



学術情報データベースの構成と利用

日本史データベース†

星野 聰†

1. はじめに

研究者が過去の事柄を説明するために仮説を立て、これを現存する不完全な史料をもとにして証明し、あるいはあり得ることを推定し、または他の史料と矛盾しないことなどを検証する。これが、歴史研究の主なパターンである。

さて、歴史学は、文学、地理学、考古学、建築学、美学、法学など、多くの分野と関連している。さらに、天文学や数学などの理工系の諸分野も、歴史研究に必要な知識情報を提供するだけでなく、それらの歴史は科学史として、重要な研究分野である。実際には、これらの分野の研究は、従来縦割りで行われてきた傾向があるが、歴史研究者の推論の過程では、関連分野の知識情報が必要である。また、残存する史料が不十分ならば、文学や絵画などの資料も参照しなければならない。そこで、ここでは歴史関連の分野も、必要に応じて考察の対象に含めることにする。

ところで、情報処理技術が、日本史研究とその関連分野にも、有効に利用されるようになってきた。残念ながら、その全貌を把握するのははなはだ困難である。歴史研究は元来個人研究の色彩が濃く、コンピュータ利用についての、情報流通が十分でない。また歴史研究の論文に、その研究に利用した情報処理関係の技術や、利用時の問題点などは、研究内容と直接関係がないとして、記されない傾向が強い。これはちょうど、研究に使用した辞書や索引類を参考文献に記載しないのと似ている。そこで、情報処理の立場から、日本史の分野へのコンピュータ利用について、実態や課題などを論ずるには、散在する資料や、シンポジウム報告などによることが多くなる。以下において

は不完全ながら、日本史研究へのデータベース利用に対して概説を試みた。

2. 古典テキストの処理

2.1 データの準備

2.1.1 古典テキストの特徴

古典テキストの形態や取扱いは、通常の図書や論文とは異なる面がある。まず、文献に相当するのが、一枚の古文書であるかもしれない。表題は付いているとは限らない。内容がすでに解読されているとは限らないので、キーワードを付加するのが困難なことが多い。また、表題に相当するものがあったとしても、「乍恐以書付奉願上候」や「差上申一札之事」の類の決まり言葉で、文書を識別するには役立たないことも多い。表題が必要ならば、適当に付けなければならない。

また、ある著作（たとえば万葉集）に対して、多くの写本（図書にあたる）があることが多い。そして、一つの著作に複数の書名があり、一人の著者が複数の著者名を有することがある。そこで、書名や著者名などに対して、一定の表記を与えるための典拠コントロールを行う必要がある。国文学研究資料館における古典籍総合目録のデータベースは¹⁾、古典籍の所在目録を収めるが、書誌データ（著作、図書の二つのファイルからなる）、著者名・書名典拠、所蔵者などのファイルを設け、それらの間をリンクさせて、一つのデータベースにしている。叢書・合綴・合刻・合写により、複数の著作がまとめられて、一つの著作とみなされることがある。また、異本や異版も多いことがあり、設計では、これらに対処する必要がある。

さて、テキストの校正は十分行う必要がある。この点、速報性が要求される科学技術文献データベースの場合と異なっている。歴史的なテキストは、急いでコンピュータに入力して誤りを含むよ

† Databases on Japanese History by Satoshi HOSHINO (Data Processing Center, Kyoto University).

† 京都大学大型計算機センター

りは、遅くとも慎重に作成すべきである。テキストの校訂作業は容易ではない。外注に出す際には、できれば照合（ベリファイ）させるべきである。類似の文字（たとえば會と曾）は、テキストを繰り返しチェックしても、見逃してしまう。テキストが不鮮明なことがよくあるが、もし研究者であれば、容易に誤りが分かることがある。たとえば、位田の田の字がかすれて、「位出」と入力されていたことがあった。位田は日本史で使われる熟語であるが、位出は意味が通じないことを知らなければ、この誤りに気付かなかったかもしれない。あるいは、テキストに「甲甲朔」とあれば、この甲甲は干支であり、甲申がかすれたものと推理できよう。さらに言えば、使用するテキストが出版されてから、すでに誤りが判明していることもあり、テキストの誤りを修正すべきである。これには研究論文を調べる必要があり、研究者でないと実行が困難である。

