柔軟な構造を持つデータベース管理システムを用いた万葉集検索システムの構築法

On Constructing a Retrieval System of MANYO-SYU based on a Semi-structured Database Management System

中田 充[†], 桑本龍也[††], 葛 崎偉[†], 吉村 誠[†]
Mitsuru NAKATA[†], Tatsuya KUWAMOTO[††], Qi-Wei GE[†], and Makoto YOSHIMURA[†]
[†]山口大学 教育学部, 山口県山口市吉田 1677-1
[††]山口大学大学院 教育学研究科, 山口県山口市吉田 1677-1
[†] Faculty of Education, Yamaguchi University, 1677-1 Yoshida, Yamaguchi-shi, Yamaguchi
[††]Graduate School of Education, Yamaguchi University, 1677-1 Yoshida, Yamaguchi-shi, Yamaguchi

あらまし：本稿では，柔軟な構造を持つデータベース管理システムを用いた万葉集検索システムの構築について述べる．近年，国文学研究においてもコンピュータが積極的に利用されており，古典文学作品データベースに関する研究も数多くなされている．筆者らは，日本最古の和歌集である万葉集の検索システムを構築し公開している．しかし，このシステムには，異同，異訓，注などの検索が出来ないという問題がある．これは，万葉集が多くの相違を持つ複数の本によって伝えられ，その中に多種多様な注が付されているために，万葉集の完全なモデル化が難しいためである．本研究では，万葉集のデータを構造の特定が困難な半構造データと捉え，半構造データを管理可能なデータベース管理システムを用いることにより，この問題の解決を図る．まず，万葉集の特徴とその検索システムに必要とされる機能について述べる．その後，半構造データのためのデータモデルを用いた万葉集検索システムの構築について述べる．

Summary: In this paper, we discuss on constructing a retrieval system of MANYO-SYU based on a Semi-structured Database Management System. Recently, computer technologies have been used actively in the research of Japanese literature. And many kinds of database systems for Japanese classical literature have been studied and realized. We have made a retrieval system of MANYO-SYU, which is the oldest Japanese Poetry collection and is bequeathed by many books that have differences between each other. However, currently our system cannot retrieve the information about notes and the differences in kanji characters or kana characters, because many differences between each books and variety notes prevent us from defining a complete model of MANYO-SYU. In this paper, we treat MANYO-SYU's data as semi-structured data and adopt a semi-structured data model to solve this problem. Firstly, we give an outline of MANYO-SYU and then discuss required functions for a retrieval system of MANYO-SYU. Finally, we explain our method of modeling MANYO-SYU.

キーワード：万葉集，万葉集検索システム，半構造データ，データモデル
Keywords：manyo-syu, manyo-syu retrieval system, semi-structured data, data model

1. はじめに

近年，国文学研究においてもコンピュータが積極的に利用されている．その利用には，作品の電子化としてのツール，すなわち，データベース技術を用いた検索システムとしての利用や，情報の加工・分析のためのツールとしての利用がある．古典文学研究では，研究対象の文学作品はもとより，関連する作品を含む多くの作品

を参照する必要がある．しかし，一つの古典文学作品に対して内容が異なる複数の本が伝わっていることが多く，作品の参照に時間がかかり研究者の負担となっている．また，電子化された作品に対してコンピュータを用いた分析を行うことで新たな知識の発見も期待される[1],[2]．このような背景のもと，古典文学作品のデータベース化とその利用に関する研究[3]-[6]が多くなされており，DVD-ROM 等の形式の古典文学作品検索システムも提供されている．

筆者らは，数多くの古典文学作品のうち，万葉集を対象とした検索システムを構築し，インターネット上で公開している．後述するこのシステムは，原文や仮名などの情報から和歌を検索可能であるが，万葉集に付された様々な注や異訓，異同の検索機能は十分に提供できていない．これは，万葉集が多くの相違を持つ複数の写本により伝えられていることと，様々な形式の注が付されていることにより，そのデータを完全に表現可能なデータベースの構造（スキーマ）の定義が困難なためである．このような特徴を持つ万葉集のデータは，特定の構造を定義できないか，できたとしても困難である半構造データであるといえる．

そこで本研究では，半構造データを表現可能なデータモデルを用いて万葉集をデータベース化し，それを用いて万葉集検索システムを実現することを目的とする．これにより，従来のシステムの問題点を解決でき，より専門的な検索機能を必要とする研究者の要求に応えることが可能となる．なお，本研究では万葉集を対象とした検索システムについて論じるが，このシステムの枠組は万葉集だけでなく，ほかの多くの古典文学作品にも適用可能である．

以降，2 節では，万葉集の概要とそのデータの特徴について述べる．3 節では，万葉集検索システムに求められる機能について説明し，いくつかの検索システムの事例についてその特徴と問題点を述べる．4 節では，半構造データのためのデータモデルである DREAM モデルについて概説した後に，DREAM モデルによる万葉集のデータベース化について述べる．5 節はまとめである．

