

日本古典系図データベースの構築

相田 満

国文学研究資料館

aida@nijl.ac.jp

Construction of a Japanese classic genealogy database

AIDA,MITSURU

NIJL

概要

人物典拠情報のデジタルライブラリを構想するためには、系図情報を取り込むことは不可欠であろう。そこで、本研究では、「日本古典系図データベース」の構築に取り組み、以下の課題の克服を試みた。

- ・ データの入力・記述手法の開発
- ・ 系図表示のビジュアライズ化
- ・ 人物情報管理のための汎用的な器の構築
- ・ 文字セットの限界の克服

In order to conceive of the digital library of person information, it is indispensable to take in genealogy information. Therefore, I conceived of construction of "a Japanese classic genealogy database." And it tackled solving some subjects generated in case of the time of data formation work.

- Input data and develop the technique for describing.
- In order to manage a person's information, build a general-purpose form.
- Realize a genealogy display on a screen.
- Conquer the limit of the type of letters of the kanji character which can be used.

1. はじめに

長大な系図は、通覧・視認性がよくない。

冊子体は、血族線が錯綜してわかりづらくなり、卷子形態では、その系図自体を広げる場所に困る始末である[図1]。しかも、見出し人物以外の詳細な記述を検索可能なインデックスは、皆無というのが現状である。

かかる状況に鑑み、本研究では、系図表示・検索の実現と、それに連動する詳細な人物情報の双方を検索可能とするデータベースの構築に取り組んだ。



[図1] 『本朝皇胤紹運録』の活字版

2. データの入力・形成手法について

2.1. 入力典籍

本データは、5種類の板本・写本から採録した[図2]。『群書類従』所収本や『諸家大系図（尊卑分脈）』には、活字本も存在するが採らなかった。これは、著作権上の配慮ということもあるが、何よりも、原本から直接入力した方が、血族線の視認性が良く、データの構造が把握しやすかったからである。

確かに、翻刻を行いながらの入力になるため、その精度も心配された。しかし、入力業者はその要求に十分応えることができた。

また、このことにより、当時実際に使用された典籍情報が提供されることになったのは、享受史や史料批判研究の面でも有益であった。

データ名	人物テーブル	細目テーブル
諸家大系図（尊卑分脈）	39,752	91,558
諸家知譜拙記	3,648	11,465
諸家知譜拙記絶家伝	1,487	3,555
群書諸家系図（群書類従本）	4,938	8,904
本朝皇胤紹運録（群書類従本）	1,830	7,422
合計	51,655	122,904

[図2] 収録系図とレコード数

2.2.原稿作成手法

大量データの入力プロジェクトは、巨大艦船の操縦にもたとえられ、データ構造を適切に把握し、効率的で確実な入力が行えるような手順と工程管理を案出することが不可欠となる。

そこで、本プロジェクトでは、あらかじめ入力原稿となる原本のコピーに直接データタグを書き込んでおき、入力業者がそこに記述された文字列をトレースして入力するという手法でデータの形成を行った[図3]。

原稿自体が、タグや付加文字列により構造化されているので、最終段階で一括処理による整形作業を行えば、ほぼ規則的な、正規化されたデータベースデータが出来上がることになるのである。

