

『古事記』学術支援データベースの構築 -神名検索システムのための内部構造について-

齋藤晋*, 柴田みゆき*, 生田敦司*, 杉山正治**, 宮下晴輝*
大谷大学文学部人文情報学科*, 立命館大学理工学部**

1 はじめに

我々は、人文科学研究者の求めるところに応じて『古事記』学術支援データベースを構想した^{[1][2]}。特に、本文の検索に関しては以下の5点が求められた。

- (1) 神や人の名前検索ができること。別名を有する場合は、両方を同じ個性として認知して同一の検索結果が得られること。
- (2) 検索結果を見出しとして表示できること。
- (3) 見出しから本文へ簡単にアクセスできること。
- (4) 本文中では、神名の強調表示や別名の明示がされること。
- (5) 神名を足がかりにした系図の表示が出来ること。

本研究では、これらを実現することを目指してデータ構造の設計を行い、操作性を確認した。それらを踏まえてデータ構造の特徴とその問題点などについて述べる。

2 データベース構築に必要なテーブル

今回のデータベースを構築するに当たり、以下のよう
にテーブルを設計し準備した。

主に神・人の名前を保持する「神名テーブル」(表1)、神・人の名前のシソーラスとして機能するための「別名テーブル」(表2)、入力が想定される語句と検索対象となる神・人名との関連を保持する「入力語句テーブル」(表3)、主に本文テキストの内容を保持する「本文テーブル」(表4)、本文の見出しを保持する「見出しテーブル」(表5)、本文に付随する振り仮名に関する情報を保持する「振り仮名テーブル」(表6)、本文に差し挟まれる註釈に関する情報を保持する「註釈テーブル」(表7)の7つである。各テーブルの詳細については次章以降で述べる。

3 神名検索システムにおける処理手順

3.1 検索ワードによる検索結果表示

検索ワードを入力すると、まず「入力語句テーブル」

[†]KOJIKI Knowledge Assistant Database System for Academic Usage: About an internal structure for searching names of gods / goddesses

*Susumu Saito, Miyuki Shibata, Atsushi Ikuta and Seiki Miyashita: Department of Humane Informatics, Otani University

**Seiji Sugiyama: Faculty of Science and Engineering, Ritsumeikan University

の[入力が想定される単語]フィールドと照合して、合致する神・人名を探索する。想定される入力語句を幅広く保持させておくことが望ましい。合致する神・人名がある場合は、「本文テーブル」の[語句]から、その神・人名が含まれるレコードを探索し、対応する見出しを[見出しID]を介して「見出しテーブル」の[見出し]フィールドから探索し、表示する。また同時に、「別名テーブル」を介して、その神・人名の別名(以下、神・人別名)についても探索し、同様の処理を行う。

3.2 見出しから本文へのアクセス

見出しを選択すると、「本文テーブル」から、その見出しに該当する[見出し ID]をもつレコードを抽出する。抽出されたレコードの[語句]フィールドから、該当する本文テキストを再構成し、画面に表示させる。このとき、検索対象となっている神・人名と[語句]フィールドを照合して、合致する語句部分は強調表示をする。神・人別名についても同様である。また検索対象以外の神・人名も明示するためには、「本文テーブル」の[語句の種類]フィールドに記録されているデータを利用する。

3.3 神名を足がかりにした系図表示

系図表示を選択した上で、本文中の神・人名を選択すると、「神名テーブル」から該当するレコードを探索する。そのレコードが持つ[X座標][Y座標][倍率]フィールドのデータがハッシュ情報に変換された上で、系図表示システム MaSSRiDGe^[3]にURLハッシュとして送信される。これにより神名を足がかりとした系図表示が可能になる。

4 振り仮名・註釈の問題

『古事記』本文テキストには、神名などに振り仮名が用意されていることがある。また、本文途中で註釈が存在する。元となるテキストデータでは、これら振り仮名や註釈が、本文に割り込む形で存在していることがある。これをそのまま検索対象のテキストとすると、神名などが正しく検索されない可能性がある。そこで、これらの情報は、本文とは別のテーブルに保持するように設計した。

