

『万葉集』諸本集成画像データベースシステムの 構築と意義

吉村 誠・中田 充

The Construction and Usefulness of an Image Database System for Manyo-Shu Books

Makoto YOSHIMURA and Mitsuru NAKATA

(Received September 26, 2003)

1. はじめに

国文学研究においてもコンピュータを利用する時代に入ってきたと認識してよいであろう。現在普及しているコンピュータにおいては、文章解析にまで迫る能力を持ってきてはいるものの、研究者のスキルや使用方法によって様々な形態になっている。もっとも国文学の方法に迫る利用形態の考察はそれ自身で新しいジャンルの開発につながるものとして期待出来るが、一方で従来の作業を改良する手段としても期待されなければならない。情報処理や情報通信といったエンドユーザーとして持つコンピュータの機能は、研究支援機器として多く利用する手段は残されているからである。

データを活用した情報処理支援機器としての扱いは、具体的なデータともども詳しくまとめられたものもあり（『電脳国文学』漢字文献情報処理研究会 2000.10 好文出版）、また『万葉集』を中心としては坂本信幸氏がまとめておられるが（『国文学別冊』No.55 02.11 学燈社）、まだ報告者の利用程度と認識も様々なレベルにあるために、方法的な基本ラインの一致を見ない。

現在の所その利用方法を整理してみると以下のような点にしばられる。

- ①手作業では不可能な大量のテキスト処理を行い、検索や類似表現の抽出などを行う。
 - ・テキスト検索 SQL文による本格的なデータベース作成から文字列一致を中心としたプログラム作成を基本としたもの。プログラムやデータは汎用性があるために広く公開されるものが多い。
 - ・ngram や XML を用いたデータベース化により、類似表現を比較する。（竹田正幸「古典和歌からの知識発見」『情報処理』Vol 43 No.9 02.09）
 - ・構文解析や計量的なテキスト解析のために、特別のプログラム作成によって行う検証方法を用いる。
- ②原本を画像データやテキストデータとしてインターネットを通して公開する。
 - ・諸写本を画像データとして公開する。
 - ・機械可読テキストを作成して公開する。

基本的には、インターネットの発達とともに、②のように情報支援機器としての利用が多くなってきているというのが現状である。

国文学において基本となるものは原本テキストである。従って依拠本文が極めて重要なもの

となることは言うまでもない。そのために原本を手軽に確認することの出来る原本画像データベースシステムの作成公開は極めて意義深いものと言える。しかし、インターネット上で公開されている多くの既存のシステムは画像の羅列のみにとどまっており、データベース技術を利用したものは少なく、検索機能などにおいて問題がある。『万葉集』に限っても、上記坂本信幸氏が中心となって作成・公開されている万葉集画像プロジェクトもそれ自身は意義深いものであるが (http://ut-raq.manyou.gr.jp/SMAN_1/)、和歌番号から検索出来るだけであり、語句検索等の機能を持たない。より充実した検索機能を提供するなど、コンピュータの機能を最大限発揮する形態を考える必要があると思われる。また国文学研究資料館の中村康夫氏が中心となって岩波書店よりその成果を公開されている原本データベースの作成事業は、国文学研究上で重要な写本とその翻刻文をリレーションして検索機能をつけた新しい視点に立つものであるが、特定の写本を対象としたものであり、同一作品で複数にまたがる写本間の比較が出来ない。

そこで本稿では、従来の国文学研究の基本的作業でもある原本確認をコンピュータで支援するための画像データベースシステムに関して、その目的や具体的な実現手法を述べて今後のあり方を検討したいと思う。なお本来、システムの対象は広く国文学の古典作品全般におよぶが、本稿では「万葉集」を対象とする。「万葉集」の本文は漢字のみの原文（表題、題詞、歌、左注、割り注、脚注）と、歌本文は平仮名で記された付訓を持ち、複雑なテキスト構成を持っているために、コンピュータの能力を利用するのに適していると思われるからである。

2. 画像データベースシステム構築の目的

画像データベースシステムを用いることによって、従来のように本を開いて見るという以上の付加価値を付けることが出来る。単に写本画像ファイルの羅列では、机上で原本を開いてみるということと大きな相違はないからである。それはコンピュータの持つ特性を十分に活用したシステム構築を行うことによって実現出来る。付加価値的な具体的形態としては、

- ①特定の語を写本間で比較する。
- ②特定の原文文字に対して、同一本の他の箇所や別本の文字を比較する。
- ③修正文字やミセケチなどの細かい箇所を同一本または他の写本によって比較する。

