

## 網羅性を重視した古事類苑データベース

古瀬蔵<sup>†1</sup> 相田満<sup>†1</sup> 山田太造<sup>†2</sup>

『古事類苑』は日本の前近代の諸事象を記した類書形式の百科事典である。本稿では、『古事類苑』の全30部を検索対象とするために構築した、全文テキストと抜粋テキストを併用した全文・抜粋検索版データベースについて述べる。抜粋テキストは総目録、書名、解説、索引から成る。全文・抜粋検索版データベースは、2012年8月にインターネット上で公開され、現在、『古事類苑』の26部を検索対象としている。

### *Kojiruien* Database with Broad Search Scope

OSAMU FURUSE<sup>†1</sup> MITSURU AIDA<sup>†1</sup>  
TAIZO YAMADA<sup>†2</sup>

*Kojiruien* is an encyclopedia, which contains cultural, political, commercial, and other numerous matters of pre-modern Japan, citing plenty of books and documents. Aimed at searching all 30 parts of *Kojiruien*, we implemented a new database system by integrating extracted data into our previous full-text database. This paper describes our new database system, especially concerning extracted data, which consists of catalogue table of contents, book titles, abstracts, and index terms. Our new database has been available on the Internet since August 2012, and its search scope currently covers 26 parts of *Kojiruien*.

#### 1. はじめに

『古事類苑』は、日本の前近代の制度や文物などの事項を30部に分けて類書形式で編纂された百科事典である。歴史学など人文科学の研究には重要な資料であり、本文1,000巻、洋装本で51冊(60冊版もあり)、和装本で350冊にのぼり、完成から約1世紀が経とうとしている今なお、百科事典として日本最大規模を誇っている。『古事類苑』の30部全体を網羅する画像データベースは、国際日本文化研究センター(以下、日文研と呼ぶ。)や国立国会図書館から既に公開されている[3][4]。

一方、テキストデータベースの構築も、国文学研究資料館(以下、国文研と呼ぶ。)と日文研の共同で進められてきた[1][6]。しかし、『古事類苑』に内在する、膨大な分量、多様な書式、規格外の文字の頻出といった要因のため、30部すべての全文テキスト化には膨大なコストと時間を要し、現在までの作業の進捗のままでは『古事類苑』全体の網羅は見通しが立たない状況である。

本稿では、全文テキストと抜粋テキストを併用することにより、テキスト検索において30部の網羅を実現させる、新しい古事類苑データベース(以下、全文・抜粋検索版データベースと呼ぶ。)について述べる。このデータベースは、2012年において全文テキスト作成完了あるいは作成完了予定の8部について、全文検索の対象とする。残りの22部については、作成の効率化を図った抜粋テキストによって、『古事類苑』全体のテキスト検索を可能にする。テキスト検索の結果から、テキスト閲覧が可能な画像データベ

スやテキストデータベースを参照するための索引的な役割を演じる。テキスト検索に目的を重点化することにより、『古事類苑』全体を網羅することを可能にする。全文テキストと抜粋テキストの作成は未完了であったものの、『古事類苑』の26部を検索対象とした全文・抜粋検索版データベースを2012年8月にインターネット上で公開した。

以下、2節で、国文研 web サイトから公開している全文テキストのデータベース(以下、テキスト版データベースと呼ぶ。)について、3節で、テキスト版データベースを全文検索するためのデータベース(以下、全文検索版データベースと呼ぶ。)について、4節で、22部で作成を開始した抜粋テキストについて、5節で、全文検索版データベースを更新して、抜粋テキストの検索もできるようなシステム更新を行った全文・抜粋検索版データベースについて、6節で、知識ベースとしての『古事類苑』の役割について、述べる。

#### 2. テキスト版データベース

2006年4月に国文研サイトでテキスト版データベースの試験公開を開始し、その後正式公開に移行した、通読に適するよう構築したインターフェースと、総目録の部立情報から該当箇所を引くことができるなどの機能により、『古事類苑』のテキストが閲覧できる。2012年9月時点で、「天部」、「歳時部」、「地部」、「稱量部」の全文テキストを公開済である[2]。「帝王部」、「方技部」、「飲食部」、「植物部」についても、公開のための作業を進めている。洋装本『古事類苑』原本の高精度な全文テキストを作成するために、3回の校正作業に加えて、規格外文字を入力するための外字の導入作業を行っている。また、原本の書式を再現するための、以下のような特色がある。

<sup>†1</sup> 国文学研究資料館  
National Institute of Japanese Literature  
<sup>†2</sup> 人間文化研究機構  
National Institutes for the Humanities