この手法の採用により、データシート転記などで起こりやすい誤謬の回避と、入力作業の効率化が果たせた。

2.3.データ構造の基本的考え方

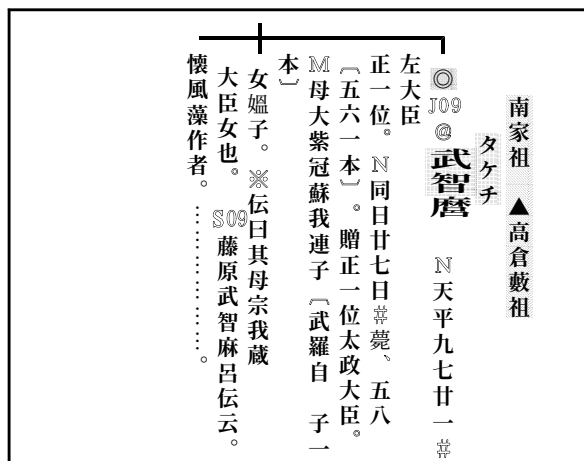
2.3.1入力用原稿の特徴

入力対象とした史料の記述形態には、以下の特徴が指摘できる。

- ①人物と付帯情報の記述位置が不規則・不定形
- ②記載順序が不規則で錯綜していること
- ①異体漢字・漢字字種が多いこと



[図3]入力用原稿（諸家知譜拙記）



[図4]入力原稿の翻刻イメージ

系図史料は、本文の全文の電子テキスト化という観点で考えると、最も複雑で重層的な記述を有する注釈書とも見なせる。テキスト記述法の面から考えてみても、今回の研究で案出された手法は、幅広い応用が可能と言えよう。

2.3.2記述構造の把握

本データベースでは、系図上で通例大字で書かれる人物名を主キーに設定し、それぞれに ID をあてた。また、その周縁に小字で記載される記事を、その人物の履歴・記事情報と規定した。

さらに、頭注・脚注のように、離れた箇所に関連する記事がある場合[図 3]、当該範囲を枠で囲い、主キーの ID で戻り値を示して指示するとともに、「分書」「頭注」のように記述位置を属性値として付与した。

2.3.3人物関係情報の入力・管理

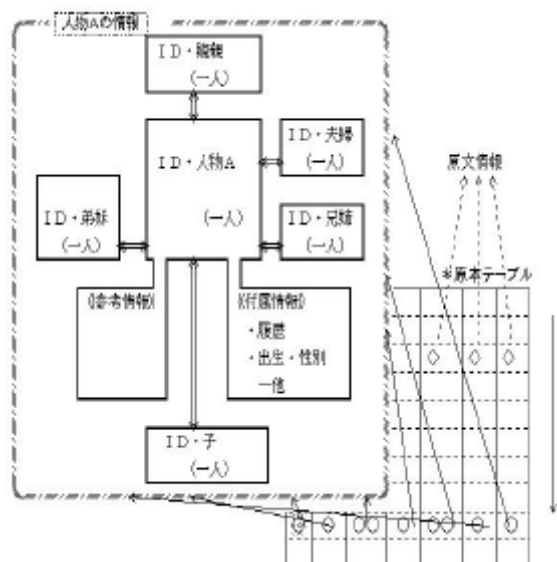
人物関係の情報、すなわち、系図の線を引いて表示させるための情報は、各人物の前後左右に隣接する人物の ID のみを指定することで管理されている。これは、後述の GEDCOM などとは異なり、兄弟姉妹や子供の総数の値は持たせない形式となっている。

したがって、長兄・始祖・末子・末裔などのように、隣に値が存在しない場合は空欄 (NULL) となっているので、階層ツリー表示のために、検索用マスタ形成時には、兄弟数・世代数などの値

を自動計算で発生させる手法でデータを発生させているのである[図5]。

処理は一手間かかるが、入力原稿の冊子体の系図などは、血族線が錯綜し、時には数頁(丁)も隔てて最も親等の近い親子兄弟が登場することも少なくないため、データ入力・採録時のエラーを回避するためである。

また、本データベースは、原本に記載される前文や奥付、凡例、序文なども合わせて入力されている。この部分のデータにも、ID と親子関係を与えることにより、データの読み順をナビゲートすることも可能にしている。この工夫については、今後さらに採録データのバリエーションを増やして、取り組んでみたい。



[図5] データ概念図

2.3.4 データテーブルの作成

データ入力は、2段階に分けて進めた。すなわち、第一段階は、資料原本の翻刻作業を主目的とする①プレマスタデータと仮称する段階。第二段階は、検索用の付属情報を充実させるための②マスタテーブルを作成する段階である。以下、各段階に分けて説明する。