2. 万葉集の概要とそのデータの特徴

万葉集は全二十巻からなる現存する最古の日本の歌集である．その中には，合計 4516 首の歌が収められているが，最も新しい歌で 759 年，一番古い歌の年代は不明である．成立過程は複雑で不明な部分も多いが，数々の編集作業を経て，ほぼ現在の形に整えられたのは奈良時代末期と推定されている[7]．また，編者についても様々な説があり明らかにされていないが，皇族や貴族を中心としながら庶民の歌までも含むという，後の勅撰和歌集には見られない特徴を持っている．万葉集二十巻は，巻一から巻十六までと，巻十七から巻二十までとに大きく分類される．巻十六までの歌は，歌が詠まれた場やその内容，表現技法，形態の違いなどから部立が行われている．巻十七からの四巻は部立が行われておらず，大伴家持の歌日記というべきものである．このように，一言で万葉集といっても，その和歌には様々な特徴や背景がある．

他の多くの古典作品と同様，万葉集の原本は現存せず，書写された写本という形で伝えられている．今日に伝えられている写本は，平安時代から室町時代にかけて書写されたものを中心に約 20 種類ほどある．また，複数の写本から原本を推定する校訂作業を経て作成された校訂本，漢字表記の原文に対して後世に付された訓読をもとにした注釈書，さらには版木で刷られた版本などが併せて 30 種類ばかり作られている．本稿では，これら様々な万葉集をまとめて諸本と呼ぶ．諸本には，約 3400 箇所にわたる文字異同（以降，単に異同とも表記する）や 3500 首あまりにわたる異訓が存在する．文字異同とは「舟と舩」のように意味は同じであるが異なる字が充てられていることであり，異訓とは同じ漢字に対して異なる読みが複数存在することである．万葉集の成立は仮名表記の成立以前であり，原文は漢字のみで記述されている．これに対して後世に複数の人が読みを振っていったため異訓が発生したのである．

万葉集の和歌は，5 音または 7 音からなる句を基本単位とした 5 句からなる短歌（約 4200 首）と複数の長句からなる長歌（約 300 首）によって構成される（以降，これらをまとめて“和

歌”と総称する）．和歌は国歌大観番号と呼ばれる明治期に付された和歌番号によって整理されており，研究者はこれを用いてそれぞれの和歌を識別している．

図 1は写本における和歌の例である（画像修正有り）．図中左は桂本の複製本[8]の 530 番の和歌の部分であり，右は西本願寺本の複製本[9]の 268 番の和歌の部分である．図 2は校定本[10]における 530 番と 307 番の和歌の部分である．多くの和歌の右側には，詠作の場や主題，作者などを漢文体で記した“題詞”が書かれている．さらに，和歌の左側に題詞と同様のことを補足的に説明した“左注”が付けられていることも少なくない．これら題詞や左注は，和歌と多対多の関連にある．また，図 1のように写本には，和歌の漢字表記に対する読みが書かれており，校訂本，注釈書にもそれに習ったものが多い．これら以外に，文字異同や異訓，校訂の違い，注釈などを表現するための，割注（図 2右），細注（図 2下），書き入れ（図 1右端の薄く小さい文字），朱書き，補注，見せ消ち，付箋などの様々な情報が存在する．以降，本稿ではこれらをまとめて“注”と呼ぶ．特に写本に多く含まれるこれらの注は，写本間の詳細な関連をとらえるための重要な情報となる．