- ・IE や Lunascape ブラウザでの縦書表示
- ・ルビ、返り点、割注の表記
- ・『古事類苑』収載図版の挿入

テキスト版データベースは、原本と同等に『古事類苑』の情報を漏れなく伝えることができる。しかし、全文テキストの作成に相当な費用と時間を要する。2006年に公開を開始したテキスト版データベースは、6年を経過した2012年9月時点で4部門の公開に留まっている。

### 3. 全文検索版データベース

テキスト版データベースの公開に続いて、全文検索版データベースを2008年12月に公開した。当時の検索対象は、「天部」、「歳時部」、「地部」(洋装本3冊のうち1冊のみ)であったが、テキスト版データベースの拡張に伴い、その後、「地部」の残りと「稱量部」も検索対象となった。全文検索版データベースは、国文研 web サイトから公開後、人間文化研究機構の統合検索システム(現在のシステム名は nihuINT[7])の検索対象データベースに加わった。統合検索システム nihuINT は、人間文化研究機構の所属機関などが持つ100以上のデータベースを横断検索できる。

図1に、「歳時部」和巻十九1397ページの引用書『守貞漫稿』に関する全文検索版データベースの検索結果と、テキスト版データベースとの対応を示す。全文検索版データベースの検索結果が該当する、国文研のテキスト版データベースと日文研の画像データベースの箇所を参照することができる。

## 4. 抜粋テキスト

### 4.1 22部での抜粋テキストの作成

2011年1月に国立民族学博物館で、「人間文化研究情報資源と知識ベース」というテーマで人間文化研究情報資源共有化研究会が開催された。その機会に、古事類苑データベースの構築に関するアンケートを実施したところ、30部全体の網羅と全文検索が重要との回答が多いという結果が得られた。

『古事類苑』全体30部を網羅するために、全文テキストの作成に未着手の部について、抜粋テキストの作成を2011年11月に開始した。抜粋テキストを作成するのは、「神祇部」、「官位部」、「封禄部」、「政治部」、「法律部」、「泉貨部」、「外交部」、「兵事部」、「武技部」、「宗教部」、「文学部」、「禮式部」、「樂舞部」、「人部」、「姓名部」、「産業部」、「服飾部」、「居處部」、「器用部」、「遊戯部」、「動物部」、「金石部」の22部である。抜粋テキストは、閲覧用のデータでなく、情報検索のためのデータとして位置づけられる。抜粋テキストの検索結果からは、テキスト版データベースへの参照はないが、画像データベースで原文への参照は可能である。すなわち、抜粋テキストを検索して、画像データベースで原文を閲覧するという役割分担になる。『古事類苑』全体30部を網羅するデータベースを早期に実現するために、抜粋テキストにおいては、作成の効率性を重視し、マークアップ言語のタグは使用せず、CSVによる単純なデータ形式を採用する。

**テキスト版データベース**

**全文検索版データベースの検索結果詳細表示**

項目	内容
ID	<<前の本文へ 020190140 次の本文へ>>
部	歳時部十九
篇	節分
解説	節分の、古来立春、立夏、立秋、立冬ノ前日ヲ云ヒシガ、後世ハ専ラ立春ノ前日ノミノ稱トナレリ、此後每戸懸豆ヲ禮シ、悪鬼ヲ驅逐ス、蓋シ追儺ノ経風ナリ。
種、付属文字	厄拂
引用書名	【◆V守貞漫稿-N二十七】
引用書本文	厄拂 京極ハ節分ノ夜ノミ来ル、江戸モ古ハ節分ノミナリシガ、文化元年以來、大卅日、正月六日、十四日ニ来ル、追儺ノ豆、大坂ハ年數ヲカシヘ一掃ヲ加ヘ、白紙ニ包ミ與フ、江戸ハ十二掃ヲ添ルナリ、又京極ハヤクハライマシヤウト云、江戸ハオンヤク——ト、御宇ヲ付ル、厄拂ノ聲ノミ、音節及ビ文句トモニ三都相似タリ、蓋文句ハ年々種々アリ、アハラ、メデタイナ、メデタイナ、ダンナ住吉御參詣、ソノ横カヲ西ナガムシ、七福神ノ稻アソビ、中ニモ美ト云人ハ、命長納ノ種ヲモチ、メキスオキスノ糸ヲツケ、金ト銀トノ針ヲタシ、釣タル横カヲ懸小瓶、カホドメデタイナカラニ、イカナル悪魔ヲ来ルトモ、此厄ハ七ツツラヘ、西ノ海トハオモヘドモ、チクラガ中ヘサラリ、或ハ役者名處シ、魚盡、何處ナド種々ヲ云、
ページ	1397 ページ テキスト : <a href="http://base1.nijl.ac.jp/~kojiruien/sajibu/frame/F001397.html">http://base1.nijl.ac.jp/~kojiruien/sajibu/frame/F001397.html</a> 画像 : <a href="http://shinku.nichibun.ac.jp/kojiruien/html/saj_1/saj_1_1397.html">http://shinku.nichibun.ac.jp/kojiruien/html/saj_1/saj_1_1397.html</a>

