

本論文について

この論文は、もともと以下に発表したものです。

渡邊隆弘「BSH（基本件名標目表）による OPAC の主題検索」

『学術情報センターセミナーレポート』平成9年度, 1998.3. p.33-134

学術情報センター（当時）の研修「学術情報センターセミナー」の最終レポートです。

ただし、ページ付けが原版とは異なります（本ファイルはp 1からはじまる）

ポストスクリプト（ps）ファイルの原版を PDF 化しています。ひょっとするとレイアウトの崩れなどがあるかもしれません。

連絡先：

渡邊隆弘（帝塚山学院大学人間科学部）

watanabe@hcs.tezuka-gu.ac.jp

BSH(基本件名標目表) による OPAC の主題検索

神戸大学附属図書館
渡邊隆弘

平成 10 年 2 月 27 日

Abstract

1998 年刊行予定の BSH4 版は、参照関係をシソーラススタイルの関係構造に再編成するなど、大幅な改訂となっている。

BSH4 版の編集用機械可読ファイルを使用して、関係構造を利用したナビゲーションによって図書館利用者の主題検索を支援する実験システムを作成した。BSH をもとにした件名典拠ファイルを介して検索を行ない、関連する標目や階層構造表が提示される。件名典拠ファイルには、固有件名や細目つき件名など、件名標目表では個別に管理していない件名標目をも収め、関係構造の中に位置付けた。

作成したシステムについて、2 通りの評価実験を行ない、概ね肯定的な評価を得るとともに、問題点も明らかになった。

また、件名典拠ファイルの構築作業、特に検索システムに使用する関係構造データの整備についても詳細に述べた。

目次

1	はじめに	5
2	主題検索、件名検索、BSH	8
2.1	OPAC の主題検索と典拠管理	8
2.2	件名標目表と主題検索	11
2.3	BSH4 版改訂と BS4mr ファイル	14
2.4	NACSIS-CAT における件名	18
3	検索システムの作成とその機能	22
3.1	件名検索に必要な機能	22
3.2	検索システム BSH-OPAC の概要と作成環境	25
3.3	BSH-OPAC の DB 構成	26
3.3.1	件名典拠 DB	28
3.3.2	階層パス DB	28

3.3.3	件名キーワード DB	29
3.3.4	典拠書誌 DB	30
3.3.5	書誌 DB	30
3.3.6	タイトルキーワード DB	30
3.4	BSH-OPAC の機能	31
3.4.1	典拠検索機能	31
3.4.2	関連語表示機能	34
3.4.3	階層系一覧機能	40
3.4.4	書誌表示機能	42
3.4.5	書誌フィードバック機能	43
3.5	対照用タイトル検索版 TT-OPAC	44
4	検索実験とシステムの評価	47
4.1	与えられた検索質問に対する有効性評価	47
4.2	対人実験による評価	50
4.2.1	細目を伴わない件名へのアクセス	52
4.2.2	細目を伴う件名へのアクセス	59
4.2.3	被験者の主観的評価	63
4.3	評価のまとめ	65
5	件名典拠ファイルの構築	67
5.1	作業の目的、概要、限界	67
5.2	情報源となる付与件名データ	70
5.3	付与標目データの分類	71
5.4	各種件名標目レコードの作成	74
5.4.1	普通件名標目レコード	74
5.4.2	個人名件名標目レコード	75
5.4.3	家族名件名標目レコード	75
5.4.4	団体名件名標目レコード	75
5.4.5	統一書名件名標目レコード	76
5.4.6	地名件名標目レコード	76
5.4.7	細目つき件名標目レコード	79
5.5	典拠ファイル中のその他のレコード	82
5.5.1	細目レコード	83
5.5.2	参照語レコード	83
5.5.3	補助地名レコード	83
5.5.4	主題+地名共出現レコード	83
5.6	標目間の関係構造データ	84
5.6.1	BSH4mr の階層構造	84

5.6.2	地名件名間の階層構造	86
5.6.3	固有名と普通件名との事例関係	87
5.6.4	細目つき件名の関係構造	89
5.6.5	連想関係	92
5.6.6	年代的关系	92
5.6.7	直接参照関係	93
5.6.8	参照的注記(仮称)と説明つき参照	93
5.7	最終的な件名典拠ファイル	94
6	まとめ	95
	謝辞	99
	参考文献	100

表目次

1	BSH4mr(編集中的数据 1997.12 現在)	17
2	CAT 和図書レコードの件名付与状況	19
3	出版年別件名付与状況	20
4	CAT 書誌レコードの NDC 分類別統計	21
5	BSH-OPAC の検索対象レコード数等	25
6	件名典拠ファイルの各種レコード	28
7	BSH-OPAC が優位にたった質問	49
8	TT-OPAC が優位にたった質問	49
9	対人実験予備調査	51
10	対人実験の検索課題一覧	53
11	対人実験被験者の主観的評価	63
12	対人実験被験者のシステムへの意見(抜粋)	64
13	件名典拠ファイルフィールド一覧	69
14	CAT 書誌レコード中の BSH 標目	70
15	付与件名の種類別統計	73
16	地名件名の種類別統計	78
17	BSH4mr 中の階層関係	84

図目次

1	件名標目「遺跡」に関する関係構造図	23
2	BSH-OPAC 初期画面	26
3	BSH-OPAC の DB 構成	27
4	検索画面(「大学」の検索結果)	32
5	件名典拠にはノーヒットの場合	33
6	「大学」の関連語表示	35
7	「日本 農業」の関連語表示(スクロール後)	38
8	「フランス」の関連語表示(スクロール後)	40
9	「河川」の関連語表示(スクロール後)	41
10	「工芸」の階層系一覧	42
11	「ヨーロッパ」の階層系一覧	43
12	書誌表示	44
13	書誌フィードバック機能	45
14	タイトル検索版 TT-OPAC	46
15	件名典拠ファイル構築作業概要	68
16	付与標目データの分類と統計	72
17	細目つき件名標目レコード確定作業	80

1 はじめに

わが国の図書館目録における主題検索機能の貧弱さは古くから指摘されている。多くの図書館の「分類目録」は函架目録に近いものであったし、件名目録はごく限られた図書館でしか編成されてこなかった。

カード目録から OPAC になって、書名などからのキーワード検索が可能になった。また MARC や NACSIS-CAT の書誌レコードを自館にダウンロードして目録作成すれば、件名標目や分類記号が自然に取り込まれてくるし、検索対象とすることも難しくない。しかしこれらをもって「主題から検索できる」と称するのは、大いに疑問がある。書名等の書誌記述から切り出したキーワードと、利用者が想起し入力した検索語とが一致するかどうかはわからない。件名標目などが一貫性をもって付与されていたとしても、その標目と利用者の検索語とが一致するかどうかは、やはりわからない。ましてや、NACSIS-CAT 等のレコードをそのまま受け入れていれば、付与されている主題標目に一貫性がなかったり、標目の変更が古い書誌レコードに反映されてなかったりするのである。

これでは結局のところ偶然に頼った検索というほかない。検索結果の的確さは、近似した言葉をどれだけ多く入力したか、といったような利用者側の知識と創意工夫と忍耐に委ねられてしまっているのである。このような不十分な「主題検索機能」を提供しつづけるのでは、むしろ利用者の目録に対する信頼を失わせることになる。「主題からの検索」というからには、ある程度の網羅性をもって適合データが検索できるシステムを構築しなくてはならないと考える。

本研究の概要

本研究では、件名標目表を利用した主題検索について追求し、プロトタイプ検索システムの作成、評価を行なう。目指すところは、「利用者の知識や忍耐をあてにせず」的確な主題検索ができるシステムであり、もう少し具体的にあげると次のようになる。

● 関係構造を利用したナビゲーション

- 件名標目表のもつ情報のうち、収載された用語間の関係構造を重視する。用語間の関係としては、優先語・非優先語関係 (USE, UF) と標目同士の上位語・下位語・連想語関係 (BT, NT, RT) がある。
- 中でも件名標目間の階層構造に重点をおき、利用者が一つの検索語入力をてがかりにして、提示された関係構造から、最も目的にかなう標目を見つけたり、他の標目に探索の網を広げたりすることを可能にするシステムを考える。

- つまり、心の中にある検索要求と正確に一致しなくても、多少とも関係のある言葉を思いついて入力すれば、あとはナビゲーションによって自然に目指す標目にたどりつけるシステムである。
- また、階層構造を利用して、検索の幅を広くしたり狭くしたりといった調節ができるようにする。

● すべての種類の標目からなる件名典拠ファイル

- したがって、アプローチとしては、書誌レコードに付与された件名から切り出した索引語をダイレクトに検索するのではなく、標目表そのものをシステムに搭載して提供することとなる。すなわち件名典拠ファイルを構築して、それを媒介に書誌レコードを検索する。件名標目表のもつ情報から、検索に有効な典拠ファイルをどのように作るかを考える。
- 書誌レコードに付与される件名標目には、件名標目表には採録されていない人名、団体名、地名といった固有名件名や、「フランス 経済」のような細目つき件名があるが、こうしたものも無視せず、典拠ファイルに収録してシステムに組み入れる。
- 特に地名件名や細目つき件名については、関係構造データを独自に作成し、普通件名と同様のナビゲーション機能が提供できるようにする。件名標目表で直接管理されない標目をも組み込むことで、より現実的なシステムとなる。

使用する件名標目表は基本件名標目表 (BSH) であり、現在改訂中の第 4 版 (1998 刊行予定) 編集用機械可読データ¹が素材となる。また、NACSIS-CAT データベースより抽出した書誌データを使用する。

最終的には、約 16 万件の書誌レコードと、約 3 万件の件名典拠レコードが相互にリンクされたシステムとなった。この規模で現実に動作するシステムを作成し、検索実験による評価を行なう。

なお、筆者は以前、日本十進分類法について類似の試みを行なったことがある [30]。このときには僅かのテストレコードでしか行なえなかったのに対して、本研究では件名標目表データが利用できるため、かなり大規模なプロトタイプが作成できる。それにより、作成したシステムの評価が可能になるほか、索引ツールの問題点などもより明確となる。

¹2.3参照。日本図書館協会より研究使用の許可をいただいた。

本稿の構成

この後の2章では、OPACにおける主題検索、件名標目表の利用について概観したあと、本研究の素材となるBSH4版について紹介する。また、NACSIS-CATデータベースの件名データについても現状をまとめる。

3章から本論に入る。3章では主に、作成したシステムBSH-OPACに実装した諸機能について述べる。

つづいて4章では、BSH-OPACに対する評価について述べる。本研究では2通りの実験を実施してシステム評価を行なっている。

5章では、件名典拠ファイルの構築作業を扱う。BSH-OPACの機能を実現するのにどのような処理が必要で、どこまでが機械的に行なえるのか、を述べる。

最後に6章で、全体のまとめを行なう。

2 主題検索、件名検索、BSH

本章ではまず、OPACにおける主題検索、典拠管理、件名の利用について、先行研究にもふれながら簡単にまとめる。つづいて、本研究の素材であるBSH4版とその機械可読ファイル(BSH4mr)について説明し、最後にNACSIS-CATデータベースにおける件名データの現状について述べる。

2.1 OPACの主題検索と典拠管理

OPACにおける主題検索

OPAC導入後の米国では、トランザクションログ分析などによる目録利用調査がさかんに行われているが²、しばしば指摘されるのは、カード目録時代と比べて主題検索の割合が大きくなっていること、そしてその主題検索の成功率が低く利用者を満足させていないということである。米国では主題検索といえば一般的にはLCSH(Library Congress Subject Headings) 件名標目が念頭におかれるので、その失敗の最大要因は利用者が適切なLCSH標目を選択できないことだとされ、自然語から索引語へ導く仕組みや関連する標目を参照する機能などの必要性が主張されている。

わが国でも数は少ないがOPAC利用調査が行なわれ³、主題検索機能の必要性や問題点も指摘されている。しかしながら、これまで大多数の図書館では(函架目録に近い「分類目録」以外には)きちんとした主題目録を編成してこなかったため、OPAC時代になっても図書館員の問題意識はやや薄いといわざるをえないように思う。

OPACでは書名などからのキーワード検索が可能になり、またMARCやNACSIS-CATの導入で件名標目データなども自然に取り込まれて検索対象となる。それで主題からのアクセスもある程度可能になったとしているが、これはきわめて中途半端な状態である。このような不十分な「主題検索機能」を提供しつづけるのでは、むしろ利用者の目録に対する信頼を失わせることになる。「主題からの検索」というからには、ある程度の網羅性をもって適合データが検索できるシステムを構築しなくてはならないと考える。

件名・分類による主題検索

カード目録時代には、「主題からの検索」といえば件名、分類であったが、OPAC時代にあってはもはや専売特許ではない。書誌レコードのあらゆるフィールドからキーワードを切り出すことができるし、さらに目次や巻末

² 齋藤 [23] pp. 79-80 に概観されている。最近の調査には、Drabenstott & Vizine-Goetz [5] pp. 123-240 などがある。

³ 国立国会図書館 [13] がログ分析と質問紙を併用して、かなり詳しい調査となっている。公共図書館では早い時期に、佐川 [22] の調査(杉並区立図書館)がある。大学図書館ではアンケートによる調査を吉田 [27] が行なっている。

索引を入力してキーワードを増やすといった試みも行われている。また最近の WWW 版 OPAC をみると「同義語検索」といった機能を実装しているものもある(実用レベルに達しているかは疑問だが)。

しかしながら件名・分類という伝統的な統制語による主題検索には、自然語によるアプローチでは代替できない点があると考えられる。とりわけ、上位語、下位語、連想語⁴を体系的に関係づけして提示し、利用者の検索語選択を助けることができる点は重要である。自然語間にももちろん階層関係等はあるが、語間の意味の重なりや微妙なずれを考えると、統制語のように関係を管理することは難しいのではないか。

文献レベルの情報検索システムとは異なり、OPAC で主題検索を行う利用者は一般にその主題にそれほど明るくはないと想定すべきである(専門家は専門書誌か専門データベースで主題検索を行い、OPAC の利用は所在調査が中心であろう)。そのような利用者は自己の検索要求を明確に表現できるとは限らず、したがってシステム側から用語間の関係づけや構造体系を提示することが求められている。

また、OPAC も大規模なものでは百万件のオーダーとなり、主題検索においては、利用者の一覧能力を超える大量ヒットの問題も指摘されている。ゼロヒット防止のためにキーワードを豊富にすることは、逆に大量ヒットの問題を拡大してしまう側面がある。件名や分類は、扱われる主題の特定性に対応した標目、記号のみを与え、上位、下位概念などとの関係は索引語間の参照構造、関係構造に委ねられる。これによって検索の適合率をあげる一方、必要に応じて関連語を探索することで再現率への要求にもこたえることができる。

主題検索と典拠システム

件名、分類などの統制索引語による主題検索にかかわる問題は、概ね次の 3 種類に大別することができよう。

- 検索システムの問題
- 索引付与作業の問題
- 索引ツールそのものの問題

このうち、使用する索引ツールそのものの有効性や問題点は、前の二つの問題を考察する中で浮かび上がってくる側面もあるのでひとまずおいて、他の二つについて考えたい。

⁴階層構造上の上位、下位ではないが密接に関連している標目間の関係は Associative relationship と呼ばれ、シソーラス上では RT(Related Term) で示される。本稿ではこれらを「連想関係」「連想語」と呼称する。「関連関係」「関連語」という訳語も使用されるが、RT 関係の意味では用いない。本稿中、作成したシステムの「関連語表示機能」などで使用する「関連語」は、階層関係や連想関係を含んだ「関連する語」という幅広い意味で用いる。

まず検索システムの問題であるが、現在のように件名、分類フィールドから検索キーを切り出して提供するだけでは、有効な主題検索は期待できない。自然語や参照語から索引語に導く仕掛けと、上位語、下位語等の関連語を参照してナビゲーションする仕組みが提供されなくてはならない。となると書誌レコードに情報を追加するといったやり方ではなく、件名表や分類表というツールそのものを何らかの見えかたで提供することが必要である。

ただ、現在内外で提供されている機械可読版の件名表、分類表に収録されているのは、基本的に印刷体に収録されている範囲の標目・参照である(追加・更新は別にして)。実際に書誌レコードに付与される標目はこれよりはるかに多い。分類にはナンバービルディングによる合成記号があり、件名には細目つき標目や人名等の固有名がある。こうしたものを無視して件名表や分類表のデータをそのまま提供するのは、明らかに不十分である。また逆に、個別の目録における書誌レコード集合を考えると、全く使用されていない標目もたくさんあり、やはり考慮が必要かもしれない。

次に、付与作業の問題である。適切な件名なり分類なりをすべての書誌レコードに付与していく作業はそれ自体大変な労力であるが、付与作業の一貫性を確保するには典拠管理作業にもコストをかけなくてはならない。また、ツールの改版等による標目変更にも多大な労力を要するが、これに目をつぶってしまうと有効な検索は保障されない。OPAC時代になり、ある程度は更新作業を自動化する道が開けていると考えられるが、それを実現するためには主題典拠ファイルをシステム内にきちんと位置づけ、書誌レコードに付与された標目とのリンクがはかられていなくてはならない。

以上のことから、有効な主題検索のためには、対象となる書誌レコード集合に付与された全標目と参照語よりなる主題典拠ファイルが必要だと考える。その典拠ファイルは、索引作業と目録検索の両方に使用されるものであって(索引作業と検索システムでは見えかたは異なるべきかもしれないが)、標目間の関係構造や注記など豊富な付加情報を含んだものでなくてはならない。

件名と分類

1章でも述べたが、筆者は以前 NDC を利用した検索システムについて発表したことがある [30]。今回は件名標目表を扱うわけであるが、主題検索システムにおいて件名と分類ではどのような違いがあるだろうか。

OPAC に搭載する主題検索ツールとして考えてみると、両者にはむしろ共通する点が多い。最大の違いは、件名がコトバでそのまま検索できるのに対して、分類は記号を介するという点であるが、検索システムに分類表を搭載する場合には当然相関索引なども収めてコトバからの検索を保証する必要

があり、それをシステム側でサポートすれば結局何らかのコトバから索引語が導かれるということになるから、もはや決定的な違いとはいえない。また列挙型分類の特徴は階層構造の明示にあったが、LCSH⁵も BSH4 版(予定)もシソーラススタイルになり、階層系が提示できるようになった。さらに、多くの件名表ではシソーラス一般と異なり、細目つき件名を採用するなど、かなり複雑な複合主題でも事前結合方式で一標目として表現しようとする場合が多く、その点でも分類との差異は薄い。

とはいえ、件名と分類(列挙型分類)の構造にはいくつかの違いがある。まず、分類表が知識の全分野を単一の階層構造として細分していくのに対し、件名は比較的多くの階層系の集合よりなる。次に、分類は記号がそのまま配列位置を示すので、一つの標目は一つの上位項目しかもたないが、件名では一標目が複数の上位項目をもつことがある。また、それと関連するが、件名が複数の階層文脈で扱われる用語を一標目で表すことが多いのに対して、分類では同一事象も分野が異なれば別標目となる。さらに、これは階層構造の問題ではないが、件名のほうが固有名や新主題を標目としやすいという違いもある。

一般的には、件名が特定主題の検索にすぐれ、分類は関連主題の探索に有効だという評価がよくなされる。従って件名と分類は補完しあっており双方を提供すべきだといった論もしばしばある。ただし、実際にわかりやすくシステム構築するのは容易でないと思われる。

2.2 件名標目表と主題検索

OPAC における件名標目の利用方法には、次のような形が考えられる。

- 件名フィールドから(ふつうは単語単位で)索引を切り出して、キーワード検索する。検索結果としては、直接書誌レコード集合が表示される。
- 件名標目を音順に配列し、入力した検索語の周辺をリスト表示してブラウジングさせる。多くは当該件名標目を付与された書誌レコード数とあわせて表示する。
- 件名標目表を OPAC に搭載し、参照構造や注記などの情報も利用者に提供する。

件名フィールドからのキーワード検索は、わが国の OPAC では最もよく見られる形である。典拠管理の問題はさておいても⁶、これでは件名標目表を知らない利用者が的確な検索をするのは困難だろう。件名として使用されて

⁵1988 年の 11th ed. から参照構造がシソーラススタイルになった。

⁶実際には、わが国の OPAC の多くは典拠管理に無関心なまま、MARC レコード等の件名フィールドから機械的に検索キーを切り出しているだけであり、事態はもっと深刻である。[11]

いる用語を想起できなければ検索できないし、ヒット数が多すぎる際の絞りこみも自分で工夫するしかない。

件名標目の音順表示は、要するにカード目録の模式であり、利用者には比較的わかりやすい。米国の OPAC にはこの形が多いようである。

この方式は、形の類似する件名や細目つき件名などが画面上で一覧できるので、利用者の選択を援助することができる。とりあえず「建築」で検索した利用者が「建築音響学」「建築計画」や「建築 便覧」「建築 フランス」などの一覧から自分の真に探したい件名標目を見つけることは十分考えられるし⁷、付与された書誌レコード数があらかじめわかれば、それも検索戦略に役立てられるであろう。

しかし、音順表示で提供されるのは「形の類似する件名」リストであって、「意味の類似」ではない。利用者が欲している「類似の件名」一覧は、やはり意味の類似する関連件名の一覧であろう。カード目録時代の件名目録は音順に配列するしかないのであるから、件名標目表の側では、転置形の標目を採用するなど⁸、なるべく意味の類似するものが形も類似(特に冒頭の形が)するように配慮してきた。とはいえ、ダイレクトな検索を考えれば人工的な語形は好ましくないから、結局中途半端な形にならざるをえない。OPAC 時代になってなお、音順を考慮して関連標目の集中を考えなくてはいけないのか、大いに疑問である。別途、標目間の関係をきちんと整理して持たせ、それを頼りに関連する件名標目を提示するべきであろう。

件名標目表そのものを OPAC に搭載し、標目間の関係構造を提供して利用者支援を行なうというのが、本研究のとり道である。

LCSH の機械可読版⁹を OPAC に搭載する実践は、既に 1980 年代半ばに Ohio State University で行なわれている (Ludy[15])。LCSH の注記や「を見よ参照」「をも見よ参照」(LCSH がシソーラススタイルになる前である)を画面に表示する仕組みである。

最近では、Dravenstott ら [5] がトランザクションログによる利用調査の結果に基づいて、LCSH を用いた検索システムのモデルを示している。このシステムの主眼は、入力した検索語と LCSH とのマッチングの度合などによって検索処理動作を変えて (“Search Tree” と名付けた処理フローを用いる)、利用者に最適と思われる結果表示をするという点にあり、なるべく多くの情報を利用者に提供して支援しようという本研究とはやや方向が異なる。ただ、細