①プレマスタテーブルの作成[図6]

第一段階では、1枚の表で作成される。この段階では、人物IDの確定、原本複写物に書き込まれた付属情報の簡略な分類符号(データ分類タグ)の確認と、簡略な情報付加を行った。

②マスタテーブルの作成[図7]

検索用テーブルは、前項のデータを基に、IDを主キーする「人物マスタ」と「細目マスタ」の2つの表で構成した。その際、各人物の詳細情報は、句点(". ")をデリミタ(区切り)としてさらに細分化した。そして、年月日、官職履歴、親族や記述内容を分類した属性精度、履歴の暦日情報などを、検索項目用に別項目に切り出した。ただし、暦日情報はデータ採録者の解釈による情報でもあるので、原文表記の情報は、表示用フィールドにそのまま残している。

この結果、「人物マスタ」を主表とし、人物のIDを主キーとして「細目マスタ」に接続される、1対多の関係表が作成された。

H	I	K	J	L	M	N	O	B	D	A	C	E
データID	嫡位号	右脇書	名	名レビ	分書1	分書2	参物註	遺書名	編次	編格名	部立	資格
え 07. 03	〇〇〇		中臣@常盤大連公		FX膳田大連公一男 BS始膳任連御物別親			群書類従	巻第六十二	校校稿保己一集	系譜部三	中臣氏系図
え 07. 14	〇	一	@頼基		KX祭主 KX位下 KX少祐 KX大祐 KX少祐 KX大親 NE天曆10年////天曆十年#辛 七十三		NK延喜22年////延喜廿二年#正六位上 任少祐 NK延長5年/1月////延長五年正月#頼大祐 NK承平3年/1月////承平三年正月#頼少祐 NK天曆3年/10月/7日//天曆三年十月七日#頼注	群書類従	巻第六十二	校校稿保己一集	系譜部三	中臣氏系図

①	②	う	え	お	か	あ	い	き	く	け	こ
第1階層	第2階層	嫡親(1人前) ID	子(一人後) ID	兄弟(1人前) ID	兄弟(1人後) ID	兄弟(1人後) ID	兄弟(1人後) ID	性別	姓見出	名見出	氏系図
		え 07. 03	え 07. 04						中臣	常盤大連公	中臣
		え 07. 13	え 07. 15						中臣	頼基	中臣

[図6] 第一段階で作成されるプレマスタテーブル

H	I	K	J	L	B	D	A	C	E
データID	嫡位号	右記号	名	名ルビ	義書名	編次	編者名	部立	資料名
え 07. 03	〇〇〇		中臣@常盤大連公		群書類定	巻第六十二	樹校部集	系譜部三	中臣氏系図
え 07. 14	〇	一	@頼基		群書類定	巻第六十二	樹校部集	系譜部三	中臣氏系図

①	②	う	え	お	か	あ	い	き	く	け	こ
第1階層	第2階層	嫡親(1人前) ID	子(一人後) ID	兄姉(1人前) ID	弟妹(1人後) ID	戻大婦 ID	行大婦 ID	性別	姓見出	名見出	氏系図
		え 07. 03	え 07. 04						中臣	常盤大連公	中臣
		え 07. 13	え 07. 15						中臣	頼基	中臣

《人物マスタテーブル》

H
データID
え 07. 03
え 07. 14

	H	I	K	J	L					M
連番	データID	和暦年	閏	月	日	西暦	小連番	位置	種別	分書1
10	え 07. 03								FX XX	黒田連公一男
20	え 07. 03								BS XX	姫御(1人後)の嫡祖
30	え 07. 14						10		KX XX	祭主
40	え 07. 14						20		KX XX	正四位下
50	え 07. 14						30		KX XX	少祐
60	え 07. 14						40		KX XX	権大祐
70	え 07. 14						50		KX XX	権少祐
80	え 07. 14						60		KX XX	大副
90	え 07. 14	延喜12年				0922	70	頭書	NK XX	延喜十二年#正六位上 任少祐
100	え 07. 14	延長5年	1月			0927	80	頭書	NK XX	延長五年正月#権大祐
110	え 07. 14	承平3年	1月			0933	90	頭書	NK XX	承平三年正月#権少祐
120	え 07. 14	天慶3年	10月	7日		0940	100	頭書	NK XX	天慶三年十月七日#補祭主
130	え 07. 14	天曆10年				0956	110	分書	NE XX	天曆十年#卒 七十三

