

「琉球家譜」の情報化と漢字処理

桶谷猪久夫：大阪国際女子大学人間科学部

1. 漢字（外字）処理の意義と概要

「琉球家譜」や「琉球王国評定所文書」[1-7]など古典文献のデータベースを構築し、柔軟な検索を可能にし、標準的な情報発信機能を備えることで急速に普及しているインターネットのWWW(World Wide Web)を利用して、研究者が手軽なパーソナルコンピュータを利用し、どこからでも利用できるようになれば、研究支援システムとして大いに役立つ。しかし、長年の歴史の間に蓄積され収集された、これらの史料には、外字や異体字の問題、解読不可能な文字や欠字の出現など多くの問題が存在する。さらに、これらの文字に対する入出力や検索機能の効果的な実現法など解決すべき種々の問題が存在するのが現状である。

ここでは、インターネットのWWWを利用した文書検索システム[8]上で、開発した外字検索機能と表示機能について述べる。

2. 文献に出現する外字と入力方法

ここで取り扱う文献は、一部に変体カナなどが出現するが、ほとんど漢文で記述され、漢字字種が多く外字や不明個所が混在するという特徴を持っている。このため「琉球家譜」等のデータ入力の字体については、ある規則[4]を取り決めて入力されている。たとえば、不明箇所は に置き換えている。また、俗字や別体については、JISコードに存在しないときは正字に置き換える。さらに、外字については、たとえば **琨** は 王 昆 に、**莘** は 艸 辛 のように、漢字通しの組み合わせを で囲んで外字の分解要素として入力されている。

以下に、「琉球家譜」、「琉球王国評定所文書」に出現する外字数とその外字固有数（種類）を記述する。

文献名	出現外字数	外字固有数
那覇市史資料編第一巻六「家譜資料(二)久米村系」	4077個	455種類
那覇市史資料編第一巻七「家譜資料(三)首里系」	1275個	261種類
那覇市史資料編第一巻八「家譜資料(四)那覇・泊系」	1835個	227種類
琉球王国評定所文書第一巻	240個	54種類
琉球王国評定所文書第二巻	406個	36種類
琉球王国評定所文書第三巻	1297個	17種類
琉球王国評定所文書第四巻	1310個	31種類
総合計	10440個	890種類

3. 外字検索機能と表示機能

古典文献の情報検索システムでは、いかにして外字を入力し、出力(表示)するのか、どのように検索するのか、さらにここで利用環境としているWWWでは、いかにして転送するのが問題になる。

外字を何らかの作成ツールを利用して作成した場合は、通常の漢字と同等に編集や検索が可能である。しかし、これは利用者の使用している機器やOSに依存するため、ここで提供基盤として想定するWWWによる情報検索や転送では利用できない。そのため、本文書検索システム[8]では、外字処理に対して検索機能と表示機能を分けて考えた。

外字入力には、漢字を部分品として分解し、分解した文字列として 印(一種のタグの役割)で囲んで入力した。この外字入力方式[9]を利用し検索機能を実現する。

外字を含んだ単語や外字の検索には、外字を 分解された文字列 の形式で入力するか、本文書検索システムで実現した外字一覧表の表示画面から選択し入力する。上記の表から外字数(890字)が多いので、分解された文字列での検索結果の表示も可能である。たとえば、部首が「艸」ならば、外字選択のテキスト入力フィールドに「艸」を入力し、外字選択をクリックすれば、分解された文字列に「艸」が含まれる外字一覧を表示する。分解された文字列のボタンをクリックすることで、その該当外字をキーワード入力フィールドに設定できる。

検索結果の外字の表示については、テキスト中で囲まれた外字に対応する文字列が現れたとき、分解された文字列とGIF形式ファイルの対応付けファイルを参照することによって、GIF形式の外字フォントに置き換えられる。外字フォントとして、京都大学人文研究所、勝村哲也教授が開発・作成した24×24ドットの漢字フォント[10]を利用した。その漢字フォントファイルは、今後の展開も考慮し、ユニコード(JIS X0221)に対応させた。ユニコード(JIS X0221)の漢字フォントが存在しない外字に対しては、Photoshopで外字を部分品として作成し対処する。 図1に、外字