といった作業が可能となることが挙げられる。特に『万葉集』では、歌本文の場合、原本の文字だけでなく訓の問題も派生するため、上記に加えて、

- ・同一の文字（文字列）に対する訓の異同
- ・同一の訓に対応する原文文字（原文文字列）の異同と写本間の異同などの比較検討が要求されるところである。

画像データベースシステムによる付加価値の一例として「冬こもり」という枕詞に関する検討を掲示する。「冬こもり」の用例と写本間の異同は以下のような状況にある。近世の古注及び現代注釈書の一部も掲げる。

巻 / 歌番号 [諸本略称] 原文 訓読

01/0016 [元][西][冷][細][温][矢][京][寛]冬木成フユコナリ

[紀]冬木成フユキナリ「木」の左に朱「コ」

[代初]冬木成ふゆきなす

[僻]冬木成ふゆこもり

[考]冬木盛ふゆこもり

[攷][お]冬木成ふゆこもり

- 02/0199 [全註釈][大系]冬木成ふゆごもり
 [西][細][類][矢][京][寛]冬木成フユコナリ
 [紀]冬木成フユキナリ
 [温]冬木成フユコキ
 [矢][京]冬木成フユコナリ(リ青書)
 [代初]冬木成ふゆきなす
 [考]冬木盛ふゆこもり
 [童][お]冬木成ふゆこもり
 [全註釈][大系]冬木成ふゆごもり
- 03/0382 [西][細][温][類][矢][京][寛]冬木成フユコナリ
 [紀]冬木成フユキナル(左朱フユコナリ)
 [代初]冬木成ふゆきなす
 [考]冬木盛ふゆこもり
 [お]冬木成ふゆこもり
 [全註釈][大系]冬木成ふゆごもり
- 06/0971 [元][細][温][類][寛]冬木成フユコナリ
 [紀]冬木成フユキナリ
 [西]冬〇成フユコナリ(〇木右書、コ朱書)
 [矢][京]冬木成フユコナリ(コ青書)
 [考]冬木成ふゆこもり
 [略]冬木盛ふゆこもり
 [お]冬木成ふゆこもり
 [全註釈][埴]冬木成ふゆごもり
- 07/1336 [元][紀][西][細][温][類][矢][京][寛]冬隠フユコモリ
 [お]冬隠ふゆこもり
 [全註釈][大系]冬隠ふゆごもり
- 09/1705 [紀][西][細][類][矢][京][寛]冬木成フユコナリ
 [藍][壬]冬木成フユキナリ
 [西][細]冬木成フユコナル
 [温]冬木成フユコナリ(成左ナル)
 [代初]冬木成ふゆきなす
 [童]冬木成ふゆこもり
 [略]冬木盛ふゆこもり
 [お]冬木成ふゆこもり
 [全註釈][大系]冬木成ふゆごもり
- 10/1824 [紀][西][細][温][類][矢][京][寛]冬隠フユコモリ
 [元]冬隠フユコモリ(右赭くれて)
 [お]冬隠ふゆこもり
 [全註釈][大系]冬隠ふゆごもり
- 10/1891 [紀][西][細][類][矢][京][寛]冬隠フユコモリ
 [お]冬隠ふゆこもり

- [全註釈][大系]冬隠ふゆごもり
 13/3221 [元][紀][西][天][細][温][矢][京][寛]冬木成フユコナリ
 [西]冬不成フユコナリ
 [類]冬木盛フユコナリ(右墨書モリ)
 [童]冬木成ふゆごもり
 [考]冬木盛ふゆごもり
 [お]冬木成ふゆごもり
 [全註釈][大系]冬木成ふゆごもり

この例は、それぞれの写本の訓異同と現在の訓の確定過程を見ることが出来る。『万葉集』の全使用例を比較出来、「冬隠」と原文のある用例から「フユコモリ」の訓の正しいことが知られるが、「冬木成」とある原文のために「フユキ(コ)ナス(ル)」という訓が付され、誤字説を経過して現代の訓に到達していった様子を詳細にとらえることが出来る。

対象伝本は略号で示している。写本の全てを掲げているわけではないが、対象写本に異動があるのは、抄本も含まれているからである。写本の書誌的詳細はここでは省略するが、名称は以下のとおりである。

- [藍] 藍紙本
 [元] 元暦校本
 [紀] 紀州本
 [西] 西本願寺本
 [天] 検天治本
 [壬] 壬生本
 [冷] 冷泉本
 [細] 細井本
 [温] 温故堂本
 [矢] 大矢本
 [京] 京大本
 [寛] 寛永版本

