

Ⅱ - 2 - 2

小口雅史編「日本古代史関係研究文献目録データベース」

(法政大学国際日本学研究所データサービス内)の構造について

小口 雅史

1、はじめに

現在、「木簡庫」での検索結果表示の下部に、「研究文献情報」というボタンがあり、当該木簡を研究対象として取り上げている研究文献一覧を表示する仕組みになっている。現時点ではその引用元は、筆者が編集および公開責任を負っている「日本古代史関係研究文献目録データベース」⁽¹⁾(法政大学国際日本学研究所データサービス〈通称、電子図書館〉⁽²⁾内)と国立情報学研究所公開の CiNii Articles の2カ所である。

木簡庫で法政大学国際日本学研究所公開データを利用するために、「独立行政法人国立文化財機構奈良文化財研究所と法政大学国際日本学研究所との学術交流・協力に関する基本協定書」⁽³⁾が、松村恵司奈文研所長と私との間で2016年10月に締結され、正式運用が始まり現在に至っている。

本稿では、木簡庫で利用していただいているものをはじめとして、国際日本学研究所が公開しているデータベースの構造について簡単に紹介させていただこうと思う。

2、法政大学国際日本学研究所公開データベースの概要

法政大学国際日本学研究所データサービスにおいては、その設置以来様々なデータを公開してきているが、基本的にはそれらは研究所内に設置された専用の Windows Server によって運用されている。

この Server 上には、国際日本学研究所のスタッフが独自に設置した、全文検索システム Namazu、および FileMaker 社製の FileMakerServer などがあり、のちに PHP によってプログラミングされた在欧日本仏教美術データベース JBAE も搭載された。それぞれのシス

(1) <http://aterui.i.hosei.ac.jp/oguchi/kodai/index.htm>

(2) <http://hijas.hosei.ac.jp/database-services-top>

(3) 本データベースをめぐっては、同様の協定を国立歴史民俗博物館(館長:久留島浩)とも結んでいる。「国立歴史民俗博物館と法政大学国際日本学研究所との学術交流・協力に関する基本協定書」2016年7月。

テムの相性の問題が生じるようになり、現在では Windows Server を2台に分けて、トラブルを回避しているが、将来的にはシステムを再検討し、サーバーを1台にまとめる方向で進んでいる。

そうしたデータベースのなかで、文字のみで構成されている研究文献目録を扱っているのが全文検索システム Namazu である。Namazu とは、手軽に使えることを第一に目指した日本語全文検索システムで、CGI として動作させることにより小中規模の WWW 全文検索システムを構築することが比較的容易にできる⁽¹⁾。国際日本学研究所では、WEB 上で自由にキーワードを入力して検索が出来るというこの柔軟性を生かし、どのフィールドに検索対象語句があっても、それを明示した形で結果を表示できる、「蝦夷(エミシ・エビス・エゾ)を中心とした東北～北方史関係研究文献目録データベース(稿)」（通称は・古代北方史関係研究文献目録データベース）をまず公開した。データベースには必ずフィールド分けがつきもので、原則として一つのデータを定義にしたがっていくつかのフィールドに分けることが要請される。しかしそのことは、自分が探している研究論文を見つける際に障害となることがある。国際日本学研究所で公開している研究文献目録データベースでは、著者・論文名・掲載誌・刊行年月・再録書などのフィールドに区切られているが、通常の検索システムでは、全てのフィールドに対していちいち検索語句を打ち込んで、繰り返しチェックを続けることになる。もちろんマクロ的なものをその都度組めばこうした作業を自動化することもできるが、一般ユーザーにはそれは易しい仕組みではない。Namazu を用いてデータベースを構築すれば、どのフィールドに自分の探しているキーワードが含まれようとも、一度の操作でそれをすべてチェックできるようになっている。これによって、初心者であってもデータベースの操作性は飛躍的に高まる。さらにこの研究文献目録データベースで特徴的なことは、通常文献目録データベースで必要とされる、上記の著者・論文名・掲載誌・刊行年月・再録書といったフィールドの他に、対象地域というフィールドを設定して、北方史の中でとくにどの地域を扱ったものであるかを明示するようにしたことがあげられる。一口に北方史といっても各論文の視点は様々で、また逆に検索する側でも北方史の中で特定地域を集中的に検索したいという要望は強い。そうした要望に何とか応えられるものになった。

国際日本学研究所では、この Namazu を用いて、中世北方史関係研究文献目録データベース、アイヌ史関係研究文献目録データベース、北方史関係考古学研究文献データベースの蓄積作業を継続中であり、またさらに研究者の利便性を高めるために、それらを横断的

(1) 詳しくは<http://www.namazu.org/>参照

に検索できる仕組みも準備されつつある。

実はこれらのデータベースのいくつかは、20万件を超える巨大な「日本古代史関係研究文献目録データベース」から一括処理で目的別に抽出したものである。当時のデータ処理能力に合わせたこともあるが、他にもいくつかの理由があって、大もとの「日本古代史関係研究文献目録データベース」は関係者のみの内部公開にとどめている。ただし木簡庫での利用に当たっては、すべてのデータの参照を許可している。公開を控えている一つの大きな理由は Namazu の検索システムである Kakashi の語句の切り分けに際しての最適化の問題が未解決であるからである。学界から強い公開要望が寄せられているが、いましばし時間をいただきたいところである。そもそもこの規模のデータベースになると、私が個人的に維持するのはすでに限界を超えているとも言われて久しいが⁽¹⁾、いずれしかるべき国家機関にこれまでのデータを預けることもやぶさかではないと考えている。

一方、FileMakerServer については、Windows 版でも Mac 版でも完全に同一ファイルを扱うことできるスタンドアロン版の FileMaker で基礎データを入力し、それを FileMakerServer が標準で備える簡易 Web 公開機能を用いて公開した。これも大変柔軟性があり、かつ Namazu では不可能な画像付データも処理できるので、“Digital Library”ではフィールドが固定されていて処理できない画像付データベースをこれで公開することにしてきた。現時点では具体的には、国際日本学に深く関係する「在ベルリン・吐魯番文書関係データベース」などをこれで処理している⁽²⁾。これはベルリンの国立図書館その他に架蔵されている、トルファン出土の漢文世俗文書を整理して全文テキスト・画像付で公開しようとするものである。これは日本古代史とも深い関係を持ち、国際的な協力によって成り立ったもので、国際日本学のモデルケースとして公開に踏み切った経緯がある。

また PHP によってプログラミングされた JBAE については、拙稿が別途用意されているので、詳細はそれを参照されたい⁽³⁾。

(1) 現実問題として私の校務負担の増加に応じて、残念なことに更新頻度は大幅に下落している。

「木簡庫」にもご迷惑をおかけしている次第である。

(2) 詳細は拙稿「在ベルリン吐魯番漢文文書とその電子化—その現状と課題・展望—」『漢字文献情報処理研究』6、2005年。田口昌弘・小口雅史「『在ベルリン吐魯番出土漢文世俗文書総合目録』のその後—FileMaker による Database の Web 公開の一例として—」『漢字文献情報処理研究』8、2007年、他参照。