⁷この際、画面上の配列が問題になる。ALA(American Library Association)では 1992 年に、OPAC 上での件名標目配列方法に関する現状と議論をまとめた報告書を出している [2]。

⁸例えば、BSH には「図書館(公共)」などがある

⁹正式には 1986 年に LCSH-mr として頒布されたが、それ以前から “Library of Congress Subject Authority File(LCSAF)” として利用されている。

目つき件名をアルファベット順に並べるのではなく、カテゴリー分けをして表示すべきである、といった提言などは、本研究を進める上で参考になった。

また、Bates[3][4]は、索引語(LCSH 件名標目)は予測不能である、索引語付与の一貫性は保たれ得ない、などの前提に立ち、利用者が適切な検索語に到達するための支援機能が必要であるとする。これまでのシソーラスはインデクサ用だったとして「スーパーシソーラス」¹⁰を提唱している。その要件の中心はより大規模な語彙であり、とりわけ口語などの採用で参照語を大幅に増やすべきだとしている。その他、ツリー構造による語間の関係提示、伝統的なシソーラスよりもより広く関係を設定、といった様々な提言がなされている。

海外の研究を詳細に調査することはできなかったが、調べ得た範囲では、細目つき件名や固有名も含めて統合的に関係構造を表示しようという方向のものは見当たらなかった。

わが国では、そもそもOPACの主題検索を扱った文献があまりなく、この分野の探求が進んでいるとはいえない。

上田[26]は、「構造を持たず、事前結合索引法を使用し、細目や限定子を使用している件名標目表」はOPACに適した索引語彙にはなりえず、シソーラスを開発すべきだとしている。一方吉田[29]は細目つき件名を用いた検索に一定の利点を認めながら、「オンラインにおける事前結合形利用の問題については、理論的にも実践的にもまだまだ解明されていない」としている。吉田はまた、LCSHのシソーラス化の不十分さや、BSH、NDLSH(国立国会図書館件名標目表[12])の現状について、オンラインでの有効性に疑問を呈している。

北、芝[11]は、わが国のOPACにおける主題検索機能の不十分さを指摘し、分類、件名による主題検索システムのモデルを提示している。このモデルでは、主題典拠ファイルをまず検索させる点や、階層構造をビジュアルに提示してナビゲーションさせるところなど、根幹の部分で本研究のシステム目標と同じであり、参考になった。また最後に、典拠ファイル構築の必要性も指摘されている。なお、北、芝は[10]で、BSH、NDLSH、NDC(日本十進分類法) 相関索引という3つの語彙による索引ツールを統合化する試みを行っており、各ツールの問題点にも言及している。

その他では、1980年代後半に件名標目表のシソーラス化に関する論がいくつかある¹¹。また、国立国会図書館における件名作業の詳細な報告[24]があり、各種件名の実態などについて参考になる。

¹⁰ “Superthesaurus” は [4] で使われており、[3] では同様のものが “End-user thesaurus” と呼ばれている。

¹¹ 例えば、豊田[25]、吉田[28]

2.3 BSH4 版改訂と BS4mr ファイル

『基本件名標目表』(以下、本稿では「BSH」)は、日本図書館協会によって出版されている件名標目表で、1956年に初版、1971年に改訂版、1983年には現行の第3版([17]以下、「BSH3」)が刊行された。間もなく第4版(以下、「BSH4」)が刊行予定である。BSH3の「標目の選定方針」¹²では「わが国における公共図書館、高等学校の図書館、大学の教養課程を対象とする図書館の件名目録編成に必要な基礎的件名標目を採録」とされている。特徴としては、学術語よりも一般に常用される語を使用していること、関連標目を示す連結参照(「をも見よ」参照)が豊富につけられていること、などがあげられる。

わが国で広く使われている件名標目表には、もう一つ『国立国会図書館件名標目表』([12]以下、「NDLSH」)がある。MARCをとってみても、TRC/MARC等の民間MARCはBSHを使用しているが、JAPAN/MARCではNDLSHが使われている。BSHは、同じく日本図書館協会の発行する『日本目録規則』(NCR)、『日本十進分類法』(NDC)と並んで整理技術の「3大ツール」と呼ばれるが、他のツールほど確固とした標準の地位にはない。

BSH4は現在まだ編集作業中であるが(1998中に刊行予定)、ほぼデータは固まっている。4版の編集作業にあたっては、BSH3版(1983)のデータをコンピュータ入力するところからスタートし、その後もコンピュータ処理を重ねて改訂作業を行っている。最終の印刷用原稿も電算出力の予定であり、機械可読版の頒布も検討している。本研究では、日本図書館協会ならびに日本図書館協会件名標目委員会の許可をいただき、BSH4版の編集用機械可読ファイル(1997年12月現在データ。以下、本稿ではBSH4mrと呼称する)を件名典拠ファイルの基本データとして使用する。

なお、本研究でのBSHの扱いについて、2点をお断りしておかなくてはならない。一点は、使用したBSH4mrがあくまで編集中のデータファイルだということである。編集作業は最終段階に入っているとはいえ、刊行までにはなお変更があるので、本稿における数値や実例はすべて暫定的なものである。また、機械可読ファイルとしてのデータ内容やフォーマットも、頒布時に再検討され、変更も予想される。もう一点、筆者は現在日本図書館協会件名標目委員として、改訂作業に携わっているが、本稿でのBSHに関わる見解はすべて個人的なものである。

さて、BSH4での改訂と、BSH4mrの内容について述べる。なお改訂に関しては、委員会名の公式報告として、「改訂の基本方針」[18]と「機械可読フォーマット案」[19]が出されている。また、冊子体BSH3の機械入力、シ

¹²[17] pp. 3

ソーラススタイルの仮作成、NDLSH からの取り込み、といったデータ処理の詳しい報告が、北、芝によってなされている [9] [10]。

3 版から 15 年ということもあり、大幅な改訂となる。機械可読になったという点は別にして、内容面での主な変更は次の通りである。

- 標目数の大幅な増加
- 例示件名、省略件名方式の見直し
- シソーラススタイルの関係構造
- 細目についての見直し

まず、標目数は約 4250 から約 7450 へと 1.7 倍余りになった (参照語等をのぞく)。NDLSH が国立国会図書館の蔵書を背景にしているのとは違い、BSH は特定の文献集合にリンクされた件名典拠ファイルではないので、新設件名の選定ソースは一つではない。文献的根拠の観点から、選定図書目録等の調査を行ったほか、図書館流通センターより TRC/MARC の新設件名候補データの提供を定期的に受けている。また、NDLSH から使用頻度の多い件名を中心にかなりのデータを取り込んでいる。NDLSH と形が異なる場合はなるべく参照として取り込む努力もしており、まだ完全ではないが相互リンクを目指している。

BSH は従来、先に述べたように公共図書館から大学教養課程の図書館までを対象としており、「特殊コレクション、大規模参考図書館の集書に必要な標目までは採録していない」として、標目の特定性のレベルを抑制してきた。その看板は必ずしもおろしたわけではないが、BSH4 では扱った図書があればかなり特定性の高い標目も採用している。

例示件名・省略件名の問題も、標目増加の一因である。従来 BSH では¹³、歴史事件名・言語名など 30 余りのカテゴリーを「比較的多く用いられるもののみを採録し、他は省略した」例示件名群とし、また、個人名・地名など 20 余りのカテゴリーを「表現形式の選択上の基準および参照関係の設定例を示す場合のほか、本表中に採録していない」省略件名群としてきた。すなわち標目として付与される語と認識しながら本表に掲載してこなかったものが多数ある。「業種名」「文学・芸術上の流派・主義」などもあげられ、いわゆる固有名詞の範囲を超えて様々なカテゴリーに及んでいる。BSH4 では、対象文献があるなら個別の名称を積極的に採用することとした (個人名、団体名、地名などは例外)。ただし、それぞれのカテゴリーについて全面的に図書出版の実績を洗い出して見直したわけではないため、採録・未採録にいささか不均衡も生じている。

¹³[17] pp. 3-6

関係構造のシソーラススタイルへの変更は、今回改訂のもっとも大きな点といえる。

BSH3の参照構造は例えば次のようになっていた。

人権	基本的人権	
	: 国家と個人, 自由	: 憲法, 国家と個人
基本的人権	人権	

で示されているのが「直接参照(を見よ参照)」「直接参照あり」で、件名標目と参照語との関係を示している。をそれぞれタグ USE、UF に変更した。これはタグの変更だけで、意味内容は同じである。

一方 :と: で示されているのが「連結参照(をも見よ参照)」「連結参照あり」であり、件名標目から関連する他の件名標目への参照である。連結参照は、一つ一つの参照関係が、上位・下位の階層関係にある場合には上位語・下位語を示すタグ NT, BT に、上下ではない連想関係にある場合には、連想語を示すタグ RT とに変更した。こちらは意味内容の点でも大きく変わっている。手法としては、従来あった「下位語から上位語への連結参照は作らない」という原則を手がかりに、「一方的な連結参照は NT、相互参照は RT」という仮説をたてて BSH3 を再構築し、これに逐一検討を加えて階層構造に整えている。ただし、シソーラススタイルになったといっても、一般的なシソーラス理論に基づいた階層構造にはなっていない。種類関係・全体部分関係等の関係構造概念は厳密ではなく、シソーラスではふつう階層と見なされない例も多い。

「人権」の例では次のようになる。

人権	憲法
UF:基本的人権	NT:人権
BT:憲法	
NT:自由	自由
RT:国家と個人	BT:人権
基本的人権	国家と個人
USE:人権	RT:人権

多くの件名標目表と同様に BSH には、「フランス語 辞典」「日本 対外関係 中国」のように複合主題を主標目と細目との組合せで表現した「細目つき件名」がある。細目つき件名はすべてを件名標目表にあげることはできないので、細目として使用できる語と使用範囲を規定している。BSH4 では細目の種別を整理して、従来あいまいさのあった適用範囲をかなり明確化した。また、地理区分について、従来特に指定した標目にのみ適用可能としていた点を改め、全標目に使用できることとした。なお OPAC 環境下では、

表 1: BSH4mr(編集データ 1997.12 現在)

件名標目	7,450
参照語	2,810
説明つき参照	40
細目	162

細目を使用した事前結合方式は不要 (ばらばらに付与して事後結合検索させればよい) との見直し論もあるが、この点は従来通りとしている。

BSH4mr は次の 4 種のレコードよりなる。

- 件名標目レコード
- 参照語レコード
- 説明つき参照レコード
- 細目レコード

表 1 に本研究に使用した時点でのレコード数を示す。

「件名標目レコード」は、実際に標目として付与される語である。階層関係、連想関係にある標目への参照を有している。

ID:BSH300230000
 HD:会社
 YM:カイシャ
 TYPE:10
 UF:私企業 BSH300266500
 BT:企業 BSH300116600
 NT:株式会社 BSH300097300
 NT:合弁会社 BSH300220200
 NT:有限会社 BSH300644800
 NT:会社法 BSH300065500
 NDC:335.4
 SANOTE:個々の会社名も標目となる

「参照語レコード」は、標目として使用されない語から件名標目への参照である。同義語のほか、下位語から上位語を参照するケースもある。また、複数標目への参照も存在する。なお、BSH4 では、件名標目ほどは増えていない。

ID:BSH300000900
 HD:あいさつ
 YM:アイサツ
 TYPE:20
 USE:式辞・あいさつ BSH300266800

「説明つき参照レコード」は、BSH4 で初めて現れる用語であるが、特定の標目への参照ではなく、注記によってある種の標目群に導くものである。BSH3 では《法律》という見出しのもとに法律関係の標目の使用法を総括説明するなど、「作業上注意を要する標目群」と呼称されるものがあったが、BSH4 では説明つき参照というカテゴリーで取り扱うことになった。現状では、もともとの経緯から、凡例的性格のものが多い。

ID:BSH300285600

HD:《社会》

YM:《シャカイ》

TYPE:30

SNOTE:(1) 各国の一般社会事情には、それぞれの国名のみをあたえる。

SNOTE:(2) 各国の社会史は、それぞれの国名のもとの主題区分 - 歴史であらわす。(例:日本 - 歴史)

SNOTE:(3) 一般社会史には、社会 - 歴史をあたえる。

「細目レコード」は、細目としてのみ使用される語で、「一般細目」「分野ごとの共通細目」「言語細目」「主題のもとの地理区分」の4種類に分かれている。なお、「歴史」「商業」のように標目としても同形が使用される場合でも、細目レコードを別レコードとしている。

ID:BSH300070900

HD:[会話]

YM:[カイト]

TYPE:40

TYPE2:言語細目

SNOTE:各言語のもとに細目として用いる。

2.4 NACSIS-CAT における件名

本研究では、学術情報センターの目録所在データベース (NACSIS-CAT) から抽出した書誌レコード集合をデータ処理の出発点とする。わが国で唯一の書誌ユーティリティである NACSIS-CAT は、参加館、レコード件数とも着実に増やし、近年では大学図書館ばかりでなく公共図書館や専門図書館の参加も目立つようになっている。各図書館で構築する OPAC も NACSIS-CAT の書誌レコード (以下、CAT 書誌レコードと呼ぶ) に大きく依存しており、その品質水準に左右されているといつてよい。したがって、CAT 書誌レコード上での件名を素材とすることは意味あることと考える。

CAT 書誌レコードはすべて参加館の目録担当者によって共同分担作成されているが、その作成負担を緩和するために、各種 MARC が「参照 MARC」として用意され、コピーカタログの材料となっている。和図書においては、国会図書館の JP/MARC と、図書館流通センターの TRC/MARC が用意されている。

CAT 書誌レコードの作成基準では、件名や分類は入力必須の項目ではない。わが国の大学図書館ではもともと書架分類以外に主題情報を組織すること

表 2: CAT 和図書レコードの件名付与状況

書誌レコード件数	1,318,744
件名付与あり	784,069
件名種類別	
BSH	179,893
NDLSH	641,256
MESH	9,208
LCSH	4,615
その他、種類なし	11,468

注: 1レコードに複数種類の件名が付与されていることもある

がほとんどなされていないので、CAT 書誌レコードにおいても、件名や分類が積極的に入力されているとはいえない。一方で、参照 MARC の JP/MARC、TRC/MARC には基本的に件名、分類が入力されているので、コピーカタログギングの場合には CAT 書誌レコードにそのまま取り込まれることになる。

参照 MARC の使用している件名標目表は、JP/MARC が NDLSH、TRC/MARC が BSH である。もっとも、BSH はこれまで「基本」の標目を掲載した標目表で、各機関が必要に応じて追加新設することを前提にしているので、TRC/MARC の BSH にも独自に追加された件名標目が多数存在する。両 MARC とも機関内では件名典拠管理を行っており、典拠ファイルに基づいて件名付与がなされている。ただし、変更や訂正があった際に、MARC 機関での書誌レコードが修正されても、参照 MARC まで修正情報は反映されないため、そうした不整合は存在する。

なお、2.3 で述べたように、BSH4 版編集にあたって TRC/MARC の新設件名候補データの提供を受けているが、この採否は図書館流通センターに連絡されており、結果的に TRC/MARC は BSH4 を一部先取りしている側面がある。

さて、本研究で使用した CAT 書誌データは、1997 年 8 月 22 日時点での図書書誌レコード全件である。うち、和図書レコード (言語コードが日本語のもの) は、1,318,344 件であった。

なお、CAT 書誌レコード中の主題情報に関しては、刊年 1985-1994 の和図書レコードに対する芳賀、松井の調査 [7] がある。本研究で出した統計数値も、基本的に芳賀らのものと近似した結果を示しているが、より広い対象年次での数値も採取したので、以下に示すことになる。

件名 (SH) フィールドが存在しているレコードは 784,069 件で、全体の 59.47% であった。件名表の種類別の数値とともに表 2 に示す。やはり、BSH と NDLSH が大半を占めている。

表 3: 出版年別件名付与状況

出版年	書誌総数	件名あり	付与率	BSH	BSH /件名あり	NDSLH	NDSLH /件名あり
1944 以前他	197,765	37,983	19.2%	1,804	4.7%	34,549	91.0%
1945-1949	20,476	7,396	36.1%	186	2.5%	7,032	95.1%
1950-1959	84,754	43,039	50.8%	900	2.1%	41,737	97.0%
1960-1969	127,994	76,412	59.7%	1,162	1.5%	74,276	97.2%
1970-1979	219,790	153,050	69.6%	1,340	0.9%	150,588	98.4%
1980-1984	154,425	109,733	71.1%	1,674	1.5%	107,494	98.0%
1985-1989	182,512	130,287	71.4%	34,962	26.8%	107,308	82.4%
1980-1989 計	337,589	240,226	71.2%	36,641	15.3%	214,994	89.5%
1990	40,127	27,838	69.4%	13,556	48.7%	18,907	67.9%
1991	42,491	29,801	70.1%	14,632	49.1%	20,479	68.7%
1992	44,150	30,064	68.1%	15,608	51.9%	18,833	62.6%
1993	44,882	30,792	68.6%	17,112	55.6%	18,068	58.7%
1994	46,050	31,797	62.3%	19,821	62.3%	16,033	50.4%
1995	48,239	33,378	69.2%	22,993	68.9%	14,425	43.2%
1996	44,894	30,521	68.0%	24,509	80.3%	9,047	29.6%
1997	17,575	11,468	65.3%	9,544	83.2%	2,048	17.9%
1990- 計	329,976	225,963	68.5%	137,860	61.0%	118,080	52.3%
全体	1,318,344	784,069	59.5%	179,893	22.9%	641,256	81.8%

「1944 以前他」には、出版年が未入力のものも含まれている。

「1980-1989」は、「198-」までの推定を含むので上 2 行の和にならない

次に、出版年別にみた数値を、表 3 に示す。

件名付与率は、1970 年ごろまで着実に増えたが、その後は、約 7 割で定着している。1990 年代に入ってから、むしろ微減を示している。参加館数増加とともに新刊書の書誌レコードを作成する時期が早まり MARC からの流用が減っているのか、MARC にのらないパンフレット類や AV 資料、コンピュータファイル等の登録が増えているからか、理由はよくわからない。

BSH と NDSLH の付与数を年代をおってみると、かなり劇的な変化がみられる。NACSIS-CAT システムがサービス開始される 1980 年代半ばまでのものほとんど NDSLH が付与されているが、その後急に BSH を付与されるレコードが増加し、1994 年刊行分で NDSLH との付与数が逆転した。その後も増え続け、1996、97 年には件名を付与された書誌レコードの 8 割以上に BSH が使用されている。これは、CAT システムでの参照 MARC 検索順序と、JAPAN/MARC における書誌レコード作成のタイムラグの問題により、新刊書については TRC/MARC が、遡及分には JP/MARC が、流用されることが多いからであろう。現在の参照 MARC 状況が続けば、この傾向は変わらないと考えられる。

もう一つ、NDC 分類別の統計を表 4 に示す。書誌レコード全件、件名付

表 4: CAT 書誌レコードの NDC 分類別統計

NDC 分類	全件 (1,318,344) 中		件名あり (784,069)		BSH あり (179,893)	
	書誌件数	比率	書誌件数	比率	書誌件数	比率
00-09(総記)	36,344	3.7%	27,705	3.6%	11,229	6.3%
10-15(哲学)	29,205	3.0%	22,356	2.9%	6,353	3.6%
16-19(宗教)	29,949	3.1%	24,630	3.2%	4,847	2.7%
20-28(歴史)	89,907	9.2%	80,529	10.6%	13,609	7.6%
29(地理)	22,887	2.4%	18,335	2.4%	4,194	2.3%
30-39(社会科学)	269,717	27.7%	238,048	31.3%	59,233	33.2%
40-48(自然科学)	54,363	5.6%	47,076	6.2%	15,934	8.9%
49(医学)	58,928	6.1%	54,060	7.1%	13,906	7.8%
50-58(工学)	87,718	9.0%	77,554	10.2%	20,421	11.4%
59(家政学)	10,831	1.1%	8,322	1.1%	2,120	1.2%
60-69(産業)	67,972	7.0%	61,189	8.1%	11,042	6.2%
70-77(芸術)	62,296	6.4%	47,760	6.3%	10,965	6.1%
78(スポーツ)	15,382	1.6%	11,599	1.5%	3,430	1.9%
79(諸芸娯楽)	2,588	0.3%	2,026	0.3%	474	0.3%
80-89(言語)	25,719	2.6%	20,595	2.7%	6,189	3.5%
90-99(文学)	139,218	14.3%	44,714	5.9%	11,365	6.4%
NDC あり総数	973,739	100.0%	759,413	100.0%	178,629	100.0%

「NDC あり総数」は NDC の付与があるレコード件数 (NDC の版は問わない)。
複数分類がついていることがあり、分類別の総和は「NDC あり総数」と一致しない。

与あり、BSH 付与あり、のそれぞれに付与されている NDC 分類記号を集計した。文学作品には通常件名が付与されないため、「件名あり」では文学分野の比率が低くなっているが、他はほぼ同傾向にある¹⁴。一方「BSH あり」では、総記の比率が高いなど少し違いが見られる。BSH の付与が比較的新しい図書に偏っているためではないかと思われるが未確認である。

いずれにせよ、それほど極端な違いはなく、BSH を付与された書誌レコードは全分野に渡っている。本研究では、BSH を付与された書誌レコードのみを対象とするが、分野による偏りを考慮に入れる必要はないといえるであろう。

¹⁴ 分類の付与されているレコードには、件名も付与されていることが多いので、これは当然の結果である。分類付与のないものも含めた全体の分野分布は確認のすべがない

3 検索システムの作成とその機能

本章では、件名典拠ファイルを利用した検索システム BSH-OPAC の作成と機能について述べる。

まず、本システムで目指した機能についてまとめ、次いで作成環境や DB 構成について述べる。その後、画面サンプルをあげて、実装した機能を一つ一つ説明していくこととする。

なお、処理の時系列からいえば、件名典拠ファイルの構築作業が先にあるわけだが、これについては 5 章で詳述する。典拠ファイル中の関係構造データ等に関して、本章では画面に直接現れる部分についての最小限の説明にとどめ、必要に応じて 5 章の関係する節番号を示すこととするので、詳細はそちらを参照していただきたい。

3.1 件名検索に必要な機能

1 章の「本研究の概要」で、本システムの目指す機能の概要を述べた。繰り返せば、特に重視するのは「関係構造を利用したナビゲーション」と「すべての種類の標目からなる件名典拠ファイル」である。ここでは具体例もあげて、もう少し詳しく説明することとする。

2.3 で述べたように、BSH4mr 中の各標目にはシソーラススタイルで他標目(または参照語)との関係データが示されている。

```
ID(レコード ID): BSH300020600
HD(標目形):      遺跡
BT(上位語):      考古学 BSH300204100
NT(下位語):      貝塚 BSH300068100
NT(下位語):      経塚 BSH300138000
NT(下位語):      巨石記念物 BS8900043300
(以下、省略)
```