テキスト内に異体字が混在していると、検索もれを生ずる。たとえば、拗音を示す「ヨ」と「ヨ」でも同じである。技術的対策は、データ内の、すべての「ヨ」を「ヨ」にするか、両者のOR集合を検索するようにすればよい。

表記が異なるために検索できない他の例として、麿が麻呂と二文字になって入力されていることがある。また、糸という人名は、検索時に久米でないことを知らないで検索できないのが普通である。さらに、菩薩を井と書いていることがある。これは、二字に書き直すのが一案であるが、この種の省略字も多い。われわれ、図書館のことを下と書くことがあるが、この文字はJISに定義されていない。また、古典辞書では、たとえば「俗」を「谷」と略記することがある。これは他の文字種に一致する場合である。

人名に用いる異体字は、これを正字に置き換えると慣用に反することになる²⁾。たとえば、持麿(かせまろ)を弄麿とするのは、文字としては同種である。しかし各文字には主として用いられる意味に差があることがあるから、置き換えるのは注意を要する。

筆者が、六国史の一つである続日本紀テキスト³⁾について、総索引(用例索引, concordance)の冊子⁴⁾作成でとった方針では²⁾、用例に使用する漢字は、なるべく底本とした国史大系本に近い字体

を用いた。しかし、パーソナルコンピュータ(PCと略す)で、このテキストを検索する際には、なるべくJISで定義された文字に変換して、検索もれを防ぎ、かつ多くの文字が表示できるようにしている。東京大学史料編纂所では、異体字を統一化する方針がとられている⁵⁾。しかし、異体字の処理は開発者によって異なるのが現状である。

奈良時代に造られた百万塔に関するデータベースが、奈良国立文化財研究所で作成された⁶⁾。百万塔は、ロクロを使って作られた木製の小塔で、底面にロクロの爪形が付いている。このデータベースには、各塔の作成者や、日付、爪の形状コードなどが格納されていて、当時の工房組織、ロクロの台数、作業の実態などを分析する手掛かりとなる。このように、分析の手掛かりとなる情報も含まれるのが、歴史関連のデータベースの一形態である。分析には、たとえば1日から月末までの各日ごとに作成された塔の個数分布を調べるなどの、計数操作が必要になる。つまり、単に検索すればよいのではない。また、爪の形状のコード化のように、図形の分類をいかに行うべきかも問題である。

2.1.2 分類と排列

通常、各図書に図書分類を付して、分野を大別し、それを利用して図書を配架する。書店や図書館に行くと、分野別に本が並んでいるので、書籍をアクセスするのに便利である。このように、図書内容によって、物理的に分散配置している。

古文書を扱う機関では、資料を扱う仕方が、図書館とは違っている。古文書では⁷⁾、史料が所蔵されていた出所(たとえば、某家文書)ごとに物理的にまとめるほうが便利である。これを出所原則(Principle of Provenance)という。古文書は一枚の紙のことも多いが、物理的に分散してしまうと、某家文書を調査したい利用者には不便である。他の原則として、原秩序(原配列)尊重の原則(Principle of Respect for Original Order, Registry Principle)があり、もともと出所で排列されていた順序を尊重する。たとえば、所蔵されている箱を特定し、また特定の箱内の文書を特定するための番号付けをする。個々の史料には、そこに収録されるに至るまでに、どのような収蔵者を経たかなど、履歴情報も必要とされる。最近各地で、公文書館が置かれ、その地域の古文書を保存するよ

うになったが、これには、会社や官庁で、ある年限が過ぎた非現用文書で、保存すべきものも公文書館に送られるのである。

キーワードによる検索には、問題がある。歴史の分野では、一般に容認されるキーワードの設定は困難とされている。ある術語を共に使用している複数の研究者集団があるとき、その術語が意味する内容や範囲が同一とは限らないし、地域による伝統的な慣習もある。図書分類にあたる分類体系も、収蔵機関で異なっている。そこで、多くの機関でデータベースを共用できるようにすることは、大きな課題である⁹⁾。キーワードは互いに読み代える必要があるが、分類方式が機関によって異なると、単純に置き換えるのが困難である。また、現存する史料群のなかで活字化されたものは、わずかに一部であるから、この種の研究は、古文書の内容が読解できないと困難である。