4.1 振り仮名の処理

本文を表示する際に、[振り仮名フラグ]フィールドから

振り仮名があると判断された語句に対しては、[本文 ID]を介して「振り仮名テーブル」の[振り仮名番目数][振り仮名開始位置][振り仮名終了位置]フィールドを参照し、該当箇所には振り仮名が表示されるようにする。ひとつの語句に対して複数の振り仮名が付帯する可能性があるので、「本文テーブル」と「振り仮名テーブル」は1対多結合とする。

4.2 註釈の処理

註釈についても、「註釈テーブル」(「本文テーブル」と[本文 ID]を介した1対多結合)を用いて、振り仮名とほぼ同様の処理を行う。本文とは別データ扱いなので、本文とは異なる表示方法(たとえば、小さめに表示するなど)が実現できる。なお、註釈が振り仮名を途中で分断することもある。これに対応するために、「註釈テーブル」には、その語句の何個目の振り仮名の何文字目を分断するのかについての情報も記録する([振り仮名番目数][註釈の場所(振り仮名)]フィールド)。

5 おわりに

以上のようなテーブル設計により、求められる機能は実現できた。今後は以下のような課題を検討していく予定である。

「本文テーブル」の[語句]フィールドに保持される、1レコードあたりのテキストの分割の程度は、最も大きな場合、1レコードが1つの見出しに該当するテキスト部分すべてになる。最も小さい場合は、1レコードが1文字になる。今回は、神名検索に重点を置いていることもあり、人文科学研究者の判断に基づき、神・人の名前については、名詞単位での分割を行った(ただし、それ以外の種類の単語については厳密ではない)。データの容量や読み込み速度などから、適切な粒度を検討する予定である。

参考文献

- [1] 生田敦司, 齋藤晋, 柴田みゆき, “『古事記』学術支援データベースの構築—基本機能の検討—”, 情報処理学会研究報告, 人文科学とデータベース, 第12回公開シンポジウム-5, pp.47-54, 2006
- [2] 生田敦司, 齋藤晋, 杉山正治, 柴田みゆき, 宮下晴輝, “『古事記』学術支援データベースの構築—系譜の図像化とインターフェイスの検討—”, 情報処理学会研究報告, 人文科学とデータベース, 第13回公開シンポジウム-2, pp.9-16, 2007
- [3] 杉山正治, 齋藤晋, 生田敦司, 柴田みゆき, “『古事記』学術支援データベースの構築—系譜史料の表示形式に関する検討—”, 情報処理学会・第75回人文科学とコンピュータ, 2007-CH-75(7), pp.47-54, 2007

表1 テーブル名:神名

データ項目名	データの種類
神の名 ID	数値
神の名	文字
X座標	数値
Y座標	数値
倍率	数値

表2 テーブル名:別名

データ項目名	データの種類
別名セットの ID	数値
神の名 ID	数値
別名 ID	数値

表3 テーブル名:入力語句

データ項目名	データの種類
入力語の ID	数値
神の名 ID	数値
入力が想定される単語	文字

表4 テーブル名:本文

データ項目名	データの種類
本文 ID	数値
見出し ID	数値
段落 ID	数値
語句 ID	数値
語句の種類	数値
語句	文字
振り仮名フラグ	数値
註釈フラグ	数値

表5 テーブル名:見出し

データ項目名	データの種類
見出し ID	数値
見出し	文字

表6 テーブル名:振り仮名

データ項目名	データの種類
振り仮名 ID	数値
本文 ID	数値
振り仮名番目数	数値
振り仮名開始位置	数値
振り仮名終了位置	数値
振り仮名	文字

表7 テーブル名:註釈

データ項目名	データの種類
註釈 ID	数値
本文 ID	数値
註釈の場所	数値
振り仮名番目数	数値
註釈の場所(振り仮名)	数値
註釈	文字