件名標目「遺跡」の上位は「考古学」であり、下位には「貝塚」「経塚」等があることがわかる。

各標目につけられた BT, NT データを総合すると、次のような「階層構造表」を導くことができる¹⁵。

```
考古学
考古学/遺跡
考古学/遺跡/貝塚
考古学/遺跡/経塚
考古学/遺跡/巨石記念物
考古学/遺跡/古墳
考古学/遺跡/古墳/前方後円墳
考古学/遺跡/祭祀遺跡
```

¹⁵ BSH における階層構造表の作成は、北、芝 [9] が最初に行なっている。本研究のものはややフォーマットが異なるが、基本的な手順は同じである。

考古学/遺跡/洞穴遺跡
 考古学/遺跡/住居址
 考古学/遺跡/チャシ
 考古学/鏡
 考古学/原始社会
 (以下、省略)

この階層構造表を OPAC で提示すれば、「考古学」を最上位語とする件名標目群を一覧することができる。

一方、この階層構造表には表現されない関係構造も存在する。図 1 に、件名標目「遺跡」をめぐる様々な関係構造の例を示す。

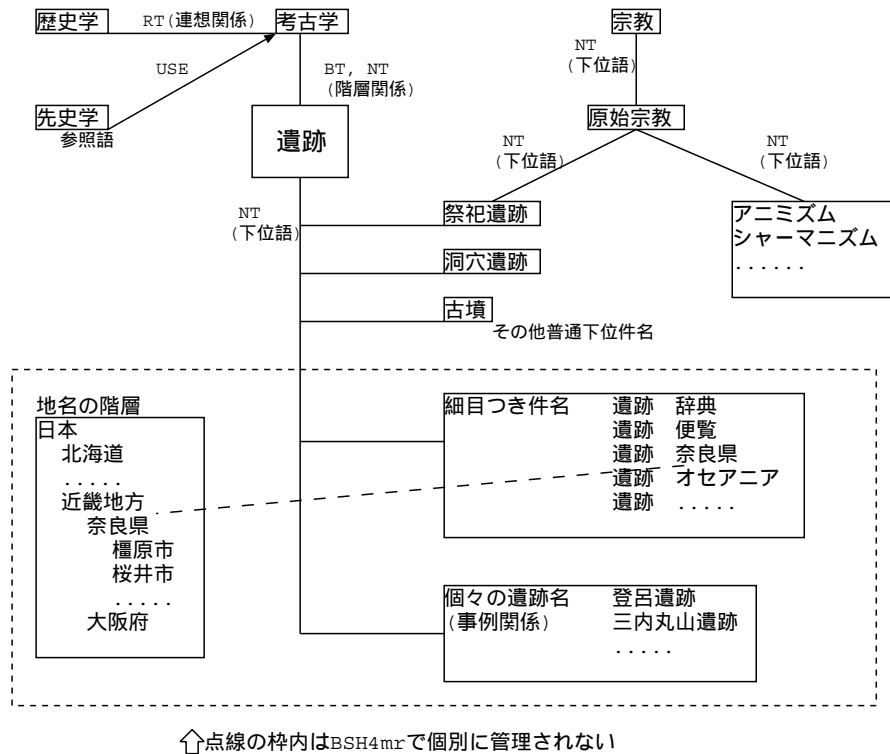


図 1: 件名標目「遺跡」に関する関係構造図

図 1 の上半分は、BSH4mr 中に存在する関係構造データである。これらは BSH4mr 中のレコードをデータ処理して件名典拠ファイルにおさめることで、提供が可能になる。

- 「歴史学」と「考古学」のような連想関係 (RT 関係) がある。階層関係ではないが、関連する標目である。

- 「先史学」は参照語であり、標目「考古学」が使用される。直接参照関係 (USE 関係) である。
- 「祭祀遺跡」は「遺跡」の下位語であると同時に、「原始宗教」の下位語でもある。別の言い方をすれば、「考古学」と「宗教」という2つの階層構造表に「祭祀遺跡」が出現することになる。BSH4mr 中の階層関係づけは、シソーラスの一般理論からするとかなり広い基準で行なわれているので (5.6.1参照)、このような重合階層関係が頻繁に存在する。

また、BSH4mr 中には存在しない件名標目、関係構造も存在する。図 1 では、下半分の点線で囲われた部分である。

- 「遺跡 辞典」「遺跡 奈良県」などの細目つき件名標目が多数存在する。細目つき件名は、件名標目表中では細目展開のルールのみが示されており、例示等を除いて BSH4mr 中には存在しない。これらは当然「遺跡」からみて下位にあたる標目であり、関係づけが必要である。また、主標目「遺跡」からのみでなく、「奈良県」等の細目部分からもアクセスできる必要がある。
- 「登呂遺跡」のような固有名件名標目も多数存在する。個人名、団体名、地名、書名といった固有名¹⁶は、例示等を除いて BSH4mr には収録していない。このような固有名件名も当然アクセスの対象となる必要がある。
- 普通件名「遺跡」に対して、「登呂遺跡」等の個々の遺跡名称を示す固有名件名標目は下位語であるといえる。このような普通件名と固有名件名との関係は「事例関係¹⁷」と呼ばれるが、この関係も視野に入れる必要がある。
- 「奈良県」は「近畿地方」の下位であるというように、地名件名標目間には独自の関係構造が存在する。これも関係づけが必要である。

「関係構造によるナビゲーション」といっても、このような関係のすべてを包含した件名典拠ファイルを構築し、検索システムを作成する必要がある。

またその他に、件名典拠レコードのキーワード検索やヒット書誌レコードの表示など、当然必要とされる機能が考えられる。

¹⁶ 「固有名」の定義は様ではない。歴史事件名や動植物名、言語名、病気名などの個別名称まで含んで用いられることもある。本研究では、個人名、家族名、団体名、地名、書名 (作品名) の範囲で用いる。これは IFLA の件名ガイドライン [8] で典拠標目の種類として普通件名と別にあげられているものと一致する。

¹⁷ [1] pp. 48

表 5: BSH-OPAC の検索対象レコード数等

書誌レコード件数	164,305	BSH を付与された NACSIS-CAT 書誌レコード 179,893 の 91.3%
付与件名標目延べ数	215,854	179,893 件の書誌レコードに付与された 247,390 標目の 87.3%
件名典拠標目数	21,071	うち、付与実績があるのは約 18,500 件。179,893 件の書誌レコードに付与された標目種類数 27,337 件のうち、約 7 割をカバーしている。
件名典拠レコード数	29,387	件名典拠ファイルのレコード数。件名典拠標目の他に、参照語等を含む

本システムでは、以上に述べた機能を極力満たすように設計した。ただし、手作業以外に手段がなくモデルを提示するにとどまった部分や、インターフェースの限界で不十分に終わったところもある。

3.2 検索システム BSH-OPAC の概要と作成環境

件名からの検索システム BSH-OPAC(図 2 が初期画面) の検索対象レコード数などを表 5 に示す。

詳細は 5 章で説明するが、BSH を付与された NACSIS-CAT 書誌レコード 179,893 件をもとに件名典拠ファイル構築を試みた。その結果、書誌レコードに付与された件名標目 27,337 件のうち 7 割ほどを件名典拠レコードとして確定した。

件名典拠ファイルをまず検索する仕様のシステムなので、この時点でカバーされなかった書誌レコード、件名標目は除去することとし、最終的には 164,305 件の書誌レコードに付与された延べ 215,854 標目を扱う検索システムとなった。

本システムは、WWW の CGI インターフェースを用いて作成している。プログラミング言語には Perl5 を用いている。データベース部分には GNU 版の dbm ライブラリである gdbm を使用し、Perl5 のハッシュ変数と連動させて処理を行なう。

この構成を選択した理由としては、筆者が HTML, CGI, Perl に多少の経験があったことのほか、gdbm の利用によりかなり大量のデータでも実効的な速度で動作すること、WWW は機種を問わず、また遠隔地からでも容易にアクセスできるためシステム評価の際の利便が大きいこと、などがある。

一方で、WWW によるインターフェースには制約条件もある。一回ごと

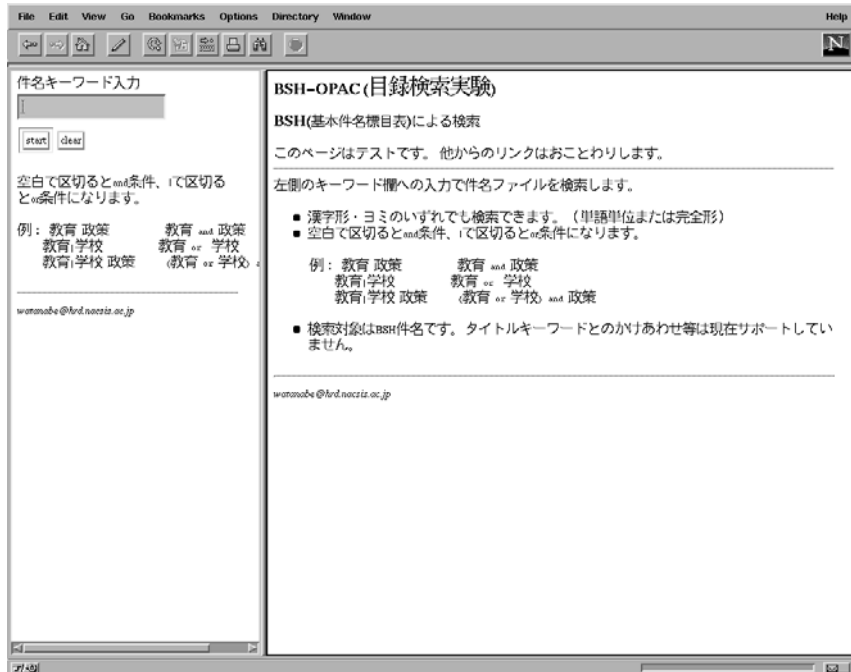


図 2: BSH-OPAC 初期画面

に接続が切れるので、履歴を保存しておくことが困難であるし、操作に応じて画面のごく一部だけを書き換えようとするロスが大きい。また、複数のウインドウ(フレーム)は開けても、一度に複数ウインドウを操作したり、複数ウインドウからの入力を同時に受け付けて処理する、といったことができない。こうした制約をクリアできる環境を選択すれば、操作性や直観的なわかりやすさの面で、なお改善ができると思われる。

なお、サーバマシンとしては、Sun SPARCstation5(OSはSolaris 2.4、メモリは64M)を使用している。

3.3 BSH-OPAC の DB 構成

本システムのデータベースは、6つのファイルからなる。「件名典拠DB」(典拠IDから典拠データを導く)、「階層パスDB」(階層パスIDからパスデータ)、「件名キーワードDB」(キーワードから典拠ID)、「典拠書誌DB」(典拠IDからリンク書誌ID)、「書誌DB」(書誌IDから書誌データ)、「タイトルキーワードDB」(タイトルキーワードから書誌ID)である。DB間の相関関係を、図3に示す。

この構成は、gdbmがキーと値をペアにして格納するだけの機能しか持た

注：模式的に示したので、フィールドの桁数などは実際と異なる。
で囲ってあるのが、各DBのキー項目。

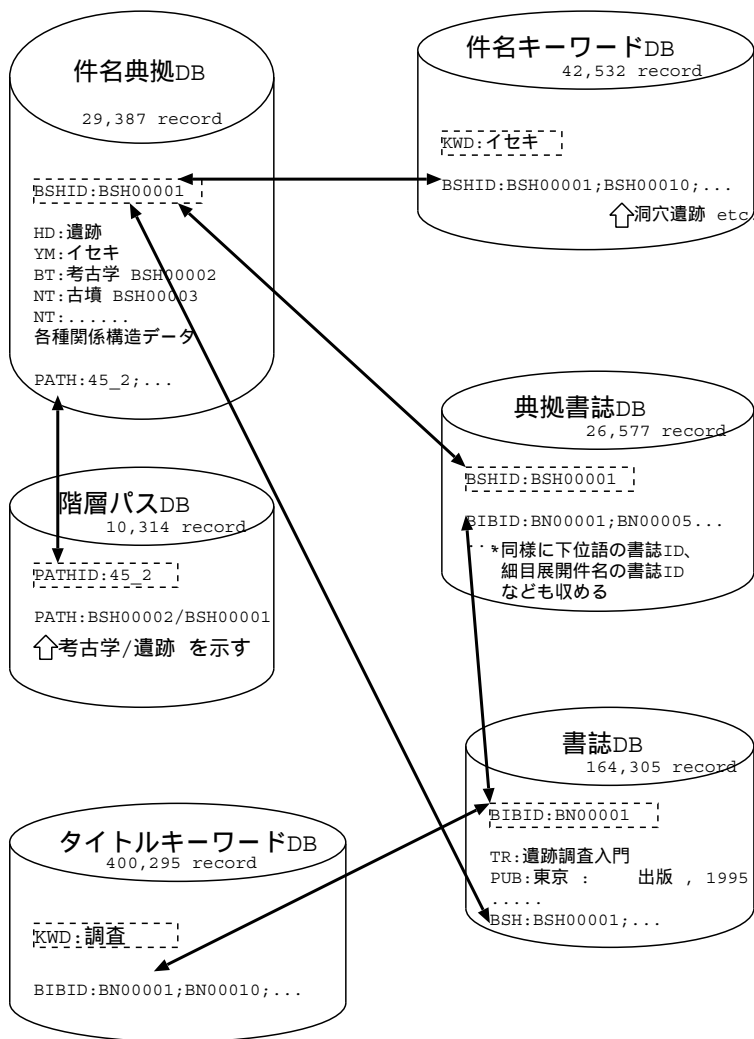


図 3: BSH-OPAC の DB 構成

表 6: 件名典拠ファイルの各種レコード

	主な情報源	件数	比率
普通件名標目レコード	すべて BSH4mr	6,789	23.1%
地名件名標目レコード	付与標目と NDC	1,575	5.4%
個人名件名標目レコード	付与標目と CAT 典拠	3,976	13.5%
団体名件名標目レコード	付与標目と CAT 典拠	464	1.6%
統一書名件名標目レコード	すべて BSH4mr	23	0.1%
固有名標目合計		(6,038)	(20.5%)
細目つき件名標目レコード	付与標目	8,244	28.1%
標目合計		(21,071)	(71.7%)
細目レコード	すべて BSH	161	0.5%
参照語レコード	すべて BSH	2,810	9.6%
補助地名レコード	NDC 本表	225	0.8%
主題+地名共出現レコード	付与データ	5,120	17.4%
総合計		29,387	100%

ないことからきており、RDBMS など他のデータベースシステムを使用すればまた違った設計となるだろう。

以下、各データベースファイルを順に説明する。

3.3.1 件名典拠 DB

件名典拠 DB は、件名典拠ファイルそのものを DB 化したもので、件名典拠 ID から件名典拠レコードを導くものである。

表 6 に、件名典拠ファイル (典拠 DB) に収録した各種レコードの件数と情報源をあげる。各種類のレコードの意味等はここでは述べないので、5 章を参照していただきたい。

21,071 件の件名標目レコードがあるが、BSH4mr 中のものは無条件で納めるなどしているので、リンクしている書誌レコードが存在しないものもある。また、標目以外のレコードもあり、典拠ファイル (典拠 DB) のレコード総数は 29,387 件である。

標目形やヨミなどに加えて、上位語、下位語をはじめとする関係データも、件名典拠ファイル中にもっている。5 章の表 13 に、フィールド一覧を掲げている。

3.3.2 階層パス DB

3.1 で、BSH4mr 中にある標目間の BT, NT 関係から「階層構造表」を導くことができると述べた。階層パス DB は、この階層構造表データを格納し

ている DB である。

階層構造表の 1 行に相当するデータをパスと呼称する。パス ID には階層系に採番した番号 (例では 45 が「考古学」の階層系を示す) と階層内での行番号を組み合わせたものを使い、ランダム位置からの順アクセスを可能としている。

パス ID	データ	(意味)
45_1	BSH300204100	考古学
45_2	BSH300204100/BSH300020600	考古学/遺跡
45_3	BSH300204100/BSH300020600/BSH300068100	考古学/遺跡/貝塚
45_4	BSH300204100/BSH300020600/BSH300138000	考古学/遺跡/経塚

(以下、省略)

ある件名標目レコードがどのパスに属しているかは、件名典拠 DB が記録している。各典拠レコードが自分の属する (自分がパスの最後となる、という意味。例えば「遺跡」は「考古学/遺跡」に対応する”45_2”を持ち、”45_3”などは持たない) パス ID を持っている。

階層パス DB は BSH4mr 中の階層関係 (5.6.1参照) と、地名件名間の階層関係 (5.6.2参照。NDC に準拠した階層構造をつけている) の両方をカバーしている。

レコード件数は、10,314 件である。

3.3.3 件名キーワード DB

件名キーワード DB は、件名典拠 DB をキーワード検索するために、キーワードから典拠 ID を導くものである。本システムでは、漢字形とヨミ形をそれぞれ単語単位で検索できることをめざした。これは、対照評価用のタイトルキーワード検索版 (3.5参照。NACSIS-CAT 書誌レコードのタイトルキーワードによる) と合わせるためである。

件名典拠ファイル中のレコードのうち、BSH4mr に存在したものは、わかち書きされていない漢字形と、わかち書きされたヨミをもっている。付与標目データよりとったものは、ヨミもわかち書きされていないし、時にはヨミのないものもある。個人名、団体名は NACSIS-CAT の著者名典拠ファイルから作成している (5.4.2, 5.4.4参照) が、これらは著者名典拠レコードにつけられたキーワードが利用できるため、そのまま機能要件を満たしている。

個人名、団体名以外のものは、日本語形態素解析システム JUMAN ver.3.4 を使用して、わかち書きやヨミふりを行なった。JUMAN のデフォルトの設定に従い、特にオプションの指定や辞書のメンテナンスなどは行っていない。わかち書きに問題のあるものの中にはあるが、そのまま使用することとした。

JUMAN の結果をもとに、各典拠レコードにキーワードを付した。単語単位と完全形の双方で切り出す。「漁業調整委員会」なら漁業、調整、委員会、漁業調整委員会、となる。ヨミも同様である。

なお、特別な操作としては、細目つき件名について、各要素の参照語から切り出した語を加えた。主標目だけのときはキーワード「運命」から「運命判断 易・占い」という参照語レコードが導かれるので「易・占い」に「運命」というキーワードはなくてもよいが、「易・占い 中国」があった場合には(「運命判断 中国」という参照語レコードはないので)、「運命 中国」という and 検索でもヒットするように「運命判断」からのキーワードを加えたのである。

各典拠レコードのキーワードが完成したら、これを転置してキーワードごとにまとめてDB化する。29,387件の典拠レコードから切り出されたキーワードは延べ141,044件で、キーワードDBのレコード数(異なりキーワード数)は42,532件である。

3.3.4 典拠書誌DB

典拠書誌DBは、典拠IDからリンクしている書誌IDの集合を導くものである。

単純なファイルであるが、本システムでは検索速度を上げるため、下位語レコードにリンクしている書誌IDや、細目展開されたレコードにリンクしている書誌IDなどもあわせて収めている。

レコード件数は、件名典拠ファイルから参照語レコードを除いた件数、26,577件である(ただしリンク書誌IDをもたない空のデータも含む)。

3.3.5 書誌DB

書誌DBは書誌情報をおさめており、書誌IDから書誌事項を導く。

本システムは書誌レコードの操作を主目的とはしないので、書誌レコードはNACSIS-CAT書誌レコードのうち、最小限のフィールドに限っている。すなわち、書誌ID、タイトル及び責任表示、版表示、出版事項、その他のタイトル、内容著作タイトル、書誌構造情報(PTBL)、BSH件名、である。

なお、BSH件名は付与件名そのままではなく、リンクしている件名典拠レコードの標目形と典拠IDを格納している。

レコード件数は、164,305件である。

3.3.6 タイトルキーワードDB

タイトルキーワードDBは、書誌DBをタイトルキーワードで検索するために、キーワードから書誌IDを導くものである。

件名典拠レコードがヒットしなかったときの補完に使用するほか、対照評価用のタイトルキーワード検索版にも必要である。

キーワードは NACSIS-CAT 書誌レコードのタイトルキーワードを使用する。書名のほか、その他のタイトルや内容著作タイトル等からも切り出される。

164,305 件の書誌レコードに対するキーワードは延べ 2,095,602 件で、キーワード DB のレコード数 (異なりキーワード数) は 400,295 件である。

3.4 BSH-OPAC の機能

ここでは、BSH-OPAC の機能について詳しく述べる。本システムは次の 5 つの機能よりなっている。

典拠検索機能 件名典拠ファイルをキーワード検索する。

関連語表示機能 ある典拠レコードに対して、上位語、下位語などの関連語一覧を表示する。

階層系一覧機能 ある典拠レコードを含む、階層構造表の全体を表示する。

書誌表示機能 ある典拠レコードとリンクした書誌レコードを表示する。

書誌フィードバック機能 書誌レコードに付与された件名標目からさらに探索を続ける。

3.4.1 典拠検索機能

本システムでは、入力された検索語で件名典拠ファイルをキーワード検索し、ヒットした件名典拠レコードの一覧を表示する。図 4 は、「大学」で検索した場合の画面である。ヒットした典拠レコード件数 (図 4①¹⁸) に続けて、各典拠レコードが表示される。

検索の様子は次の通りである。

- 漢字形、ヨミ形の単語単位、または完全形で検索できる (3.3.3 参照)。
- 複数の検索語を入力して、and, or 検索ができる (組合せもある程度は可能)。
- ヒットしたレコードにはランクをつけ、入力キーワードに「完全一致するもの」「冒頭が一致するもの」「それ以外」の順に表示する。ヒット件数が多くても完全一致するものが紛れることはない。なお、「完全一致」には、検索語「価値」に対する典拠「価値 (経済学)」なども含む (つまり、括弧で示される限定語は無視する)。