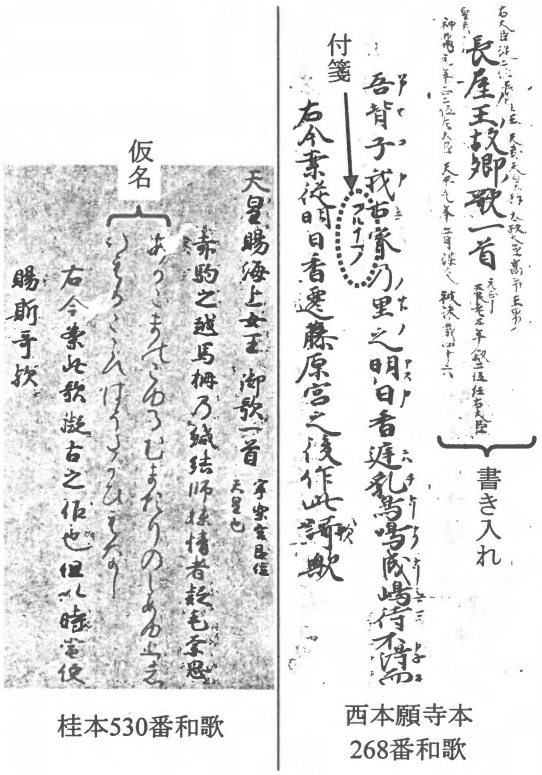

図1：写本における和歌の例

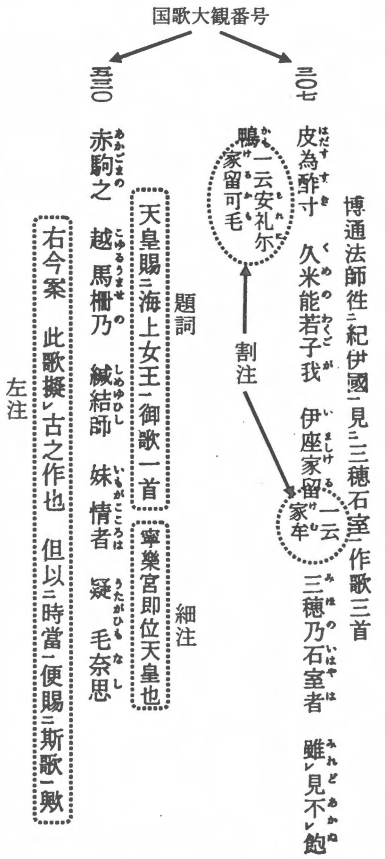

図2：校定本における和歌の例

3. 万葉集検索システムについて

万葉集は前節で述べたような特徴を持つため，一つの作品であるにもかかわらず，その和歌に関連する情報を参照することは煩雑で研究者の負担となっている．また，万葉集は最古の和歌集であり，他の文学作品に引用されるなど多くの影響を与えているため，他の古典文学作品の研究においても参照される．そのため，万葉集の研究分野のみならず，広く国文学研究の分野において，万葉集の和歌情報を簡単に検索できるシステムが強く望まれている．本節では，検索システムに求められる機能について説明したあと，現時点で提供されている検索システムの特徴や問題点を述べる．そして，本研究における万葉集検索システムで実現する機能について論じる．

3.1. 検索システムに求められる機能

以下に，万葉集検索システムに求められる機能を挙げる．これらの機能は，万葉集の表現に関連する機能（機能 1〜3），データの検索に関連する機能（機能 4〜6），そして，その他の機能（機能 7〜9）に大別される．

機能 1） 万葉集の種々のデータを格納する機能

万葉集のデータには，基本的なものとして，和歌，題詞，左注の漢字表記の原文（以降，単に原文），読み（以降，仮名），漢字仮名混じり文（以降，訓読）があるが，これら以外に，万葉集に付された様々な注とそれによって表現された異訓，異同などが挙げられる．検索システムは，これら全てのデータを文字データとして格納可能であることが求められる．さらに，全ての諸本について同様のデータを格納することも求められる．

機能 2） 和歌と題詞・左注との対応を表現する機能

和歌と題詞・左注には，多対多の対応がある．“詠作の場が同じである”などの観点から和歌を分類する際に，これらの関係が表現されていることが求められる．

機能 3） 異体字の表現

万葉集に書かれた漢字には，標準字体ではない，いわゆる異体字が数多く含まれる．これらも含めて万葉集を表現可能とする機能が求められる．

以降，機能 1 の種々のデータと機能 2 における対応関係をまとめて“万葉データ”と呼ぶ．

機能 4） 万葉データの検索機能

格納した万葉データから，所望のデータを検索する機能が求められる．この機能は，単に原文や仮名等から対応する和歌，題詞，左注を検索する機能にとどまらず，注などのその他の万葉データを基にして，和歌，題詞，左注や関連する万葉データの検索を可能とする機能を含む．

機能 5） 異訓，異同を考慮した万葉データの検索機能

万葉集の和歌に影響を受けた古典文学作品では，異同，異訓の存在を考慮せずに和歌を引用，参照している．そのため，これらの古典文学作品の研究分野における万葉集の和歌の検索では，異訓や異同を意識することなく全ての諸本の和歌が検索可能であることが求められる．さらに，注によって表現された原文，仮名からでも和歌を検索できなければならない．同様に，ある和歌に対する全ての異同，異訓も検索可能であることが求められる．

機能 6） 高度な分析機能，または，その機能を実現するための検索機能

「特定の漢字の出現回数や出現箇所」，「ある読みをする全ての漢字の出現回数，箇所，ならびにそれらの漢字の他の読み」などの分析的な処理が行える機能，あるいは，別途プログラミング言語等と組み合わせてその機能を実現するための検索機能が求められる．

機能 7） 原本画像データの格納と提示の機能

国文学の研究分野において基本となるものは原本である．従って依拠本文が極めて重要なものとなる．そこで，写本を中心とした原本を手軽に確認することを可能とするために，原本画像データの格納・提示機能が求められる[11].

機能 8） 使用目的に応じたグラフィカル・ユーザ・インタフェース（GUI）

検索システムには万葉データを検索するための GUI が必要であることは当然であるが，その GUI は使用目的ごとに様々な形態が考えられる．例えば，万葉集を研究対象とする研究者が用いる GUI には，複雑な検索条件の指定と詳細な万葉データの提示が行える検索機能が望まれる．一方，他の文学作品を研究対象としている研究者にとっては，高度な検索よりも簡単に全ての諸本にまたがった検索が行えることが重要である．GUI の提供においては，このような使用目的を考慮する必要がある．