以下注釈書

- [代初] 代匠記初稿本 契沖
 [童] 童蒙抄 荷田春満(門下生)
 [考] 万葉考 賀茂真淵(門下生)
 [略] 万葉集略解 加藤千蔭
 [全註釈] 万葉集全註釈 武田祐吉
 [大系] 岩波古典文学大系「万葉集」

校訂本

- [お] 万葉集 おうふう 鶴久、森山隆

なお、ここに掲げている諸本は全てではない。例示の都合上割愛した。これらの比較検討は、異本間の文字異同を確認する従来の作業に加えて、同一本の同一文字の異同やその特定の文字の異本間での確認も行うことになり、立体的なテキスト情報を短時間で得られるという特性を持つことになる。それは単に文献学や書誌的な研究目的ばかりでなく、テキストを読んでいく上で本文定位の基礎的なシミュレーションを容易に行うことが出来、より緻密なテキスト

作成に重要なプロセスを与えることになり、解釈上の重要な手がかりを得る可能性もある。

3. 画像データベースシステムの仕様

周知のように『万葉集』の本文と訓読文は、二十種ばかりの写本（抄本も含む）と江戸初期の普及本である『寛永版本』（『宝永再刻本』も含む）を原本とし、中世以降現代までの数十冊に及ぶ注釈書による考証を経たものが存在する。本稿では『寛永版本』を含む写本におけるテキストを対象とするが、検索の至便性を考慮する上で検索対象の本文は、現代までの注釈書における校訂本文や訓読文も含める。このことにより現代の訓読を手がかりにして、写本間のテキスト、訓読の異同を確認することが可能となる。以下は、画像データベースシステムが対象とするデータをその種類別に挙げたものである。

検索対象は、すべての諸本（写本、版本、校訂本、注釈書）

検索対象とするデータ

歌番号

原文（表題、題詞、歌本文、割り注、脚注、左注の全て）の原字

訓読平仮名（歌本文、割り注、脚注、左注の異伝歌の訓読）の語句

翻案訓読表記の語句

検索結果として提示するデータ

写本の画像データ（複数）

翻刻のテキストデータ（複数）

検索対象とするすべてのデータは、異訓データファイルと呼ぶ一つのファイルに格納する。そして検索結果として提示するデータは、各諸本の歌番号を単位として一つのHTMLファイルとして格納する。

具体的に検索対象とするデータの一例を以下に記す。これらのデータの構成やその内容は、拙作の「manyou.txt Ver4.3 R 1.6」(<http://yoshi01.kokugo.edu.yamaguchi-u.ac.jp/manyou/>)をベースにしている。

#[番号] 01/0010

#[題詞] 中皇命往于紀温泉之時御歌

#[原文] 君之齒母 吾代毛所知哉 磐代乃 岡之草根乎 去来結手名

#[訓読] 君が代も我が代も知るや岩代の岡の草根をいざ結びてな

#[仮名] , きみがよも, わがよもしるや, いはしろの, をかのくさねを, いざむすびてな

#[左注] (右檢山上憶良大夫類聚歌林曰 天皇御製歌 [云々])

#[対象] [元][冷][古][寛][お][塙][全註釈][注釈]

#[校異] 磐 [元][冷][古] 盤

#[訓異] きみがよも [お][塙][全註釈][注釈], [元][寛] きみかよも, [冷] きみかはも, わがよもしるや [お][塙][全註釈][注釈], [元][冷] わかよもしらす (右朱しれや), [寛] わかよもしれや, いはしろの [元][冷][寛][お][塙][全註釈][注釈], をかのくさねを [元][寛][お][塙][全註釈], [西][文][細][注釈] をかのかやねを, いざむすびてな [お][塙][全註釈][注釈], [元][冷][寛] いさむすひてな,

#[対象]というのは、#[校異]、#[訓異]における参照諸本である。例示しているのは現代注釈書も含めて数本に過ぎないが、実際には数十種類に及ぶものである。また『万葉集』には全巻にわたっての歌の記されていない抄本も存在するので、個々の歌データは、どの本に含まれているのかというリストが必要になる。例えば、『桂本万葉集』は巻四の一部しか伝わっていないので、ここでは対象外となる。また#[訓異]における各写本の訓読体裁は一様ではない。上記例においても[元]（『元暦校本』）においては、「をかのくさねを」訓の原文「草根」の右に朱筆で「かやね」という訓がほどこされていたり、[西]（『西本願寺本』）においても、原文「草」の右訓字は消した跡があり、さらに原文の左側に「カヤイ」という訓が書かれている（図1）。しかし検索対象文字列の目的は写本の翻刻ではないので、文字情報を忠実に写し取るのではなく、訓読の相違が明確に現れるように記載することを主眼とする。そうした写本の細かい体裁は画像で直接確かめればよいであろう。