(3) 小口雅史「在欧美術館・博物館所蔵の日本仏教美術を訪ねて（1）—フィンランド・ヨエンス—美術館の巻—」『法政史学』87、2017年。

3、Namazuによる具体的なプログラム内容

1) 作業の目的と使用したソフトウェアについて

ここではすでに一般公開されている「蝦夷(エミシ・エビス・エゾ)」を中心とした東北～北方史関係研究文献目録データベース(稿)、通称「古代北方史関係研究文献目録データベース」を全文検索システム Namazu を使用してウェブ上で処理した過程を紹介したい。本簡庫で利用している、一般には未公開の「日本古代史関係研究文献目録データベース」もシステム上は全く同じ構造で処理されている。

また2)で述べるように、データ処理に当たっては、スクリプト言語 Perl を補助的に使用した。なおこのデータベース(以下「論文DB」と表記)の公開に当たっては、INBUDS(インド学仏教学論文データベース⁽¹⁾)のシステムを参考にしている。同システムの構築に当たった師茂樹氏に、この場を借りて謝意を表したい⁽²⁾。

a 作業の対象とした論文DBの性質

論文DBの特徴は以下の通りである。

- ・管理工学研究所の日本語データベースソフト「桐V9」(当時)によって作成された。
- ・ただし公開用にデータを抜き出した時点では実質的にはリレーショナルデータベースとしての機能は使用されておらず、カード型データベースとして扱うことが可能な形式となっている。
- ・入力されている項目は21項目以上。
- ・項目内での区切り記号は|・|・||の3種類。通常は|(全角)・|(半角)が使用される。半角区切りはデータが数字である項目内で使用されている。||は大区切りで、必要のある場合のみ、|・|に加えて使用されている。
- ・使用されている文字コードはJISベースで独自外字とメーカー外字(IBM外字やWindows外字と呼ばれるもの)も使用している。外字を使用した場合はデータの「漢字注記」の項に、今昔文字鏡番号が&M型式で記入されている。

b 全文検索システム Namazu⁽³⁾について

全文検索システム Namazu は、もともと高林哲氏によって開発され、version 2.0からは

(1) <http://www.inbuds.net/>

(2) 法政大学国際日本学研究所において、当時、プログラミングの具体的作業に従事したのは田口昌弘氏である。なお、本章の記述は、「ワーキングプロジェクト「成果と情報化の活用」『日本学の総合的研究』(法政大学国際日本学研究所、2007年)をベースにしている。

(3) <http://www.namazu.org/project.html.ja>

Namazu Project によって共同開発されている日本語全文検索システムである。特徴は以下のようなものである。

- ・高速な検索が可能。
- ・実行できる複合検索の種類が豊富。
- ・検索結果表示のカスタマイズ性が高い。
- ・多くのプラットフォームに対応。
- ・フリーウェアである。

c スクリプト言語 Perl について

スクリプト言語 Perl はラリー・ウォール氏によって開発されたインタプリタ方式のプログラミング言語。しばしばウェブ上で動作する CGI の作成に使用される。また Namazu の動作にも必要とされる。

2) 作業の流れについて

作業内容は以下の4つの段階に分けられる。

(1) Namazu と Perl (とそれらを補助する) ソフトウェアをサーバーにインストールする。次にデータを置くためのフォルダを設定し、あわせてウェブ上で公開するためのフォルダを作成する。→ 3)

(2) 作成されている論文 DB のデータはそのままの形では Namazu で扱えないため、Perl によって適切な形式に変換する。→ 4)

(3) Namazu の動作に必要な設定を行う。→ 5)

(4) Namazu での検索結果をウェブ上で公開するための形式に Perl によって加工する。→ 6)

3) ソフトウェアのインストールと各種フォルダの作成

作業に必要なソフトウェアは aterui サーバーの C ドライブにインストールし、データフォルダとウェブ公開用のフォルダは F ドライブに置くことにした。

a どのサーバーを使用するか

マイクロソフト社の Windows2000 Server (当時) で運用している aterui サーバーを使用した。

Namazu も Perl も多くのプラットフォームに対応しており、扱うデータもテキスト形式であるので、使用するサーバーの OS については特に制限がない。よって実務担当者が作業に慣れているマイクロソフト社の Windows 系列のサーバーを選択した。サーバー室

を管理している SE によれば、今回の用途には aterui サーバーが適しているとのことであった。

b ソフトウェアのインストール

Perl を C ドライブの perl フォルダにインストール。

Kakasi (Namazu の動作に必要。後述) を C ドライブの kakasi フォルダにインストール。

Namazu を C ドライブの Namazu フォルダにインストール。

システム置かれている C ドライブはもともと 4 ギガバイトしかなく空き容量に乏しかった。ソフトウェアのインストールは避けたかったのだが、これ以外の場所では挙動に問題が生じる可能性があり、上記のようにした。

インストールに使用したパッケージはそれぞれ、Perl : ActivePerl-5.6.1.635-MSWin32-x86.msi、kakasi : kakasi-2.3.4.zip、Namazu : nmz2014beta.exe。Perl は安定している 5.6.1 系の最新版 (当時) を、kakasi は最新版 (当時) を、Namazu はベータ版ではあるがセキュリティ関連のバグフィックスがなされた最新公開版 (当時) を用いた。

c データフォルダの作成

F ドライブに oguchi_local というフォルダを設定し、さらに下位に ezo_db というフォルダを作成して、今回の論文 DB 関連のファイルを置くことにした。

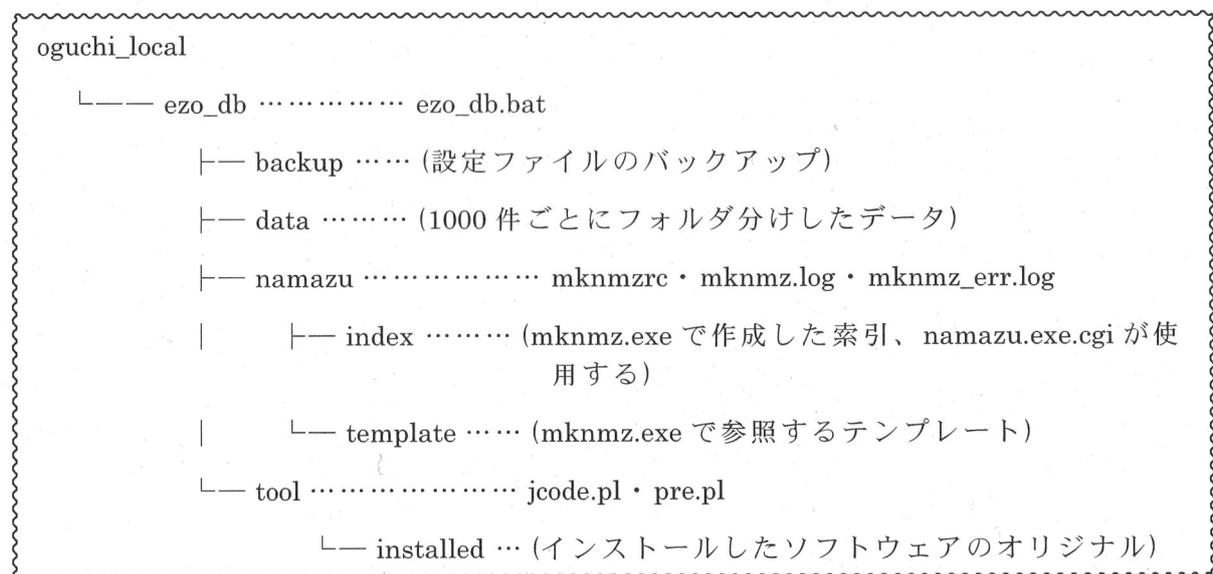


図44 データフォルダ配置

ezo_db フォルダの下位には、data フォルダ (データを 1000 件ごとにフォルダ分けして置く) ・ namazu フォルダ (Namazu の設定ファイル等を置く) ・ tool フォルダ (データ加工用に Perl で作成するスクリプト等を置く) ・ backup フォルダ (上記関連ファイルのバックアップを入れる) を設置した。