機能 9) ユーザが独自にデータの追加，変更，加工を行える機能

既存のデータに反映されていない新事実や，利用者自身の仮説や考えを検索システムに独自に追加したいという要求がある．さらに，データの誤りの修正や，データを独自の手法で加工したいという要求もある．このような要求を満たすためには，利用者が独自に管理可能なスタンドアロン形態の検索システムが必要である．また，インターネットへの常時接続可能な環境が一般的に普及した現在でも，ネットワークの負荷やサーバ等の要因により検索が行えない状況を避けるため，スタンドアロン形態の検索システムを望むユーザも少なくない．

3.2. 既存のシステムの特徴と問題点

ここまでに述べてきたように，万葉集検索システムには大きく分けて 9 つの機能が求められる．これらの全ての機能を一つのシステムで実現することはきわめて困難であり，現実的ではない．既存の検索システムは，これらの機能のうちいくつかに主眼を置いて実現されている．

万葉集校本データベース作成委員会（委員長坂本信幸氏）が作成・試験公開している万葉集校本データベース[12]は，版本の一つである寛永版本の画像データを基軸として，各種写本の句部分の画像データを引用することにより，本文校異の比較研究を可能とするデータベースの試作版である．このシステムは，Web ベースのシステムであり，ブラウザを用いてインターネットを介した利用が可能である．データとして，写本（元暦校本，広瀬本，紀州本，神宮文庫本，西本願寺，京都大学本など）の画像データと，和歌の原文，仮名あるいは訓読，さらに，一部注釈書の注釈も格納している．これらの特長により，機能 1，3，7 の多くを実現している．しかし，原文等の情報による和歌検索の機能は有しておらず，和歌番号として付されたリンクを辿ることで，所望の和歌のデータを表示する形式である．

国文学資料館[1]の本文データベース検索システム[13]は，インターネットを介して日本古典文学本文データベースを利用するための試験的なシステムである．日本古典文学本文データベースとは，岩波書店刊行の旧版「日本古典文学大系」の全作品の本文をデータベース化したものであり，万葉集のみならず約 580 作品もの検索が可能なシステムである．万葉集検索システムとしては，機能 1，4，6 を概ね提供している．しかし，特定の写本を対象としており，同一作品で複数にまたがる写本間の比較などは対象としていない．

筆者の一人である吉村は，1995 年よりインターネット上にて万葉集テキストファイルを配布してきた．このファイルには，万葉集の原文，仮名，訓読，一部の注，諸本間の異訓や異同の情報，さらには，和歌の作者や関連する地名，詠作の場などの事項を表すキーワードが含まれる．単なるテキストファイルであるため，機能 9 は容易に実現できるものの，そのデータの検索には別途 grep 等のテキスト処理プログラムが必要であり，国文学研究者には利用が難しい面があった．そのため，Windows 上で動作する検索用アプリケーションを作成し，テキストファイルと共に配布している[14]．また，このファイルのデータをリレーショナルデータベース管理システムで管理し，インターネットを介して万葉データを検索可能なシステムも構築している[15]，[16]．これらのアプリケーションやシステムでは，異体字（機能 3）や画像データ（機能 7）は対象とせず，基準となる一つの諸本（底本と呼ぶ）のテキストを全て格納し，異訓や異同の情報は差分の形式で格納している．このデータ格納方式のため，現システムは機能 1，2，4 の一部を提供するにとどまっている．

そのため，機能 4，5，6，8 を提供するための検討[17]-[19]を行っているが，その際の問題が万葉集に付された多数の注の表現とその検索である．これらの注は，その意味や性質，形式，記入されている位置などが異なり，一様に扱うことが出来ない．例えば，同じ形式の割注が存在する場合でも，それが持つ情報が句に係るのか，和歌や題詞全体に係るのかはそれぞれの注ごとに異なる．さらに写本には，見せ消ち，朱書き，付箋などの多種多様な注が存在する．このような注を含めた全ての万葉データのデータ

[1] http://www.nijl.ac.jp/index.html

ベース化には，万葉データを完全に表現可能なデータベースの構造（スキーマ）の定義が必要である．そのためには，これら全ての注について検討し，その名称や表現形式を決定しなければならない．現状では，テキストエディタや表計算ソフトウェア，Perl 等のスクリプト言語による注の整理分類を行っているが，この作業法では限界がある．そこでデータベース技術を用いることが考えられるが，これではデータベース化のためのデータベース化という状況に陥る．

この問題を解決する一つの方法として，OEM[20][21]やエッジ付きラベルグラフ[22]などの半構造データのためのデータモデルを採用したデータベース管理システムを用いる方法が挙げられる．半構造データとはスキーマレスでデータの構造がデータの中に埋め込まれているデータのことである[23]．本研究では，中田らが提案している半構造データのためのデータモデルである DREAM モデル[24]を用いて万葉データを表現しデータベース化することで，より多くの機能をもつ万葉集検索システムを構築する．なお，本システムでは上で述べた機能 1～9 の全てを実現することを目標とするが，さしあたり，機能 1，2，4，5 の実現を目的とし，それ以外の機能は今後の課題とする．