《細目マスタテーブル》

[図7]第2段階で作成されるマスタテーブル

3.検索・表示系

3.1.暦日情報

人物履歴に記載される暦日情報には、旧暦が使用されるため、「閏月」用の項目を立てた。また、それに対応する西暦は、今のところ、一般的な和暦・西暦対照表を使用して、年のみが宛てられている。これは、学校教育の歴史教科書や歴史年表と同様の扱いである。日本国内の利用だけを考えれば、これで事足りよう。

ただし、今後、歴史を世界的視野（特に中国）での利用・データ構築を目指すためには、共通のキーとなる単位系の採用も必要だろう。

現在、天文学では、1973 年の国際天文学連合（International Astronomical Union:IAU）において、1858 年 11 月 17 日 0 時（UT）JD=2400000.5 日を新起算日とする準（修正）ユリウス日（Modified Julian Day:MJD）が採択されている。しかし、実際にはユリウス日（Julian Day:JD）が使用されることの方が、今なお一般的である。

そこで、本データベースでは、ツール類も整いつつある現状を鑑みて、ユリウス通日やユリウス・グレゴリオ暦との併用も考慮したい。

3.2.漢字の記述・表示法

3.2.1 大漢和・文字鏡収載字の場合

入力文字種は、UCS2.0（いわゆるユニコード）日本語版である JIS X0221-1995 日本語面を使用し、表示・検索系は UTF-8 によった。

その結果、入力不可能な文字種は極端に減少したが、それでもなお覆いきれないものは残った。

それについては、以下の方法で対処した。まず、既存の情報交換用符号文字に存在が認められない文字種は、いわゆる諸橋大漢和番号と上位互換性を持つ「今昔文字鏡」収載の文字番号によるマーキングを施した。当該文字には、文字鏡ホームページ（<http://www.mojikyo.org/>）から提供されるフォントにリンクを貼った。

リンクのアドレスは、今昔文字鏡の 96 ドットのビットマップ画像で、たとえば、文字鏡番号 10456 の文字では、

<http://www.mojikyo.gr.jp/gif96/010/010456.gif>

から、

佑

が画像で表示されるようにした。

【G】内に、その原形推定情報を記述する。方法については、その漢字の組み合わせ方を示す記号を使用する。組み合わせ用記号の意味は下記の通り。

※※組み合わせ記号の演算方法※※

- | | | | | |
|------|--------------|-------------|---|--------------------------------|
| ・ | ：右と左の部位を接合する | 〈例〉【G日・月】 | → | 明 |
| ／ | ：上下に組み合わせる | 〈例〉【G敬／言】 | → | 警 |
| < | ：挿入 | 〈例〉【G門<口】 | → | 問 |
| － | ：右要素を削除 | 〈例〉【G時－寺】 | → | 日 |
| {S*} | ：上下逆 | 〈例〉【G {S了}] | → | 𠂔 ^{①かける。
②男子の陰部。} |
- ・括弧の優先順位は問わず、{ } を繰り返し使用する。
記述例：【G {付一寸}・{術一行}】 → 楯<sup>ジュツ、お
それる。</sup>
- ・全角文字列を使用する。