**雑載**

廿五日、○大正六年十一月節分の夜、○中略京には没落して、  
 て、かまふれば我八十の御事録やくとていかにおしやるべき

**対応**

帯巻手記  
 節分ハ昔ノ節分ニ似テ、今亦節分ト云フ、厄ハ昔ハ二部四段ノ人儀ヲ懸置キ、同じ横カヲ懸カシ、節分ノ夜ハ、  
 間、(口ニ時ノ種)ありきとし、文化より以來、冬ニ降夜正月の年越毎に来る。

守貞漫稿 二十七  
 厄拂 京極ハ節分ノ夜ノミ来ル、江戸モ古ハ節分ノミナリシガ、文化元年以來、大卅日、正月六日、十四日ニ来ル、追儺ノ豆、大坂年數  
 ヲカシヘ一掃ヲ加ヘ、白紙ニ包ミ與フ、江戸ハ十二掃ヲ添ルナリ、又京極ハヤクハライマシヤウト云、江戸ハオンヤク——ト、御宇ヲ付  
 ル、厄拂ノ聲ノミ、音節及ビ文句トモニ三都相似タリ、蓋文句ハ年々種々アリ、アハラ、メデタイナ、メデタイナ、ダンナ住吉御參詣  
 ソノ横カヲ西ナガムシ、七福神ノ稻アソビ、中ニモ美ト云人ハ、命長納ノ種ヲモチ、メキスオキスノ糸ヲツケ、金ト銀トノ針ヲタシ、  
 釣タル横カヲ懸小瓶、カホドメデタイナカラニ、イカナル悪魔ヲ来ルトモ、此厄ハ七ツツラヘ、西ノ海トハオモヘドモ、チクラガ中ヘ  
 サラリ、或ハ役者名處シ、魚盡、何處ナド種々ヲ云、

現代之史乘不可誦歎也、我輩嘗  
 於于江家(安)第、其時(節)分、未  
 幾角如刺、惡鬼如鬼也、又種

図1 全文検索版データベースとテキスト版データベースの対応



#### 4.5 書名と編目位置

総目録の事項のそれぞれの内容には、『古事記』や六国史など古代の文献から明治以前までの様々な文献を対象に関連箇所を引用した資料が収載されている。これらの資料は、引用書と、引用書に関連する参考書の二階層があり、引用書と参考書は、書名と資料本文から構成される。

『古事類苑』は類書の範疇に属し、どのような資料が引用されているか、他の文献では探すことが難しい資料が存在していないかなどの情報を、書名（図名の場合もある。）のデータによって提示することが期待できる。書名は、書名の補足情報である編目位置と共に、『古事類苑』の原本において縦書きの角括弧で表示されており、抽出作業は困難ではない。図5に、抜粋テキストにおける、官位部和卷四洋装本1冊41ページの「神服部」條の書名データを示す。例えば、「神服部」條の最初に、編目位置が「二十九天武」の『日本書紀』が引用されていることを示す。次行に記述されている、編目位置が「中編四加」の『倭訓栞』は、『日本書紀』の引用箇所に関連する参考書であり、編目位置が「十加」の『姓氏考』、編目位置が「十二錦綺」の『倭名類聚抄』、編目位置が「二十二」の『古事記傳』は、『新撰姓

氏録』に関連する参考書である。

2012年9月時点で、抜粋テキストの書名データが完了しているのは、「神祇部」、「官位部」、「政治部」、「法律部」、「文学部」、「禮式部」、「人部」、「姓名部」、「産業部」、「器用部」、「金石部」の11部である。11部での書名の異なり数は、厳密な名寄せ作業は行っていないものの、約8,000に達した。残りの11部についても、2012年度中に、書名のデータ化を完了させる予定である。

表1は、データ化が完了した11部それぞれについて、引用書または参考書として出現した件数が上位の書名を示す。「法律部」での『御仕置例類集』、「産業部」での『萬金産業袋』などのように特定の部において高頻度で出現する文献がある一方、『日本書紀』などの六国史、『延喜式』などの法典、『倭名類聚抄』などの辞書などの文献は各部に共通して高頻度で引用されている傾向がある。

#### 4.6 索引

書名に後続して多様な様式で記述されている資料本文は、『古事類苑』のテキストの大部分を占め、全文テキスト作成で労力が最もかかる部分である。資料本文を全く含めず、総目録、解説、書名だけで抜粋テキストを構成すると、