古文書の内容を解読して、キーワードを付けることは容易でないから、文書の書き出しや書き止め文言だけをデータベースに格納することがあり⁵⁾、これによって、文書の性格を推し量るのである。さらに、文書の差し出し人・受取人を読みとって、いつの文書か、あるいは文書の性格などを知る手掛かりとする。また、表題には文書内容の識別能力に乏しいことが多く、通常の図書検索のような、書名による検索が有効でないことが多い。従来は、所蔵機関の所蔵目録を利用して調べるのである。所蔵目録は各所に散在しているので、その調査が行われた⁹⁾。

2.2 データの入力

2.2.1 入力的方式

古典テキストを扱うには、通常のワードプロセッサ (wp と略す) や pc では不便なことがある。これに関して述べると、

(1) 外字の種類制約

古典テキストでは、多くの文字種を必要とする。しかし、大きな漢和辞典 (たとえば、大漢和辞典¹⁰⁾) に収められた、5万種にのぼる文字は必要でない¹¹⁾。用例がなく、実際にテキストに現れない文字が多いからである。

古典テキストは、一般に幾度も筆写されたもので、残存する諸本には異同がある。そこで研究者は、ある本 (底本という) を基準にして、他の本との異同を調べ、その研究者が妥当と考えるテキ

ストを作る。これが校訂である。当然、でき上がったテキストは校訂者によって異なる。研究の対象になる異本の数は、百数十にのぼることがある。

さて、各本での文字を忠実に表現しようとする、必要な文字種は増加する。そこで、同じ種類の文字は、一つの字形で代表させて、文字種を減らすことが考えられる。これに対して、あくまでテキストに忠実であるべきだと主張もある。これは、研究の目的によることが多いので、一概に是非を言うことはできない。筆者は、意味上で差がない微細な字形の差を有する文字は区別していない。通常の歴史研究では、この方針でよいと考えている。個々の字形が問題になる場合は、原テキストを参照する必要があり、コンピュータ利用では、その画像表示をすべきである。

筆者の経験では、続日本紀テキスト (約 30 万字) のファイルで、作成した外字は、108 字種に止まっている⁹⁾。ただし、コンピュータは、Fujitsu M780 システムを使用し、作成した外字は、JIS 定義漢字と、富士通提供外字になかった漢字である。作字する必要がある漢字の数は、史料によって異なるが、この例では多くない。ただし、六国史でも、pc で数種類の史書を一度に扱うと、外字定義できる個数以上の外字が必要になる。

外字の種類を増すには、JIS 漢字の追加や、すでに外国で行われている 3 バイトコード¹²⁾ のような広いコードスペースの確保が考えられる。ただし、そのテキスト特有の文字もあるので、ユーザが自由に外字を定義するためのスペースを確保できる必要がある。もし、定義したい外字が多くて、定義用のスペースが不足する場合には、外字がディスプレイに表示されている間だけ、そのスペース内に動的にコードと文字パターンを定義することによって、實際上、多くの外字を表示する方法がある。筆者も pc では、この方法を採用している。

なお、古典辞書をコンピュータで扱う際や、言語翻訳を行う際には、多数の文字種を必要とするから、コード系をどうすべきかという問題は避けられない。

(2) 漢字のみの文の入力

日本古代においては、中国から伝来した漢字文化を受容し、史書も漢文で記述された。これは以

後も引き継がれ、公式文書などは漢字のみで記されている。しかし、これを通常の wp で仮名漢字変換を用いて入力すると、通常、送り仮名を消去する必要があるから不便である。ディスプレイに表示する直前に、仮名の部分を取り去るだけでよいから、この機能が一般化されることを期待したい。

2.2.2 データ入力の問題点

2.2.2.1 読みの付与

漢字を、読みで検索するには、仮名を付けなければならないが、読みは一通りとは限らない。たとえば、正四位は、「おほいよつのくらののみかみつしな」と読まれたが、われわれは、通常、「しょうしい」と読んでいる。また「読み癖」と呼ばれる読みもあり、検索もれを生じる原因になる。そこで、各漢字に一意の音読みを与える方式がある⁵⁾。ところが、たとえば、人名を検索するために、人名をこの音読みで排列すると、慣用されている読みによる排列とは異なる順序になることがある。しかし、歴史上の人物が、当時どのように呼ばれていたか不明のこともあるので、ある決まった順序に排列できることには便利な面がある。