4. DREAM モデルを用いた万葉集データベースの構築

本節では，DREAM モデルの概要とそれを用いた万葉データの表現例について述べる．そして，それらのデータを検索する際に必要となるデータベースの構造に関する情報を提供するシェイプとシェイプグラフについて述べたあと，万葉集データベースの概要について説明する．

4.1. DREAM モデルの概要と万葉データの表現

DREAM モデルは集合論を基盤としており，データエレメント，名前付きエレメント，視点，オブジェクト，バンドルの 5 つの要素を持つ．これらの要素は 3 つ組であり，データエレメント以外の要素の第 3 項は集合である．データエレメントはデータの実体（データ値）を格納する要素であり，名前付きエレメントは一つの属性を表す要素である．また，視点は対象の一つの側面を表し，オブジェクトは一つの対象物を表す要素である．バンドルはオブジェクトの集合を表現する．それぞれの要素の詳細は以下のとおりである．

データエレメント：
(deid, type, value) 但し，*deid* は識別子，*type* はデータ型，*value* は値．
名前付きエレメント：
(neid, name, {deid または oid})
但し，*neid* は識別子，*name* は名前．
視点：
(peid, name, {neid}) 但し，*peid* は識別子．
オブジェクト：
(oid, name, {peid}) 但し，*oid* は識別子．
バンドル：
(bndlid, name, {oid または bndlid})
但し，*bndlid* は識別子．

万葉データを DREAM モデルで表現した一例を図 3に示す．この例では，和歌，題詞，和歌の句をそれぞれ個別のオブジェクトと考え，万葉集の各巻と諸本はバンドルで表現している．図右上の「句 1」を表すオブジェクトは，*(obj1, “句 1”, {pe1})* である．そのオブジェクトに含まれる視点は *(pe1, “和歌”, {ne1, ne2})* であり，その視点に含まれる名前付きエレメントは *(ne1, “原文”, {de1})*，*(ne2, “仮名”, {de2})* の二つである．また，句 1 の原文や仮名の文字列を格納するデータエレメントは，*(de1, “string”, “皮為酢ず”)*，*(de2, “string”, “はだすすき”)* である．

同様に，「307 番の和歌」は，*(obj307, “307”, {pe2, pe3, pe4})*，*(pe2, “ORG”, {ne11, ne12, ne13 ne14, ne15})*，*(pe3, “適用”, {ne21,ne22, ne23, ne24, ne25})*，*(pe4, “題詞”, {...})*，*(ne11, “句 1”, {obj1})*，*..., (ne13, “句 3”, {obj3})*，*..., (ne15, “句 5”, {obj5})*，*(ne21, “句 1”, {obj1})*，*..., (ne23, “句 3”, {obj3’})*，*..., (ne25, “句 5”, {obj5’})* の要素で表現される（一部略記）．ここで，*obj1, obj2, ..., obj5, obj5’* は句 1～句 5 に対応するオブジェクトの識別子である．また，“*ORG*” という名前の視点 *pe2* は諸本に書かれた本来の原文や仮名を表し，“適

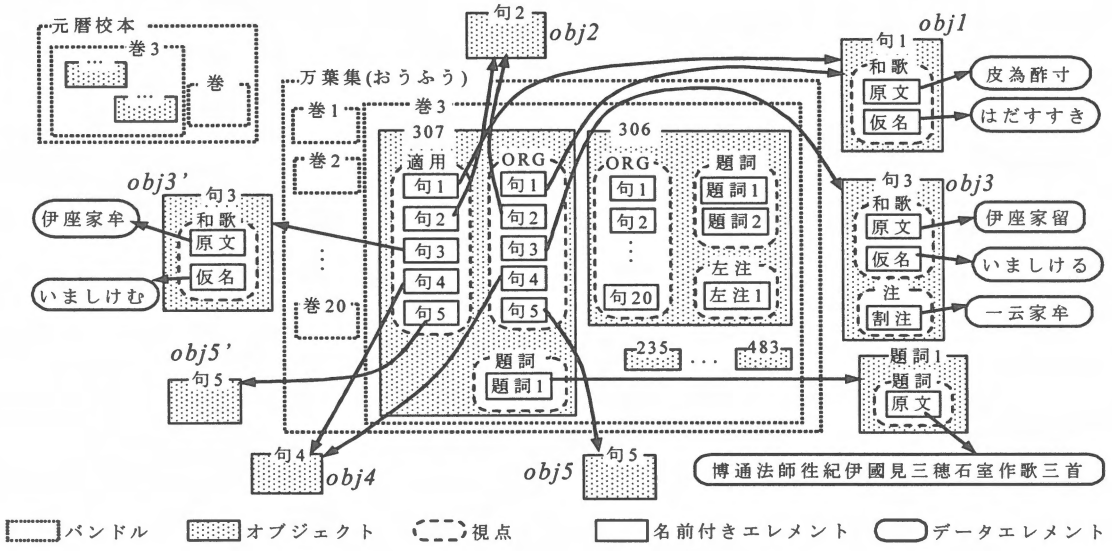

図3：万葉データの表現例（一部の要素は略記している）

用”という名前の視点 *pe3* は図 2右の注“一云家牟”と“一云安礼尓家留可毛”によって記された漢字や仮名を適用した原文や仮名を表す（図3左の“句3”，“句5”）．