¹⁸ 図 4 中、① で示されたところ。以下、このように画面例を参照するものとする。

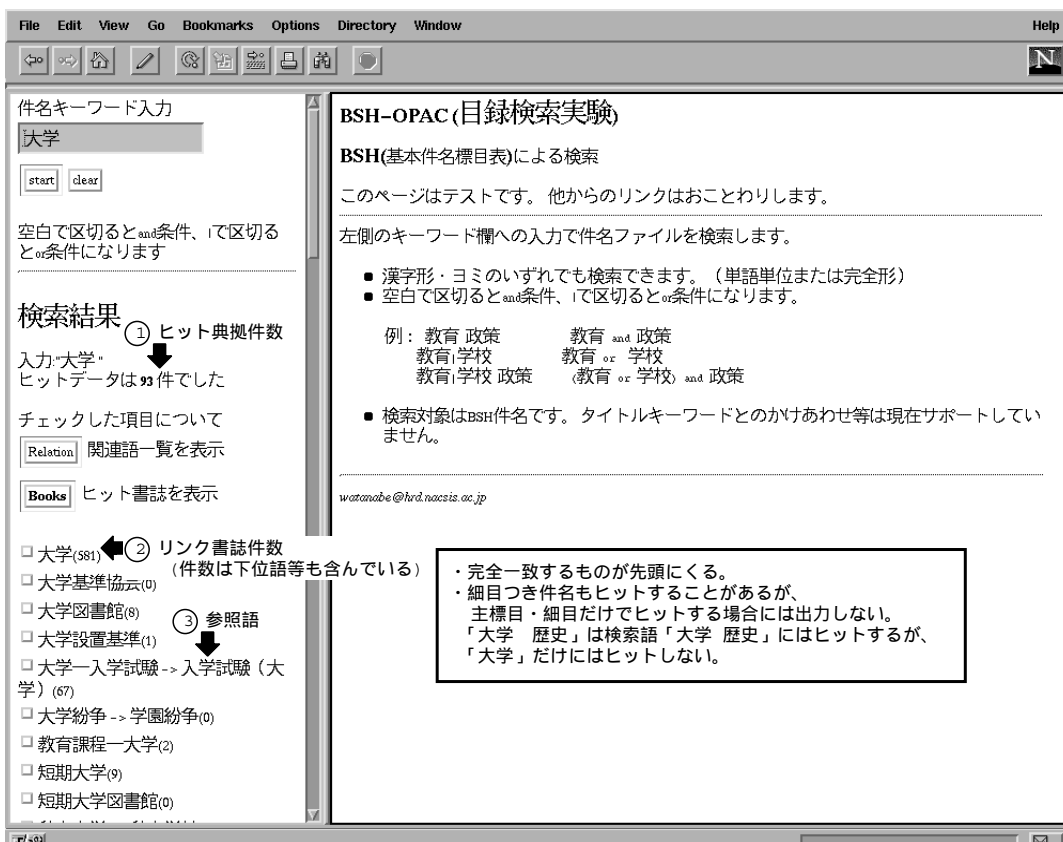


図 4: 検索画面 (「大学」の検索結果)

- 同ランク内では、固有名以外、地名、その他の固有名、の順に表示する。図 4では見えないが、続きには個別の大学名が並んでいる。(なお、「大学基準協会」は普通件名扱いとなっている)
- 参照語レコードにヒットした場合は、「参照形 標目形」となる(図 4③)。たまたま参照語と行き先標目が両方ヒットしたときは、参照語の方が一致度ランクが高ければ参照語を表示、そうでなければ標目形のみ表示する。
- 細目つき件名もヒットすることがあるが、主標目または細目だけでもヒットする場合には、細目つき件名は出力しない。「大学 歴史」に対しては「大学 歴史」がヒットするが、「大学」だけ、もしくは「歴史」だけの場合は「大学 歴史」を表示しない。(なお、「教育課程 大学」は細目つき件名ではなく全体で主標目)

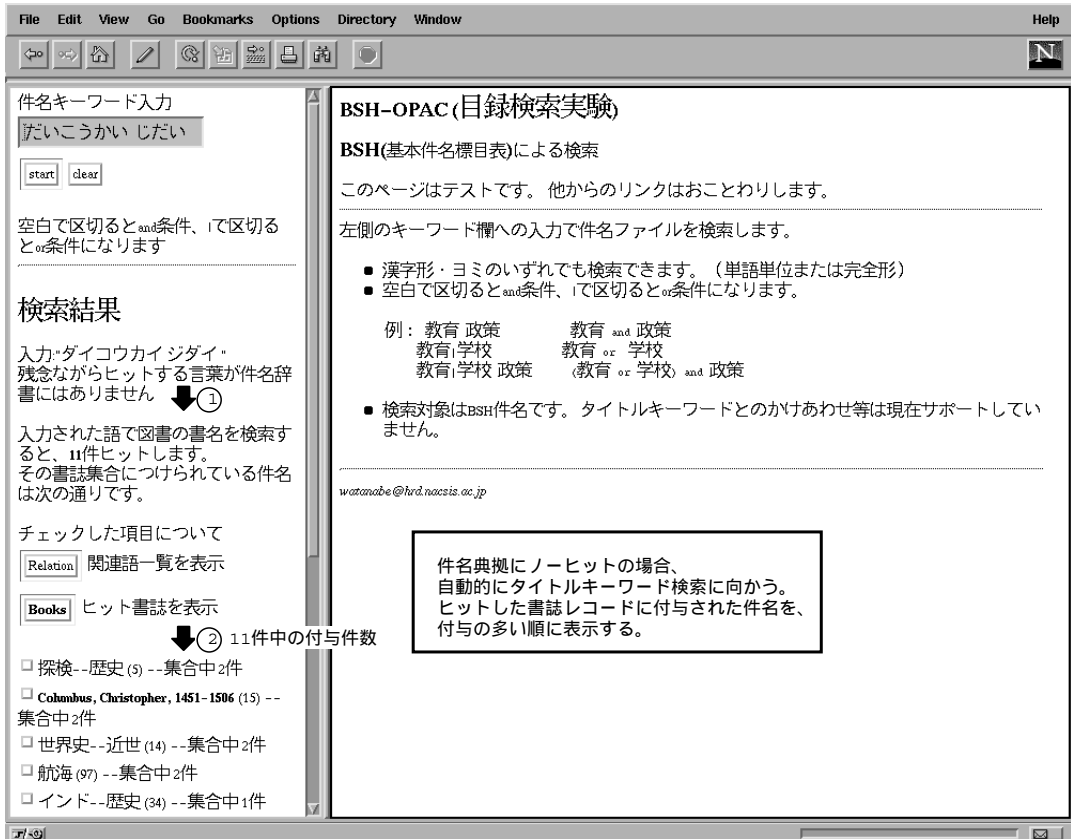


図 5: 件名典拠にはノーヒットの場合

- 「総合開発+北海道」のような「主題+地名共出現レコード」(5.5.4参照)の扱いも細目つき件名と同じである。

なお、ノーヒットの場合には、自動的に同一の検索語でタイトルキーワード検索を行ない、ヒットした書誌レコード(図5①)に付与された件名を、付与の多い順に表示する。図5に、「大航海時代」を検索した結果を示す。「探検 歴史」や「Columbus...」という関連件名が発見でき(図5②)、リンクされている書誌を見にいけば、書名に「大航海時代」とないものも含めて関連する図書を見つけることができる。

検索結果には、標目形の後ろに()でリンク書誌件数を表示している(図4②)。なお、この件数は当該標目形のつけられたものだけでなく、下位語や細目展開等を含んだ件数(詳しい範囲は3.4.2で述べる)である。

検索結果表示の次の動作としては、「検索し直し」のほか、「関連語一覧

を表示」(関連語表示機能)と「ヒット書誌を表示」(書誌表示機能)が選択できる。各ヒットレコードにはチェックボックスがつけられているので、選択したのち、それぞれのボタンを押下する。関連語一覧の場合は一つしか選択できないが、書誌表示では複数選択を受け付ける。

3.4.2 関連語表示機能

標目間の関係構造データを提示するのが「関連語表示機能」と「階層系一覧機能」である。「関連語表示機能」では他に、書誌レコード表示の範囲を調節する機能も提供する。

図 6 に、「大学」の関連語表示画面を示す。

関連語表示画面の表示内容は、次の通りである。

上位語と階層経路 上位語標目一覧(図 6①)と、その階層経路(階層構造表へのリンク 図 6②)を表示する。

ヒット書誌件数内訳 「現在位置」という見出しで、当該標目のリンク書誌レコード件数の内訳を示し(図 6③)、書誌レコード表示の調節を可能にする。

下位語 下位語標目一覧(図 6④)を表示する。

その他関連語 連想関係にある標目一覧を表示する。

細目展開 細目を付加して展開された標目一覧(図 6⑤)を表示する。

個別名称 普通件名に対して、事例関係にある固有名標目一覧を表示する。

参照、注記 典拠レコードに含まれる注記、参照、参照的注記を表示する。

ここにあげた順序が、画面での表示順である。ただし、細目つき件名の場合は、「細目展開」を「下位語」の上に行っている(細目つき件名の下位語は「準下位語」(5.6.4参照)が中心となり、「細目展開」のほうが重要度が高いため)。

以下、それぞれの項目について説明する。

「上位語」という見出しで、当該標目の上位語とされている標目を表示し(図 6①)、あわせて階層構造表へのリンクをはる(図 6②)。

- 上位語標目とは BT とされているもの、つまり直上位にある標目である。図 6 のように複数存在することもある。
- あわせて、各上位語標目にいたる階層構造表(5.6.1参照)上の階層経路を表示する。これにより当該標目への階層的文脈を明示するが、さらに階層構造表そのものへのリンク(「階層系一覧機能」)でもある。な

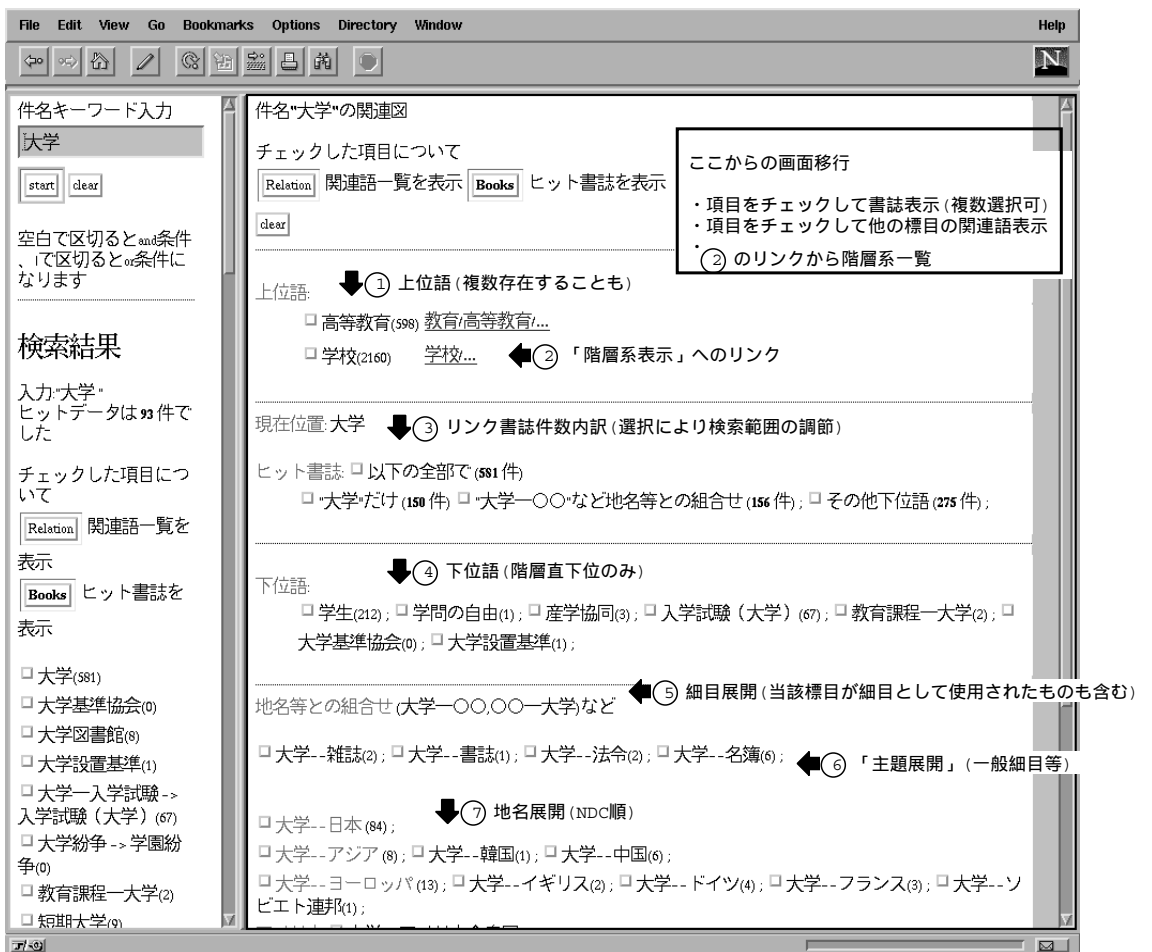


図 6: 「大学」の関連語表示

お、上位語標目も複数の上位語をもちうるから、一つの上位語に複数の階層経路が示されることもある。

- 自らが最上位語の場合には、上位語は示されないが、階層構造表(自らをトップとする表)へのリンクがはられる。
- 細目つき標目の場合には、細目を順次除去した形と、主標目・細目の各要素をすべて上位語とする。たとえば「日本 貿易 中国」の場合、「日本」「貿易」「中国」「日本 貿易」の4つが上位語となる。「直上位」という定義からはずれるが、細目つき件名には階層構造表にあたるものがないので、まとめて表示してしまう。
- 標目の後ろに示されたリンク書誌件数は、他の箇所と同様、下位語等を含んだ件数である。

「現在位置」(あまりいい名称とは思えないが、上位と下位の間なので)という名称で、当該標目に対するリンク書誌レコード件数の内訳を示す(図6③)。示されるのは「当該標目だけ」「細目展開」「下位語」「個別の名称」4分類である。

- 「“大学”だけ」といった見出しで、下位語や細目展開などを含まない、当該標目だけに限った件数を示す。非常に件数の多い場合に、これだけに限定すると、当該主題の一般的な図書が探しやすくなる。
- 「“大学 ” など地名等との組合せ」といった見出し(地名標目では「地名等」が「主題」になる)で、細目展開された標目にリンクする件数を示す。「細目展開」標目一覧(図6⑤)に対応しており、当該標目が細目として使用されるものなども含んでいる。
- 「その他下位語」という見出しで、下位語標目にリンクする件数を示す。「下位語」標目一覧(図6④)に対応している。
- 「個々の河川の名」といった見出しで、事例関係にある標目にリンクする件数を示す。
- それぞれにチェックボックスをつけており、各項目に対応するリンク書誌レコードを表示させることができる。もちろん複数選択も可能である。

「下位語」という名称で、当該標目の下位語とされている標目一覧を示す(図6④)。下位語には「準下位語」を含む。

- 下位語標目とは NT とされているもの、つまり直下位にある標目である。なおその下に下位語がある場合も多いが、それは表示されない。なお、下位語標目の配列規則は、特にない。
- 細目つき件名に対しては、「準下位語」(5.6.4参照) というものを設定し、下位語として一覧に示している。これは、細目つき標目を分解した各要素のいずれかが、その下位に位置する標目に置き換わった形式の、細目つき標目である。たとえば、「日本 農業」に対して、「北海道 農業」がそうである。また、「農村」が「農業」の下位にあると、「農村 日本」「農村 北海道」なども「日本 農業」の「準下位語」とみなす。

図 7 に「日本 農業」の例を示した。下位語のうち、先頭の「農本主義」は普通の NT 関係にある標目である。その後ろにやや間隔をあけて「準下位語」標目一覧を表示する。主題を表す部分ごとにまとめ、その中は地名の NDC 順に配列されている。

なお、「農村 岩手県」や「札幌市 農業」などは「日本 農業」の「直下位」というには疑問があるが、「準下位語」に対しては階層構造表を作成したり、ミッシングリンクを補正したりという作業を行っていないので、便宜的ではあるが、すべてまとめて表示する。「札幌市 農業」は「日本 農業」にも「北海道 農業」にも「準下位語」として現れることになる。

「その他関連語」という名称で、当該項目と RT 関係 (連想関係) にある語を表示する (ここにあげた画面例にはない)。

「地名等との組合せ」「主題との組合せ」といった名称で、細目展開された標目一覧を表示する。「細目展開」には、「細目使用」「準細目展開」(5.6.4 参照)「主題+地名共出現」(5.5.4 参照) などを含む。

- 典拠ファイル中では、当該標目が細目展開されたもの、および当該標目 (細目レコードも含む) を細目として使用したものを、「主題展開」「地名展開」「時代展開」「細目使用」に分けて関係づけている (5.6.4 参照)。検索システムでは、「細目使用」も含めて「細目展開」とみなす。特に地名と主題を組み合わせる場合に、地名を主標目とする (「フランス 商業」) が細目とするか (「博物館 フランス」)、は少なくとも OPAC の利用者に必要な区別とは思えないからである。
- 本システムでは「北海道+総合開発」のように主題と地名を”+”で結んだ「主題+地名共出現レコード」を件名典拠ファイルに多数収めている (図 8① 詳しくは 5.5.4 参照)。現行 3 版までの BSH において地理区分の可能な標目が限定されており、主題と地名を別々に付与することが

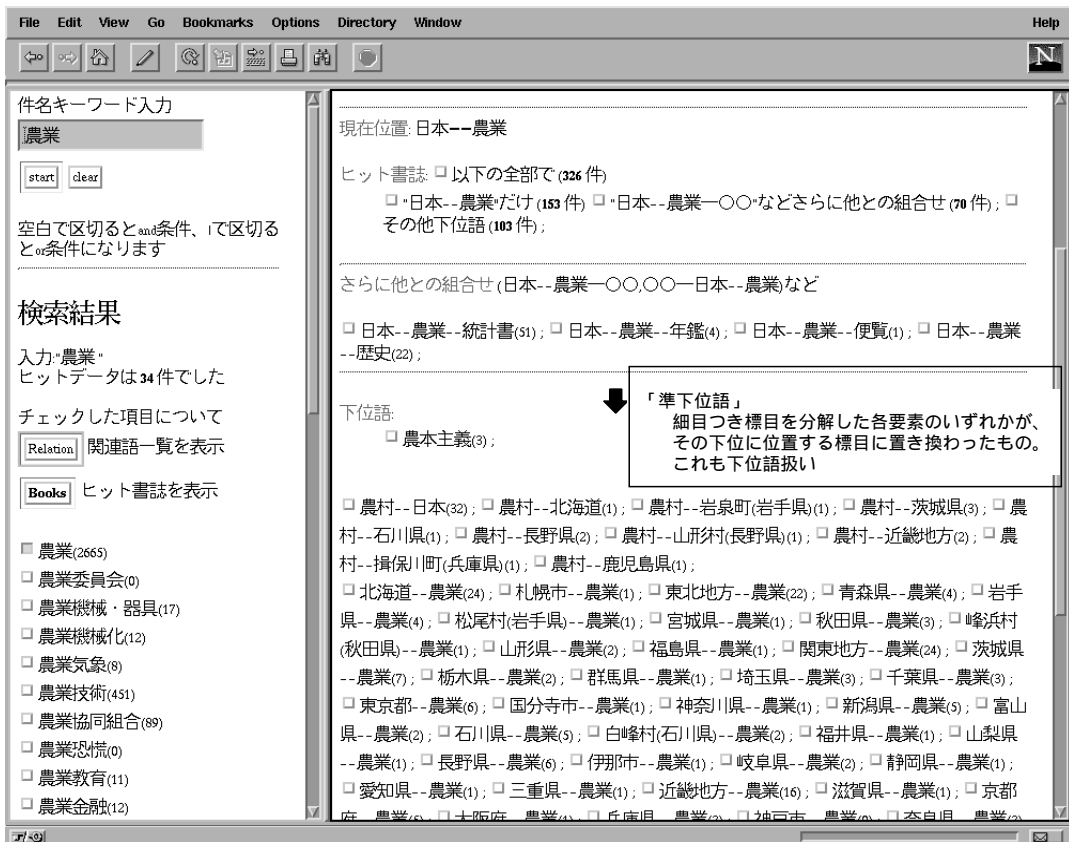


図 7: 「日本 農業」の関連語表示 (スクロール後)

頻繁に行なわれてきた。地理区分による細目展開と整合がとれず、検索に支障が出るため、機械的に主題と地名を合体させて作成したのが「主題+地名共出現レコード」である。地理区分による細目展開を補助する性格のものであるから、通常細目展開とまとめて表示する。

- これらすべてを「主題展開」「地名展開」「時代展開」の3つに分類し、以上の順に分けて出力する。図6でいえば、「大学 雑誌」の行が主題展開(形式区分も含んでいる)であり(図6⑥)、その後に地名展開がまとめて出力される(図6⑦)。
- それぞれのブロックの中がランダム配列では、多数にわたる時に見づらいため、ソート処理を加えている。
図6では地名展開が、地名階層構造のもとにしたNDC9地理区分に従った配列になっている。
また、図8は地名標目「フランス」に対する細目展開一覧である。ここ

では階層構造表の単位にまとめ、それを階層構造表の最上位語の NDC 順 (BSH4mr では標目レコードに NDC 記号をもっている) に並べてある (先頭の「紀行・案内記」等は細目レコードで、NDC 記号がない)。各行の先頭には階層構造表の最上位語が来る (実際の画面では文字色を変えている)。「哲学」のようにチェックボックスがない語 (図 8②) は、見出しとしてのみ示した最上位語 (「哲学 フランス」という標目はない) である。複数の階層構造に属する場合の処理ができていないなど不十分ではあるが、類似の標目がなるべく集まるように表示した。

- 本システムでは、細目つき件名の細目部分について、「歴史」のように単独の標目としても使用される場合は普通件名「歴史」にリンクし、「対外関係」のように細目としてのみ使用される場合は「細目レコード」として作成された「対外関係」にリンクしている (5.4.7, 5.5.1 参照)。このため、細目部分から検索すれば「歴史」や「対外関係」といった典拠レコードにヒットし、その関連語表示画面では、ちょうど図 8 とよく似た形で、その語が細目として使われている細目つき件名標目を一覧することができる。
- 細目つき件名において、「日本 歴史」に対する「日本 経済 歴史」等を「準細目展開」と設定している (5.6.4 参照)。これらは「準下位語」と同様、その他の「細目展開」とは別にし、後ろにまとめて表示する。表示順序はやはり階層構造表の単位にまとめた NDC 順である。

「個々の の名」という名称で、事例関係にある固有名標目の一覧を表示する (5.6.3 参照)。固有名標目が BSH4mr 中で管理されていないので、普通件名との関係づけは手作業でしか行なえず、事例関係はごく一部の地名標目にしか実装していない。

図 9 に「河川」の例を示す。階層構造表のもととなった NDC9 地理区分に従った配列順序である。

典拠レコードに注記、参照、参照的注記 (5.6.8 参照) があれば、「注記」または「参照」という名称で表示する。これらの中には、他の関連標目を示すものがあるが、何の処理もしておらず表示しているだけである。