検索対象とする文字列は、原文の場合は、#[原文]の原字文字列と#[校異]の原字一字単位、訓読文の場合は、#[訓異]の句の範囲内による語句文字列になる。このことにより、『万葉集』の全ての写本、校本、注釈書の校訂本文の原字及び、訓読が含まれることになり、どのような原文訓読からでも検索が可能となる（ただし、現在のところ原字異同の場合の文字列、および異訓の句をまたがった文字列検索は出来ない）。検索に用いるインターフェースの仕様は図2である。ここでは「ひとごと」という文字列を検索した例で、Webベースのものを示す。

そして、検索結果1次ファイルに、画像データをハイパーテキスト的にリンクした内容のものを提示する。図3である。#[対象]に写本の略号を掲げ、それぞれにリンクしている。#[原文]、#[仮名]は、現在公開中の拙作 manyou.txt Ver4.3 R1.6 による校訂テキストである（公開URLは前掲）。

それぞれの写本における一例を図4～5に示す。この図は、図3に表示されている「04/0538番歌」の#[対象]に[桂]とある、『桂本万葉集』の当該歌と[西]とある『西本願寺本』を掲示したものである。それぞれの写本のファイルは、歌一首を基本単位とし、翻刻本文を右に掲げ、画像を置いているものである。縦置きを原則とする。1行目には冊子体の場合は、所在の丁数と行数、国歌大観番号を整理番号として記述する。翻刻本文は、ミセケチ、訂正、墨色（黒以外のもの。朱、藍等）、付箋等の情報もなるべく原本通りの位置に書き込む。例示した西本願寺本の場合は、朱の返り点1箇所（「不」と「相」の間）を示している。そして「子」は別筆の可能性があるので（）で囲む。また原文文字で標準漢字コード以外のは、今昔文字鏡（文字鏡研究会篇 製品化したものは紀伊国屋書店より検索ソフト付で販売。freeのものは、『パソコン悠々漢字術』文字鏡研究会 02.04 エーアイネット等に収載されている）によるコードで記述する。そしてファイル名は、写本略号アルファベットと歌番号で示す。今の例で述べると、ファイル名は「katura_04_0538.html」、或いは、「nisi_04_0538.html」というファイルを開覧していることになる。