図3の「巻3」を表すバンドルは，*(bndlid3, “巻3”, {obj235,...,obj483})* となる．ここで，*obj235* ～ *obj483* は235～483番の和歌に対応するオブジェクトである．同様に，「万葉集（おうふう）」を表すバンドルは，*(bndlid21, “万葉集（おうふう）”, {bndlid1, ..., bndlid20})* である．この他にも，データベース中には同じような構成のバンドルが存在し，それぞれの諸本の情報を格納している（図左上のバンドル“元暦校本”など）．

この様な形式で全ての諸本に関して万葉データを格納することで機能1，2が実現される．さらに，和歌を句ごとのオブジェクトに分解し，注が付された句については注の内容を表す別のオブジェクトを別途作成して，“*ORG*”と“*適用*”という別々の視点からそれらの句のオブジェクトを参照している．これにより，諸本に書かれたもともとの原文や仮名だけではなく，注により記された原文や仮名からでも和歌が検索可能となり機能5が実現できる．

4.2. シェイプ，シェイプグラフ

これまでに述べたように，DREAM モデルを用いることで，複雑な万葉データをデータベース化することが可能となる．しかし，図 3で示した例では，オブジェクトや視点等の構造（属性の数や名前，それが持つデータの型など）は，対応する和歌や題詞，注などによって異なる．さらに，DREAM モデルを用いた万葉データの表現法はこの方法以外にも考えられる．従って，データベース中には構造が不揃いな要素が大量に存在することになる．このことは，データの格納においてはモデルの柔軟性を示す利点となるが，データの検索や操作を行う際には対象のデータベースの構造の把握を難しくする原因となる．また，異なる表現形式の万葉データの統合や相互変換の際にもデータベースの構造に関する情報が必要となる．

そこで，DREAM モデルが提供するシェイプ及びシェイプグラフと呼ばれる情報を利用する．シェイプは，バンドル，オブジェクト，視点の構造を示す情報であり，シェイプエントリと呼ばれる基本単位から構成される．シェイプエントリは名前付きエレメントの構造を表す情報である．シェイプグラフは，バンドルやオブジェクトがどのような構造のオブジェクト及び視点を持っているかを表す情報であり，自身のシェイプとそれらが含む視点やオブジェクトのシェイプで構成される．シェイプエントリとシェイプ，シェイプグラフは，データエレメントやオブジェクトなどの挿入，削除，更新によって動的に変化する．これらの構造は以下のとおりである．

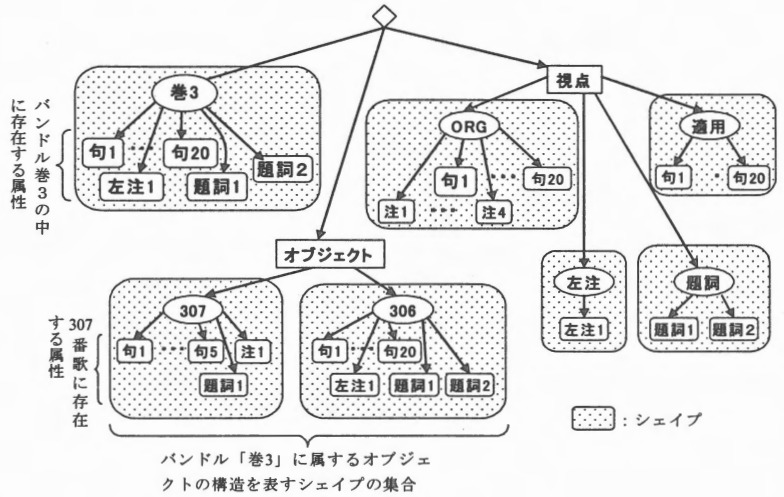

図4：「巻3」バンドルのシェイプグラフ

シェイプエントリ：

(seid, name, DT) 但し，*seid* は識別子，*name* は名前付きエレメントの名前，*DT* は名前付きエレメントが持つ要素のデータ型の集合．

シェイプ：

(sid, name, S) 但し，*sid* は識別子，*name* は対応する要素の名前，*S* はその要素が持つ名前付きエレメントのシェイプエントリの集合．

シェイプグラフ：

オブジェクトのシェイプグラフ：

(osgid, obj_sid, {per_sid}) 但し，*osgid* は識別子，*obj_sid* はオブジェクトのシェイプの識別子，*per_sid* はオブジェクトに属する視点のシェイプの識別子である．

バンドルのシェイプグラフ：

(bsgid, bndl_sid, {obj_sid}, {per_sid}) 但し，*bsgid* は識別子，*bndl_sid* はバンドルのシェイプの識別子，*obj_sid* はバンドルに属するオブジェクトのシェイプの識別子，*per_sid* はバンドルに属する視点のシェイプの識別子である．

例えば，図3の「句1」を表すオブジェクトと「巻3」を表すバンドルのシェイプとそのシェイプエントリは次のようになる．これらの，シェイプを参照することにより，「句1」を表すオブジェクトと「巻3」を表すバンドルにどのような属性が幾つ存在するのかが明らかになる．