関連語表示の次の動作としては、適当な項目をチェックして「関連語一覧を表示」で他標目の関連語表示に向かう、項目をチェックして (複数可能) 「ヒット書誌を表示」で書誌表示機能へ向かう、階層構造表へのリンクから階層系一覧機能へ向かう、が考えられる。もちろん検索語入力画面のフレームも残っているので新規検索もできる。

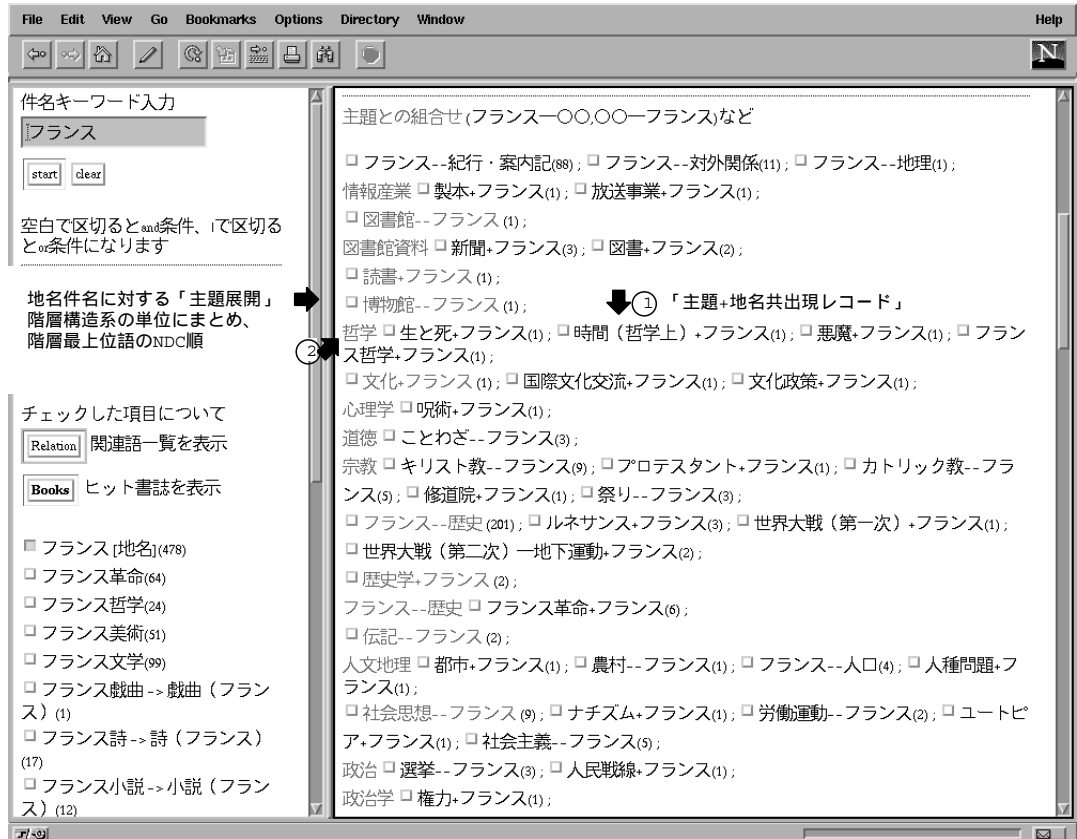


図 8: 「フランス」の関連語表示 (スクロール後)

3.4.3 階層系一覧機能

「階層系一覧機能」では、階層構造表の全体像を提示する。

図 10に、「工芸」を最上位語とする階層構造表を示した。各標目が、他の上位語をもっている場合は、その他の階層経路を後ろに示し、当該階層構造表へのリンクをはっている (図 10①)。ただし、上位語が同じになる場合は、他の階層表に属していても示していない。例えば図 10で「紙工芸」は「手芸」の階層表にも属していると示されている。ということは、その下位語である「折紙・切紙」等も「手芸」に属しているわけだが、下位語すべてに示すと煩雑なので「紙工芸」にのみ示している。

図 11には、地名標目の階層表の例として「ヨーロッパ」を示した。地名の階層構造は、NDC9 版に準じて作成している (5.6.2参照)。「日本」「アジア」「ヨーロッパ」といった単位の階層構造表であり、すべての標目を入れ

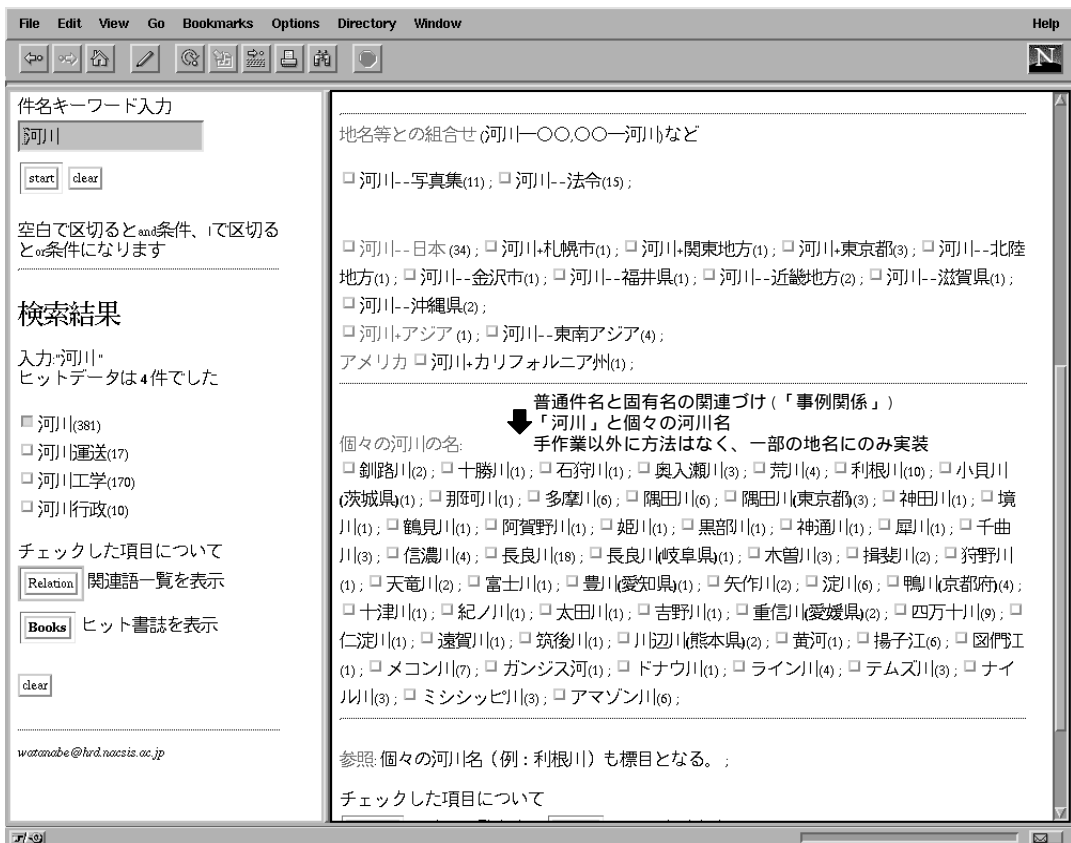


図 9: 「河川」の関連語表示 (スクロール後)

ると巨大になりすぎるので、NDC9 分類記号のある範囲 (NDC 記号の指し示す範囲と一致する標目) にとどめ、それより下位の項目は「関連語一覧」で見ってもらうようにしている (図 11①-③)。

階層構造表は往々にして大きなものとなり、限られた画面スペースでは全体が把握できるとはいいがたい。本システムでは、画面にとにかく表示する以上のことはできなかった。

階層系一覧の次の動作としては、関連語表示とほぼ同様で、選択した標目の関連語表示へ向かう、書誌表示へ向かう、他の階層系一覧に向かう、が考えられる。

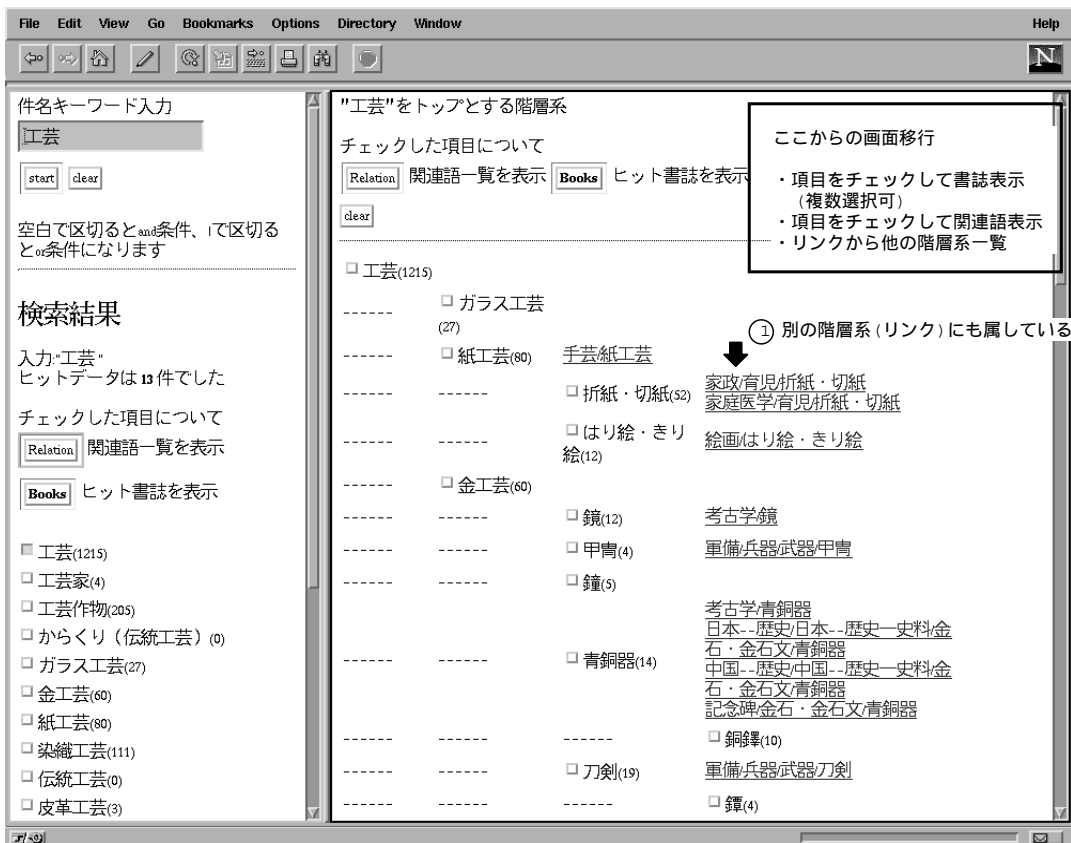


図 10: 「工芸」の階層系一覧

3.4.4 書誌表示機能

本システムでは、各画面に表示された件名標目の一つまたは複数選択して、リンクされた書誌レコードを表示できる。通常は、選ばれた標目に対して下位語や細目展開をも含んだ結果となるが、関連語表示画面では表示の範囲を調節することも可能である。

図 12 に、「玩具」に対する書誌表示画面を示す。この場合は下位語等を含んでいるが、その場合でも、“玩具”だけ、細目展開、下位語の順で表示されるので、比較的一般的なものが最初にくることが期待できる。

各書誌レコードに表示されている付与件名は、それぞれの件名標目の関連語表示画面へのリンクとなっている(図 12③)。書誌表示から新たな件名標目を発見して探索することができる。

書誌表示からの次の展開としては、書誌レコード中の件名標目を選択して

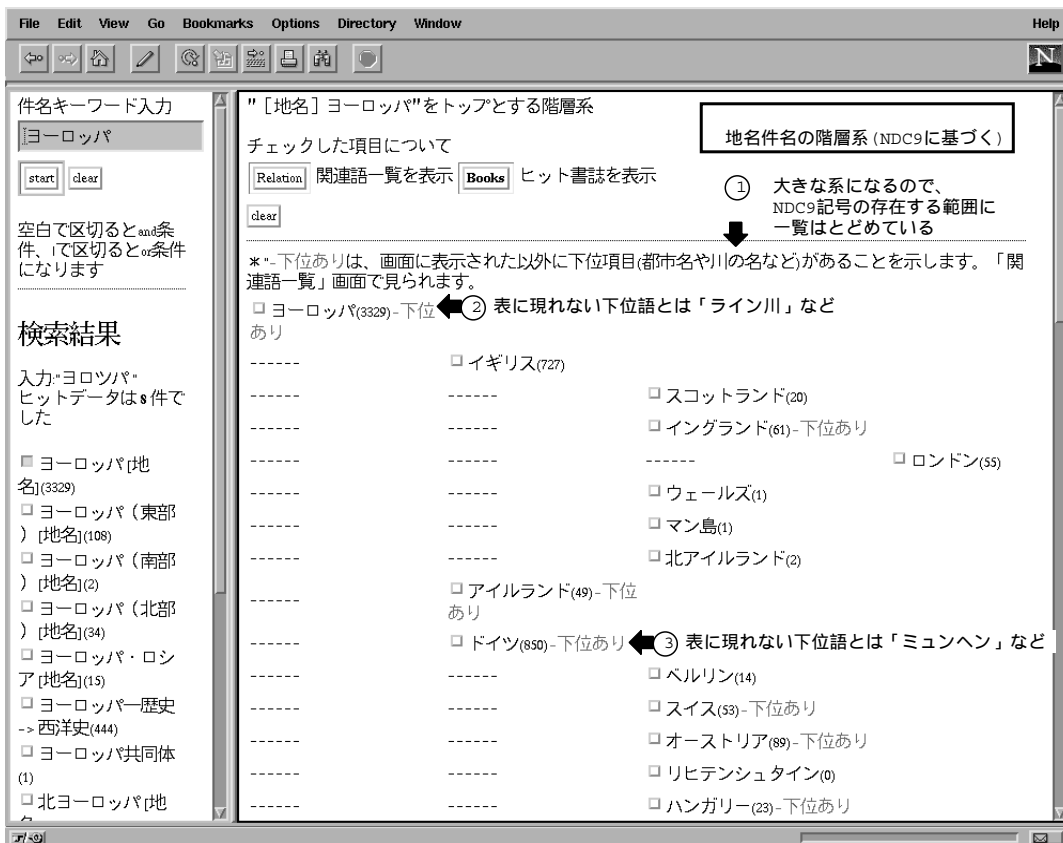


図 11: 「ヨーロッパ」の階層系一覧

関連語表示画面へ向かう、「この一群につけられた件名のリスト」(図 12②)で書誌フィードバック機能を使用する、といったことが考えられる。

3.4.5 書誌フィードバック機能

書誌フィードバック機能は、書誌表示画面上の書誌レコード集合に付与された件名標目を集計したリストを示すものである。本システムではタイトルキーワードでの検索機能がなく(ノーヒットの場合の自動処理を除いて)、書誌表示は常に件名標目を介して行なわれるので、もう一度件名標目を集計しても有意な結果が出ることは多くないが、それでも関連語表示等ではみつからなかった関連標目が発見される可能性もある。もし、タイトルキーワード等での検索を併用するなら、こうした機能は必須であろう。

図 13に、結果の例をあげる。

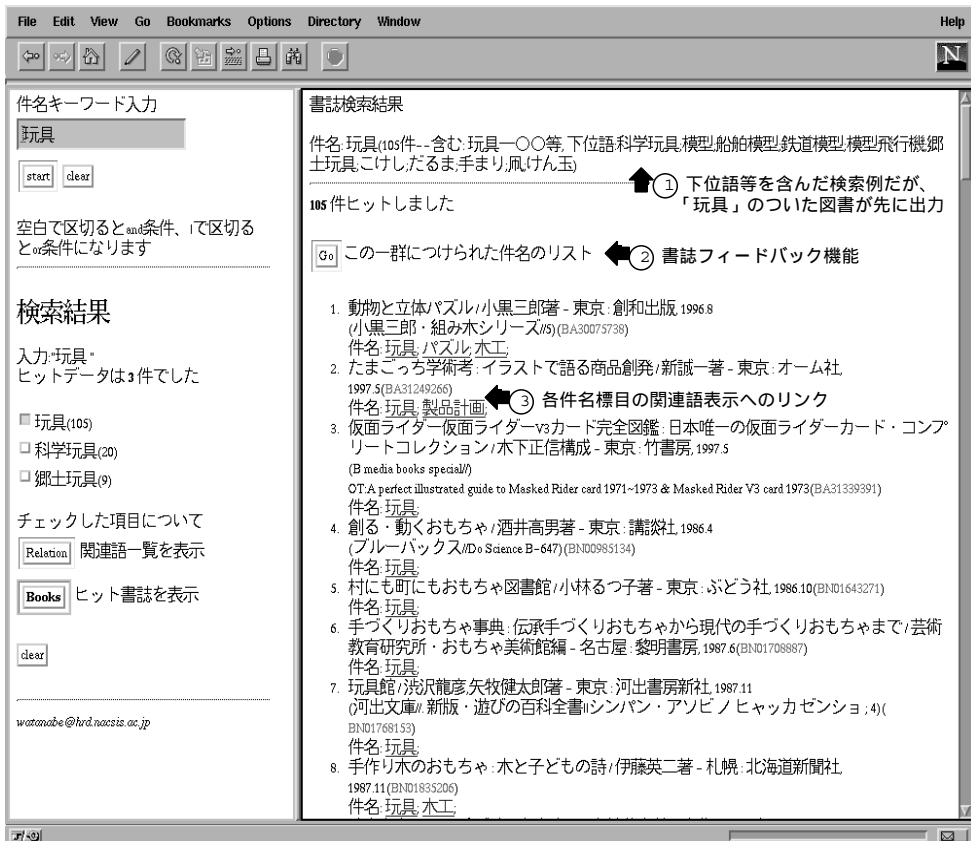


図 12: 書誌表示

3.5 対照用タイトル検索版 TT-OPAC

検索システムの評価実験(4章)を行なう際の対象用システムとして、同一の書誌DBに対するタイトル検索版を作成した。タイトルキーワードから該当書誌レコードを表示するだけの単純なシステムである。検索語入力方式はBSH-OPACと統一し、and, or 検索も可能である。ヒット書誌レコードの表示には、件名標目を含まない。

図 14に、画面例を示す。

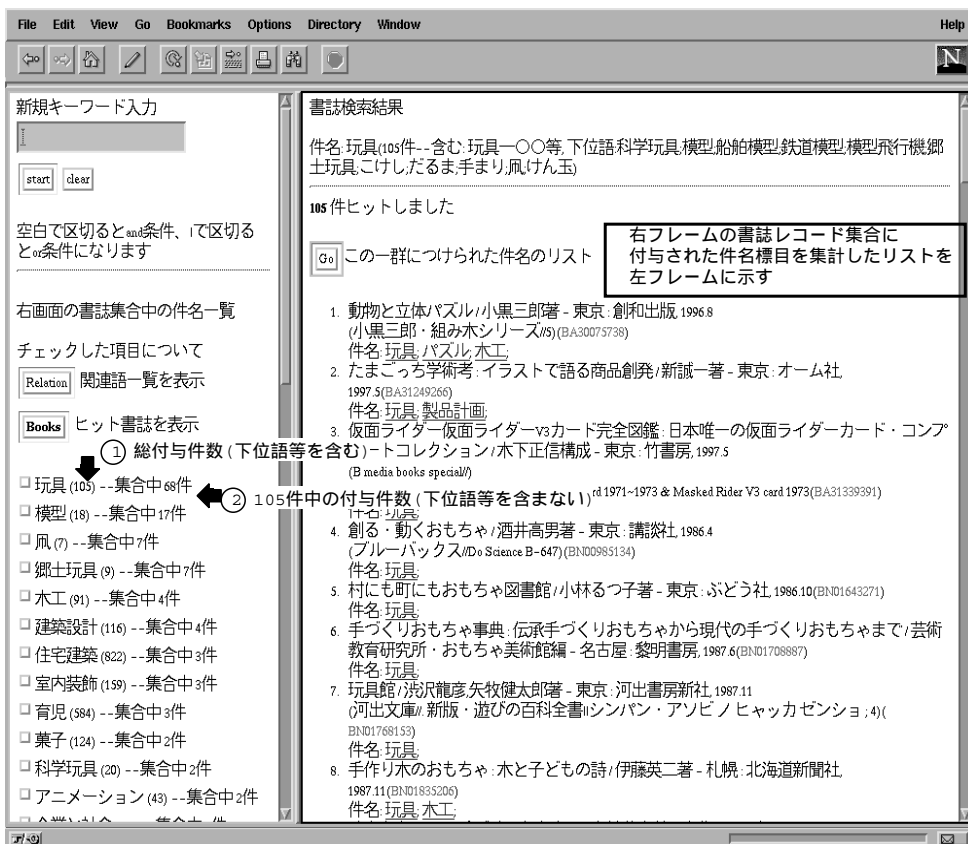


図 13: 書誌フィードバック機能

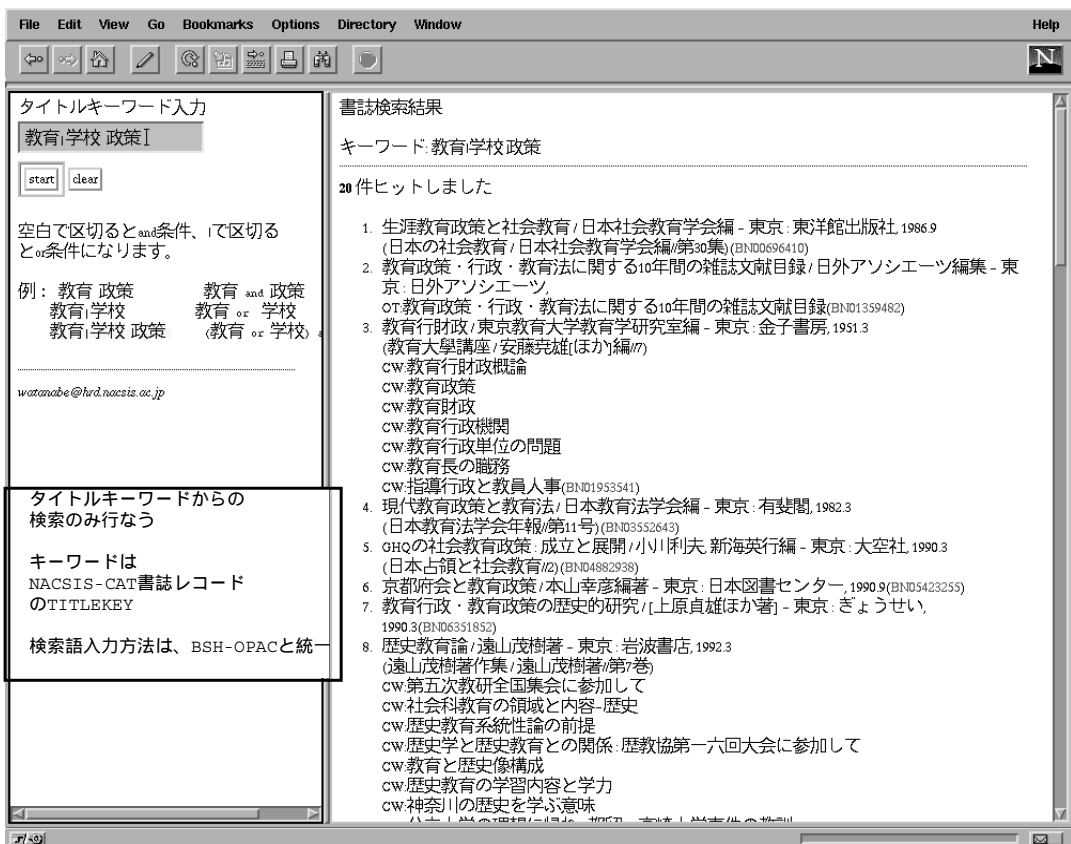


図 14: タイトル検索版 TT-OPAC

4 検索実験とシステムの評価

本章では、3章で作成した検索システム BSH-OPAC に対する検索実験を行ない、システムの有効性を評価する。

本システムにおいて、実験評価すべき点は2つあると考える。一つは、「与えられた検索質問に対して、BSH-OPAC の検索にどれだけ有効性がみられるか」である。なるべく利用実態に近く、偏りのない検索質問集合に対して行なわなくてはならない。いま一つは「BSH-OPAC の機能が利用者にとどの程度使いこなせ、受け入れられるか」という点である。本システムは、典拠ファイル上にある様々な情報を画面に提示するが、選択は常に検索者に任せるというスタイルであるから、この点の評価も大変重要である。