ページ	部	篇	條	解説
189	神祇部四			
189		神體		
189				神體ハ、正體又ハ御形トモ、靈體トモ云フ、神體ニハ鏡ヲ以テスルアリ、玉ヲ以テスルアリ、石ヲ以テスルアリ、兵器ヲ以テスルアリ、影像ヲ以テスルアリ、兵器ニハ弓アリ、矢アリ、劍アリ、矛アリ、影像ニハ木像アリ、畫像アリ、佛説ノ之ニ混ジテヨリ以後ハ、佛、菩薩、沙門ノ像ヲ以テ、神體トスルアリ、此外ニ鈴ヲ以テスルアリ、笏ヲ以テスルアリ、釜ヲ以テスルアリ、或ハ神名ヲ記シテ神體トスルガ如キハ、影像ニ近キモノナリ、而シテ幣帛ヲ以テ神體トスルハ、特ニ後世ノ事タリ、權代及靈宮ハ、神體ヲ納ムル具ニシテ、船代ハ權代ヲ納ムル具ナリ、並ニ此ニ併載ス、
189			名稱	

図4 解説データ

ページ	部	篇	條	項	引用書名	編目位置	参考書名	編目位置
41		神服部						
41					日本書紀	二十九天武		
41							倭訓栞	中編四加
41					新撰姓氏錄	和泉國神別天孫		
41							姓氏考	十加
41							倭名類聚抄	十二錦綺
41							古事記傳	二十二
41				長幡部				
41					古事記	中開化		
41							古事記傳	二十二
41					常陸風土記	久慈郡		
42							類聚國史	五十四人
42							延喜式	二十四主計
42		神麻績部						
42					古語拾遺			
42							倭訓栞	前編五乎

図5 書名データ

表1 各11部での頻出書名

部	高頻度の書名 (左から右に高頻度順)
神祇部	延喜式, 三代實録, 日本紀略, 百練抄, 吾妻鏡, 中右記, 日本書紀, 續日本紀, 玉海, 箋注倭名類聚抄
官位部	續日本紀, 延喜式, 吾妻鏡, 吏徴, 令義解, 憲教類典, 日本書紀, 吏徴別録, 三代實録, 明良帶録
政治部	續日本紀, 延喜式, 令義解, 三代實録, 吾妻鏡, 類聚三代格, 日本書紀, 日本紀略, 令集解, 日本後紀
法律部	御仕置例類集, 御定書百箇條, 吾妻鏡, 續日本紀, 令義解, 科條類典, 御仕置裁許帳, 律疏, 百練抄
文学部	日本教育史資料, 延喜式, 先哲叢談, 袋草紙, 古今著聞集, 日本紀略, 江談抄, 源氏物語, 三代實録
禮式部	玉海, 日本書紀, 榮花物語, 日本紀略, 中右記, 續日本紀, 吾妻鏡, 三代實録, 台記, 令義解
人部	日本書紀, 伊呂波字類抄, 倭訓栞, 書言字考節用集, 類聚名義抄, 倭名類聚抄, 箋注倭名類聚抄
姓名部	續日本紀, 日本書紀, 三代實録, 新撰姓氏錄, 續日本後紀, 寛永諸家系圖傳, 日本後紀, 古事記
産業部	延喜式, 倭名類聚抄, 和漢三才圖會, 倭訓栞, 日本書紀, 箋注倭名類聚抄, 萬金産業袋, 類聚名義抄
器用部	延喜式, 倭訓栞, 倭名類聚抄, 江家次第, 嬉遊笑覽, 和漢三才圖會, 箋注倭名類聚抄, 守貞漫稿
金石部	本草綱目譯義, 倭名類聚抄, 紀伊續風土記, 箋注倭名類聚抄, 大和本草, 重修本草綱目啓蒙, 佐渡志

『古事類苑』のテキスト検索の入力にヒットする件数が少なくなる恐れがある。そのため、資料本文の中で、『古事類苑』の索引に該当する箇所について、抜粋テキストに含めることにした。索引は調べたい項目が存在する箇所を効率的に参照するために利用され、索引に記載される項目（以下、索引（索出事項）と呼ぶ。）は、データベースの情報検索の入力となる可能性がある。『古事類苑』全30部が完成した直後に著作された旧版の索引が記す索出事項に対して、『古事類苑』原本から索引（本文）の抽出を行い、索引データを作成している。旧版の索引は、旧仮名遣いで配列され索出事項を探しにくく、誤りが多い。そのため、索引（本文）の箇所の決定には、新仮名遣いで配列し直し、多くの誤りを修正した新版の索引[5]を主に利用している。

