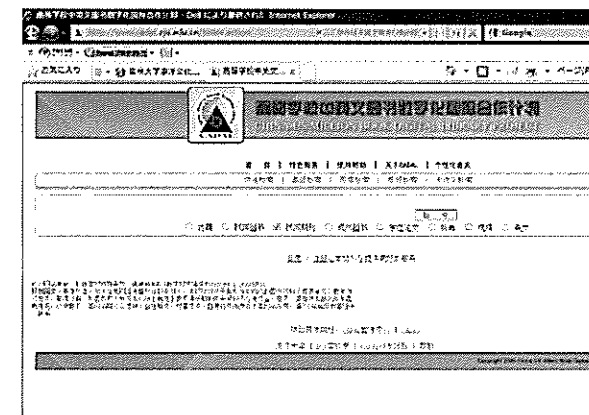


[http://foreign.dibrary.net/aboutDibrary.do?lang=ja\\_JA](http://foreign.dibrary.net/aboutDibrary.do?lang=ja_JA)

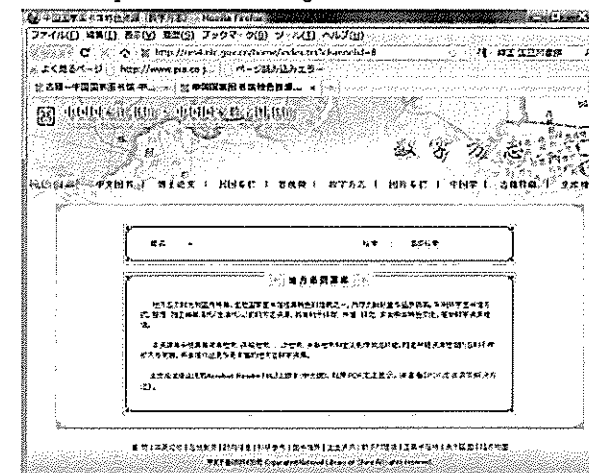
資料には書きませんが、韓国の国立図書館でも同じようなプロジェクトを進めていて、韓国の図書館にある日本の明治時代の本が、実は完全に画像データで公開されています。『吉原竹枝』という、明治時代の文人が吉原の様子を詠み込んだ漢詩集があるんですが、それをいろいろと探しておりましたら韓国の国

立図書館にあって、しかもその画像が公開されているのにぶち当たって、大変喜んでダウンロードし、全部印刷いたしました。

それから、中国の浙江大学が中心になり、アメリカの大学などとも共同して行っている「高等学校中英文図書数字化（国際）合作計画」（CADAL（カダル）：China-America Digital Academic Library）というのがあります。このデータベースもかなり威力を発揮していると思います。ただ実は、私が最初教えてもらって使い始めたときにはかなりの部分を見ることができたんですが、最近これで本を見ようとすると、極めて不安定な状態になっている気がいたします。漏れ聞くとところによると、要するに課金システムを考えているようで、結果的にただ見られる部分が小さくなっているのではないかという気がいたします。



<http://www.cadal.zju.edu.cn/Index.action>



<http://res4.nlc.gov.cn/home/index.trsf?channelid=8>

国家図書館（北京図書館）には中国地方志のデータベースがあって、画像でみんな見られます。これは最近かなり重宝しております。

それから、インターネットではありませんが、北京の愛如生（アイジョセイ）というデジタル関係の会社から『全四庫』というのが出ています。『四庫全書』は大分前にデジタル化されましたけれども、『四庫全書』に加えて『四庫存目書』とか、あるいは『四庫未収書』なども含まれた、膨大なデータベースです。ここに収められた本が、全部画像のCD-ROMになって、見ることができるんです。ただしこれはいわゆるデジタルテキストではないので、検索はできない。しかし、本を読む上では、とにかくこの『全四庫』があれば、相当強力なものになると思います。ただ、ちょっと使ったことはありますが、やっぱり膨大な書物を画面上だけで読んでいくのは結構大変でした。本であれば何冊か借りてきて開いておくとか、あるいは指を挟んでおくとか、紙を挟んでおくとかいろいろできるわけですが、やはりそのあたりが苦勞のしどころだろうと思います。

(4) 最後にデジタルテキストです。語句を検索する上で一番よく使っているのが、『四部叢刊』のデジタルテキストです。『四部叢刊』は、経・史・子・集の四部にわたる中国古典の基本的な書物を収めた叢書ですが、これが全部デジタルテキストになっており、検索することができます。