大きな規模で行なえるなら、被験者に自由に検索してもらい、トランザクションログを解析するといった手法で、両方の点からの調査が可能であろう。しかし、本研究では、時間も十分でなく、それほど多くの被験者も集められないので、2通りの評価実験を行なってカバーすることにした。

まず、筆者の勤務館の OPAC から採取した検索質問集合に対して、BSH-OPAC と対照用のタイトルキーワード検索版 TT-OPAC の比較検索実験を行ない(筆者自身で行なう)、与えられた検索質問への有効性を主に調査する。

一方でボランティアの被験者(図書館職員と学生)に協力をあおいで対人実験を行なう。いくつかの検索質問を与えて、BSH-OPAC の機能がどれだけ正しく理解されるか、を主に調査する。あわせて、TT-OPAC での対照実験も行なって、両システムにおける検索行動の比較も行なう(筆者自身による比較実験では、利用者の検索行動を模して行なうが、それはあくまで仮想的であるから、実際の被験者での比較によって検証する必要がある)。

4.1 与えられた検索質問に対する有効性評価

外部より与えられた OPAC の主題検索質問に対して、BSH-OPAC にどれだけ有効性が認められるかの評価実験を行なう。

素材となる検索質問は、筆者の勤務館である神戸大学附属図書館の WWW 版 OPAC からログを採取して得た。神戸大学では図書館内での OPAC には専用ソフトウェアを使用しているため、WWW 版は主に研究室や自宅からの利用、学外者の利用ということになる。ログ採取期間は 1998 年 1 月 29 日夕方から、2 月 2 日朝までの間である。

ログから機械的に和図書の主題検索にあたるものが拾えるわけではない。OPAC では検索対象ファイルや検索対象フィールドが指定できるものの、初期値がいずれも ALL になっているので、指定して検索する利用者はほとんどいない。結局はログを目で見て判断するが、特定の書名を検索しているのか、主題検索なのかは明確ではない。便宜的ではあるが、単語、熟語を入力

している場合には、主題検索とみなして抽出した。また、どこからどこまでが一連の検索なのか。その都度目で判断し、一連と思われるものは一つにまとめた。

主題検索として抽出された検索質問は 45 件、うち「生物」「司法」「ロシア」など、あまりにも一般的な語一つだけで検索が終了していて意図の判断できない 6 件を除き、最終的に 39 件の検索質問を得た。もっとも、他にもややあいまいな質問もあり、こちらで意図を仮定したものがある。

これらに対して、それぞれ BSH-OPAC と TT-OPAC の双方で検索を行った。なるべく一定のレベルの検索を行なうため、次の基準をあらかじめたてて行った。

- 検索に使うキーワードは概ね検索質問にある語のとおりとし、同義語は考えない。
- まず、利用者が最初に入力したキーワードから検索を開始する。
- BSH-OPAC での探索は、関連図等で常識的にたどれる範囲とする。
- 最終検索結果を一覧して、適合文献を数える。ただし、30 件を超える集合となったときは、先頭の 30 件のみを数える。
- 数えた適合文献の多い方を優位におく。つまり最終検索結果が 30 件以下なら検索できた適合文献の総数が、30 件を超えるときには先頭 30 件での適合数が問題になる。

適合文献の判断基準などあいまいな部分は残るが、とりあえず以上の基準で検索実験を行った。

実験結果は、39 件中、BSH-OPAC が優位にあるものが 18(46%) 件、TT-OPAC が優位にあるものが 11(28%) 件、同じ結果のものが 1(3%)、双方ともノーヒットのものが 9(23%) 件であった。同一結果のものを除くと、6 割強で BSH-OPAC のほうが優位にたつてはいるが、圧倒的な差ではなく、TT-OPAC のほうが良好な結果がでる質問もかなり存在している。

表 7、表 8 に、それぞれのシステムが優位にたつた検索質問をあげた。

二つの表を見比べてまず明らかなことは、「件名?」欄に示した「一致する件名標目の有無」が結果に大きな影響を与えていることである。

入力した検索語にほぼ意味内容の一致する件名標目がある 16 件のケースでは、「新交通システム」「日本語教育」「児童福祉法」以外の 13 件で、BSH-OPAC が優位を示している。一致する件名標目があれば、書名に同義語等が使われていても検索できる、下位項目も包含できる、という点で当然な

表 7: BSH-OPAC が優位にたった質問

最初の検索語	その他の語等	件名?	備考
水理学			書名に「水工学」等の語でも現れる
公共経済学			書名に「公共選択」等の語でも現れる
審査	貸出 (銀行業務と仮定)	×	件名「銀行貸付」書名に「融資判断」等の語
ローマ法			
社会教育			
海藻			件名「海そう」「わかめ」等個々の海そうのものもある
牛乳			書名に「ミルク」「乳」等の語でも現れる
権力	(政治学と仮定)		書名からではあまりに多様な集合
鶏	ブロイラー		件名「にわとり(鶏)」
国際金融			
ドイツ語史			件名「ドイツ語 歴史」
貨幣 哲学		×	件名「貨幣」一覧の手間はどちらも同じだが、BSH-OPAC のほうが余分なものが少ない
日本近代建築		×	件名「建築 日本」「建築 歴史 近代」
成人病			書名からでは無駄が多くなる
日本 食習慣	伝統的	×	件名「食生活 歴史」「食生活+日本」
博物館			書名では比喩的にも使われる
大航海	大航海時代と仮定	×	件名「探検 歴史」「Columbus, Christopher」等
グラフ	統計図表のグラフと仮定		BSH 参照語「グラフ 統計図表」

件名?: =意味の一致する件名標目あり ×=一致する件名標目なし

表 8: TT-OPAC が優位にたった質問

最初の検索語	その他の語等	件名?	備考
短距離 交通	新交通 交通システム		件名「新交通システム」あるが使われず、付与されているのは「鉄道」
一人っ子	ライフスキル 対人関係スキル 自尊心 セルフエスティーム	×	件名「児童心理学」等まで広がるので探せない。
しょうちょうは	象徴派	×	件名「彫刻」などを探すしかない。
リゾート		×	件名「観光事業」「レクリエーション施設」「レジャー産業」どれもびったりではない。
つうかきき	通貨危機	×	件名「国際通貨」
日本語教育			件名「日本語教育(対外国人)」あるが、付与されているのは「日本語」が多い。
長期波動	長期成長 景気循環論	×	件名「世界経済」「景気」
児童福祉法			件名「児童福祉 法令」だが、「児童福祉」だけの付与がある。
財務会計	連結財務諸表(こちらが検索主題と仮定)	×	件名「財務諸表」だが付与数非常に多く、連結だけ探すのは無理
TQM		×	件名「品質管理」付与数多すぎる。
アジア 民主主義		×	

件名?: =意味の一致する件名標目あり ×=一致する件名標目なし

がら件名に分がある。もっとも、「適合しているのに違う件名がつけられている」図書も存在し、それが甚だしいと「日本語教育」のような結果になるのであるが、概ねは件名検索の優位をくつがえすには至っていない。

一方、一致する内容のBSH 標目がない13件のケースでは、8件でTT-OPACの方が優位にあった。多くのケースでは、上位の概念の件名が(件名になればタイトルキーワードから検索するから)見つかるのであるが、どれも付与件数の多いもので、30件程度をブラウズしても目的のものには行き当たらない。一致する内容のものがないでもBSH-OPACが優位にたったケースには、「審査」「日本 食習慣」のようにある程度近いレベルの標目が存在するものが多い(「大航海」が例外)。

以上を要するに、主題を十分特定できる件名標目が存在しないと、件名からの検索は弊害の方が大きい。十分特定できる件名標目があれば、ほとんどはタイトルキーワードより良好な結果が得られる。したがって、どちらかといえば、件名付与作業の品質よりも、件名標目表の良否の与える影響の方が大きい、ということになるのか。

次に、BSH-OPACに実装してきた様々な機能が、有効に働く場面がどれだけあるか、という点にふれる。

以下のような機能が有効な場面があった。

- ノーヒットの際に自動的にタイトルキーワード検索を行ない、ヒットした書誌レコードの付与件名を表示する仕組みは、大変有効である。BSH-OPACが優位にたったもののうち、「審査」「日本近代建築」「日本 食習慣」「大航海」は、この機能がなければたどりつけない。
- 「グラフ」の例では参照語にヒットしている。
- 「海藻」の例では、「わかめ」といった下位語も拾ってきている。
- 「日本 食習慣」では「食生活+日本」という「主題+地名共出現レコード」(5.5.4参照)が役立っている。

今回収集した検索質問には、残念ながら複合主題のものがあまりないので、細目展開を提示するといった機能はそれほど生きる場面がない。これは、複合主題の要求がないわけではなく、入力する検索語となって現れにくいものと思われる。実システムのログを利用した場合の問題点である。

4.2 対人実験による評価

筆者の勤務館の図書館職員19名と学生4名を被験者として、対人実験を行った。いくつかの検索質問を与えて、どのようにシステムが使われるかをみるものである。

表 9: 対人実験予備調査

BSH について	職員	学生	WWW の使用状況	職員	学生
ある程度仕組みがわかる	5	0	毎日一度は	6	0
存在を知っている程度	10	0	週に 2, 3 回	1	0
はじめて聞いた	4	4	週に 1 回くらい	5	1
			あまり使わない	7	3

まず予備調査として、被験者に BSH に関する知識と、WWW の活用状況を尋ねた。結果は、表 9 の通りである。BSH については、職員でも「存在を知っている程度」が過半数を占めており、学生では全員「はじめて聞いた」とのことであった。同じ主題索引ツールでも、分類が目録作業の一部として行なわれ、書架配列を通じて利用者にも多少は周知しているのとは大きな違いがある。

実験にあたっては、職員の被験者 19 人を 10 人 (A) と 9 人 (B) の 2 群に分けた。BSH に関する知識と WWW 使用度が、システム操作の理解度に影響すると考えられるので、この 2 点で均等になるようグループ分けしている。検索課題は全部で 13 題を用意し、これも 2 群に分けた。それぞれの課題について BSH-OPAC と TT-OPAC での検索結果を得るが、同じ課題で同一人が両システムを検索すると入力語彙等で互いに影響しあうことが考えられる。そこで、被験者群 A に BSH-OPAC で検索するよう与えた課題群を、被験者群 B には TT-OPAC で検索してもらおう、という方式をとった。もう一つの課題群はその逆である。なお、学生被験者 4 人については、人数が少ないことと、それほど時間もとれないことから、13 題から選んだ 6 題を BSH-OPAC で検索してもらおうことのみの実験となった。

実験の手順は次のようなものである。

- システムの一般的な動きについて説明する。遠隔地なので口頭ではなく、若干のマニュアルを送付して読んでもらった。
- マニュアルを読んだあと、順次課題を検索してもらおう。
- 検索の過程 (入力キーワード、画面展開など) を用紙に記入してもらおう。これにより、最終結果だけでなく、途中経過の全てが把握できる。別途、検索ログも採取したので、不明瞭な場合はそちらを参照した。
- 最終的な検索結果集合はブラウズしてもらい、課題に適合すると思われる件数を数えてもらおう。大量にわたる場合は、30 件を下限として、ブラウズ数は被験者にまかせた。

表 10 に、検索課題一覧を掲げた。検索課題はなるべくいろいろな機能、状

況が試験できるよう作成した¹⁹。

それぞれの課題について、入力キーワード、その後の画面展開、最終的に見つかった適合文献数、が明らかになった。ただ、適合文献数は、集計してみると個人差がかなり激しいことがわかった。同じ結果集合に対しても、適合とみる数に数倍の開きがあることが珍しくない。また、BSH-OPACとTT-OPACでは、同じものがヒットしていても適合評価に違いがみられる(TT-OPACのほうが厳しくなる)ようである。ほとんどの場合BSH-OPACのほうが適合文献数が多く出ているが、それをもってBSH-OPACの優位性を唱えることには無理がある。

このようなシステムでは、最初に入れたキーワードからどのような画面展開を通じてどこにたどりついたか、という経過が重要な評価対象になる。したがって以下では、各課題の検索経過を中心に述べ、最終的な適合文献数は補完的なデータとして扱うことにする。また、学生被験者のデータは上述のように限定的なので、まず職員被験者の結果をまとめ、必要に応じて学生被験者の結果を付記する形とする。

4.2.1 細目を伴わない件名へのアクセス

用語のバリエーション 「建築物の防音」

〈課題：建築物の防音〉建築物の防音に関する図書を探したい。

この主題は典拠ファイル上の件名「建築音響学」が使用され、参照語レコード「防音装置」からもアクセスが可能である。リンク書誌レコードは22件である。「建築音響学」は少し広い主題なので、若干不適合なものも混じるが、概ねは防音に関係した図書である。

- BSH-OPAC(10名)における最初の検索語²⁰は、「防音」3名、「建築」5名、「建築物」「防音*建築物²¹」各1名、であった。
- 「防音」なら上述の参照語から、「建築」でもヒットした中に「建築音響学」がある。「建築物」での検索者も検索をやり直すなどして、最終的には全員が「建築音響学」に到達し、22件の書誌レコードを選択した。
- 適合文献と判断した書誌レコード数の平均は、12.9であった。
- 学生被験者4名も、全員「防音」か「建築」から「建築音響学」に到達した。

¹⁹ 課題作成にあたって、国立国会図書館のOPAC利用調査[13]中に紹介された実例をもとにアレンジしたものがある

²⁰ 以下、漢字形とヨミ形の違いは無視し、すべて漢字形で示す

²¹ 防音 and 建築物。本システムではスペース区切りが論理積だが、以下”*”で表記する

表 10: 対人実験の検索課題一覧

細目を伴わない件名へのアクセス		学生
建築物の防音	建築物の防音に関する図書を探したい。	
邦楽	邦楽(伝統的な日本音楽)についての図書をさがしたい。これはたくさんあると思うが、雅楽・琵琶歌など分野に特化したものよりは一般的なものがよい。	
外科学の入門	外科学について、概説的・教科書的な図書をさがしたい。	
バルト三国	バルト三国(ラトビア・エストニア・リトアニア)について扱った図書を探したい。一国に特定したものでもかまわない。	
脱酸処理	図書館資料の劣化を防ぐ手段として「脱酸処理」というものがあるらしい。これについて書かれた図書を探したい。	
ドライフラワー・リース	ドライフラワーやリースなど、花を使った装飾品について書かれた図書を探したい。	
日本の玩具	「こけし」など日本古来の玩具を扱った図書をさがしたい。特に個々の玩具を扱ったものでもよい。	

細目つき件名へのアクセス		学生
稲の病害	稲作における病害防除に関する図書をさがしたい。	
自然保護に関する法律	自然保護、動植物保護に関する法律を扱った図書をさがしたい。	
ベトナムの紀行記・旅行記	ベトナムの紀行記・旅行記の類を探したい。	
アジア映画	香港をはじめとするアジア諸国(日本は対象外)の映画に関する図書を探したい。	
九州地方の経済	九州地方の経済・産業を扱った図書をさがしたい。九州全体を扱ってなくてもよい。	
日本と東南アジアとの交渉史	日本と東南アジア諸国との対外関係史、交渉史に関する図書を探したい。東南アジアの各国に特定したものでもかまわない。	

「学生」の は学生被験者へも課題としたもの

- TT-OPAC(9名)では、適合文献数の平均は3.9であった。4名が「建築物」にヒットする32件をブラウズしているなど、検索に対する努力の跡がみえるが、結果的には「防音」の含まれない適合文献は、1名が「遮音」から2件を得ただけであった。

その他の課題でも明らかだが、TT-OPACでの被験者の多くはキーワードを様々に変えて検索を繰り返すものの、タイトルに現れる用語のバリエーションに対応することは難しい。検索課題に対して一致する、もしくは近似する内容の件名標目がある場合には、BSH-OPACにおいて良好な結果が簡便に得られる。

一般的な内容の図書を検索「邦楽」「外科学の入門」

「邦楽」も「外科学」もかなり広い主題であり、そうした語での検索は、網羅的というよりは一般的、概説的な図書を求めていることが多い。この両課題は、そうした検索を念頭において作成した。

〈課題: 邦楽〉邦楽(伝統的な日本音楽)についての図書をさがしたい。これはたくさんあると思うが、雅楽・琵琶歌など分野に特化したものよりは一般的なものがよい。

件名「邦楽」のリンク書誌件数は108件であるが、関連語表示画面のヒット書誌件数内訳(3.4.2参照)から「邦楽だけ」を選ぶと16件である。他に、細目展開として邦楽 辞典、邦楽 歴史等がある。

- BSH-OPAC(10名)における最初の検索語は、「邦楽」5名、「音楽」2名、「日本*音楽」「日本」「民謡」各1名、であった。
- 最終的に「邦楽だけ」を5名が、「邦楽だけ」に細目展開「邦楽 など」を加えた集合を4名が選んだ。
- 1名は件名「音楽 日本」を選んだ。「音楽 日本」は地理区分であり、「邦楽」は様式という使い分けであるが、わかりづらいのは確かである。
- 適合文献数の平均は、10.6であった。
- TT-OPAC(9名)では、適合文献数の平均は4.7であった。2名が「音楽*日本」から適合文献を見つけたが、他は「邦楽」からのみである。

〈課題: 外科学の入門〉外科学について、概説的・教科書的な図書をさがしたい。

件名「外科学」は476件のリンク書誌レコードを持つが、「外科学だけ」では42件、他に「外科学 雑誌」等の細目展開がある。

- BSH-OPAC(9名)における最初の検索語は「医学」が1名、他8名は「外科」か「外科学」であった。
- 全員が「外科学だけ」を選択した。
- 適合文献数の平均は、14.0であった。
- 学生被験者4名も「外科学」まで到達し、2名が「外科学だけ」を、2名が加えて細目展開「外科学 など」も選択した。
- TT-OPAC(10名)では、適合文献数の平均は8.4であった。「入門」や「概説」との組合せ検索も試みられているがあまり功を奏さず、結局「外科学」や「外科」の結果をブラウズしている場合が多い。いきおい「整形外科学」なども混在するので、効率が悪くなる。

リンク書誌件数の多い一般的な主題の場合、関連語表示画面で「 だけ」を選択すると下位語等が排除されて概説的な図書が見つかりやすくなることは、理解されているようである。

ただ、この際細目展開された標目を含むかどうかで、被験者の間に差がみられる。地理区分や地名のものと主題区分は主題の細分であり、「 だけ」とは別にした方がよい。しかし「 辞典」「 便覧」といった一般細目の多くは、その主題に対して一般的知識を得るのに有用で、「 だけ」に加えてもよい性格を持つ。一般細目を別扱いにすることも考えられるが、同じ一般細目でも「 法令」などは主題の細分である。細目に形式区分と主題区分が混在しているところに問題があり、これを整理するとより適切なシステムができると思われる。

下位項目の包含 「バルト三国」

〈課題: バルト三国〉バルト三国(ラトビア・エストニア・リトアニア)について扱った図書を探したい。一国に特定したものでかまわない。

典拠レコード「バルト3国」は、NDCより作成した補助地名レコード(5.5.3参照)である。下位の「ラトビア」「リトアニア」「エストニア」に、あわせて7件のリンク書誌レコードがある。

- BSH-OPAC(10名)における最初の検索語は、「バルト」ないし「バルト3国」が8名、「ヨーロッパ」「リトアニア」各1名であった。
- 全員「バルト3国」に到達したが、5名はあわせて各国別の標目も選択している。
- 適合文献数の平均は5.9であった。
- TT-OPACでの実験(9名)での適合文献数の平均は5.6であった(スペルミスで0件だった1名を除く)。「バルトの歴史」という図書に

「社会主義 歴史」だけがつけられているなど、不適切な件名付与によって件名からたどれないものがあり、各国別も検索する努力をすると、BSH-OPACより成績がよい。ただし、「バルト」では人名にも多くヒットするので、一覧に労力を要する。

BSH-OPACのほうが検索の手間が少ないのは確かであるが、半数が各国別の標目も選択しようとしていることは、階層表示の考え方が完全には浸透していないことを示しており、インターフェースに改善の余地がある。

上位項目の参照 「脱酸処理」

〈課題: 脱酸処理〉図書館資料の劣化を防ぐ手段として「脱酸処理」というものがあるらしい。これについて書かれた図書を探したい。

件名「脱酸処理」があるが、リンクされている書誌レコードが存在しない。階層上位が「資料保存」であり、こちらには15件がリンクされている。

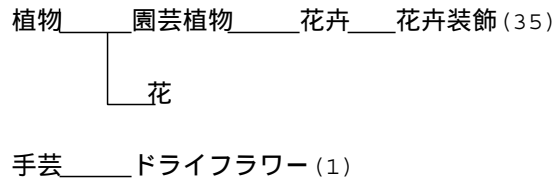
- BSH-OPAC(9名)における最初の検索語は、「脱酸」ないし「脱酸処理」が8名、「図書館*資料*劣化」が1名であった。
- 全員が「資料保存」に到達した。1名が「資料」によるキーワード検索で見つけているが、8名は「脱酸処理」から上位をたどっている。
- 適合文献数の平均は、7.1であるが、他の課題にもまして適合判断の個人差が大きい。
- TT-OPAC(10名)では、適合文献数の平均は4.2であった。「資料*防止」「資料*保護」「劣化」「酸性紙」など20種類近い検索がなされ、一人当たりでも4個以上の検索語を入力している。「資料*保存」での検索は4名であった。