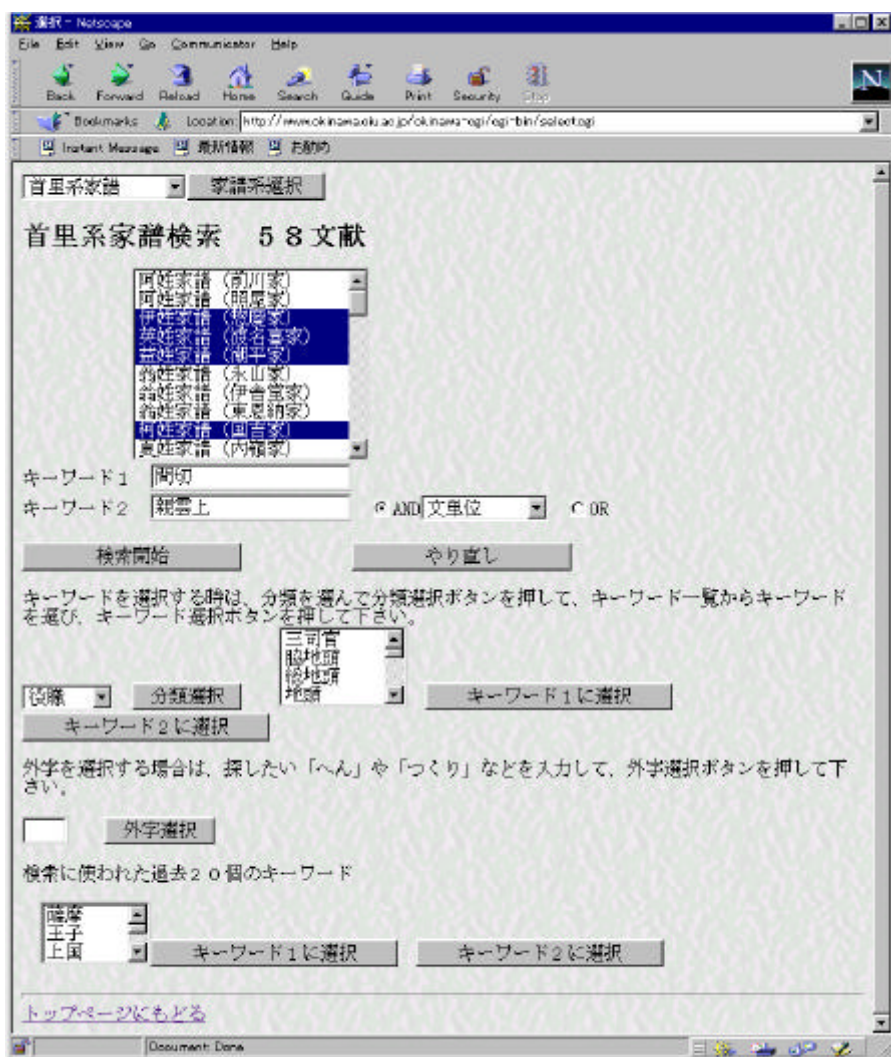


図1. フォーム画面での外字検索例

- [2] 那覇市史編集委員会編、『那覇市史資料編第一巻八「家譜資料(四)那覇・泊系」』、那覇市企画部市史編集室、P.1 - 821、人名索引、P.1 - 36,1983.3.31
- [3] 那覇市史編集委員会編、『那覇市史資料編第一巻六「家譜資料(二)久米村系」』、那覇市企画部市史編集室、P.1 - 946
- [4] 琉球王国評定所文書編集委員会編、『琉球王国評定所文書第一巻』、浦添市教育委員会、P.1 - 612, 1988.3.25
- [5] 琉球王国評定所文書編集委員会編、『琉球王国評定所文書第二巻』、浦添市教育委員会、P.1 - 588, 1989.1.31
- [6] 琉球王国評定所文書編集委員会編、『琉球王国評定所文書第三巻』、浦添市教育委員会、P.1 - 477, 1989.3.20
- [7] 琉球王国評定所文書編集委員会編、『琉球王国評定所文書第四巻』、浦添市教育委員会、P.1 - 588, 1990.3.20
- [8] 桶谷猪久夫、『32.02「琉球家譜」の検索システム』、総括班研究報告書,P. - ,1997.03
- [9] 中村洋子、豊見山和行、『家譜入力 of 字体について』、P.1 - 5、1995.11.2
(注) 文献に出現する外字に対する入力時における規則と作字一覧表
- [10] 国際符号化文字集合(UCS) - 第1部 体系及び基本多言語面
(注) 漢字フォント(20,902文字)、京都大学人文科学研究所勝村哲也教授提供