しかし、抜粋テキストに索引データを加えることを難しくしている最大の問題は、誤りや仮名遣いではなく、索引（索出事項）と、索出事項に該当する『古事類苑』の本文の箇所（以下、索引（本文）と呼ぶ。）の表記が必ずしも一致しているわけではないことである。例えば、索引（索出事項）「男山八幡宮」は、資料本文では「石清水八幡宮」で出現することが多いし、索引（索出事項）が「太田持資志和歌」のように要約的表現である場合や、資料本文は和文であるのに、索引（索出事項）は「以長歌稱短歌」のように漢文形式の場合がある。このような場合、索引（索出事項）に対応する索引（本文）は正解が明らかでなく、抜粋データに加える索引（本文）を決めるのに時間が掛かりがちとなる。また、索引（索出事項）と一致する表記があっても、文脈情報として周辺テキストも提示しないと意味が分かりにくい場合や、情報検索の入力となる可能性がある語句が周辺テキストにあれば、それも索引（本文）に含めたい場合がある。このような場合は、索引（索出事項）に該当するテキストを広めに取って、抜粋テキストに含めるように努めている。索引（本文）を拡大すると抜粋テキストの情報は充実するが、22部を網羅させるのに時間が掛かるというトレードオフがある。

索引（本文）と索引（索出事項）が一致する例、一致しない例を、それぞれ図6と図7に示す。図6では、「人部」の「諺」篇において、索引（索出事項）に一致する「重荷に小附」が『年々隨筆』の資料本文に、「人間萬事塞翁馬」と「綸言如汗」が『下學集』の資料本文に出現することを示す。図7は、「文學部」の「和學」篇において、索引（索出事項）の「源氏物語傳授」に完全照合する表現はなく、『消閑雜記』の資料本文に、索引（索出事項）に該当すると思われる「源語秘訣といふ一卷に、源氏の大事を十五箇條あらはして、傳授することなり、」という表現があることを示している。

索引（本文）の抽出は、作業者の判断に依存するため、解説や書名のデータ化に比べ、時間を要する。2012年9月

引用書名	編目位置	参考書名	編目位置	索引（本文）	索引（索出事項）
後撰和歌集	二十頁				
		年々隨筆	四		
				重荷に小附	重荷に小附
言志錄					
歌臺雜話	一				
下學集	下言辭			人間萬事塞翁馬	人間萬事塞翁馬
朝倉始末記	三				
平家物語	七				
下學集	下言辭				
				綸言如汗	綸言如汗
平家物語	三				

図6 索引データ（本文と索出事項が一致する場合）

引用書名	編目位置	索引（本文）	索引（索出事項）
基照卿記			
消閑雜記			
		源語秘訣といふ一卷に源氏物語傳授 源語秘訣 源氏の大事を十五箇條といふ一卷に、源氏の大あらはして、傳授する事を十五箇條あらはして、ことなり、	源語秘訣 源氏の大事を十五箇條あらはして、傳授することなり、
源語秘訣			
駿府政事錄			
東照宮御實紀附錄	二十二		

図7 索引データ（本文と索出事項が一致しない場合）

時点で、22部のうち、抜粋テキストでの索引のデータ化が一通り終了しているのは、「姓名部」と「金石部」、50%以上終了しているのは、「神祇部」と「文學部」に留まっている。索引（索出事項）の一部は、総目録、解説、書名に該当箇所があり、既にデータ化されていると見なせる場合もある。それでも、索引データを22部に網羅させるには時間がかかるが、少しずつ索引データの充実化を進めていく予定である。

## 5. 全文・抜粋検索版データベース

全文テキストを検索対象とする全文検索版データベースに対して、抜粋テキストも検索対象に含めるように検索機能を拡張した、全文・抜粋検索版データベースを構築し、2012年8月から国文研 web サイトから公開した[2]。2012年9月時点で、4部について全文テキストを、22部について抜粋テキストを、検索対象としている。抜粋テキストの作成状況については、4節を参照のこと。

全文テキストと抜粋テキストでは、『古事類苑』原本から抽出している情報の密度に差がある。そこで、検索結果の単位（以下、レコードと呼ぶ。）を、全文テキストでは、引用書・参考書（書名と本文）または解説ごととし、抜粋テキストでは、ページごととしている。抜粋テキストでは、解説や資料本文で、ページ跨ぎの語句などが検索入力キーワードとなりそうな場合には、次ページに跨がる表現をレコードに含めている。

各レコードを、項目とその値の組の集合によって記述する。項目は、原本の記述である本文と、原本には記述されていないメタデータに大別される。抜粋テキストでは解説、



図 8 全文・抜粋検索版データベースの検索画面

書名、索引(本文)などが本文の項目であり、レコード ID、ページ番号、リンク先の画像データベースの URL、索引(索出事項)などがメタデータである。総目録の部、篇、條、項、綱、目(全文・抜粋検索版データベースでは、項から下位の階層は付属文字として扱っている。)はメタデータの項目であるが、部、篇、條、付属文字が本文で表記されているページレコードでは、本文の項目にもなっている。