インデックスによる検索を行うには、分かち書きを行う必要があるが、手作業によると、統一性を欠くことが多い。辞書を用いた自動分かち書きは、高橋氏により、データベース MUROMATI 開発の際に、最長一致法を用いて行われた¹³⁾。問題点として、「山城守護代」「保坂関務」を、山城守・護代、保坂関・務と切断してしまうという例が示されている。

2.2.2.2 天文の計算と年月日の表示

天文の軌道計算などは、歴史研究に活用されるべきである。日本暦日原典は¹⁴⁾天文計算を行うのに必要な、ユリウス日を求めたり、節気を知ることでもできる重要な文献で、コンピュータを使用して作成されたものである。月や星などの天文現象に関する過去の記述、たとえば日食、月による惑星の隠べいなどは、天体の軌道計算でチェックできる。そこで、史書の天文記事の日時の誤りなどが科学的に検出できるのである¹⁵⁾。ただし、pc による軌道計算ソフトウェアを、古代における計算に用いるには、精度のチェックが必要である。

史書の記録には、年月日が明確でなく、是歳、是月などと記されていることがある。また、文献

が扱っている範囲が、たとえば平安前期にわたることがある。そこで、時代コードとして、前半・中期・後期など一定期間を意味する情報を用いる必要がある。その一例として、京都大学人文科学研究所の東洋学文献類目のデータベース化で採用されているものがある。

和暦による年の表記を西暦年に変換する機能は、記事を編年順に並べるのに必要である⁵⁾。この際に是歳、是月などの表記も、適当に変換しなければならない。排列の例として¹⁶⁾、日の不明なものは月の末尾に、月の不明なものは年の末尾に、年の不明なものは年号の末尾に配するものがある。

2.2.2.3 データ入力のための工夫

史料に現れる人名などは、テキストによっては局所性があり、あまり間隔をあげないで、再びテキストに現れることがある。そこで、テキスト入力において、人名などを選んで画面の一部に一覧表示させておき、再び入力する際には、画面上でそれに付された番号を指定するだけで入力できる方式が実用されている⁵⁾。

2.3 検 索

テキスト検索では、それが編年体の場合は、日付ごとに切り分けて、おのおの文献情報のように扱うことができる。検索の対象になる語が予想される場合には、あらかじめテキストの分かち書きを行い、インデックス・ファイルを作って検索に利用する。ここで、分かち書きされた句に跨る検索に、難点を生ずる可能性があることに注意しておく。たとえば「うきよ」を「うき」と「よ」に分割してあったとして、「うきよ」で検索したい場合である。

もし、テキストを順に探索するのであれば、ボタン・マッチング手法によるテキスト検索が利用される¹⁷⁾。六国史のような古代史料は、サイズが数百キロバイトのことが多いので、pc で十分対応でき、光磁気ディスクなどを使用すると、さらに大きいテキストも扱うことができる。

さて、テキスト検索の結果は、一度に多数の結果が表示されても、ユーザがそれを認識するのに若干時間を要する。そこで、用例が検索されるごとに表示し、もし次の一件が検索されるまでの時間が長くなければ、実用上十分である。そこで、設計によっては、低速の pc でも十分なことが

ある。

筆者が採用している検索方法は¹⁸⁾、テキストの任意の文字cに対して、次にテキストに現れる同じ文字の位置を示すポインタを、文字cの直後に格納したファイルを使用する。このポインタを利用して、指定した文字列を含む用例を検索する。ポインタを辿るべき文字としては、その文字列中の文字で、生起回数が最小の文字が選ばれる。

筆者は、pcのハードディスク(容量40MB)に、六国史である続日本紀・日本後紀・続日本後紀・文徳天皇実録および三代実録の一部を格納し、この方式により検索している。中世以後の歴史研究では、データ量が増大するので、CD-ROM利用の普及が期待される。CD-ROMは、離れたセクタへアクセスする際に、シーク時間が長いのが難点である。CD-ROMを用いる際に、アクセス時間を短くする検索方式の研究が必要である¹⁹⁾。