【図8】漢字の部品組合せによる外字の記述方法

3.2.2大漢和・文字鏡にも採録されない字の場合

大漢和・文字鏡にもない漢字は、漢字の部品を組み合わせた外字表記法を採用した【図8】。

ただし、現在では（2001.6）、ほかにも同種の大規模文字セットデータベース、たとえばGTプロジェクト（<http://www.l.u-tokyo.ac.jp/GT/>）なども公開されており、さらにはJIS X 0213（第3・4水準漢字）や、USC3.1も公開されている。

本プロジェクトで文字鏡を採用したのは、現時点で最もユーザが多いことによる。しかし、今後、文字環境の社会的基盤の安定・進行状況に合わせて、適宜、他の文字セットによる補完・代替を行うことも考えている。

3.3.動態要素検索（種別タグ検索）

原文情報の保持と、検索のパフォーマンスの向上という二つの課題の両立のために、本データベースでは、記事の中にどのような事柄が述べられているかということについて、文中に認められる人間行動の動態要素を符号化して検索可能な属性情報として埋め込んだ。

具体的には、「事柄の始まり」・「事柄の終わり」・「身分」・「父」・「母」など、人間の行動様態に関わる事柄を符号で表現し、検索タグに使用している【図9】。

このタグは、複数の行動要素を組み合わせることによって、通常のデータベース項目に立項される項目と同様の値を求めることができるように体系化を試みたものである。

たとえば、「N：年月日要素」と「E：情報の終わり」という行動要素を示すタグがある。この2種が組み合わせると、薨去・卒去記事を示す記事になる。

例：NEXX 九十九年庚午七月三日#崩、百四十一。

さらに、「I：典拠・出典情報」というタグが加わると、資料引用による記事という意味になる。

例：INEX 東寺長者補任、N 治暦元年#權僧正覺源、八月十八日#卒、六十六。

機能語の組合せ次第では、より柔軟な内容検索も可能となろう。

ただし、システム運用時には、この組合せ要素を基に、選択メニュー形式によるナビゲートも行っている。

また、現段階では、この属性情報は人間の判断により一々に付加作業が行われているが、今後、本データベースのような大量データを形成する場合には、本データを基礎データとして、記事内容の表現を頻度により類型化・分類化することにより、内容を自動的に分類するためのシソーラスを作成することも可能ではないかと考える。

この方式を採用するメリットは、文章表現中に複数のデータベース項目に採録されるべき情報が含まれている場合に、データを複数発生させることなく、属性情報の識別子のみでデータの意味づけが可能であるという点である。特に、本データベースで構築されているような、原本の記述が尊重されるべきデータなどでは、有効な手法とは言えまいか。

記号の意味	1 字目 ← 《配列優先度》	4 字目	
年月日要素	N en		各 1 文 字 だ け が 使 用 可 能
居所・地名要素	A dress		
別名要素	B etumei		
法名要素（仏家）	H ounmyou (Holly name)		
著作・業績要素・業種	C opyright		
父親情報要素（含む養子・出生秘密）	F ather		
母親情報要素（含む養子・出生秘密）	M other		
頭目・師匠情報	T ounmoku (Teacher)		
その他の血縁・親族情報	L ink		
官職情報（無印データに自動発生）	K anshoku		
典拠・出典情報（続紀日…等）	I nformation		
病・けが・謀略・不幸	U nlucky		
引退情報（含出家）	R etire		
情報の始まり（含誕生）	S tart		
情報の終わり（含死亡）	E nd		
原本記載を他資料で修正したもの	W ritten-data		
上記要素以外もの	1 字目～4 字目いずれでも（複数使用可）	X	
参考頭注などの記述でJの人物に直接は関係ない情報	4 字目のみ（複数使用不可）	D ustbox	

【図 9】動態要素検索用機能語タグ一覧

- 1) 1 文字だけのタグについては 2 文字目以降に X を立てる → ○ 印
- 2) 2 文字のタグについては 3 文字目以降に X を立てる → △ 印
- 3) 3 文字のタグには 4 文字目に X を立てる → ▲ 印
- 4) ○ 参考頭注などの記述で J の人物に直接関係ない記述は 4 文字目に D を立てる。（『諸家知譜拙記』に見える。）
- 5) 4 文字目まで埋まらないタグは、X を入れて 4 文字タグ化する。