図 8 に、全文・抜粋検索版データベースの簡易検索と詳細検索の web ページ画面を示す。初期表示される簡易検索では、本文全体を対象に、全文テキストと抜粋テキストをシームレスに検索を行う。詳細検索では、メタデータの検索、範囲を限定した本文の検索も可能である。索引(索出事項)についての検索は、抜粋テキストのみが検索対象となる。

基本的に旧字体で記述されているテキストを検索するために、異体字や仮名の表記揺れを正規化して検索する機能を導入している。

図 9 に、検索結果の表示画面を示す。初期画面はスニペット形式である。「断片」という意味のスニペットは、情報検索では、検索キーワードが含まれる抜き書きのことを指す。検索入力がどのようにヒットしたかが分かるため、テキスト検索において、必要な検索結果を選択するのに有効な表示形式である。スニペット形式から表形式に切り替えることができる。表形式は、部と和巻番号の組ごとに検索

結果を分類して表示する。

図 10 に、抜粋テキストのレコードについて、検索結果詳細表示の例を示す。本文の項目を中央に表示し、その上下にメタデータの項目を表示している。

詳細表示では、引用書や参考書の本文について、書名と編目位置との組を角括弧で囲んで表記する。例えば、[倭訓栞 前編六加] は、書名が『倭訓栞』、編目位置が「前編六加」であることを表している。書名『七福神考』には編目位置の情報がなく、[七福神考]と表記する。

抜粋テキストのページ単位のレコードからは、日文研の画像データベースへのリンク情報があり、原本の画像を参照できる。全文テキストのレコードからは、検索機能の更新前の全文検索版データベースと同様に、検索結果が該当する日文研の画像データベースと国文研の全文テキスト版データベースの箇所を参照することができる。

全文・抜粋検索版データベースは、2012 年 9 月時点で公開されてからあまり時間がたっていないため、利用状況からの評価はこれからの状況である。データベースの有用性を利用者にも感じてもらうためには、多様な検索キーワードに対して、検索結果がない、あるいは、有用な検索結果が含まれていないといった事例を減少させなければならない。検索件数だけでなく、検索結果の件数、同一利用者の連続検索件数、再訪状況など様々な指標で、評価を行う予定である。

**表形式**

1 歳時部十三 // 年始雑戯 // 初夢、二日夜初夢 // 892ページ  
全文テキスト / ID : 020130126  
引用書本文 : ... (印帳ナリ民間二重丸巻ハ、七福神成) 寶龜等ヲ量久、(寶龜、下子、...

2 歳時部十九 // 節分 // 厄掃 // 1397ページ  
全文テキスト / ID : 020190140  
引用書本文 : ... 註、ノハ横カヲ西ヲナガムレバ、七福神ノ形アズ、中ニモ裏ト云人ハ、...

3 神祇部一 // 神祇總載 // 三十番神七福神 // 洋装冊数1 // 88ページ  
全文テキスト / ID : 0410088  
引用書本文 : ... 三十番神七福神 引用書名 : ... 言字考前編 十数巻 (日本七福神考 凡例(七福神考))

4 神祇部一 // 神祇總載 // 七福神動物植物爲神 // 洋装冊数1 // 90ページ  
全文テキスト / ID : 0410090  
引用書本文 : ... 七福神動物植物爲神 索引(本文) : 七福神考

5 神祇部四 // 神體 // 影像 // 洋装冊数1 // 212ページ  
全文テキスト / ID : 0410212  
引用書本文 : ... (日本七福神考) (神祇考) (新編武藏)

6 神祇部五十 // 神使 // 鼠 // 洋装冊数2 // 1835ページ  
全文テキスト / ID : 0421835  
引用書本文 : ... (倭訓栞 前編六加) (七福神考)

7 神祇部五十 // 神使 // 猪蛇 // 洋装冊数2 // 1854ページ  
全文テキスト / ID : 0421854  
引用書本文 : ... (五) (茶室漫錄 甲) (日本七福神考)

**スニベット形式**

部

No. ID / 篇 / 條 / 目 / 項 / 索引(引用書本文) / 索引(本文) / 洋装冊数 // ページ

歳時部十三 : 全文テキスト

1 020130126 // 年始雑戯 // 初夢、二日夜初夢 // (◆) 守貞漫稿(二十四) // 正月二日、今夜寶龜ノ神ヲ枕下ニシキテ寐ル也。(昔ハ節分ノ夜行ノ之、伊勢伊弉守、定非(國府)連ノ上ニ、將軍大引、其他小引、下ノ引合、末女ハ形原、皆匿名也。)今世華裏二月(玉ツハ)、身ニ米俵ヲ懸ム(國也、(印帳ナリ)民間二重丸巻ハ、七福神成)寶龜等ヲ量久、(寶龜、下子、打出娘、分撰、カクレ親、カクレ五等也、鏡、七寶、)京坂ハ近世廟堂ス、江戸ハ今モ専ラ元日二日ノ宵ニ、小民實ノ之述ル、寶龜ノ形也、道中覽ハ、(形)ヲ量久也、其日、道中覽六、オタカラノ、今夜ノ夢ヲ初夢ト云、故ニ言夢ヲ見ト云寶龜ノシゴト也。 // 892ページ