他にEドライブも空き容量が多かったが、既にインストールされているソフトウェアから判断するとCドライブの補完的な役割をするドライブとしてソフトウェアをインストールするための場所と位置づけられているようであった。よってデータフォルダの設定場所候補から除外した。

d ウェブ公開用フォルダの設定

Fドライブに oguchi というフォルダを設置し、これを国際日本学研究所のウェブサイトの置かれている mediacenter フォルダの下位に仮想ディレクトリとして OS の IIS (インターネットインフォメーションサービス) を使用して設定した。

さらに oguchi フォルダの下位に Namazu の検索エンジン部分である Namazu.cgi.exe や検索結果を加工して表示する Perl のスクリプトを置くための search というフォルダを作成した。

oguchi フォルダには HTML や画像ファイル等の静的コンテンツを配置することとし、IIS の設定は読み取り・ログアクセスのみを許可し、スクリプト等の実行アクセス権は無

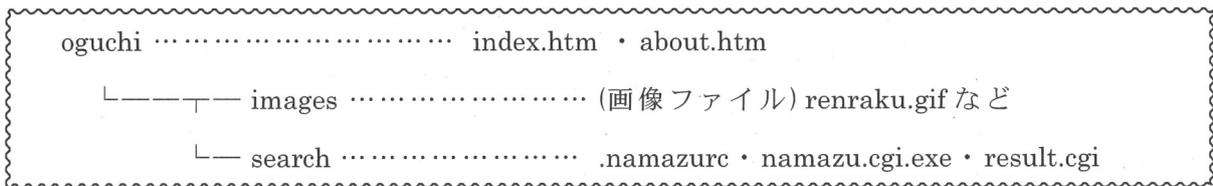


図45-1 ウェブ公開用フォルダの配置

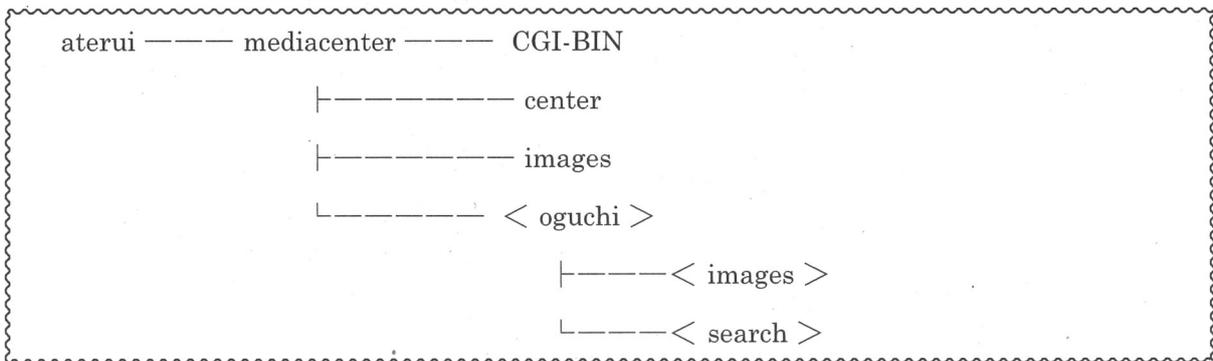


図45-2 IISで設定したサーバー上でのウェブ公開用フォルダの配置 (<>は今回作成したもの)

しとした。search フォルダには CGI 等の動的コンテンツを配置するため、IIS の設定はログアクセスのみ可とし、スクリプト及び実行可能ファイルに実行アクセス権を許可した。また同フォルダについてのみ、拡張子 cgi のファイルを動作可能とするため Perl のダイナミックリンクライブラリ (C:\Perl\bin\perlis.dll) を関連付け、Namazu の検索結果を加工するために必要な GET メソッドの動作を許可した。

e ウェブ公開用 HTMLの準備

論文 DB 公開のトップページとなる index.htm と、論文 DB の凡例である about.htm を作成した。ともに図 45-1 のように配置した。

・ index.htm の作成

index.htm からは、最新情報(リンク先未作成)・古代北方史関係研究文献目録データベースについて (about.htm) ・オンライン検索 (namazu.cgi.exe) ・法政大学国際日本学研究所ホームページへ (<http://www.hosei.ac.jp/21coe/nihon/>) へのリンクを張った。

また連絡先メールアドレスを画像で掲載した。5) a 参照。

```
< index.htm >
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">
<HTML lang="ja">
<HEAD>
<TITLE>Result</TITLE>
<META http-equiv="Content-Type" content="text/html; charset=EUC-JP">
<LINK rel="stylesheet" href="./ezo_db.css" type="text/css">
</HEAD>
<BODY>
<H1>古代北方史関係研究文献目録データベース</H1>
<P>正式名称：「蝦夷(エミシ・エビス・エゾ)」を中心とした東北～北方史関係研究文献目録データベース(稿)</P>
<HR>
<P>
<LI>最新情報
<LI><A href="/about.htm">古代北方史関係研究文献目録データベースについて</A>
<LI><A href="/search/namazu.cgi.exe">オンライン検索(テスト中)</A>
<LI><A href="http://www.hosei.ac.jp/21coe/nihon/">法政大学国際日本学研究所ホームページ</A>
</P>
<HR>
<ADDRESS><IMG src="/images/renraku.gif"></ADDRESS>
</BODY>
```

</HTML>

--- (ここまで) ---

- ・ about.htm の作成

about.htm には論文 DB の凡例を記述し、さらに index.htm に戻るリンクを張った。

< about.htm >

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">

<HTML lang="ja">

<HEAD>

<TITLE>Result</TITLE>

<META http-equiv="Content-Type" content="text/html; charset=EUC-JP">

<LINK rel="stylesheet" href="../ezo_db.css" type="text/css">

</HEAD>

<BODY>

<H1>古代北方史関係研究文献目録データベース</H1>

<P>正式名称：「蝦夷(エミシ・エビス・エゾ)」を中心とした東北～北方史関係研究文献目録データベース(稿)</P>

<HR>

<H2>凡例</H2>

 このデータベースは「蝦夷」に関する研究文献を対象としたものである。その「蝦夷」と呼ばれた人々が生活していた舞台である東北以北の地域に関する研究文献を集成することを目的としている。

 「蝦夷」と呼ばれた人々の中でもとくに「エゾ」と呼ばれた人々と密接に関わるアイヌ文化、またそのアイヌ文化の形成に影響を与えた擦文文化・オホーツク文化に関する研究文献も対象とした。アイヌ文化をめぐる諸論考については、その前史と関わるもの、あるいは北方世界の一つの特徴である交易ないし物の移動をめぐる論考を主体としている。そのためアイヌ史全般の研究文献目録とはなっていない。擦文文化期にさかのぼるとされているアイヌ語地名については採録の対象とした。