図 1

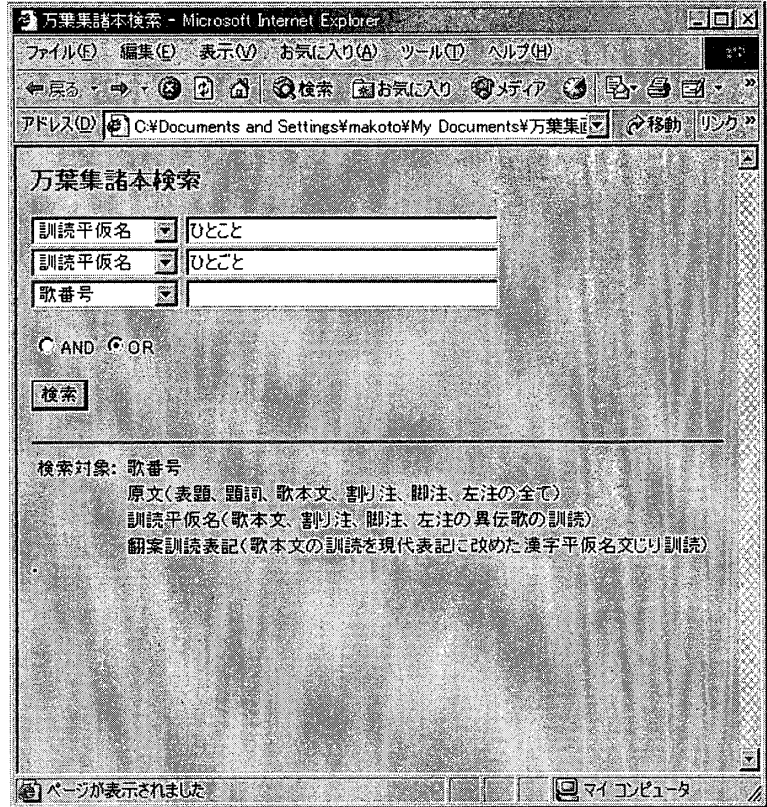


図 2

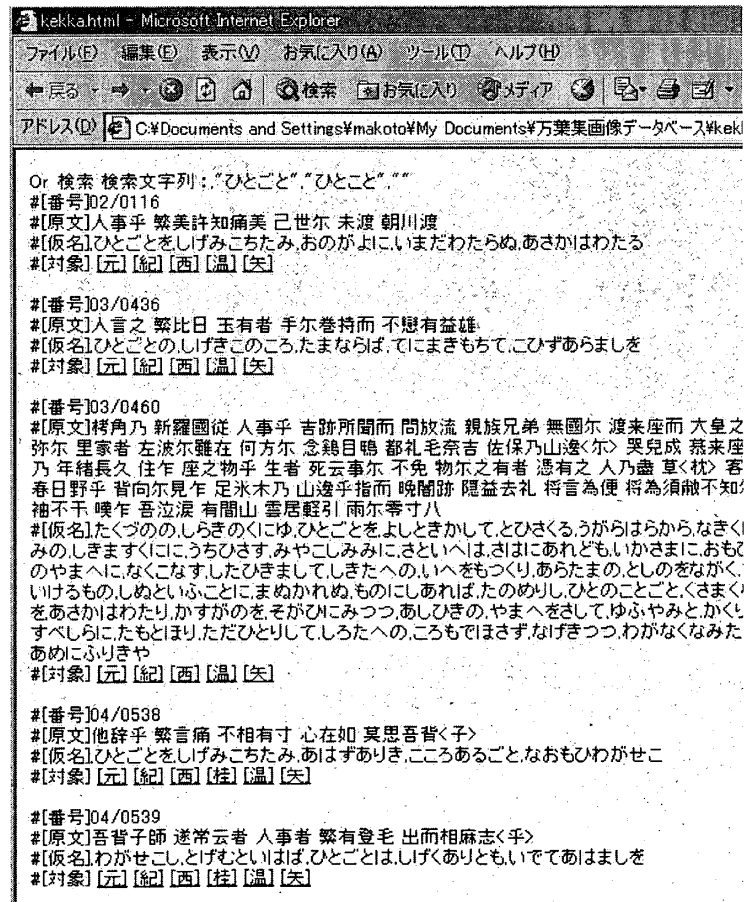


図 3

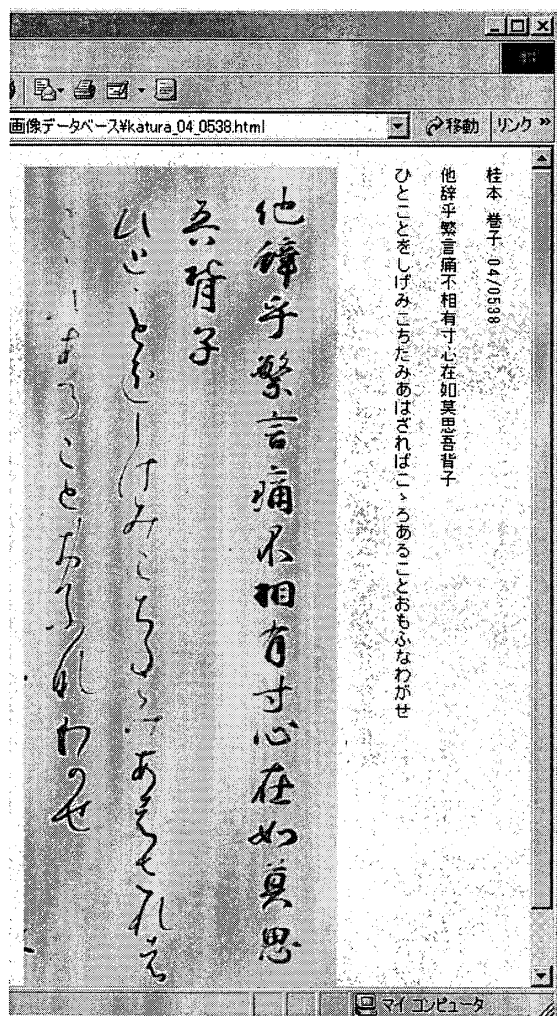


図 4

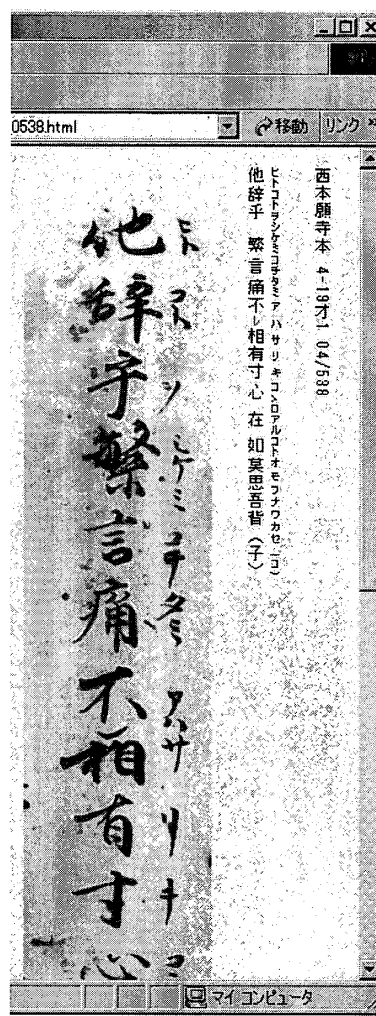


図 5

4. 利用環境と写本揭示プログラム手順

画像データベースの利用形態は、Windows 上でのアプリケーションプログラムによるものと、ネットワーク環境での Web ベースによるものとの両者の方法が考えられる。上記例では Web ベースのものを掲げたが、両者ともそれぞれ長短があるからである。Windows 上のアプリケーションの特性としては、プログラムも含めてすべてのデータを手元に置くことが出来、いつでも利用出来る。しかしデメリットとしては、供給側の問題として、改訂版が生じた場合はその再配布が二度手間になる。また画像ファイルであるので、かなり大きなファイル容量となる。鮮明に画像を見せるためには最低でも一首あたり20Kb ぐらいの大きさが必要であり、全巻にすると単純計算で90Mb が1本となる。10種類の写本があるとして900Mb の容量が必要となる。現在の所、供給方法は CD メディアによる配布とネットワークによるダウンロード方式の2方法が考えられるが、たとえ高速ネットワーク回線が普及しているとはいえ、ダウンロードには時間的にかかなりの負担になるであろう。

一方で Web ベースのネットワークによる検索システムの利用は、クライアント側(利用者側)に膨大な量のファイルを置く場所は不要であり、提供側の更新も容易であるが、ネットワークや供給サーバのコンディションに影響され、トラブルのある場合は、利用作業に影響が出る。