歳時部十九 : 全文テキスト

2 020190140 // 節分 // 厄掃 // (◆) 守貞漫稿(二十七) // 厄掃、京坂ハ節分ノ夜ノ出来ル、江戸モ古ハ節分ノ夜ナリガ、文化元年以來、大黒日、正月六日、十四日ニ来ル。道徳ノ旨、大衆ノ年数ヲカヘテ一掃ノ功加ヘ、白紙ニ包ミ、江戸ハ十二歳ノ幼キナリ、又京坂ノヤクソイマツケト云、江戸ハオノヤクソ、御子ヲ付ル、厄掃ノ旨、昔節分及文化ノモニ三都府相違ナリ、蓋文ヲイテ洋キ種々アリ、アホス、メデタメ、メデライナ、ダンナ住吉御參詣、シノ横カヲ西ヲナガムレバ、七福神ノ形アズ、中ニモ裏ト云人ハ、(形)ヲ量久也、鏡、七寶、メデタメオキス(坐ヲカ、坐ト掛ト針ヲカ、約カハ横カ掛ト掛、カネメデタメオキカカニ、イカナル惡魔ヲ取ルトモ、此厄ハウセカヒントラヘ、西ノ鬼トハオモヘドモ、テカラガキヘナリ、或ハ役者名盡シ、魚、何處ナド種々云、 // 1397ページ

神祇部一 : 全文テキスト

3 0410088 // 神祇總載 // 三十番神七福神 (勅書記) (言字考前編 十数巻) (日本七福神考 凡例(七福神考 序)) // 白紙の塊にたまるを見て、御心安くお出のされぬ、稲荷の大黒神の御事にておはす御財天長門天大黒天兼比叡壽壽布袋廣福老人 // 洋装冊数1 // 88ページ

4 0410090 // 神祇總載 // 七福神動物植物爲神 // 日本書紀 一時代(古事記 上) // 七福神考 // 洋装冊数1 // 90ページ

神祇部四 : 全文テキスト

5 0410212 // 神體 // 影像 // (日本七福神考) (神祇考) (新編武藏風土記 西二十多摩郡) (古事記 五神紀傳考) // 三寶堂 // 洋装冊数1 // 212ページ

神祇部五十 : 全文テキスト

6 0421835 // 神使 // 鼠 // (倭訓栞 前編六加) (七福神考) // 鼠を大黒の使といふは二義あり、大己食命にていへば、古事記に鼠の故事見えたり、大黒天にていへば、聖寶藏神經に左手持鼠囊と見えたり。 // 洋装冊数2 // 1835ページ

7 0421854 // 神使 // 猪蛇 // (倭訓栞 前編五) (茶室漫錄 甲) (日本七福神考) // 猪蛇神爲愛宕之靈物、御財天女傳 承日、中略 江戸竹生島、相州工島、以此爲神使者、是取理有宇賀神玉現自蛇形之事也。 // 洋装冊数2 // 1854ページ

表示切り替え

図 9 検索結果の表示形式

**全文・抜粋検索版データベース 検索結果詳細表示**

項目	内容	
ID	<<前のページへ	0421835
部	神祇部五十	次のページへ>>
篇	神使	
條	鼠	
索引(索出事項)	以鼠爲神使 一鼠 以鼠爲大黒天ノ使 一鼠を大黒の使といふは二義あり、大己食命にていへば、古事記に鼠の故事見えたり、大黒天にていへば、聖寶藏神經に左手持鼠囊と見えたり。	
索引(本文)	鼠を大黒の使といふは二義あり、大己食命にていへば、古事記に鼠の故事見えたり、大黒天にていへば、聖寶藏神經に左手持鼠囊と見えたり。	
参考書名	〔古事記 上〕	
引用書名	〔七福神考〕	
洋装冊数	2	
ページ	1835ページ	
	画像 : <a href="http://shinku.nichibun.ac.jp/kojruen.html/jing_2/jing_2_1835.html">http://shinku.nichibun.ac.jp/kojruen.html/jing_2/jing_2_1835.html</a>	

画像データベース(日文研)にリンク。

対応

抜粋テキスト

ページ	部	篇	條	目	項	引用書名	編目位置	参考書名	編目位置	索引(本文)
1835			鼠							
1835						倭訓栞	前編六加			
1835										鼠を大黒の使といふは二義あり、大己食命にていへば、古事記に鼠の故事見えたり、大黒天にていへば、聖寶藏神經に左手持鼠囊と見えたり、
1835								古事記	上	
1835						七福神考				