上位を参照すればいいというのは、直観的にわかりやすいようである。TT-OPACでの結果は、利用者が自ら上位語を想起して検索するのが容易でないことを示している。

より複雑な階層 「ドライフラワー・リース」「日本の玩具」

〈課題: ドライフラワー・リース〉ドライフラワーやリースなど、花を使った装飾品について書かれた図書を探したい。

件名「ドライフラワー」はリンク書誌レコードが1件、「花卉装飾」が35件で最も望ましいが、やや階層構造に問題があり、階層系が別々になっている。

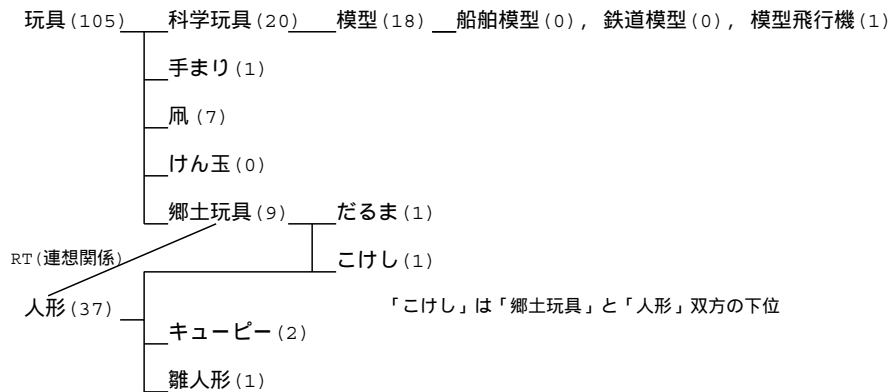


- BSH-OPAC(10名)における最初の検索語は、「ドライフラワー」3名、「花」3名、「装飾*花」「装飾」「装飾*植物」「装飾品」各1名であった。
- 「花卉装飾」に到達したのは7名である。4名は「装飾」による検索を行なった。2名は「ドライフラワー」の書誌表示(1件)に「花卉装飾」が同時に付与されているのに気づいた。1名は「フラワー*アレンジ」で検索し(件名としてはノーヒット)、タイトルキーワードでヒットしたものの付与件名として「花卉装飾」が表示された。なお、7名中4名は「ドライフラワー」をも見出しているが、3名は「装飾」から出発したため、全く別の系に「ドライフラワー」があることは気がついていない。
- 3名は「ドライフラワー」「クリスマス」等を選択した。
- 適合文献数の平均は20.4、「花卉装飾」に到達した人に限ると28.6であった。
- TT-OPAC(9名)では、適合文献数の平均は3.9であった。「花」等ではヒット数が多すぎてどうしようもなく、結局「ドライフラワー」「リース」以外から適合文献を発見した人はいない。

この種の主題はタイトルに使われる語のばらつきが大きく、件名検索の優位性が出るところである。しかし、階層系が分かれてしまっているため、「花卉装飾」「ドライフラワー」を両方発見したのは4名である。それでも、事前の予想よりは「花卉装飾」に到達した人が多く、特に書誌表示中の件名からの到達は考えていなかった。本システムでは、「書誌フィードバック機能」はあるものの、(主題と地名の場合を除いて)書誌レコード中での複数件名の共出現ということは深く考えてこなかったが、実験全体を通じて書誌表示中の他の件名を参照する人は予想外に多かった。

〈課題:日本の玩具〉「こけし」など日本古来の玩具を扱った図書をさがしたい。特に個々の玩具を扱ったものでもよい。

件名「玩具」に関する階層構造は次のようになっている。「郷土玩具」のほか、「手まり」、「凧」といった件名がある。



- BSH-OPAC(9名)における最初の検索語は、「玩具」5名、「日本*玩具」2名、「日本*玩具*伝統」「日本」各1名であった。
- ほとんどの人が「玩具」の関連語表示画面を手がかりにしているが、最終的な選択はかなり分かれた。
- 4名は「郷土玩具」を中心として選択した。「郷土玩具」だけが2名、それに「玩具 歴史」ないし「玩具」を加えたのが2名である。
- 2名は「玩具だけ」に選択した下位項目を加えた。1名は「郷土玩具」を加えて選択し、1名は「郷土玩具」「手まり」「凧」「こけし」を加えている。
- 残り3名は「玩具」全体を一覧した。
- 3名が、別途「人形」もブラウズした。「日本*玩具」の検索で件名はヒットせず、タイトルキーワードでヒットしたものの付与件名として「玩具」「人形」が表示される。3名はこれで「人形」を発見している。
- 階層系表示は利用されていない。
- 適合文献数の平均は11.9であった。
- 学生被験者4名も「玩具」の関連語表示画面から項目選択を行なっている。「郷土玩具」を中心とした選択が1名、「玩具だけ」を中心とした選択が3名である。一人平均4個の項目をチェックしている。
- TT-OPAC(10名)では、適合文献数の平均が13.2であった。「玩具」の語は比較的ゆれが少なく、「たこ」など個別名称も加えれば、一定の再現率が得られる。件名「玩具」だけが付与されたレコードが大変多く、「郷土玩具」だけでは漏れが多くなることから、適合数平均がTT-OPACの方が高くなっている。

関連語表示画面での項目選択は、学生被験者を含めてよく理解されているようで、かなりの人がきめこまかい選択を行なっている。一方で、階層系表

示は利用されていない。

4.2.2 細目を伴う件名へのアクセス

主題の細分 「稲の病害」「自然保護の法律」

〈課題: 稲の病害〉稲作における病害防除に関する図書をさがしたい。

件名「稲 病虫害」に2件のリンク書誌レコードがある。「病虫害」は「生物・農業・畜産」分野の共通細目である。

- BSH-OPAC(9名)における最初の検索語は、「稲」5名、「稲作*病害」「稲*病気」「稲*病害」「稲作*病害防除」各1名であった。
- 全員が「稲」の関連語表示画面に到達し、「稲 病虫害」を選択している。1名がさらに文献を求めて「農薬」等のブラウズを試みている。
- 適合文献数の平均は2.1であった。
- 学生被験者4名も、全員「稲 病虫害」に到達した。
- TT-OPAC(10名)では、適合文献数の平均は2.9であった。適合文献の基準を幅広くとって12件と解答した被験者が平均を押し上げており、6名は1件しか見つけていない。「稲*病害」では1件しかみつからず、あとは「イネ」の検索結果を一覧するしかない。これをやれば、「作物病虫害」など上位の件名が付与されたものも若干見つかリ、BSH-OPACを上回る結果が出る。

〈課題: 自然保護に関する法律〉自然保護、動植物保護に関する法律を扱った図書をさがしたい。

件名「自然保護 法令」に5件のリンク書誌レコードがある。うち2件は、「動物保護 法令」「植物保護 法令」が付与されたもので、「自然保護」の下位に「動物保護」「植物保護」という階層構造から、細目つき件名においても「準下位語」(5.6.4参照)として設定されている。

- BSH-OPAC(9名)における最初の検索語は、「自然保護」5名、「自然保護法」2名、「自然」「法令」各1名であった。
- 8名が最終的に「自然保護 法令」を選択した。1名は「動物保護」「植物保護」を一覧している。
- 適合文献数の平均は4.3であった。
- TT-OPAC(10名)では、適合文献数の平均は7.4であった。「自然*保護*法」では2件しか見つからず、それ以上の発見にはかなり多くの検索を余儀なくされている。なお、適合文献数平均が多いのは、3名が「環境法」方面まで網を広げて適合とみなしたからである。

主標目の関連語表示画面から細目つき件名を一覧して選択することは、問題なく行なわれている。

「環境法」一般が「自然保護の法律」に適合しているかどうかは別にして、本システムの一つの問題点が明らかになったと考える。細目つき件名の場合、「準下位語」の存在で主標目の階層構造を取り込んでいるが、「下位」に限られており、上位語や連想関係にある語については手当していない。仮に²²「自然保護」の上位語や連想語に「環境保全」等があったとしても、「自然保護 法令」の画面には現れないので「環境保全 法令」の存在には恐らく気づかない。

地名と主題「ベトナムの紀行記・旅行記」「アジア映画」「九州地方の経済」
(課題: ベトナムの紀行記・旅行記) ベトナムの紀行記・旅行記の類を探したい。

件名「ベトナム 紀行・案内記」に15件のリンク書誌レコードがある。

- BSH-OPAC(10名)における最初の検索語は、「ベトナム」6名、「紀行」3名、「文学*旅行」1名であった。
- 全員が「ベトナム 紀行・案内記」を選択した。「ベトナム」の関連語表示画面から7名、細目レコード「紀行・案内記」から2名、「東南アジア 紀行・案内記」経由が1名である。
- 2名が、「ベトナムだけ」35件もブラウズした。
- 「ベトナム」の関連語表示画面から、「紀行・案内記」以外に2名が「探検+ベトナム」を、1名が「ベトナム 写真集」を発見している。
- 適合文献数の平均は14.9であった。
- 学生被験者4名も全員「ベトナム 紀行・案内記」に到達している。1名は、「探検+ベトナム」「ベトナム 写真集」も発見している。
- TT-OPAC(9名)では、適合文献数の平均は3.3であった。4名は「ベトナム」でヒットする120件にブラウズを試みている。

「ベトナム」からだけでなく、「紀行・案内記」からのアクセスもあった。また、「探検+ベトナム」という「主題+地名共出現レコード」(5.5.4参照)も発見されることがあるなど、関連語表示機能が十分に使用されている。

(課題: アジア映画) 香港をはじめとするアジア諸国(日本は対象外)の映画に関する図書を探したい。

²²実際には「自然保護」と「環境保全」とは関係づけられてない

件名「映画 アジア」にリンク書誌レコード 31 件がある。「準下位語」として「映画 香港」なども含んでいる。

- BSH-OPAC(10名)における最初の検索語は、「アジア*映画」5名、「映画」5名、であった。
- 全員が「映画 アジア」に到達した。
- 適合文献数の平均は 29.2 であった。
- TT-OPAC(9名)では、適合文献数の平均は 11.7 であった。9名中6名が国名を入力しており、中には7か国を入力した人もいた。

「ベトナム 紀行・案内記」は地名を主題区分し、「映画 アジア」は主題を地理区分している。この次の「九州地方 経済」はまた、地名の主題区分である。本システムでは区別せず扱われるが、被験者も特に意識しているようには見えず、両者の違いに疑問の声もなかった。

〈課題:九州地方の経済〉九州地方の経済・産業を扱った図書をさがしたい。九州全体を扱ってなくてもよい。

件名「九州地方 経済」にリンク書誌レコード 24 件、「九州地方 産業」に 11 件、がある。「準下位語」(5.6.4参照)として、「長崎県 経済」なども含んでいる。

- BSH-OPAC(9名)における最初の検索語は、「九州」2名、「九州地方」1名、「九州*経済」4名、「九州*(経済|産業²³)」1名、「経済」1名、であった。
- 8名が「九州地方 経済」に到達した。1名は最終的に「九州地方」を選択してしまった(数が多く、適合文献はなかなか見つからない)。
- 上の8名中、5名は「九州地方 産業」にも到達したが、3名は「経済」だけにとどまった。
- 適合文献数の平均は 19.6 であった。
- 学生被験者4名も全員「九州地方 経済」に到達した。うち2名は「産業」をも発見している。
- TT-OPAC(10名)では、適合文献数の平均は 9.3 であった。「九州*(経済|産業)」でも8件であり、あとは各県名を入力するなどの労力を要する。

「経済」「産業」に到達すれば、準下位語も含めて効率良く検索できる。

²³経済 or 産業 を示す

ほぼ全員が「経済」までは到達したが、「産業」を発見できたのは約半数であった。最初に「九州*経済」などでダイレクトに「経済」が見つかり、関連語表示をしても「経済」と「産業」は上下関係ではないから「産業」は表示されない。また、「九州地方」の関連語表示を見ても、「産業」を発見できないケースもあった。図8にあげた「フランス」の例のように、地名標目に対する細目展開一覧は、階層構造表単位にまとまって、最上位語のNDC順に配列されるが、NDCの構造上「経済」は330、「産業」は600と大きく離れており、見落としのもとになっていると考えられる。

なお、被験者の意見に、「九州地方」を選んで書誌レコードを表示させると「九州地方」が細目展開されたものだけなのに、その内訳として列挙された「九州地方 経済」を選ぶと「沖縄県 経済」なども含まれている点がわかりにくい、というものがあつた。本システムの仕様では、「沖縄県 経済」は「沖縄県」と「九州地方 経済」の双方に含まれた形となっている。「九州地方」のリンク書誌件数内訳では「下位語」(つまり「沖縄県」)の方にのみカウントしているため、不整合が生じており、確かにわかりにくくなっている。

複雑な例 「日本と東南アジアとの交渉史」

〈課題: 日本と東南アジアとの交渉史〉日本と東南アジア諸国との対外関係史、交渉史に関する図書を探したい。東南アジアの各国に特定したものでもかまわない。

件名「日本 対外関係 東南アジア 歴史」にリンク書誌4件がある。「歴史」まで細目を重ねてない場合もあり、「日本 対外関係 東南アジア」21件中にも歴史を扱ったものがある。

細目「対外関係」と「貿易」では、後に相手先地名をつけてさらに細分する。視点により「日本 対外関係 東南アジア」と「東南アジア 対外関係 日本」は別標目となる。

「日本 対外関係 東南アジア」には「準細目展開」として「東南アジア 対外関係 日本」を、「準下位語」として「日本 対外関係 ベトナム」「ビルマ 対外関係 日本」などを含んでいる(5.6.4参照)。なお、NDC9の「中間見出し」に対応していない(5.6.2参照)ため、「インドネシア(NDC9で-24)」が「東南アジア(同-23)」に含まれなくなっている。

- BSH-OPAC(10名)における最初の検索語は、「東南アジア」2名、「日本」3名、「日本*東南アジア」「歴史*交渉」「日本*東南アジア*歴史」「交渉史」「対外関係史—交渉史」各1名であった。
- 9名が「日本 対外関係 東南アジア 歴史」か「日本 対外関係 歴史」に到達した。経路は「日本 対外関係」から、「東南アジア」か

表 11: 対人実験被験者の主観的評価

BSH-OPACの仕組みが	職員	学生	図書館利用者に提供するとしたら	職員	学生
大体理解できた気がする	6	2	マニュアルがあれば使える	11	3
少し不明の部分があった	9	2	CD-ROM検索のように個別指導が必要	4	1
基本的な流れくらいしかわからなかった	2	0	利用者には理解できないと思う	0	0
ほとんど理解できなかった	2	0			

注: 「図書館利用者に...」の設問は本人が「大体理解できた」「少し不明」と回答した人のみに尋ねた。

ら、直接ヒット、など様々である。残る1名は「日本 対外関係 インドネシア」に到達した。

- 各国別の標目や、「東南アジア 対外関係」などもあわせて一覧している場合もある。
- 適合文献数の平均は13であった。
- TT-OPAC(9名)では、適合文献数の平均は2.6であったが、「東南アジア」のわかち書き(1語ではない)に誤りが多いので、この数字は判断材料とできない。

最初の検索語が様々であっても、なんとか目的の標目に到達できている。

ただ、これだけ複雑になると「準下位語」等の表示もわかりにくくなり、どの範囲のものが含まれるのかを直観的に理解させるには至っていないようである。

4.2.3 被験者の主観的評価

最後に、各被験者に本システムに対する評価と意見(特に、システムの問題点)を求めた(表11、表12)。全体的には、自分自身においても一般利用者を想定しても、ある程度は使いこなせるという評価であったが、様々な問題点もあげられた。指摘された問題点をながめてみると、本システムについてよく理解した上で不備や限界をとらえたものが多いように思われる。

表 12: 対人実験被験者のシステムへの意見 (抜粋)

肯定的評価
<ul style="list-style-type: none"> ● 階層表示などによって件名の方がより多く探すことができる。 ● テーマを体系的にとらえることができ、関連分野の調査に手を伸ばしやすい。 ● (学) 件名のつけ方が適切なら、参考文献の探索が楽になると思う。
画面インターフェースに関する問題点
<ul style="list-style-type: none"> ● ヘルプ機能が足りない。 ● 画面上に表示される用語がわかりにくい。 ● (学) 階層系表示がわかりにくい。線でつなぐなどすべき。
システムの動きなどに関する問題点
<ul style="list-style-type: none"> ● 件名キーワードで検索した結果の相互関連を、関連図等で見せたほうがよい。 ● (学) 最初の検索結果で出たものの相互関連がわからない。 ● 「九州地方 など」を選ぶと「九州地方」に細目がついたものだけなのに、「九州地方 経済」には「沖縄県 経済」も含まれていてわかりにくい。 ● 件名どうしが離れた階層系に位置していると、同時選択する手段がない。 ● 複数の件名を同時に付与されたものを検索したい場合もある。 ● データ量が増えると、ヒット書誌件数が多くなりすぎるのでは。 ● (学) 件数が多過ぎた場合の絞り方に困ることがあった。
データに関する問題点
<ul style="list-style-type: none"> ● その主題の図書が DB 上にあるのに、件名がないものがある。 ● 「Windows」などタイトルキーワードなら見つかるが、件名では多岐にわたり検索しづらい。新しい言葉への対応に問題があるのでは。 ● (学) 「ユダヤ人の歴史」という図書が入っているのに件名「ユダヤ人歴史」で出てこないのはなぜか(注: 「ユダヤ人」が付与されていた)。 ● (学) データ分類(注: 件名の選定のことと思われる)の基準がよくわからない。大きすぎたり細かすぎたりするように思う。 ● 調べたい言葉が件名でヒットせず、他の言葉を探すのに苦労した。 ● 専門用語から件名にたどれる道ができないか?

注: 「(学)」は学生被験者の意見を示し、それ以外は職員によるものである。

4.3 評価のまとめ

限られた範囲の実験ではあるが、「与えられた主題検索質問に対して、タイトルキーワード検索と比べて優位性のある結果が得られるか」「実装した様々な機能が利用者に理解され、使いこなされるか」という2点で、概ね肯定的な評価結果が得られた。

- OPAC ログをもとにした検索実験では、ほぼ6割の検索質問で、BSH-OPACによる検索が、TT-OPACによる検索を上回る性能を見せた(双方ともノーヒットの場合を除く)。しかし、TT-OPACのほうが有効な結果の出るものも4割近くにのぼる。
- BSH-OPACが有効かどうかは、件名典拠ファイル中(つまり普通件名であればBSH中)に検索質問に一致する件名標目があるかどうか、とかなり密接に関連している。特定性の十分な件名がなく、上位概念が付与されていると、件名による検索に大きな障害がある。
- 対人実験では、関連語表示画面の諸機能はほぼ理解され、概ね目指すべき件名標目にたどりついている。
- 被験者の中心は図書館職員であったが、部分的に行なった学生被験者による実験も、ほとんど職員と変わらない検索経過をたどっており、(大学図書館であれば)一般利用者にも理解される可能性が高い。
- 同じ件名標目に到達するにしても、最初の検索語やその後の経過は多様性に富んでいる。関連語表示などで、様々な角度から関連標目を一覧させる仕組みはやはり必要である。
- 適切な件名標目が存在している場合、タイトルキーワード検索から同等の検索結果を引き出すことは困難で、件名による検索に優位性があることが、対人実験の結果からもわかる。

また、主として対人実験の結果から、本システムの機能について、次の点が明らかになった。

- 件名をキーワード検索した段階で、ヒットした件名標目群が階層構造にあるのか、そうでないのか、といった相互関係が明らかにならない点に、検索者の不満がある。
- 件名典拠ファイルにヒットしなかった場合に、タイトルキーワードによる検索を自動的に行なってその付与件名を表示するが、この機能が役立つ場面は極めて多い。もっともこれは、件名典拠ファイルに語彙(標目や参照)が不足しているということでもある。

- 細目つき件名が多くある場合など、関連語表示は長くなるが、比較的注意で見られている。
- 主標目、細目と細目つき件名の関係をはじめ、「主題+地名共出現」「準下位語」「準細目展開」など、本システムで独自に構築した関係構造も検索に寄与している。
- 階層構造は基本的に理解されているが、時に上位項目と下位項目の同時選択など不要な動作も見られる。ことに細目が複数使用されるなど複雑なケースではわかりにくいようで、インターフェースに何らかの改善が必要かもしれない。
- 関連語表示機能が十分に使いこなされているのに対して、階層系表示はあまり使用されない。画面表示のわかりにくさによるのか、別の原因があるのかはよくわからない。

5 件名典拠ファイルの構築

本章では、BSH4mr と NACSIS-CAT データから、検索システムに用いた件名典拠ファイルを構築するまでのプロセスについて述べる。典拠ファイルのデータ内容は検索機能の要件と密接に関連しており、実際の作業は検索システム作成過程で幾度もフィードバックして行なわれている。

5.1 作業の目的、概要、限界

件名標目を正しく典拠管理するには、対象となる書誌集合中に出現する件名標目をすべて含んだ件名典拠ファイルがあり、書誌中のすべての標目と典拠レコードのリンクが設定されているべきである。今回は NACSIS-CAT 中で BSH 標目を付与された書誌レコードを対象とするが、BSH4mr に存在する件名標目だけでなく、標目表においては個々に管理されない細目つき件名標目、地名件名標目、団体名件名標目、個人名件名標目についても典拠レコードを作成して書誌レコード中の標目とのリンクをはからなくてはならない。一方、標目間の関係構造を利用した検索機能を実現するためには、BSH4mr 中に含まれる関係情報を典拠ファイル中に収録するほか、その他の標目についても各種の関係を適切につけてやらねばならない。

作業手順としては、対象となる NACSIS-CAT 書誌レコード 179,893 件から抽出した延べ 247,390 件の付与件名標目を、一般件名、細目つき件名、地名件名、団体名件名、個人名件名といった種類別に分けて、それぞれに独立して処理を行ない、最後に合算した。また、標目間の関係データは、種類別の作業中にも準備作業があったが、最後に検索システムの設計にあわせて必要なデータを生成する処理を行なった。本章での説明も作業手順に従い、種類分け、種類別の処理、関係データの整備と最終統合、という順序で説明する。図 15 に大まかな処理の流れを示した。