記述例と意味	○ KXXX: K anshoku 官職位階（本データではデフォルトで K X が立てられる） ○ AXXX: A dress 住居等＝居所・地名要素 △ A E X X: A dress E nd 死亡した場所・葬られた場所 ○ BXXX: B etugo 別号 △ B S X X: B etugo S tart 別号の始まり ○ CXXX: C opyright 著作製作物＝著作・業績要素・業種 ○ EXXX: E nd 死亡等の記事 ○ FXXX: F ather 父や出生秘密等 △ F E X X: F ather E nd 父の死亡 ○ HXXX: H ounmyo (Holly name) 法名（出家名） ○ IXXX: I nformation 典拠・出店情報 例 ○ LXXX: L ink （父母以外の）血縁・親族情報 ○ MXXX: M other 母 ○ NXXX: N endata 年号月日データ △ N A X X: N endata A dress （年月日が特定できる）場所情報 ▲ N A E X: N endata A dress E nd 死亡年月日・場所が特定できる場合 △ N B X X: N endata B etugo 別号発生の年月日データ 意 △ N C X X: N endata C opyright 著作の制作年月日が特定できる場合 △ N E X X: N endata E nd 死亡等の年月日データ △ N K X X: N endata K an 官職関係年月日データ 味 △ N R X X: N endata R etire 隠退年月データ＝得度・入室した日など △ N S X X: N endata S tart 誕生年月日 ○ TXXX: T ounmoku 頭目・師匠情報 ○ UXXX: U nlucky 病・けが・謀略などの不幸な情報 ○ XXXX: その他
--------	---

【図 10】機能語タグの記述例

本研究で対象としたデータは、読解用の辞書も十分には整っていないのが現状である。したがって、このデータを活用することで、新たな史料解析・解説辞書を作り出すこともできよう。

また、このような機能語を叙述要素として埋め込むことで、叙述内容を分析可能とするデータベースを構築することも可能であろう。その際の機能語の体系は、必ずしも本研究にて採用されたものでなくてもよい。様々に独自の観点で体系化されたシソーラスを宛てるならば、よりいっそうフレキシブルで、目的にかなったデータベース検索が行えよう。

4. 検索システム[図11]