 文献に「蝦夷」と呼ばれた人々が登場し活躍する前後の時代についても、上記2.と同じく北方世界における交流の問題をみつかった論考は採録している。

 上記の目的を達成するため、文献分野に限らず考古学分野の諸研究についても対象としているが、現時点では考古学分野については網羅し切れていない。また発掘調査報告書については原則として対象としなかったが、発掘調査報告書に含まれる考察的な部分については一部採用した。とくに後に著者個人の論文集に収録された場合には優先的に採録した。

 自治体史については原則として採録していない。ただし上記3.と同じく後に論文集に収録された場合などについては配慮して一部採録している。

 辞典類については長文のもの、筆者の見解が端的に要約されているものを中心に採録した。啓蒙書については一応対象としているが、一部掲載を省略したものがある。

 以上の原則に従って、小口雅史編『日本古代史関係研究文献目録データベース』（テキストファイル版）から分類記号 D7・D8 を基準に抄出し、さらにやはり小口が作成を担当している青森市史編纂のための北方史関係文献データベースを参照した上で、あらためてさらに増補を加えた。なお論文名欄に「-」とあるものは、その書籍に含まれる論文中、当データベースに採録すべきものがすべて既刊であることを示している。

 データ項目のうち、「参照 No.」は再録書の、本データベースにおける通し番号上の位置を示す。また「既成目録」は従来この分野で利用されてきたデータ No.5448、12682 の両目録に掲載されていることを示し、前者を T、後者を K で表している。ただしその目録のデータを補正した場合にはそれぞれ t、k で表している。

 データについてはできるだけ原典と照合するようにつとめたが、一部現物にあたれなかったものがある。それらについてはとりあえず本データベースにも収録されているいくつかの先行文献目録に依存した。今後現物との照合に努力していきたい。

 データのなかには論文等の書かれた時代の認識を反映して差別的な表現がとられているものがあるが、データベースとしての性格上そのまま残している。もちろん本データベースはそうした認識を許容する立場にたつものではない。

 これだけの規模になると脱漏・正誤などは避けられない。またとくに各地域で刊行されている雑誌類については一般に入手が困難であり、貴重な論考が多数脱落しているものと推測される。利用者各位からの積極的な情報提供を期待している。

<P> 【トップに戻る】</P>

<HR>

<ADDRESS></ADDRESS>

</BODY>

</HTML>

--- (ここまで) ---

4) 論文DBの加工

元データを Namazu で扱うために論文データ 1 件ずつ (= 1 レコードごと) に切り離し、さらに検索対象とする部分と検索対象にしない部分を別ファイルとした。これらのファイルは 1000 レコードごとにフォルダ分けして作成した。

a 桐からのデータ出力

まず桐からデータをカンマ区切りの csv 形式でテキストファイルとして出力する。

ただし出力は桐に登録した自動操作により、論文 DB に入力されているデータ項目の内、で公開する項目のみとする。この工程は小口の手によるもので作業員田口氏は関与していない。

csv ファイルに出力される項目は以下の 21 項目。各項目の頭についているのは出力される順番である ([0] から始まっているのは後の作業での都合による)。各項目の後の説明は作業員が作業用に付した説明で、必ずしもデータ入力者の意図と合致するとは限らない。

表4 データ項目一覧

- | |
|--|
| [0] No. : データの通し番号 |
| [1] 執筆者 |
| [2] 論文名 |
| [3] 掲載誌等 |
| [4] 巻号 : [3] の巻号 |
| [5] 特集 : [3] の特集 |
| [6] 編・著 : [3] の編・著者 |
| [7] 別誌名等 : [3] 所収の叢書名など |
| [8] 刊行年月 : [3] の刊行年月 |
| [9] 出版社 : [3] の出版社 |
| [10] 備考 : データ全体についての備考 |
| [11] キーワード : 論文内容のキーワード (題名と重複するものは採っていない) |
| [12] 史料 : 論文内で扱っている史料 |
| [13] 参照 No. : [14] の [0] |
| [14] 所収・再録書名 |
| [15] 改題・改稿 : [14] の際の改題・改稿 |

- [16] 漢字注記：データ全体の漢字表記についての注記
- [17] ふりがな：[1]のふりがな
- [18] 既成目録：既存目録での収録状況
- [19] 対象地域：論文内で扱っている地域
- [20] 版本データ No.控：小口雅史・畠山恵美子・中野栄夫編「北方史関係文献目録（稿）」『法政大学国際日本学研究所研究報告』第4集、法政大学国際日本学研究所、2004年3月での番号

b データの切り離し

そもそも Namazu は多数のファイルに対してあらかじめ語彙索引を作成しておき、その索引に対して全文検索を行うソフトウェアである。したがって本来、今回のような多数のレコードからなる単一ファイルの DB に対して使用するものではない。さらに [0] や [4] などの項目は検索対象としても意味がない。

そこで csv ファイルを1レコードごとに切り離し、さらに検索対象とする部分としない部分(下記)を別ファイルとする。結果としてレコード件数の倍のファイルが生成されるが、これをすべて同一フォルダに置いた場合、そのフォルダを操作する際の OS の挙動が非常に緩慢もしくは不安定になるため、1000レコードごとにフォルダを作成して配置する(よって各フォルダには原則として2000のファイルが置かれる)。

表5 検索対象にする項目としない項目

<検索対象にする>

- [1] 執筆者, [2] 論文名, [3] 掲載誌等, [5] 特集, [7] 別誌名等, [9] 出版社, [10] 備考,
[11] キーワード, [12] 史料, [14] 所収・再録書名, [15] 改題・改稿, [18] 既成目録,
[19] 対象地域

<検索対象にしない>

- [0] No., [4] 巻号, [6] 編・著, [8] 刊行年月, [13] 参照 No., [16] 漢字注記, [17] ふりがな,
[20] 版本データ No.控

※ [17] については、同一の読みの別人物である場合に区別がされていないため、検索対象から外すことにした。

c データを切り離すためのスクリプト作成

前節で述べた加工を行うために検索前処理用の Perl のスクリプト pre.pl を作成した。あわせて日本語コード変換を行う jcode.pl も tool フォルダに置いた。なお jcode.pl*9 は歌代和正氏によって作成された Perl のライブラリで、JIS ベースの日本語コードの相互変換

をするために広く使用されているものである。

d pre.plの詳細

< pre.pl >

```

01: #!/usr/local/bin/perl
02:
03: # pre.pl v0.942
04:
05: $dat_folder = "F:/oguchi_local/ezo_db/data" ;
06: require "jcode.pl" ;
07:
08: while ($line = <>) {
09:   chomp ;
10:   $line = &jcode::euc ($line, "", "z") ;
11:   $line =~ s/¥|/ | /g ;
12:   @dat = split (/,/, $line) ;
13:
14:   $folder = int (@dat [0] / 1000) ;
15:   $folder_03d = sprintf ("%03d", $folder) ;
16:   $file = @dat [0] % 1000 ;
17:   $file_03d = sprintf ("%03d", $file) ;
18:
19:   if (!-d "$dat_folder/$folder_03d") {
20:     mkdir "$dat_folder/$folder_03d" ; }
21:
22:   if ("$folder_03d-$file_03d" ne "000-000") {
23:     $write_url = "$dat_folder/$folder_03d/$folder_03d-$file_03d.htm" ;
24:
25:     @out1 = (@dat [1], @dat [2], @dat [3], @dat [5], @dat [7], @dat [9], @dat [10], @dat [11],
,@dat [12], @dat [14], @dat [15], @dat [18], @dat [19]) ;
26:     $out1_join = join ("", @out1) ;
27:     open (OUT1, "> $write_url") ;
28:     print OUT1 $out1_join ;