「句1」を表すオブジェクトのシェイプ：

(sid1, “句1”, {seid1, seid2})
(seid1, “原文”, {string})
(seid2, “仮名”, {string})

「巻3」を表すバンドルのシェイプ：

(sid2, “巻3”, {seid3, seid4, ..., seid22, seid23, seid24, seid25, ..., })
(seid3, “句1”, {object}),
(seid4, “句2”, {object})
⋮
(seid22, “句20”, {object})
(seid23, “題詞1”, {object})
(seid24, “題詞2”, {object})
(seid25, “左注1”, {object})
⋮

図4は「巻3」を表すバンドルのシェイプグラフである．バンドルのシェイプグラフを参照することで，このバンドルの中に全体として句1～句20，左注1，題詞1，題詞2という属性が存在し，“307”，“306”という名前のオブジェクトと，“ORG”，“適用”，“左注”，“題詞”という名前の視点を持つことが分かる．さらに，それぞれが個々に持つ属性も把握可能となる．

4.3. 万葉集データベースの概要

DREAMモデルで表現された万葉集データを，

図5が示す構成のDREAMシステムで管理する．システムの実現には，関係データベース管理システムであるPostgreSQL[25]を用いる．DREAMモデルのデータ構造・操作をリレーショナルモデルに変換するために，API関数群(DREAM API)をJavaとPostgreSQLのJavaインターフェースを用いて実現する．実際のデータベースでは，万葉集データはオブジェクト，名前付きエレメントなどの要素に分解されて，それぞれの要素に対応するテーブルに別々に格納される．

図6はシステムのGUIの画面の例である．図中には「巻3」に対応するバンドルの内容が表示されている．また，図7にはそのバンドルのシェイプグラフが表示されている．万葉データの検索は，これらのウィンドウに検索条件を入力することで行う(3節の機能4の実現)．追加，削除，更新等の操作も同様である．このような，GUIを用いて万葉データを格納し，その後でデータを参照しつつ更新等の操作を行うことで，データの整理分類を行う．

万葉データの整理分類が終了し，多様な注に関してその表現形式や名称等が決定できれば，統一的な操作により万葉データを検索できるシステムが提供可能となる．その際に，DREAMのGUIはDREAMモデルの操作系を用いた詳細な検索や操作が可能であるが，モデルに関する知識を必要とするため操作が難しい．そのため，万葉集検索システムとしての必要最小限の検索のみが可能なアプリケーションを別途作成することが望まれる（機能8）．さらに，機能6を実現するアプリケーションやGUIを用いるとかえって煩雑になるような操作（例えば，大量のデータの挿入など）を行う専用アプリケーションなども考えられるが，このようなアプリケーションはJavaとDREAM APIを用いて実現可能である．

5. おわりに

多くの相違を持つ複数の本によって伝えられ，かつ，その中に多種多様な注が付されているために完全なモデル化が難しい万葉集を検索可能なシステムを実現することを目的に，半構造データのためのデータモデルを用いた万葉集データをデータベース化について述べた．現在