供給側の検索システムとしては、Web ベースのネットワークで利用する形態のものとしては、Unix や Linux サーバ上での Perl や Java によるプログラミングのシステムであり、アプリケーションでの供給は、VisualBasic や Java による作成が可能である。

いずれにしても検索システムを構築する上でのプログラム手順は同じであるので、以下にその手順を示す。

1. 検索文字列入力
2. 異訓データファイルを参照（ファイルの内容構成の一例は第3章を参照）
3. 検索文字列対象として、資料名を示す [] 内の文字を検索対象からはずす（現在のところ、上記ファイル内容においては、原文については、校訂本文を対象とした場合は、文字列検索が可能であるが、文字異同を対象とする場合は、一字単位の検索となる。また訓読平仮名の仮名語句においては、句をまたいでの検索は出来ない。今後の改良点となる。）
4. 検索文字列と検索対象文字列が一致した場合は、検索結果ファイルが生成され、それぞれの写本情報は、個々の写本画像とリンクされて掲示される。

という手順になる。例として、Web ベースにおける検索結果ファイルのソース（上記図3）は、次のようなものである。

```
#[ 番号 ]04/0538<br>
#[ 原文 ] 他辞乎 繁言痛 不相有寸 心在如 莫思吾背子 <br>
#[ 仮名 ],ひとごとを,しげみこちたみ,あはずありき,ころあるごと,なおもひわがせこ <br>
#[ 対象 ]
<a href="gen_04_0538.html">[ 元 ]</a>
<a href="ki_04_0538.html">[ 紀 ]</a>
<a href="nisi_04_0538.html" target="new">[ 西 ]</a>
<a href="katura_04_0538.html" target="new">[ 桂 ]</a>
<a href="on_04_0538.html">[ 温 ]</a>
<a href="ya_04_0538.html">[ 矢 ]</a>
<br><br>
```