```

```

29: close (OUT1) ;
30:
31: @out2 = (@dat [0] ,@dat [4] ,@dat [6] ,@dat [8] ,@dat [13] ,@dat [16] ,@dat [17] ,@dat [20] )
;
32: $out2_join = join (" , " , @out2) ;
33: open (OUT2, "> $write_url.dat") ;
34: print OUT2 $out2_join ;
35: close (OUT2);
36: }
37: }
38:
39: exit ;
--- (ここまで) ---

```

5) Namazuの設定

Namazu のシステムは2つのプログラムから構成される。あらかじめ全文検索の対象とする文書のインデックスファイルを作成しておくためのプログラム `mknmz.exe` と、そのインデックスを利用して検索を行うプログラム `namazu.exe` である。

a mknmzで使用するテンプレートファイルのカスタマイズ

`mknmz` は指定されたテンプレートに基づきインデックスファイルを生成する。

このインデックスファイルには、厳密な意味でのインデックス (=索引) 以外にも、`namazu` での検索結果と共に表示される HTML のヘッダやフッタをも含む。

カスタマイズすべきファイルは「`NMZ.????.ja`」(`.ja` は日本語表示用) というファイル名の6ファイルである。これらをカスタマイズ後に図44の `template` フォルダに配置する。

表6 日本語表示用の Namazuテンプレート一覧

NMZ.result.nomal.ja	: namazu での検索結果を表示するためのテンプレート
NMZ.result.short.ja	: namazu での検索結果を表示するための簡略版のテンプレート
NMZ.head.ja	: namazu での検索結果と共に表示される html のヘッダのテンプレート
NMZ.foot.ja	: namazu での検索結果と共に表示される html のフッタのテンプレート
NMZ.body.ja	: namazu の検索画面にある「検索方法」というリンクから表示される検

索方法例のテンプレート

NMZ.tips.ja : namazu での検索で該当するデータが無い場合に表示される画面のテンプレート

- ・ NMZ.result.nomal.ja ・ NMZ.result.short.ja のカスタマイズ

NMZ.result.nomal.ja からは、検索でヒットしたファイルの「著者・日付」行を削除した。

これは今回扱うデータが、ワープロ文書や PDF などのような、これらの項目をそれ自体のうちに保有しているタイプのデータではないからである。インデックス作成の対象とするのは csv を切り離して生成されるテキストファイルのため、「著者」は全て「不明」と表示され「日付」は全て切り離した日時となってしまう。NMZ.result.short.ja については手を加えていない。

< NMZ.result.nomal.ja - オリジナル >

```
< dt > ${namazu::counter}. < strong > < a href="${uri}" > ${title} < /a > < /strong > (スコア :
${namazu::score})
< dd > < strong > 著者 < /strong > : < em > ${author} < /em >
< dd > < strong > 日付 < /strong > : < em > ${date} < /em >
< dd > ${summary}
< dd > < a href="${uri}" > ${uri} < /a > (${size} bytes) < br > < br >
--- (ここまで) ---
```

< NMZ.result.nomal.ja - カスタマイズ後 >

```
< dt > ${namazu::counter}. < strong > < a href="${uri}" > ${title} < /a > < /strong > (スコア :
${namazu::score})
< dd > ${summary}
< dd > < a href="${uri}" > ${uri} < /a > (${size} bytes) < br > < br >
--- (ここまで) ---
```

- ・ NMZ.head.ja のカスタマイズ

見出しとして論文 DB の名称を挿入した。

インデックス化されている文書の数・登録されているキーワードの数・インデックスの最終更新日を表す行を 6) のスクリプトから取得できるような位置に改行コードを挿入し

た。

デフォルトのソート順を発表年度(昇順)にした。これは指示による。

データベースのトップページへ戻るリンクを追加した。

< NMZ.head.ja -オリジナル：関係箇所のみ>

<h1>Namazu による全文検索システム</h1>

<p>

現在、<!-- FILE --> 0<!-- FILE --> の文書がインデックス化され、

<!-- KEY --> 0<!-- KEY --> 個のキーワードが登録されています。

</p>

<p>

インデックスの最終更新日:<!-- DATE --> date<!-- DATE -->

</p>

<p>

ソート:

<select name="sort">

<option selected value="score">スコア

<option value="date:late">日付 (新しい順)

<option value="date:early">日付 (古い順)

<option value="field:subject:ascending">題名 (昇順)

<option value="field:subject:descending">題名 (降順)

<option value="field:from:ascending">著者 (昇順)

<option value="field:from:descending">著者 (降順)

<option value="field:size:ascending">サイズ (昇順)

<option value="field:size:descending">サイズ (降順)

<option value="field:uri:ascending">URI (昇順)

<option value="field:uri:descending">URI (降順)

</select>

</p>

--- (ここまで) ---

< NMZ.head.ja -カスタマイズ後：関係箇所のみ>

<H1>古代北方史関係研究文献目録データベース</H1>

<P>正式名称：「蝦夷(エミシ・エビス・エゾ)」を中心とした東北～北方史関係研究文献目録データベース(稿)</P>

<HR>

<P>Namazu による全文検索システム</P>

<P>

現在、

<!-- FILE --> 0 <!-- FILE -->

の文書がインデックス化され、

<!-- KEY --> 0 <!-- KEY -->

個のキーワードが登録されています。

</P>

<P>

インデックスの最終更新日:

<!-- DATE --> date <!-- DATE -->

</P>

<P> 【トップに戻る】</P>

<p>

ソート:

<select name="sort">

<option value="score">スコア

<option value="field:subject:ascending">題名 (昇順)

<option value="field:subject:descending">題名 (降順)

<option selected value="field:uri:ascending">発表年度 (昇順)

<option value="field:uri:descending">発表年度 (降順)

</select>

</p>

--- (ここまで) ---

・ NMZ.foot.ja のカスタマイズ

連絡先として表示されるメールアドレスの表示とリンクを、無用な spam を避けるため

に削除。代わりに連絡先メールアドレスを表示する画像ファイルを挿入した。

使用している Namazu のバージョンを表す行を 6) のスクリプトから取得できるような位置に、改行コードを挿入した。

```
< NMZ.foot.ja -オリジナル>
```

```
<hr>
```

```
<p>
```

この全文検索システムは

```
<strong><a href="http://www.namazu.org/">Namazu</a> <!-- VERSION --> v <!-- VERSION -->
```

```
</strong>
```