従来、研究者は、テキストごとに手作業で作成された索引類を使用して研究を行ってきた。しかし、それでは限界がある。コンピュータによるテキスト検索の有用性を示す例として、続日本紀テキストで、「蟻」を検索して、当時(8世紀)に蟻をどのようにみていたかを調べた。すると、「蟻結肆毒」(蟻は結集し害毒を欲しいままにする)「蜂屯蟻聚」(蜂は屯し蟻は集まる)「捨則蟻結」(捨ててもすぐに蟻は結集する)の三例が見つかる。ところが、従来、研究者が用いている続日本紀索引(六国史索引)では、二例のみで、最初の用例が抜けている。これは、手作業による索引作成で、用例が脱落したのであろう。日本のビジネスマンを蟻にたとえた外国の首相があったという。それで、筆者は続日本紀を検索してみたのであるが、蟻のイメージは良くないことが分かる。このように、古代史料が現代研究にも役立つこと、コンピュータ利用が学際的利用を促進することが理解されよう。また、検索もれを防ぎ、高度な研究支援に効果的であることも示唆している。

2.4 テキストの編集

テキストに対して、頭注や脚注を設けて注記することがある。注記にはそのテキストの他の箇所への参照、他のテキストの参照、読みや異本との違い、あるいは注記者の見解なども記される。読みや異本との違いは、対象となるテキストの文字

列の右または左の行間に記されることもある。

注記をテキストに追加記入する作業は、通常のエディタでは不便である。なぜなら、注記はテキストのある文字列に対して付けられるが、テキストと注記の両者が混在して、見にくくなるからである。また、注記付きのテキストを、適当な見やすいスタイルでプリントする必要がある。もし、注記の追加記入が可能であれば、プリントした結果に対し、さらに研究成果を書き込み、それを再びプリントするという操作を繰り返して、そのテキストについての研究を有効に進めることができる。筆者は、注記付きのテキストを扱うエディタを作成した。テキストは、あるコラムに縦一列に並べ、その右側の欄に注記を記入する。画面は上下にスクロールできる。また、このファイルをプリントすると、頭注付きのテキストになる¹⁸⁾。

3. 画像データベースの利用

3.1 準備と入力

歴史とその関連分野で画像情報を、データベース化して利用したいという期待や実施例は多い^{5), 22), 23)}。しかし、画像データベースの、歴史研究への寄与を示すのは、むしろ今後の課題である。

さて、画像情報に期待されている内容は一律でない。データベース化の必要性とその効果も明確にする必要がある。鮮明な画像でなければ、実際は意味がないこともある。しかし、画像を鮮明にすれば、ファイル装置に格納できる画像枚数が制約され、データベースとしての意味は薄れる。そこで、デジタル画像の圧縮技術の利用が期待される。また、ワークステーションによって、ビデオカメラなどのAV機器を制御し、画像を取り込む方法も有効である。データベース開発には、画像入力の所要時間が短かく手間がかからないこと、便利なソフトが利用できることなどが望まれる。

文字情報であっても、文字の配置関係も必要な場合がある。たとえば、木簡では文字が不鮮明であったり、折損していることもある。この場合には、推定した文字数を記入したり、他の文字との位置関係なども、コード化の対象になり得るが、複雑な記号列になる。そこで、画像表示するのが便利である。

入力の際のデータフォーマットは、扱うべきテキストの特殊な性格に対応して、個々に考案されている。たとえば、古典辞書（類聚名義抄）については文献 20)、付加情報があるテキスト（小右記）については、文献 21) を参照されたい。

3.2 編集と検索

東京大学史料編纂所の荘園史料データベースのためのシステムでは²²⁾、文献データベースによって、所要の画像を特定するためのコードを求め、これを用いて画像データをアクセスし表示する。また、画像の入力と、文献情報の付加の順序に関わらず、画像データベースを構築してゆく機能を有している。

一般に公開されているシステムとして、東京国立博物館資料館の、pc と追記型ビデオディスク装置からなるシステムがある²³⁾。ユーザは、題名・作者・フィルム番号・所蔵者などから、所要の画像写真などを検索できる。東京国立文化財研究所では、題名などをあらかじめ知ることは困難であるので、人名・地名・主題・モチーフなどから文献を検索する試みがなされている²⁴⁾。