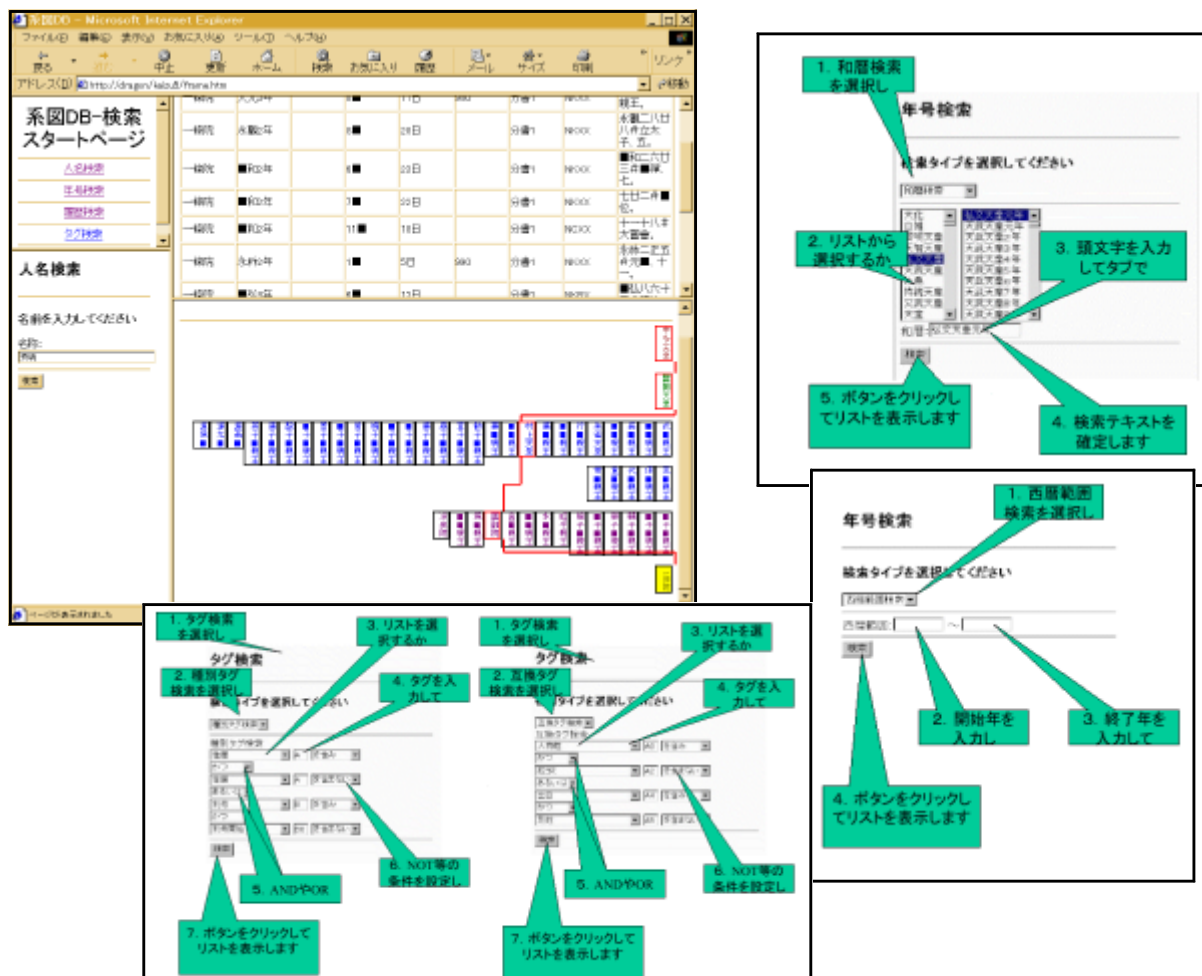
データベースを運用するエンジンについては、設計段階では OODB をはじめ、さまざまな可能性を検討した。しかし、最終的には、汎用性と漢

字セットへの対応状況を考慮して、現在最も汎用的に使用されるアプリケーション（Microsoft Access2000）を採用し、XML のスタイルシートを使用して表示が行えるようにした。

現状の動作環境は MS IE5.0 以上、縦書き表示の乱れを防ぐには IE5.5 以上で、文字のエンコードは「自動選択」にしておく必要がある。

系図表示は5世代が動的に展開すると同時に、人物情報の表示、さらには関連記述からも検索が行えるようになっている。さらに、親子関係をわかりやすくするために、現在検索作業を行っている血族線をわかりやすくインタラクティブに展開する線で表示できるインターフェイスとしたが、これには VRML を使用している。

なお、UNICODE のハンドリングのために サーバ側の OS は Windows2000 以降が要求される。



【図11】検索システムと検索インターフェイス（一部）

5.系図アプリケーションの動向—まとめにかえて—

本研究開始の頃から、人物系図をコンピュータ上で扱うための研究やアプリケーションがいくつか現れ始めた。特に、合衆国では、すでに GEDCOM がデファクトスタンダード（業界標準）的地位を占めつつあり、それをデータ交換プロトコルとしたパッケージ製品も版を重ねている。研究開始当初は、数製品が競合状態にあったようだが、現在は淘汰の段階に入ろうとしている。

一方、日本には GEDCOM に対応した製品はない。これは、日本が欧米製品とは無縁に開発を行ったことに起因するようである。

もっとも、それ以前に、採録される情報の質的差異、つまり文化的・社会的風土の差異を欧米型の現代的フォーマットに無理矢理押し込めようとする事は、本研究も含め、現段階では難しい。