によって構築されています。

```
</p>
```

```
<address>
```

```
<!-- ADDRESS -->
```

```
<a href="mailto:foobar@namazu.org">foobar@namazu.org</a>
```

```
<!-- ADDRESS -->
```

```
</address>
```

```
</body>
```

```
</html>
```

```
--- (ここまで) ---
```

```
< NMZ.foot.ja -カスタマイズ後>
```

```
<hr>
```

```
<p>
```

この全文検索システムは

```
<strong><a href="http://www.namazu.org/">Namazu</a>
```

```
<!-- VERSION --> v <!-- VERSION -->
```

```
</strong>
```

によって構築されています。

```
</p>
```

```
<address>
```

```
<!-- ADDRESS --><!-- ADDRESS -->
```

</address>

</body>

</html>

--- (ここまで) ---

・ NMZ.body ja ・ NMZ.tips ja のカスタマイズ

NMZ.body ja については、論文 DB と関連する検索例を小口が用意して入れ替えた。「フィールド指定の検索」は今回使用していないので説明をコメントアウトした。

NMZ.tips ja についても同様に検索例を用意する予定であるが、現時点では変更を加えていない。

< NMZ.body ja - カスタマイズ後：検索例関連の行のみ抜き出し >

<H3>単一単語検索</H3>

<P class="example">蝦夷</P>

<H3>AND 検索</H3>

<P class="example">多賀城 and 胆沢城</P>

<H3>OR 検索</H3>

<P class="example">徳丹城 or 志波城</P>

<H3>NOT 検索</H3>

<P class="example">末期古墳 not 陸奥国</P>

<H3>グループ化</H3>

<P class="example">(蕨手刀 or 馬具) and 岩手 not 青森</P>

<H3>フレーズ検索</H3>

<P class="example">{征夷 将軍}</P>

<H3>部分一致検索</H3>

<DT>前方一致検索<DD><code class="example">出羽*</code> (<code>出羽</code> から始まる単語を含む文書を検索)

<DT>中間一致検索<DD><code class="example">*国府*</code> (<code>国府</code> を内包する単語を含む文書を検索)

<DT>後方一致検索<DD><code class="example">*柵</code> (<code>柵</code> で終わる単語を含む文書を検索)

<H3>正規表現検索</H3>

<P class="example">/妙法?蓮[華花]経/</P>

--- (ここまで) ---

b mknmzの動作オプションの設定

mknmz の動作の設定は、設定ファイル mknmzrc とコマンドラインオプションで行う。
カスタマイズした mknmzrc は図44の namazu フォルダに配置した。

・ mknmzrc のカスタマイズ

カスタマイズしたのは以下の3箇所である。

```
# $ADDRESS = 'webmaster@foo.bar.jp'; → $ADDRESS = '';
```

これは作成されるインデックスファイルに含まれる連絡先メールアドレスの設定である。5) a 参照。

```
$DENY_FILE = ".dat"; を追加
```

これはインデックスファイル作成の対象から除外するファイルの指定である。4) d の33 ~ 35 行で述べた拡張子 dat のファイルを指定した。デフォルトのままでも除外されるが、念のため指定しておく。

```
$TEMPLATEDIR = "F:\¥oguchi_local¥ezo_db¥namazu¥template"; を追加。図44参照。
```

・ mknmz のコマンドラインオプション

作成するインデックスファイルの対象を図44の data フォルダとする。

-O オプションで、作成するインデックスファイルの置き場所を図44の index フォルダ(あらかじめ作成しておく)に指定する。

-T オプションで、使用するテンプレートファイルを図44の template フォルダのものと指定する。

-f オプションで、読み込む mknmzrc の位置を図44のものと指定する。

インデックスファイルの作成結果と作成エラーの結果を図44の namazu フォルダにそれぞれ mknmz.log・mknmz_err.log という名称で出力するように指定(1)・(2))。

カレントフォルダを F:\¥oguchi_local¥ezo_db とした場合、コマンドラインで以上のオプションを記述すると次のようになる。

```
mknmz .¥data -O .¥namazu¥index -T .¥namazu¥template -f .¥namazu¥mknmzrc 1>
```

```
¥namazu¥mknmz.log 2>¥namazu¥mknmz_err.log
```

c インデックスファイル作成までの作業の自動化

バッチファイル ezo_db.bat を作成し、ezo_db フォルダに配置した。

4) a のデータをドラッグ&ドロップすると 4) d のスクリプト pre.pl と 5) b のコマンドラインオプションを付加した mknmz を一括して実行する。

- ezo_db.bat の詳細

```
< ezo_db.bat >
```

```
1: @echo off
```

```
2: F:
```

```
3: cd F:¥oguchi_local¥ezo_db¥tool
```

```
4: perl pre.pl %1
```

```
5: cd ..
```

```
6: mknmz ¥data -O ¥namazu¥index -T ¥namazu¥template -f ¥namazu¥mknmzrc 1>
```

```
¥namazu¥mknmz.log 2>¥namazu¥mknmz_err.log
```

```
7: gcnmz -b ¥namazu¥index
```

```
--- (ここまで) ---
```

d namazu.cgi.exe の設定

使用するのが namazu.exe であれば、設定は namazurc やコマンドラインオプションで行う。しかし今回使用するのは ウェブ上で cgi として動作させるための namazu.cgi.exe であるので、設定は同じフォルダに置く .namazurc (namazurc の名前を変えて作成)で行う。

- .namazurc のカスタマイズしたのは以下の 4 箇所。

カスタマイズしたのは以下の 4 箇所である。

```
Index          F:¥oguchi_local¥ezo_db¥namazu¥index
```

検索対象にするインデックスファイルを指定した。図 44 参照。

```
Replace/F¥|/oguchi_local/ezo_db/data/¥d¥d¥d/
```

```
http://aterui.i.hosei.ac.jp/oguchi/search/result.cgi?
```

aterui の F ドライブのデータをウェブ上で表示させるための設定。図 44 のデータフォルダを図 45-2 の aterui サーバー上から読み込めるように設定。「result.cgi?」とは、検索で

ヒットしたファイル名を6) の result.cgi に渡すための仕組みである。

Langja_JP.eucJP

表示する日本語コードをシフト JIS から EUC に変更した。

```
EmphasisTags "<strong class=¥"keyword¥">" "</strong>"
```

検索でヒットしたキーワードを強調表示する設定。もともとコメントアウトしていたのを有効にした。

※以下の行のコメントアウトを外してしまうと、カスタマイズしたテンプレートよりもこちらが優先されてしまうので、変更を加えてはならない。

```
# Template C:¥namazu¥share¥namazu¥template
```