それからこれは、東大に来られる中国の先生方がよく使っておられるので、使わせていただいたり、この言葉を検索していただけないかとお願ひしたりしているものに、『国学宝典』というテキストデータベースがあります。これは中国語簡体字のデータベースですが、正統的な書物を収めた『四部叢刊』などには収められない戯曲とか小説のたぐいも含まれているので、重宝します。ただ、これもCD-ROM版、あるいはアクセス権を買う形などいろいろあるわけですが、結構高い。

台湾の中央研究院で「瀚典」というテキストデータベースを前から作っております。この「瀚典」ですが、目録を見ますと、いくつかの部分に分かれておりまして、「二十五史」とか「十三經」とか、幾つかの部分は日本でも自由に使うことができます。もちろん「二十五史」とか「十三經」とか基本的な文献ですから、検索するのにとても使い勝手がいいのですが、実はここに「漢籍全文資料庫」という膨大な文字数を持つ部分があります。これは「申請使用」と書いてあるんですけども、昔は台湾の国内の方が申請できたわけです。申請するとパスワードがもらえて、これは大きな声ではいえませんが、海外でも使えたわけです。ところが最近では、台湾の国内では自由にただで全部見られちゃうということになって、そのかわり、たしかこれはIPアドレスかなんかで制限がかかって、日本ではどうやっても全体が見られない。「漢籍全文資料庫」でも、「免費使用」という部分はあるんですけど、実はこれはほんの少しだけで、裏に隠れている部分がものすごくたくさんあるんです。でも、それは台湾に行かないと使えない。今はそういう状態です。これも恐らく将来的には、課金のシステムができて、お金の問題になってくるんじゃないかと思っています。



<http://hanji.sinica.edu.tw/>

実は私、明日から急に台湾、台北に行くことになりました。それにかからめて、「中国基本古籍

庫」についてもご紹介しておきたいと思います。この「中国基本古籍庫」は、中国古典に関する本当に膨大なデータベースで、特に明代、清代の文集が非常にたくさん入っておりますので、私たちには非常にありがたいものです。ですが、この「中国基本古籍庫」を日本で買おうとすると、ワンアクセスで1800万ぐらいという代物です。学術研究のためのものにはちがいないのですが、商業ベースで作られたという感じが強いんですね。日本でも幾つかのお金持ちの大学ではそれを買っていて、私、某大学の友達のところへ引かせてもらいに行ったのですが、そのこの大学でも、結局ワンアクセスですから、だれかが使っていると、もうだめなんです。朝早くとか真夜中じゃないとだめだということで、無駄足だったんですね。

「中国基本古籍庫」も、台湾の中央研究院あたりはかなりのアクセス数を持っていて、台湾に行ったときには、いつもこれを使わせてもらうことにしています。今回も、いくつか宿題を持って行って、調べてきます。これを日本で1800万円で買うことを考えますと、今台湾に行く飛行機代が、安ければ5万円ぐらいでありますから、仮に何泊かして一回の旅行で10万円かかったとしても、180回行けてしまう。どうやらこっちのほうがお得みたいですね。「瀚典」も全部使えますし。

以上は、私がたまたま知っていたものをご紹介したにすぎず、関連のデータベースのすべてを満遍なくカバーしているというお話では全くありません。むしろこういうのがあって便利だということを、ぜひお教えいただければと思う次第です。どうもありがとうございました。以上です。

#### 【質疑】

司会 私からちょっと補足がてら質問もあります。東文研で公開しております漢籍の全文データベースはおおよそ4000点です。基本的な貴重な本はほとんどすべて公開しております。

す。しかし台湾の場合は、さっきありましたように、世界じゅうに公開しているものはほんの一部でほとんどが公開してないわけですね。

大木 私が申し上げた「瀚典」は、中央研究院のデジタルデータのほうですけど、歴史語言研究所では画像のデータベースも作って持っているんです。これはもう、ここら辺から見られるのはほんのちょっとだと思います。

司会 あと中国でも、先ほど話にでたCADALというのがあります。これは古典籍十万点の全文データを今つくっています。これはやはり日本からは入れないんです。さらに、『中国古籍全文検索叢書』などをつくっております。

す凱希という会社は、中国の書籍およそ1000万点を画像化するという計画をやっております。どこまでできるのかはわかりませんが、しかしそれは基本的に商品ですから、高い値段で買わなきゃいけない。つまり古典籍の画像及びデータがそういう形でさまざまなアクセス制限がなされたり、つまり財産として扱われているわけです。学術的にオープンソースとして使うという状況にならずに、より囲い込みが進んでいる状況があります。それは学問にとってはあまり健康なことではないと思いますけども、現状はそういうことになっているということを補足しておきます。

ありがとうございました。