そこで、本研究以外に系図に取り組んだアプリケーション、研究について述べてみたい。

5.1.GEDCOM

欧米で現在標準的なプロトコルとしての地位を得つつある GEDCOM は、現時点でのバージョンは 5.5 を数えるが、本来、the Church of Jesus Christ of Latter-day Saints' Family History Department により開発され、Brother's Keeper, Family Edge, Family Origins, Family Tree Maker, PAF, Roots, The Family History などの製品がこの形式の入出力に対応、付加機能を競っている。

GEDCOM は、SGML などと同様に、規格先行型のプロトコルである。現在合衆国で発売される系図ソフト間でのデータ交換形式の一つとしてとして機能はしているものの、GEDCOM データとの完全互換はとれていない。その事情は、日本の住所録ソフトの交換プロトコルである J-ADDRESS の状況にも似ている。

5.2.日本の系図アプリケーション

本研究開始当初、人物系図を扱うアプリケーションは、日本では低調で、管見の限りでは中国（大陸・台湾）にも見あたらなかった。

しかし、現在、日本でもいくつか家系図ソフトと称する系図アプリケーションが登場しつつあり、それぞれ独自の展開を見せ始めている。いずれも csv 形式のデータ交換プロトコルは用意されているが、人物関係情報や、他の項目の意味レベルで交換可能なプロトコルは未確立である。

また、系図自体の表示機能は、グラフィックモードで系図を描き、それをアプリケーションの内部機能で整形する方式のものが過半を占める。整形は、アプリケーションにより行えはするが、階層関係情報自体の入出力機能はない。この点を、いかにして標準的なプロトコルにまで発展させて行かかが、今後の課題であろう。

いずれにしても、日本における系図アプリケーション、データベースは黎明期の状態に入ったといえる。だが、今後、人名典拠録に代表される人物に関するアーカイブを充実させるためには、系図・系譜についての情報は不可欠であろう。そのためにも、系図を管理するアプリケーション、データベース、そして、そのデータ交換のための標準的なプロトコルの確立が、ますます重要となる。

〔人物系図を扱う日本のアプリケーション製品〕

- ①家康 16,800 円。ソーゴ印刷株式会社制作・販売。
<http://www.sogo-printing.com/ttop.html>
- ②源 GEN.valis.to (株)ヴァリス制作。11,500 円。
<http://www.gen.valis.to/>
- ③ファミリーウェア 5.0 (家系図ソフト「親戚まっぷ 5.0」同梱) シェアウェア。(株)石川コンピュータ・センター制作。<http://www.incl.ne.jp/fam/simap/>
- ④ LifeTree Ver1.0A ライツフリー制作委員会 4,500 円。
<http://www1.newweb.ne.jp/wb/lifetree/>
- ⑤家族の記録 2 ファイルメーカー (3・4 対応) 使用のシェアウェア (500 円)。(C)Kazuyoshi Seza 制作。
<http://www2.freeweb.ne.jp/~sirius/family/>
- ⑥ Field Note アボリジニの複雑な家族・親族関係についてのフィールドノート研究の一環として開発された系図ソフト。ファイルメーカー使用。
<http://www.kusa.ac.jp/~shuhei/FNnow/>
- ⑦「どんクジラ」(土谷重幸)氏の自作系図ソフト
<http://www.tsuchiya.com/keizu/ten.html/> DOS 版、JAVA 版、JavaScript 版あり。新規入力対応。ボックス型。
- ⑧血統王 (株)エム・エス・ディ・ジャパン制作。9,800 円。種牡馬 6,500 頭以上、繁殖牝馬 25,000 頭以上(ダービースタリオン繁殖データ登録済み)、現役馬 4,000 頭以上(地方競馬の競走馬も一部登録)登録。
<http://www.msdl.com/j/products/keiba/pd97102.html>