検索条件に適合した歌番号情報を利用して、#[対象] の写本略号にリンク情報を提供するという作業を行って作成する。この部分の処理を Perl 言語で示すと以下のようなになる。

```
#!/usr/bin/perl

$no = "04/0548";          # 一例として巻4、548番歌を処理する；
$shon = "[ 桂 ][ 元 ][ 西 ][ 紀 ][ 細 ]"; # 当該歌には左記の諸本がある；

@bangou = split /\//,$no; # 歌番号を巻と番号に区分する；
$maki = $bangou[0];
$uta = $bangou[ 1 ];

@syahon = split /\N/,$shon; # 諸本情報を個々に区分する；
```

```

$kazu = split /\N/, $hon;
print "#[ 対象 ]\n";
for ($a=0;$a<$kazu;$a++) {
    if ($syahon[$a] =~ / 桂 /) {           #それぞれの諸本情報をもとにリンク先ファイル名;
        $name="katura";                 #に変換する;
    }
    elsif ($syahon[$a] =~ / 元 /) {
        $name="gen";
    }
    elsif ($syahon[$a] =~ / 西 /) {
        $name="nisi";
    }
    elsif ($syahon[$a] =~ / 紀 /) {
        $name="ki";
    }
    elsif ($syahon[$a] =~ / 細 /) {
        $name="hosoi";
    }
# リンクのタグを付ける;
    if ($a == 0) {
        print "<a href=\"\".$name.\"_\".$maki.\"_\".$uta.\".html\">\".$syahon[$a].\"</a>\n";
    }
    elsif ($a == $kazu-1) {
        print "<a href=\"\".$name.\"_\".$maki.\"_\".$uta.\".html\">[\".$syahon[$a].\"</a>\n";
    }
    else{
        print "<a href=\"\".$name.\"_\".$maki.\"_\".$uta.\".html\">[\".$syahon[$a].\"</a>\n";
    }
}

```

以上の具体例は、スクリプト型のプログラミング言語を用いた Web サーバ上での検索システム構築（利用はネットワークを介して行う）の部分例であるが、同様にサーバ上で構築するならばデータベース管理システムを用いた方が検索速度やより緻密な文字列検索を行うことが出来るであろう。そこで次にデータベース管理システムを用いた例を掲げる。

5. データベース管理システムを用いたシステム構築

前節で述べたシステム構築例では、検索対象となる和歌データを、ファイルシステムを用いて管理している。そのため、複数の検索条件を指定した検索の実現が困難であったり、実現したとしてもその検索時間が長くなるといった問題が生じると考えられる。また、特定の底本に「訓異」などの差分情報を付加することにより、すべての諸本のデータを表現しているため、

底本以外の情報を検索する際に時間がかかるという問題点もある。これらを解決する方法として、和歌データをデータベース管理システム（以降、DBMS）を用いて管理することが挙げられる。本稿では DBMS として、現在最も一般的に用いられているリレーショナル DBMS を用いたシステム構築について検討する。

基本方針として全ての諸本の和歌データを句、題詞、左注に分けてデータベースに格納する。和歌を句に分割して格納することで諸本の相違を句単位で検索可能となる。また、特定の底本の差分の形ではなく、すべての諸本のデータを同等に格納する。これにより、特定の情報のみ検索に時間がかかるといった問題点を解決することが可能となる。

さらに、3節における歌データ中の「天皇御製歌 [云々]」などに表されるような、和歌や題詞、左注に付加された様々な注を、付加対象の和歌、題詞、左注から分離して別途管理する。これにより、注に関わる検索も実現が容易になる。なお、これらの注は、全体注あるいは部分注の2種類のみで大別して格納する。これは注の種類が多様なため、全ての種類の注を個別に管理することが困難なためである。

具体的には句、題詞、部分注などのデータは、それぞれ個別のリレーショナル DBMS のテーブルに管理する。下記は全ての和歌データを管理するために必要なテーブルの一覧である。ここではテーブル「本」、「和歌」、「句」、「和歌-題詞関連」についてのみ述べる。

- (1) 本 (本 id, 名前, 備考)
- (2) 和歌 (和歌 id, 巻番号, 和歌番号, 和歌画像ファイル名, 本 id)
- (3) 句 (句 id, 番号, 原文, 仮名, 訓読, 訓読ルビ付, 和歌 id)
- (4) 題詞 (題詞 id, 至近和歌 id, 番号, 原文, 訓読, 訓読ルビ付)
- (5) 左注 (左注 id, 至近和歌 id, 番号, 原文, 訓読, 訓読ルビ付)
- (6) 全体注 (全体注 id, 至近和歌 id, 名前, データ)
- (7) 部分注 (部分注 id, 至近和歌 id, 名前, データ, 開始句番号, 終了句番号, 開始 offset, 終了 offset)
- (8) 和歌-題詞関連 (題詞 id, 和歌 id)
- (9) 和歌-左注関連 (左注 id, 和歌 id)
- (10) 全体注-和歌関連 (全体注 id, 和歌 id)
- (11) 全体注-題詞関連 (全体注 id, 題詞 id)
- (12) 全体注-左注関連 (全体注 id, 左注 id)
- (13) 部分注-句関連 (部分注 id, 和歌 id)
- (14) 部分注-題詞 (部分注 id, 題詞 id)
- (15) 部分注-左注関連 (部分注 id, 左注 id)

(1)は「本」を表すテーブルであり、本の識別子「本 id」、本の「名前」、「備考」を属性として持っている (図6参照)。「本 id」とは本の識別子であり、これにより複数存在する本を識別する。備考には本の種別、著者、書かれた年代、現代の書籍であれば出版社などの情報が含まれている。

本 id	名前	備考
3399020	万葉集注釈
3399021	西本願寺本
.....