6) 検索結果の加工表示

表5のように一つのレコードは二つのファイルに分けられている。よって検索してヒットしたデータについて結合を行う必要がある。さらにそれを HTML 形式で出力してウェブブラウザで表示させる仕組みが必要となる。そこでそのためのスクリプト result.cgi を Perl で作成し図45-1のフォルダに配置した。

a result.cgi の詳細

```
< result.cgi >
```

```
001:#!/usr/local/bin/perl
```

```
002:
```

```
003: # result_0944.cgi
```

```
004:
```

```
005: $nmzcgi = "namazu.cgi.exe"
```

```
006:                                $ n m z c g i o p                                =
```

```
"&amp;whence=0&amp;max=20&amp;result=normal&amp;sort=field:uri:ascending";
```

```
007: $nmzheadja = "F:/oguchi_local/ezo_db/namazu/index/NMZ.head.ja";
```

```
008: $nmzfootja = "F:/oguchi_local/ezo_db/namazu/index/NMZ.foot.ja";
```

```
009: $resultcgi = "result.cgi";
```

```
010:
```

```
011: $buffer = $ENV{'QUERY_STRING'};
```

```
012: $buffer =~ s/query=//;
```

```
013: @folder = split (/,,$buffer) ;
014: $read_url = "F:/oguchi_local/ezo_db/data/@folder [0] /$buffer" ; # aterui
015:
016: open (INPUT0,"<$nmzheadja") ;
017: while ($line0 = <INPUT0>) {
018:   chomp $line0 ;
019:   push @nmzhead, (grep (/<!-- FILE -->/,$line0)) ;
020:   push @nmzhead, (grep (/<!-- KEY -->/,$line0)) ;
021:   push @nmzhead, (grep (/<!-- DATE -->/,$line0)) ;
022: }
023: close (INPUT0) ;
024: $nmzhead [0] =~ s/<!-- FILE -->//g ;
025: $nmzhead [1] =~ s/<!-- KEY -->//g ;
026: $nmzhead [2] =~ s/<!-- DATE -->//g ;
027: open (INPUT1,"<$read_url") ;
028: $line1 = <INPUT1> ;
029: chomp ;
030: @input1 = split (/,,$line1) ;
031: close (INPUT1) ;
032: open (INPUT2,"<$read_url.dat") ;
033: $line2 = <INPUT2> ;
034: chomp ;
035: @input2 = split (/,,$line2) ;
036: close (INPUT2) ;
037: open (INPUT9,"<$nmzfootja") ;
038: while ($line9 = <INPUT9>) {
039:   chomp $line9 ;
040:   push @nmzfoot, (grep (/<!-- VERSION -->/,$line9)) ;
041: }
042: close (INPUT9) ;
043: $nmzfoot [0] =~ s/<!-- VERSION -->//g ;
044:
```

```

045: print "Content-type: text/html¥n¥n" ;
046: print "<!DOCTYPE HTML PUBLIC ¥"-//W3C//DTD HTML 4.01 Transitional//EN¥">¥n¥n"
;
047: print "<HTML lang=¥"ja¥">¥n" ;
048: print "<HEAD>¥n" ;
049: print "<TITLE>Result</TITLE>¥n" ;
050: print "<META http-equiv=¥"Content-Type¥" content=¥"text/html; charset=EUC-JP¥">¥n" ;
051: print "<LINK rel=¥"stylesheet¥" href=¥"../result.css¥" type=¥"text/css¥">¥n" ;
052: print "</HEAD>¥n¥n" ;
053: print "<BODY>¥n" ;
054: print "<H1>古代北方史関係研究文献目録データベース</H1>¥n" ;
055: print "<P>正式名称：「蝦夷(エミシ・エビス・エゾ)」を中心とした東北～北方史関係
研究文献目録データベース(稿)</P>¥n" ;
056: print "<HR>¥n" ;
057: print "<P><STRONG>Namazu による全文検索システム</STRONG></P>¥n" ;
058: print "<P>現在、$nmzhead[0]の文書がインデックス化され、$nmzhead[1]個のキーワ
ードが登録されています。</P>¥n" ;
059: print "<P><STRONG>インデックスの最終更新日:$nmzhead[2]</STRONG></P>¥n" ;
060: print "<HR>¥n" ;
061: print "<P>¥n" ;
062: print "<FORM method=¥"GET¥" action=¥"$nmzcgi¥">¥n" ;
063: print "<STRONG>検索式:</STRONG>¥n" ;
064: print "<INPUT type=¥"text¥" name=¥"query¥" size=¥"40¥">¥n" ;
065: print "<INPUT type=¥"submit¥" value=¥"Search!¥">¥n" ;
066: print "<INPUT type=¥"hidden¥" name=¥"whence¥" value=¥"0¥">¥n" ;
067: print "<!-- <input type=¥"hidden¥" name=¥"idxname¥" value=¥"foobar¥"> -->¥n" ;
068: print "<A href=¥"$nmzcgi¥">[検索方法]</A>¥n" ;
069: print "</P>¥n" ;
070: print "<P>¥n" ;
071: print "<STRONG>表示件数:</STRONG>¥n" ;
072: print "<SELECT name=¥"max¥">¥n" ;
073: print "<OPTION value=¥"10¥">10¥n" ;

```

```

074: print "<OPTION selected value=¥"20¥">20¥n" ;
075: print "<OPTION value=¥"30¥">30¥n" ;
076: print "<OPTION value=¥"50¥">50¥n" ;
077: print "<OPTION value=¥"100¥">100/¥n" ;
078: print "</SELECT>¥n" ;
079: print "<STRONG>表示形式:</STRONG>¥n" ;
080: print "<SELECT name=¥"result¥">¥n" ;
081: print "<OPTION selected value=¥"normal¥">標準¥n" ;
082: print "<OPTION value=¥"short¥">簡潔¥n" ;
083: print "</SELECT>¥n" ;
084: print "<STRONG>ソート:</STRONG>¥n" ;
085: print "<SELECT name=¥"sort¥">¥n" ;
086: print "<OPTION value=¥"score¥">スコア¥n" ;
087: print "<OPTION value=¥"field:subject:ascending¥">題名 (昇順) ¥n" ;
088: print "<OPTION value=¥"field:subject:descending¥">題名 (降順) ¥n" ;
089: print "<OPTION selected value=¥"field:uri:ascending¥">発表年度 (昇順) ¥n" ;
090: print "<OPTION value=¥"field:uri:descending¥">発表年度 (降順) ¥n" ;
091: print "</SELECT>¥n" ;
092: print "</P>¥n" ;
093: print "</FORM>¥n" ;
094: print "No.@input2 [0] " ;
095: print "<P>執筆者 : "; $word = @input1 [0] ; &link_word ;
096: if (@input2 [6] == "") { print " (@input2 [6]) " ; } # (ふりがな)
097: print "</P>¥n<H3>論文名 : @input1 [1] </H3>¥n" ;
098: print "<TABLE width=¥"100%¥" border=¥"2¥" summary=¥"検索結果 : No.@input2 [0] ¥">
¥n<TBODY>¥n" ;
099: print "<TR>" ; print "<TH width=¥"15%¥">掲載誌等<TD> " ; $word = @input1 [2] ;
&link_word ; print "</TR>¥n" ;
100: print "<TR>" ; print "<TH width=¥"15%¥">巻号<TD> @input2 [1] </TR>¥n" ;
101: print "<TR>" ; print "<TH width=¥"15%¥">刊行年月<TD> @input2 [3] </TR>¥n" ;
102: print "<TR>" ; print "<TH width=¥"15%¥">編・著<TD> @input2 [2] </TR>¥n" ;
103: print "<TR>" ; print "<TH width=¥"15%¥">出版社<TD> @input1 [5] </TR>¥n" ;