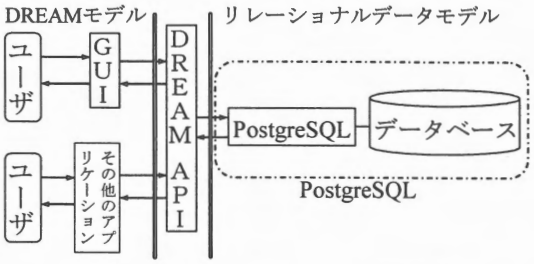

図5：DREAMのシステム構成図

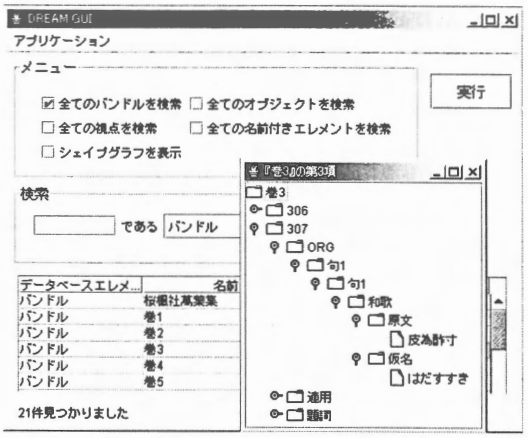

図6：DREAMのGUI 1

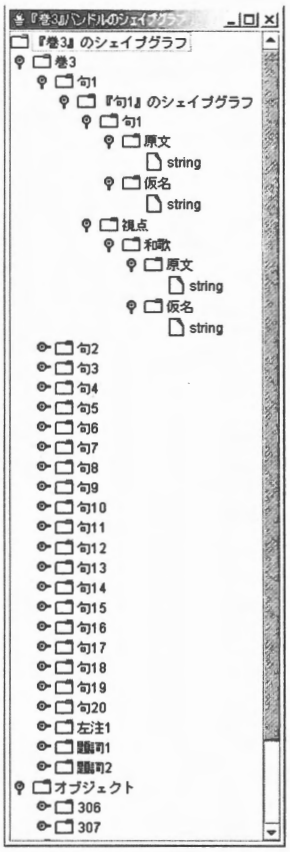

図7：DREAMのGUI 2

は，DREAMシステムのGUIの実装を行っている段階である．また，文献[14]の万葉データの基本的な部分（和歌，題詞，左注の原文，仮名など）の格納が終了している．GUIの実装完了後に，それを用いて注のデータの格納と整理分類を行うことで万葉集データベースを構築し，さらに，万葉集検索システムとしてのGUIを実現する予定である．また，文字コードの問題で現状では格納できていない漢字や異体字への対応，諸本の画像データの格納やそれに付随する著作権等への対応，さらには，システムの評価などが今後の課題である．

参考文献

[1] “N-gramが開く世界”，漢字文献情報処理研究第二号 特集2，pp49-73，好文出版，2002.

[2] 竹田正幸：“古典和歌からの知識発見”，情報処理 Vol. 43, No.9, pp.941-949，2002.

[3] 安永尚志：“国文学研究とコンピュータ”，勉誠社，1998.

[4] 佐竹昭廣，立川美彦：“重層型情報時代に対応する国文学高機能情報形成手法の開発とその実用化に関する研究”，平成7年度～平成9年度科学研究費基盤研究(A)(2)研究成果報告書（課題番号07401014），国文学研究資料館，1998.

[5] 中村康夫：“古典籍原本データベースにおけるテキストと絵図の構造的探索の研究”，平成10～平成13年度科学研究費基盤研究(A)(2)研究成果報告書（課題番号10301022），国文学研究資料館，2002.

[6] 中村康夫：“古典研究のためのデータベース”，臨川書店，2000.

[7] 神野志隆光：“別冊國文學 No.55 万葉集を読むための基礎百科”，學燈社，2002.

[8] “御物 桂萬葉集”，集英社，1976.

[9] “西本願寺本萬葉集（普及版）巻第三”，主婦の友社（発売おうふう），1993.

[10] 鶴久，森山隆 編，“萬葉集”，おうふう，1972.

[11] 吉村誠 中田充：“『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”，山口大学教育学部研究論叢 第53巻 第1部，pp. 81-93，2003.

[12] 万葉集校本データベース作成委員会：“万葉集校本データベース”，http://www.manyou.gr.jp/，1999.

[13] 国文学資料館：“本文データベース検索システム”，http://base3.nijl.ac.jp/Rcgi-bin/hon_home.cgi

[14] 吉村誠：“万葉集テキスト Ver5.00 R1.0 検索ソフト Ver2.0 for windows”，http://yoshi01.kokugo.edu.yamaguchi-u.ac.jp/manyou/man_user.html，2002.

[15] 中田充，新原久仁子，松尾圭子：“DBMSを用いた万葉集検索システムの構築”，山口大学教育学部研究論叢 第50巻 第2部，pp.73～83，2000.

[16] 中田充，吉村誠，新原久仁子，松尾圭子：“万葉集検索 Ver.1.01”，http://infws00.inf.edu.yamaguchi-u.ac.jp/MANYOU/manyou_kensaku.html，2004.

[17] Mitsuru NAKATA, Makoto YOSHIMURA, and Qi-Wei GE: “A Database System Designing Method for Japanese Poems Manyo-Shu ”, Proceedings of the ICFS 2002, pp.S5-43～48, 2002.

[18] 中田充，吉村誠，葛崎偉：“万葉集データベースにおける和歌検索の効率に関する検討”，情報処理学会研究報告 2002-CH-54，pp.9～16，2002.

[19] 中田充，大鶴仁美，葛崎偉，吉村誠：“万葉集検索システムにおける異訓情報の生成法について”，第16回 回路とシステム（軽井沢）ワークショップ論文集，pp.225～230，2003.

[20] Papakonstantinou, Y., Garcia-Molina, H., and Widom, J.: Object Exchange Across Heterogeneous Information Sources, Proc. of 11th International Conference on Data Eng., pp.251-260, 1995.

[21] Goldman, R. and Widom, J.: “DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases”, Proc. of the 23rd VLDB Conf. , pp. 436-445, 1997.

[22] Buneman, P., Davidson, S. and Suciu, D.: “Programming Constructs for Unstructured Data”, Proc. of Int. Workshop on DBPL, electronic Workshops in Computing, Springer-Verlag, 1995.

[23] 田島敬史：“半構造データのためのデータモデルと操作言語”，情報処理学会論文誌 Vol.40, No.SIG3(TOD1), pp.152-170，1999.

[24] Mitsuru NAKATA, Qi-Wei GE, Teruhisa HOCHIN, Tatsuo TSUJI: “An Extended Dynamic Schema for Storing Semi-structured Data”, Proc. of ITC-CSCC 2002, pp.301-304, 2002.

[25] 石井達夫：“PostgreSQL 完全攻略ガイド”，技術評論社，2001.