図 10 抜粋テキストの検索結果詳細表示

## 6. 知識ベースとしての役割

統合検索システム nihuINT は、人間文化研究機構の所属機関などが持つ 100 以上のデータベースを横断検索できる。

『古事類苑』の全文・抜粋検索版データベースは、全文検索版データベースを継承して、nihuINT が検索対象とするデータベースに含まれている。古事類苑データベースが nihuINT への貢献に期待されているのは、検索結果の充実だけではない。知識ベースとしての役割も期待されている。

『古事類苑』は、図 11 に示すように、前近代の日本の諸事項を秩序立てて階層的に整理・分類して、百科事典の総目録として体系化されている。この総目録の体系をシソーラスとして nihuINT が検索対象とするデータベースの情報の体系化や分類に利用できる可能性がある。また、抜粋テキストから、『古事類苑』が引用する資料の書名を容易に抽出することができる。すなわち、書名の辞書や索引を知識ベースとして構築することが期待できる。また、書名に比べると抽出に労力を要するが、人名や地名などについても、『古事類苑』は多くの情報を含んでおり、人名や地名の知識ベースの構築を支援できる可能性がある。

## 7. おわりに

全文・抜粋検索版データベースを構築する意義は、『古事類苑』全体を網羅した検索を実現することである。2011 年 11 月に作成を開始した抜粋テキストによって、2012 年 8 月に、4 部の公開に留まっていた全文検索版データベースを、検索対象が 26 部の全文・抜粋検索版データベースに更

新した。今後、全文テキストについては、残りの 4 部の作成を進め、全文検索の対象を 8 部に拡大させる。22 部の抜粋テキストについては、書名と解説データ化の完了、索引データの増強を進め、全文・抜粋検索版データベースにおける『古事類苑』全 30 部の網羅を進めていく。また、『古事類苑』の総目録の体系をシソーラスとして nihuINT の検索対象データベースや検索結果の整理・分類に利用することや、全文・抜粋検索版データベースから抽出した書名などの情報を nihuINT の検索対象データベースの情報と関連づけることなど、『古事類苑』を知識ベースとして活用する方法についても検討していく。

## 参考文献

- 1) 相田満: テキストコンテンツのポテンシャル -古事類苑データベースの取り組みから-, 人文科学とコンピュータシンポジウム論文集, pp.253-260 (2010).
- 2) 国文学研究資料館 古事類苑データベース, <http://base1.nijl.ac.jp/~kojiruie/>
- 3) 国際日本文化研究センター 古事類苑ページ検索システム, <http://shinku.nichibun.ac.jp/kojiruie/>
- 4) 国立国会図書館デジタル化資料, <http://dl.ndl.go.jp/>
- 5) 倉本一宏: 古事類苑新仮名索引, 吉川弘文館 (2010).
- 6) 山田奨治, 相田満: 古事類苑全文データベースの構築について, 人間文化研究情報資源共有化研究会報告集 1, pp.151-159 (2010).
- 7) 山田太造, 古瀬蔵: nihuINT による人文学研究資源の情報統合, 画像電子学会年次大会予稿集, Vol.40, T1-4 4 (2012).

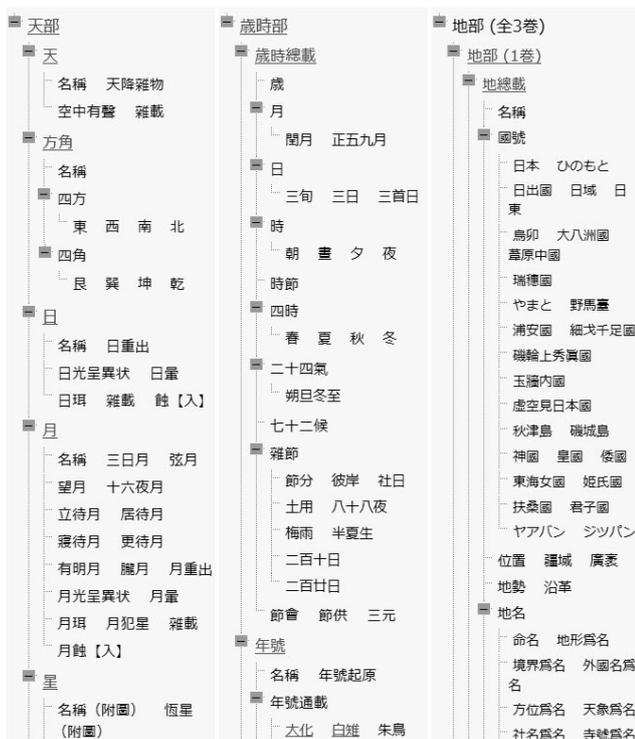


図 11 『古事類苑』の総目録の体系