3.3 地図情報の利用

地形図のスクロール表示機能は、歴史研究に有用である。すでに、地理情報のシステムは各種開発されているが、筆者は歴史地理研究を目的として¹⁸⁾、pc を用いて、国土地理院の現在の 25,000 分の 1 の地形図とともに、大正十年ごろの同一地域の地形図を同時にスクロール表示するようにした。戦前までは、工事は人力に頼っていたから、災害を除外すれば大きな地形の変化は見られない。しかし、戦後は地形や道路が開発にともなって著しく変化したので、現在の地図だけでは、研究に不十分である。

4. 発見の利用

コンピュータ利用によって、史書に明記されていない事柄を発見できれば、すばらしい。この分野の開拓は、ほとんど今後に残された課題である。ここでは、筆者が試みた例を紹介したい。

4.1 時間的な関連性の検証

歴史的な事柄について、ある仮説を設けるとき、時間的な関連性に矛盾を生ずるような仮説は認められない。したがって、この関連性の検証が大切であり、このためのツールは、仮説を立てる

ときに役立つであろう。

筆者が試みたものに、系図のチェックがある²⁵⁾。これは、結婚年齢の上限・下限、寿命の上限などを仮定して、ある人が矛盾なく与えられた系図に収まるか否かをチェックするものである。筆者が試みた例は、日本古代の皇族に見える一夫多妻制に関連した複雑な系図である。対象とする各人の生没年など不明のことも多いが、上記の諸仮定の下で、生没年などの上下限を、可能なかぎり推定するのである。多くの人の生没年などの時間的な関係を、人間がチェックするのでは、矛盾を見落とす危険がある。この種のツールは、別に系図に限らず、より一般的に時間的な関係のチェックに利用できる。

4.2 記事量の分布による休日の推定

続日本紀テキストについて、記事の件数を一日から月末まで、日付によって集計すると、どのようになるか。この疑問を抱いた筆者は、テキストファイルを入力として、この集計を行った。その結果として、六日ごとに記事が少なくなることが分かった²⁶⁾。記事が少ない日は官人の休日である。実は、後で分かったことであるが、これはすでに山田英雄氏²⁷⁾によって、手作業で発見されていた。ところが、続日本紀では巻を追うごとに、六日ごとの記事量の減少が目立たなくなる。これは、時代とともにこの記事量の変化が、どのようになるかを調べた際に気付いたもので、コンピュータ利用で見出されたものである。このような記事量の変動は、テキストを読んでいるだけでは分からない。

なぜ、記事量の落ち込みが減少してゆくのかなど、休日の実態は全く分からない。しかしこれは、コンピュータ利用が、研究の手掛かりを与えてくれる可能性を示している。

5. 今後の問題

高度な、コンピュータ利用の開発と実用化の多くは、今後に残された課題であり、データベースは、これらに関連して構築されることになる。

また、日本史関係の研究グループのネットワーク利用や、CD-ROM による多量の史料の配布などが始められることが期待される。

6. あとがき

日本史とその関連分野での、データベース関連の現状と課題について概説した。データベースは個人的に作られていることも多いので、公的機関で公開されている例に限って次表に示す。紙面の制約もあり、言及しなかった事項が多い。ご寛容をお願いする次第である。