図6

(2)は「和歌」を表すテーブルであり、識別子「和歌 id」、和歌が書かれている巻の「巻番号」、国歌大観番号である「和歌番号」、和歌の画像を格納したファイルの名前、そして、和歌が属する本の「本 id」を属性として持っている(図7参照)。「本 id」により、それぞれの本がどのような和歌を含むかを管理している。この「本 id」が存在するため、同じ「巻番号」、「和歌番号」を持つ行が複数存在することになる。

和歌 id	巻番号	和歌番号	画像ファイル名	本 id
.....
3399026	1	4	3399026.gif	3399020
3399027	1	5	3399027.gif	3399020
.....

図7

(3)は「句」を表すテーブルであり、和歌の情報を句ごとに分割して格納している。属性として和歌における句の順番を示す「番号」、そして「原文」、「仮名」、「訓読」、「訓読ルビ付き」、句が含まれる和歌を示す「和歌 id」を持っている(図8参照)。なお、「訓読ルビ付き」は「訓読」中の難解な漢字の読みを付加したものであり、結果表示の際の利便性を考慮して付けられた属性である。また「和歌 id」を持つために、「和歌」と同様に同じ和歌の同じ句でも本ごとに行が作成される。テーブル「題詞」、「左注」は属性「和歌 id」を持たない点を除いて、テーブル「和歌」と同様である。

句 id	番号	原文	仮名	訓読	訓読ルビ付き	和歌 id
.....
3400211	1	玉尅春	たまきはる	たまきはる	たまきはる	3399026
3400212	2	内之大野尔	うちのおほのに	内の大野に	内の大野に	3399026
3400213	3	馬藪而	うまなめて	馬なめて	馬なめて	3399026
3400214	4	朝布麻須等六	あさふますらむ	朝踏ますらむ	朝踏ますらむ	3399026
3400215	5	其草深野	そのくさふかの	その草深野	その草深野(ふかの)	3399026
3400216	1	霞立	かすみたつ	霞立つ	霞立つ	3399027
3400217	2	長春日野	ながきはるひの	長き春日の	長き春日の	3399027
.....

図8

(8)のテーブル「和歌-題詞関連」は、和歌と題詞の対応付けを行うためのものであり、その行は関連のある和歌と題詞の識別子の組である。これにより、和歌と題詞の関連を表現することが可能となる。(9)のテーブル「和歌-左注関連」も同様である。題詞や左注は必ずしも一つの和歌にかかるとは限らず、一つの和歌が複数の題詞や左注を持つことも考えられるが、これらのテーブルにより、そのような複雑な題詞や左注と和歌の関連も表現可能となる。

このように、和歌データを句、題詞、左注などのテーブルに分けてDBMSを用いて管理することにより、和歌の検索の際に、複数の条件を組み合わせたような複雑な問合せを実現することが容易になる。さらに、題詞や左注と和歌の関連、あるいは、割り注などの様々な注と句(あ

るいは和歌)との関連が表現可能となるため、万葉集の構成の把握が容易になるという利点が見られる。なお、現状では、先に挙げた16のテーブルを作成し、その中に約4600首の画像データを除いた和歌データを格納した段階である。

6. まとめ

以上のように技術的な問題はほぼ解決していると言ってよい。しかし画像の公開においては所蔵権の問題が生じる。上記坂本信幸氏を代表とする Web ページにおいても、画像公開権利上、会員登録を用いた閉鎖的なものであり、部分公開となっている。たとえその利用目的が研究、教育であったとしても所蔵者でない限り、インターネット上での画像公開や配布は所蔵者の許諾条件にゆだねられ、制約を受けるものとなっている。このような理由により、上記のようなシステムを構築したとしても、具体的な公開にあつては、システムと翻刻ファイルのみの公開しか出来なく、実際の画像は、個人使用目的で各ユーザーが保有している影印本等から画像ファイルを個々に作成し、それとリンクして利用するという方法しか今のところ考えられない。システム作成者が、画像作成のマニュアルとリンクページを表示するシステムまでを提供し、画像本体は、利用者個々の所有するものを使用するということである。

今後、こうした法的処置が利用者側の立場から整備されて、さらに発展的な利用が可能になることを期待したい。

なお、本稿で論じた画像データベースは、画像ファイルの権利上の都合により未公開である。

謝辞：本稿は、科学研究費補助金 基盤研究(C) (課題番号：C15500071) の研究成果の一部である。