```

```

104: print "<TR>" ; print "<TH width=¥"15%¥">特集<TD> " ; $word = @input1 [3] ;
&link_word ; print "</TR>¥n" ;
105: print "<TR>" ; print "<TH width=¥"15%¥">別誌名等<TD> " ; $word = @input1 [4] ;
&link_word ; print "</TR>¥n" ;
106: print "<TR>" ; print "<TH width=¥"15%¥">所収・再録書名<TD> " ; $word = @input1
[9] ; &link_word ; print "</TR>¥n" ;
107: print "<TR>" ; print "<TH width=¥"15%¥">改題・改稿<TD> @input1 [10] </TR>¥n" ;
108: print "<TR>" ; print "<TH width=¥"15%¥">参照 No.<TD> " ; $num = @input2 [4] ;
&link_num ; print "</TR>¥n" ;
109: print "<TR>" ; print "<TH width=¥"15%¥">備考<TD> " ; $word = @input1 [6] ;
&link_word ; print "</TR>¥n" ;
110: print "<TR>" ; print "<TH width=¥"15%¥">キーワード<TD> " ; $word = @input1 [7] ;
&link_word ; print "</TR>¥n" ;
111: print "<TR>" ; print "<TH width=¥"15%¥">史料<TD> " ; $word = @input1 [8] ;
&link_word ; print "</TR>¥n" ;
112: print "<TR>" ; print "<TH width=¥"15%¥">対象地域<TD> " ; $word = @input1 [12] ;
&link_word ; print "</TR>¥n" ;
113: print "<TR>" ; print "<TH width=¥"15%¥">既成目録<TD> " ; $word = @input1 [11] ;
&link_word ; print "</TR>¥n" ;
114: print "<TR>" ; print "<TH width=¥"15%¥">版本データ No.控<TD> @input2 [7] </TR>¥n"
;
115: print "<TR>" ; print "<TH width=¥"15%¥">漢字注記<TD> @input2 [5] </TR>¥n" ;
116: print "</TBODY>¥n</TABLE>¥n" ;
117: print "<P> <A href=¥"../index.htm¥">【トップに戻る】 </A></P>¥n" ;
118: print "<HR>¥n" ;
119: print "<P>¥n" ;
120: print "この全文検索システムは <STRONG><a href=¥"http://www.namazu.org/¥">
Namazu</A>$nmzfoot [0] </STRONG>¥n" によって構築されています。" ;
121: print "</P>¥n" ;
122: print "<ADDRESS><IMG src=¥"./images/renraku.gif¥"></ADDRESS>" ;
123: print "</BODY>¥n" ;
124: print "</HTML>¥n" ;

```

```

125:
126: sub link_word {
127:   $word =~ s/ \| / \| \| /g ;
128:   @word = split (/ \| /,$word) ;
129:   @linkw = () ;
130:   foreach $i (@word) {
131:     $str = $i ;
132:     $str =~ s/ ([^¥w]) /%!.unpack ("H2",$1) /eg ;
133:     $str =~ tr/ +/ ;
134:     if ($i ne " \| ") { push @linkw,"<A href = ¥"$nmzcgi?query=$str$nmzgiop¥">$i</A>" ; }
135:     else { push @linkw,$i ; }
136:   }
137:   $linkw_join = join (" \| ",@linkw) ;
138:   $linkw_join =~ s/ \| \| \| \| / \| \| /g ;
139:   print "$linkw_join" ;
140: }
141:
142: sub link_num {
143:   $num =~ s/ \| / \| \| /g ;
144:   @num = split (/ \| /,$num) ;
145:   @linkn = () ;
146:   foreach $i (@num) {
147:     $folder = int ( $i / 1000 ) ;
148:     $folder03d = sprintf ("%03d",$folder) ;
149:     $file = $i % 1000 ;
150:     $file03d = sprintf ("%03d",$file) ;
151:     push @linkn,"<A href = ¥"$resultcgi?query=$folder03d-$file03d.htm¥">$i</A>" ;
152:     $linkn_join = join (" \| ",@linkn) ;
153:     $linkn_join =~ s/ \| \| \| \| / \| \| /g ;
154:   }
155:   print "$linkn_join" ;
156: }

```

157:

158: exit ;

--- (ここまで) ---

7) 補助的バッチファイルの作成

ezo_db_backup.bat を作成し、図44の tool フォルダに配置した。ダブルクリックすれば、図44の backup フォルダに今まで述べてきた各種設定ファイルのバックアップを取るようになっている。

```
< ezo_db_backup.bat >
@echo off
F:
cd F:\oguchi_local\ezo_db\backup
copy ..\namazu\template*.ja
copy ..\namazu\mknmzrc
copy ..\namazu*.log
copy ..\tool\pre.pl
copy ..\tool*.bat
copy ..\ezo_db.bat
copy ..\*.bat
copy ..\..\oguchi\search\amazurc
copy ..\..\oguchi\search\result.cgi
copy ..\..\oguchi*.htm
--- (ここまで) ---
```

また、その逆の動作をする、ezo_db_resore.bat も同じフォルダに作成して配置した。

```
< ezo_db_resore.bat >
@echo off
F:
cd F:\oguchi_local\ezo_db\backup
copy *.ja ..\namazu\template
copy mknmzrc ..\namazu
```

```
copy pre.pl ../tool
copy ezo_db_backup.bat ../tool
copy ezo_db_restore.bat ../tool
copy ezo_db.bat ..
copy .namazurc ../oguchi/search
copy result.cgi ../oguchi/search
copy *.htm ../oguchi
--- (ここまで) ---
```

8) 未処理の課題

- ・外字処理…今昔文字鏡番号での画像リンクを挿入する機能を result.cgi に追加する⁽¹⁾。
- ・データ項目の変化…4) a の段階で各データに固有 ID を付加し、さらにこれを参照 No. と同期させる。pre.pl と result.cgi 双方に手を入れる必要がある。
- ・その他いくつかのスキプトのバグを解消する。
- ・HTML のレイアウトやデザインを調整する。スタイルシートの使用を検討中である。
- ・またデータベース上の書籍類の自動検索先として法政大学図書館と国会図書館を設定して、それらのボタンをクリックすると所蔵情報が表示されるようにしてあるが⁽²⁾、この仕組みの改良も検討中である。
- ・本稿 1 で触れたように Kakashi のデータ上の語句の切り分けの設定が最大の難関である。どう設定しても一長一短という状況で、悩ましいところであるが、とりあえず現状の設定でユーザーから特段の苦情は来ていない。今後の課題としたい。

4、おわりに

以上、木簡庫の研究文献情報のリンク先である法政大学国際日本学研究所におけるデータ処理について紹介してみた。

データベースは更新されてこそ意味があるので、引き続き作業を継続していきたい。な

(1) その後、対応済みである。

(2) これまでリンク先の検索システムが変更になって、こちら側で急にトラブルが生じることもあったが、今のところ正常に動いているようである。

おサーバー上の複数のソフトの相性の問題もあって、このまま Namazu を使い続けるのかどうかも今後検討の俎上になってこよう。将来的には専門的な技術を持った業者との関係も深めていく必要がある。

また木簡庫との連携では、どの研究論文がどの木簡とつながるのかという大前提が、マンパワーによるしかないという状況下で、いかに効率よく結果を出せるかが勝負になるろう。

論文データを透明テキスト付 PDF にして蓄積することができればいいのであるが、これはもはや私どもが対応できるレベルを超えている。今後このデータをどのように拡充していくべきかについても検討が必要になってきている。