公開されているデータベース例

データベース	公開機関
維新史料綱要網文	学術情報センター
続日本紀	京都大学大型計算機センター
室町幕府奉行人奉書	同上
旧高旧領取調帳	国立歴史民俗博物館

参考文献

- 1) 古典籍総合目録データベースの構築と出版, 国文学研究資料館報告, No. 12 (1991).
- 2) 星野 聰: 異体字をどう考え, どう処理するか?, しにか, Vol. 3, No. 2, pp. 23-29, 大修館書店, 東京 (1992).
- 3) 新訂増補国史大系 続日本紀, 前後編, 吉川弘文館 (1984-1985).
- 4) 星野 聰, 村尾義和 (編): 続日本紀総索引 上下巻, 高科書店 (1992).
- 5) 永村 真: 日本史史料データベースとデータ処理に関する研究, 国立歴史民俗博物館研究報告, No. 30, pp. 109-144 (1991).
- 6) 金子祐之: 百万塔データベース, 東洋学研究支援データベースの研究, 科学研究費研究成果報告書 (代表者 星野 聰) 所収, No. 60300015 (1987).
- 7) 大藤 修, 安藤正人: 史料保存と文書館学, 吉川弘文館 (1986).
- 8) 山中秀夫: 近世地方文書の検索の機械化及びその利用, ビブリア, No. 87, pp. 161-182 (1986).
- 9) 近世・近代史料所在情報の収集及びその体系化に関する基礎的研究, 科学研究費研究成果報告書 (代表者 安澤秀一), No. 60301048 (1988).
- 10) 諸橋徹次: 大漢和辞典, 大修館書店 全13巻, 東京 (1955-1960).
- 11) 星野 聰, 今城一夫: 計算機による漢字データベース開発研究の現状と将来, アジアクォーターリー, Vol. 13, No. 2-3, pp. 22-38 (1981).
- 12) たとえば, Huang, J. K. T. and Huang, T. D.: An Introduction to Chinese, Japanese and Korean Computing, World Scientific, Singapore (1989).
- 13) 高橋康夫: 私とデータベース: 都市・建築史の立場から, 人文科学データベース研究, No. 2, pp. 26-36 (1988).

- 14) 内田正男 (編著): 日本暦日原典, 雄山閣, 東京 (1975).
- 15) 齊藤国治: 古天文学, 恒星社, 東京 (1989).
- 16) 古瀬奈津子, 山口英男: 西宮記勸物等データベース, 儀式書を中心としてみた平安時代政治機構の総合的研究, 科学研究費研究成果報告書所収 (代表者 早川庄八), No. 0130143 (1991).
- 17) たとえば, 有川節夫, 篠原 武: 文字列パターン照合アルゴリズム, コンピュータソフトウェア, Vol. 4, No. 2, pp. 98-119 (1987).
- 18) 星野 聰: 歴史学研究支援システムの開発, 90-CH-5, No. 5-7, 情報処理学会研究報告 (1990).
- 19) たとえば, Cooper, L. K. D and Tharp, A. L.: Inverted Signature Trees and Text Searching on CD-ROMs, Inf. Process. Manage., Vol. 25, No. 2, pp. 161-169 (1989).
- 20) 類聚名義抄の注釈的研究—電算機利用による—, 科学研究費研究成果報告書 (代表者 田島誠堂), No. 63301058 (1990).
- 21) 永村 真: 日本史史料全文テキスト・データベースの構築と句切処理について—大日本古記録本 小右記を素材として—, pp. 57-68, 第28回研究セミナー報告, 京都大学大型計算機センター (1990).
- 22) 京畿内荘園の史料データベース作成と歴史地理学的研究, 科学研究費研究成果報告書 (代表者 石上英一), No. 62410008 (1989).
- 23) 高見沢明雄: パーソナルコンピュータとビデオディスクによる画像データベース, 国立歴史民俗博物館研究報告, No. 30, pp. 31-47 (1991).
- 24) 島尾 新, 長岡龍作, 中村節子: 「キーワード辞書」を用いた美術史関係文献検索システム, 国立歴史民俗博物館研究報告, No. 30, pp. 67-86 (1991).
- 25) 星野 聰: 歴史学に於ける系図の分析手法とその拡張, 情報処理学会 (昭和62年前期) 全国大会, 5L-2 (1987).
- 26) 星野 聰: 歴史研究と計算機利用—古典テキストの処理—, 情報の科学と技術, Vol. 40, No. 1, pp. 21-26 (1990).
- 27) 山田英雄: 律令官人の休日, 続律令国家と貴族社会所収, 吉川弘文館.

(平成4年2月3日受付)



星野 聰 (正会員)

昭和6年生。昭和29年京都大学工学部電気工学科卒業。昭和34年同大学院博士課程修了。工学博士。昭和56年図書館情報大学教授。昭和59年京都大学大型計算機センター教授。東洋学とくに歴史学および文学への計算機利用の研究に従事。「続日本紀総索引 (高科書店)」、「数値計算の技法」(コロナ社), 電子情報通信学会, 続日本紀研究会, 情報科学技術協会各会員。