

CHINA3 for WWW

守岡 知彦*

2001 年 10 月 4 日

1 はじめに

CHINA3 for WWW (図 1) は「東洋学文献類目」を WWW 上で利用できるようにしたもので、2000 年 10 月に最初の版の公開を始めました。それ以前から大型計算機センターでの検索サービスが存在しましたが、大型計算機センターを利用していたために利用登録が必要であり、利用料金も必要でした。また、速度も遅く繁雑で使いにくいものでした。これに対し、‘CHINA3 for WWW’ は漢字情報研究センターの発足に伴って導入された UNIX ベースのサーバーを利用して、WWW で自由に利用できるようにしています。従来サービスに比べ高速であり、利用登録は不要であり、無料で自由に利用することができます。

2 使い方

2.1 URL

CHINA3 for WWW は現在

<http://www.kanji.zinbun.kyoto-u.ac.jp/db/CHINA3/>

で公開しています。IE を使っている場合、「アドレス」にこの URL を打ち込んでください。そうすると、図 1 のような画面が出て来ます。もし、日本語環境以外で使っている場合などでこのような画面がうまく表示されない場合は

<http://www.kanji.zinbun.kyoto-u.ac.jp/db/CHINA3/index.html.ja>

と打ってください。

なお、この頁は漢字情報研究センターの先頭頁

<http://www.kanji.zinbun.kyoto-u.ac.jp/>

からも迎えることができます。

2.2 検索の仕方

「検索文字列」の入力窓 (図 2.2) に検索したい文字を入力し「検索開始」ボタンを押せば検索が実行されます。

*京都大学 人文科学研究所 附属 漢字情報研究センター



図 1: CHINA3 for WWW



図 2: 検索文字列の入力

検索結果は図 2.2 のように表示されます。外字は画像で表示されます。

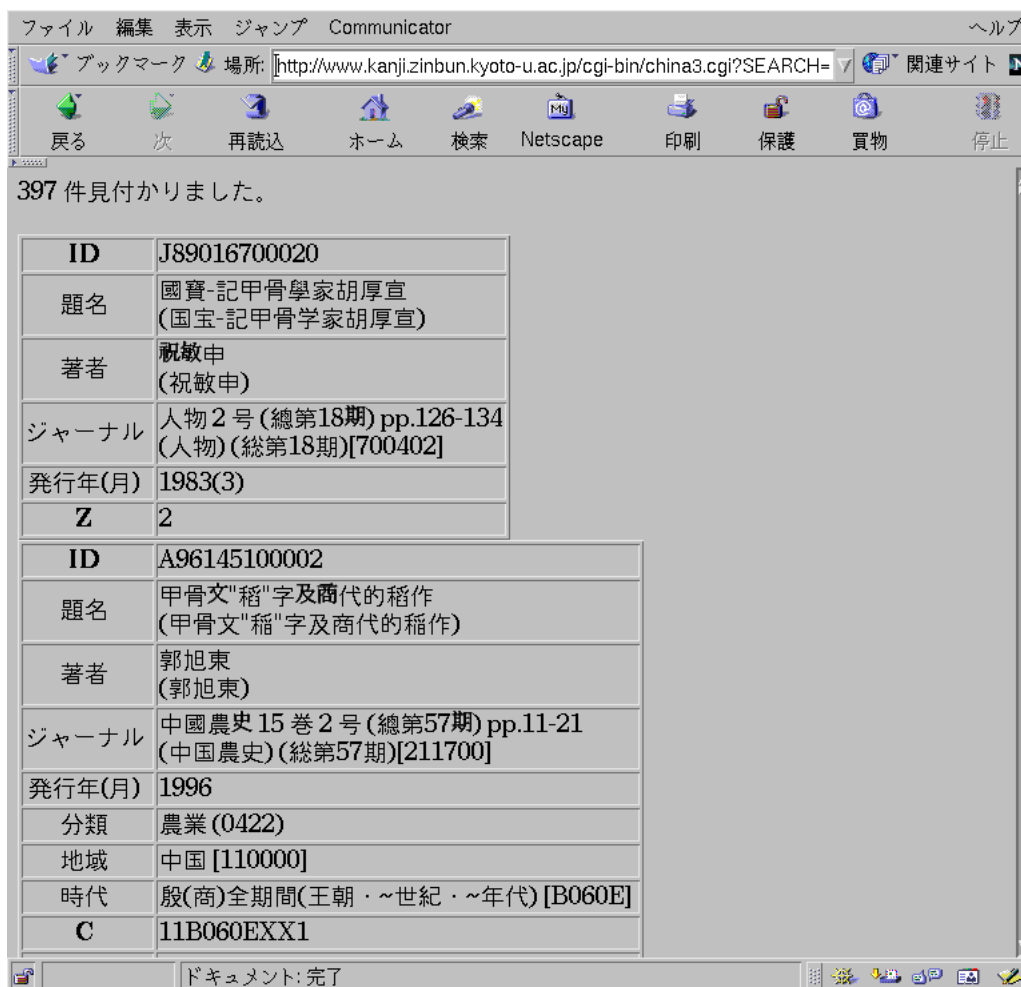


図 3: 結果の表示

なお、現在のシステムではキーワードはできるだけ常用漢字ないしは JIS X 0208 の漢字に正規化して登録されているので、「検索文字列」に入力する文字列は繁体字ではなく常用漢字にする必要があります。

2.3 文字符号

検索の結果を表示する文字符号を指定することができます。ISO-2022-JP と Shift_JIS の場合、JIS X 0208 の範囲の文字が通常のテキストとして送信され、それ以外の文字は画像として送信されます。EUC-JP の場合、JIS X 0208 に加え JIS X 0212 の範囲の文字も通常のテキストとして送信されます。

3 将来計画

- 「著者」などでも検索できるようにする
- 複合検索ができるようにする
- 文字データベースと連係して、繁体字や各種異体字も考慮して検索できるようにする
- 文字符号として GB, Big5, UTF-8 など使えるようにする
- 英語版、中国語版の頁を設ける
- 他の WWW 頁とリンクできるようにする

といったことを計画しています。