最終的な典拠ファイルのレコード件数は、3章のはじめの表 6 に示した通りである。また、典拠ファイルのフィールド一覧を、表 13 に示す。

なお、典拠ファイル完成までの処理には、多くの部分に D-Series コマンドを使用した。D-Series はテキストデータベースを処理するコマンドツールで、射影、選択、結合等のデータベース演算が容易にできる。繰り返しフィールドをもつデータを処理することができるので、書誌データや件名データには極めて有用である。D-Series を使用しない部分は主に Perl5 で処理を行った。

NACSIS-CAT 中の付与件名標目は典拠管理がなされていないので、それ

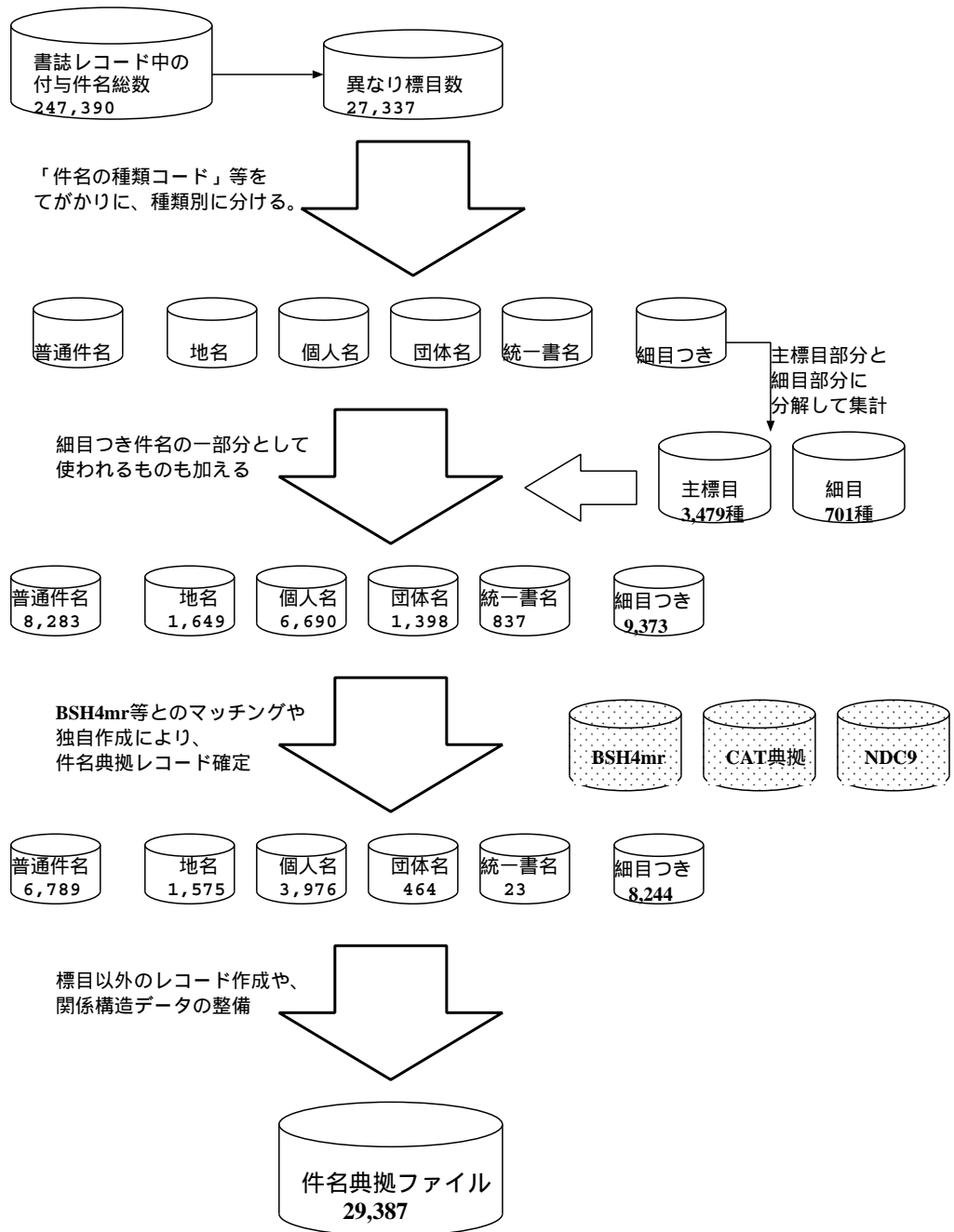


図 15: 件名典拠ファイル構築作業概要

表 13: 件名典拠ファイルフィールド一覧

name	フィールド名	長	繰	存在数	
ID	レコード ID	(12)		29,387	全レコードに存在。主題+地名共出現レコードは 25 桁
TYPE	レコード種別	2		29,387	全レコードに存在。10(細目つきを除く標目) 11(細目つき標目) 20(参照語) ...
TYPE2	固有名種別	1		6,263	a(個人名) c(団体名) g(地名) u(書名) 細目つき件名には持たない
HD	統一標目形	V		29,387	全レコードに存在
PATH	階層パス ID	7	R	7,703	階層パス DB へのリンク ID。複数階層系に属していると繰り返す。
EMATCH	完全一致キーワード	V		24,267	標目形やヨミに正規化を加えて収める。検索の際の一致度判定に使用。
YM	ヨミ	V		11,328	別途キーワード作成したものがあるので全レコードには存在しない。BSH-OPAC には不要。
SNOTE	BSH 注記	V	R	261	BSH4mr 中の注記
SANOTE	BSH 参照的注記	V	R	305	BSH4mr 中の参照的注記
NDC	NDC 分類記号	V	R	7,450	BSH4mr 中の標目レコードに存在。現時点では 8 版
PART	分解要素 ID	V		8,244	全細目つき件名に存在。主標目、細目に分解したそれぞれの ID を連ねたもの。
NDCPL	NDC 地理区分	V		1,614	地名標目に存在。
PLTP1	地名種別	2		1,575	10(国名) 13(外国都市名) 41(山岳名) ...
PLTP2	NDC 一致フラグ	1		501	地名標目が、NDCPL の示す分類項目とレベルが一致することを示す。
NCNO	CAT 典拠 ID	10		4,414	個人名、団体名標目に存在。
SEEFM	から見よ参照	V	R	2,649	CAT 典拠レコードの SF
SEEFM	からも見よ参照	V	R	101	CAT 典拠レコードの SAF
STYPE	細目種別	V		161	細目レコードに存在。一般細目、共通細目等の種別。
これより以下が関係構造に関わるフィールド					
BT	BSH4mr 上位語	V	R	6,896	BSH4mr の関係データは相手先の標目形と ID を両方もつ。
NT	BSH4mr 下位語	V	R	1,839	
RT	BSH4mr 関連語	V	R	492	
USE	直接参照	V	R	2,810	全参照語レコードに存在。
UF	直接参照あり	V	R	2,171	BSH-OPAC では不要。
BTPL	地名上位 ID	V	R	1,606	地名標目に存在。これ以下は独自作成の関係データで ID のみもつ。
NTPL1	地名下位 ID1	12	R	96	下位の地名標目中、NDC 一致フラグがたっているもの。
NTPL2	地名下位 ID2	12	R	176	それ以外の下位地名標目。
SUBGEN	主題展開 ID	12	R	3,314	当該標目に対し、一般細目等によって展開された細目つき件名。
SUBGEO	地名展開 ID	12	R	219	地理区分によって展開された細目つき件名。
SUBPRD	時代展開 ID	12	R	25	時代区分によって展開された細目つき件名。
SUBUS	細目使用 ID	12	R	505	当該標目を細目として使用した細目つき件名。
PTUP	細目上位 ID	12		8,244	細目つき件名に存在。最後の細目を除去してできる標目の ID。
KOBETU	事例関係 ID	12	R	13	事例関係にある個々の固有名件名。
IPPAN	一般件名 ID	12		697	固有名から見て事例関係上位の普通件名。
KYOSH	共出現 ID	25	R	2,412	当該標目を使用した主題+地名共出現レコード。
NT2	準下位語 ID	12	R	1,881	細目つき件名に存在。
SUB2	準細目展開 ID	12	R	386	細目つき件名に存在。

「長」はフィールド長 (V は不定長)。

「繰」が R のものは繰り返しあり。

「存在数」は当該フィールドの存在するレコード数。

なお、BSH-OPAC で使用しないフィールドで、若干省略したものがある。

表 14: CAT 書誌レコード中の BSH 標目

書誌レコード件数	179,893
延べ BSH 標目数	247,390
1 書誌あたり標目付与数	1.38
異なり標目数 (正規化前)	29,742
異なり標目数 (記号類正規化)	27,337
1 標目あたり付与書誌数	9.05

をすべて適切に位置付けるには相当の手作業を要し、研修という性格上不可能である。この際、位置付けえなかったものをそのまま残しておいては、件名典拠ファイルを前面に出した検索システムの評価があいまいになってしまう。そこで、今回は典拠レコードとして確定できなかった標目は最終的に除外した。また、TRC/MARC との関係である程度反映されているとはいえ BSH4mr は刊行前のデータであるから、現在の付与標目との不一致があるのは当然である。したがって、不一致データの中身を深く追及することは行っていない。

5.2 情報源となる付与件名データ

典拠ファイル構築作業の出発点になる情報は、BSH4mr のもつ確定した件名標目データと参照語等 (2.3を参照)、それに CAT 書誌レコード中の BSH 件名標目データである。その他、地名件名の処理に NDC-MRDF9(NDC9 機械可読版 [21]) を、団体名件名と個人名件名の処理に CAT 著名名典拠データを使用しているが、局所的な使用なので各々の箇所で言及する。

表 14に、CAT 書誌レコード中の BSH 件名標目に関する諸元を示す。

CAT 書誌レコード中の件名 (SH) フィールドは、件名表の種類コード、標目形、ヨミ、件名の種別コード、の 4 要素よりなっている。

SH:BSH:人間工学||ニンゲンコウガク//L

件名表の種類コードはすべて”BSH”であるから、残りの 3 要素を抽出して処理を行なう。表 14中の「異なり標目数 (正規化前)」は、標目形で単純に名寄せした結果である。この段階では、”日本 歴史”のような細目形での記号や空白の置き方、あるいはハイフンと長音の違いなどによるゆれが相当あり、記号類や空白をとった形の正規化データで名寄せし直したのが「記号類正規化」である。以下、この 27,337 件が処理の起点となる。

なお、最終的には書誌レコードの付与標目と典拠レコードをリンクするのであるから、いかなる統合処理を行なう場合も、付与されたもとの形との対応をとっておく必要がある。その都度記述はしないが、この先の処理すべてにわたって、もとの形を保管している。

5.3 付与標目データの分類

27,337 件の付与件名を各種類別に分ける作業を図 16 に示した。実際に行なった処理順序とは少し異なるが、やや単純化して示した。

まずは「件名の種類コード」をもとに分類するが、これでは不完全である。コピーカタログに利用されている参照 MARC が細かく分類していない(「個人名以外の件名」というコードがある)し、現在のシステムでは特に利用されていないのでいきおい入力にもあまり注意が払われていない。

入力時に種類コードを記述しないと自動的に K(普通件名)となるので、団体名や地名が普通件名に混じり込んでいる恐れは高いが、その逆はあまり考えられない。そこで、普通件名とされた 19,441 件を中心に、さらに処理を行なう。

- BSH4mr にヒットする標目を除く。BSH4mr には固有名も若干含むので、抽出して振り分ける。
- ”-” を含むもの²⁴を細目つきデータとして別にする(細目の区切りは一樣でないが、先の正規化段階でなるべくそろえてある)。これは、地名など他の種類のものも集めて「細目つき」でまとめる。
- 残り 5,502 件は一覧して手作業で分類する。約半数が普通件名でなく、特に団体名が多い。
- 5,502 件のうち、普通件名は 2,715 件である。これに BSH4mr ヒット分(細目つきを除く)を加えて、普通件名は 8,032 件となる。
- BSH4mr ヒット分のうち 363 件を「細目つき」に繰り入れる。
なお、この時点での 363 件の算出は、人間による多少のチェックを必要とする。なぜなら BSH には「特殊細目」(「経済学 古典学派」のように、細目による合成ではないのに細目つきと同じ形式をとるもの。「古典学派」という細目はないので、全体を主標目とみないと処理できない)というカテゴリーがあるので、単純に形だけでは「細目つき」を抜き出せないためである。

ここまでの作業で、付与件名の種類別分布が明らかになるわけだが、手作業で分類した 5,502 件中には、一見しただけでは判然としないものも多かった(どうしても不明のもので、普通件名としたものが若干ある)。特に統一書名件名(音楽作品や映画等も含む)らしきものに、著者名や限定語がほとんど含まれておらず識別困難の例が目だった。詳細に調査するわけにもいかず仮

²⁴ BSH における細目の区切り記号はダッシュ” ” であるが、本システムの DB 中では処理の便宜上”-” を細目区切りとして使用している。ただし、本文中に例をあげるときは” ” で表記する。

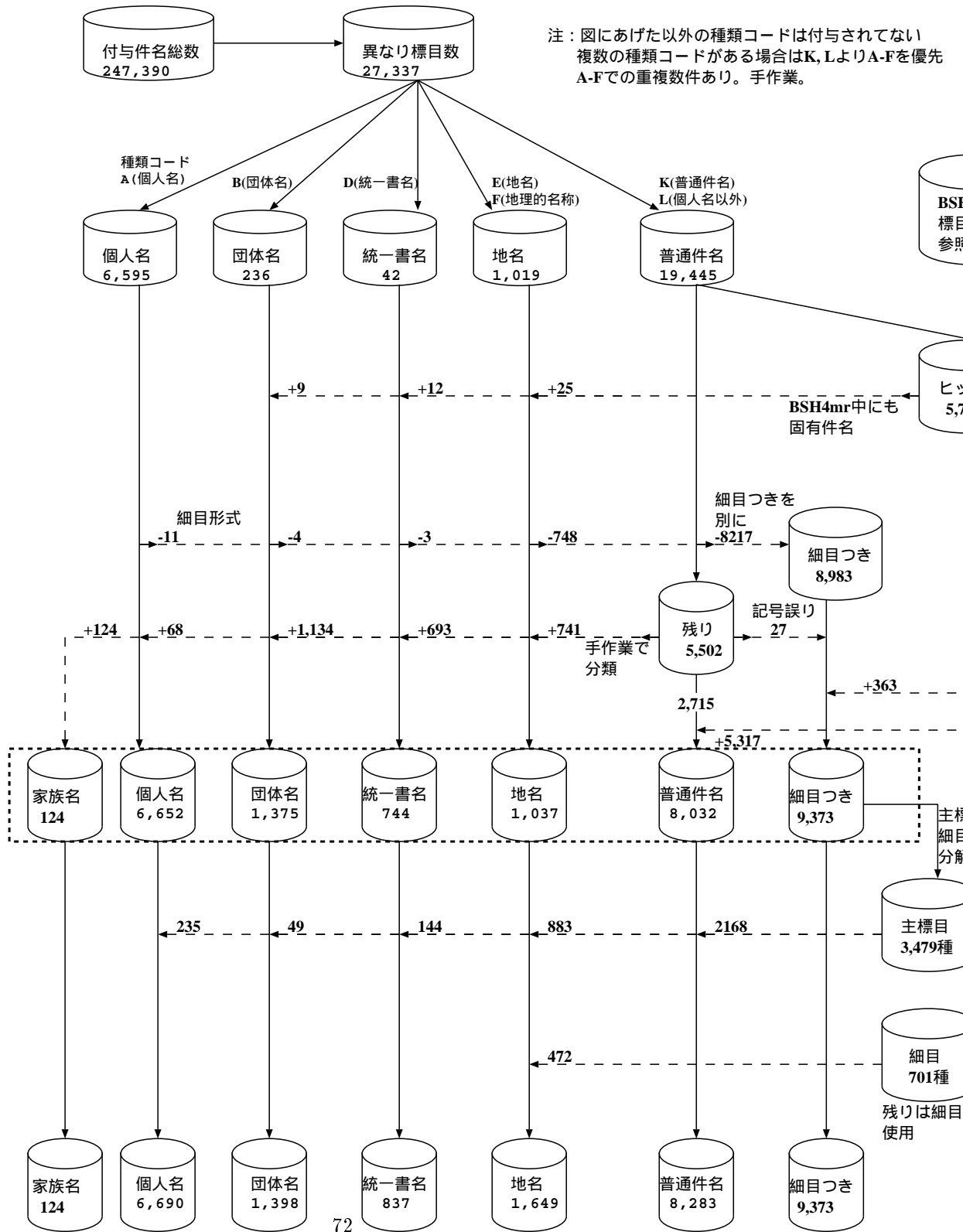


図 16: 付与標目データの分類と統計

表 15: 付与件名の種類別統計

種類	件数	比率	細目つきを 主標目に繰り入れ	比率
普通件名	8,032	29.4%	14,306	52.3%
個人名件名	6,652	24.3%	6,903	25.3%
家族名件名	124	0.5%	124	0.5%
団体名件名	1,375	5.0%	1,430	5.2%
統一書名件名	744	3.9%	918	3.6%
地名件名	1,037	3.8%	3,656	13.4%
固有名合計	(9,932)	(36.3%)	(13,031)	(47.7%)
細目つき以外合計	(17,964)	(65.7%)		
細目つき	9,373	34.3%		
総合計	27,337	100%		

に位置づけたので厳密な数字とはいえない。また、その後の作業で種類に誤りを発見したものもある。

とはいえ、概ねの分布は示していると思われる。種類別に比率を算出すると、表 15 のようになる。表の右半分は、約 3 分の 1 を占める細目つき件名を、主標目の種類別で振り分けた数値である。

なお、地名件名などはその範囲が必ずしも自明といえないが、各種別処理の項で分類の基準にふれる。

ところで、細目つき件名標目への柔軟なアクセスを実現するため、本システムでは細目つき標目レコードと、主標目レコード・細目レコードとの間にリンクをはるという手法をとる。そのため、主標目・細目が(単独では使用実績がなくとも)単独で典拠ファイルに位置付けられていなくてはならない。このため、各種別の典拠レコード作成作業に入る前に以下の処理を行なった。

- 細目つき件名標目を、主標目、細目に分解して集計する。主標目 3,479 種、細目 701 種を得る。
- 主標目を BSH4mr へのマッチングや手作業で分類し、それぞれに振り分ける。
- 細目のうち、地理区分は地名件名に加える。他は一般細目など細目としてのみ使用されるものである(5.4.7参照)。
- 主標目、細目から振り分けたものの大半は、単独標目としても存在しているので、重複は除く。

結果、図 16 の最下段が典拠レコードとして作成すべき種類別件数となる。

5.4 各種件名標目レコードの作成

本節では、各種類別の標目レコード作成を、普通件名、固有名件名、細目つき件名の順で行なう。普通件名標目は件名標目表の管理するところなので、BSH4mr とのマッチング作業が主である。固有名件名標目は BSH4mr で管理されていないため、どうやって統一標目形を決定するかという問題が生じる。細目つき件名標目は、主標目・細目に分解してそれぞれの典拠レコードと関連づける。

5.4.1 普通件名標目レコード

細目つき件名を除く普通件名は、BSH4mr にヒットするもの(参照形へのヒットを含む)が 5,317 件、ヒットしないものが 2,715 件である。他に、細目つき件名の主標目としてのみ使用されるものが 251 件あり、合計 8,283 件が普通件名標目の候補である。

ヒット分のうち、286 件は BSH4mr の参照語レコードにヒットしている。これらは参照先標目形に置き換えて確定する(参照先標目が既であれば統合する)。書誌とのリンクも参照先 ID とするので、最終的にははじめから正しい標目形が付与されていたかのように検索される。なお、BSH には「青少年 児童 青年」のように複数標目への参照が存在する。このような参照へのヒットは、行き先が自動的に決定できないので、確定できない(17 件)。

ヒットしないものの一部を調査したところ、文字種や限定語など微妙なところで BSH4mr 標目と食い違っているもの、動植物や言語などの個別的名称(BSH3 版で省略件名、例示件名扱いとしていた種類のもの)が付与されているが BSH4mr に未採録のもの、などが多い。これらを精査して適切に位置づけることは、時間的制約の中ではできなかった。限定語を無視するなどすれば、もう少しノーヒット分を機械的に取り込めるが、検索精度に影響を与える可能性もあるので行なっていない。

今回普通件名として実装したのは、結局 BSH4mr に存在するデータだけである。8,283 件のうち、参照へのヒットも含めて確定できたのは 5,499 件(65.8%)である。データ内容も、この段階で BSH4mr レコードに付加した項目はない。

なお、BSH4mr 中の普通件名標目 6,789 件の中には、今回の書誌集合内に全く付与実績のないものが約 1,600 件あるが、典拠ファイルにはこれらも収録した。本システムでは、上位語をはじめとする関係構造から関連標目をたどれるので、それ自体に付与実績がなくとも、探索に有益だからである。

5.4.2 個人名件名標目レコード

個人名件名標目レコードの候補は 6,690 件であるが、生没年などの付記事項やイニシャル、字体等のゆれで、同一人が分かれてしまっていることも多く、実際にはいくらか下回る数になる。

個人名件名標目確定のツールとして、NACSIS-CAT の著者名典拠ファイルデータ (1997.8.22 現在 952,678 件) を使用した。個人名や団体名を件名標目とするときには、著者標目としての形と一致させ、典拠管理も一括して行なうのが、一般的な方式だからである。6,690 件のうち、付記事項を含めて著者名典拠ファイルの統一標目形と一致したものが 1,191 件、付記事項を含めて「から見よ参照」(SF) に一致したものが 418 件、付記事項をとった形で統一標目形に 1 件のみヒットする (つまり著者名典拠ファイル中には同姓同名がない) ものが 2,590 件となった。4,199 件 (62.8%) を確定したことになるが、最初は付記事項の有無で分かれていたものが、結局まとまるケースがある。最終的には 3,976 件の典拠レコードができた (BSH4mr 中にあった 9 件も加える)。

データ内容としては統一標目形のほかに、「から見よ参照 (SF)」と「からも見よ参照 (SF)」を収録している。なお、SF へのヒットなどの場合、統一標目形は、著者名典拠レコードの統一標目形で置き換えた。

ここまでで確定できなかった 2,491 件については、典拠レコードを作成していない。

5.4.3 家族名件名標目レコード

「件名の種類コード」による分類で普通件名とされた中に「ロスチャイルド家」「今川氏」「三井家」といった家族名件名標目が 124 件あった (図 16 参照)。

典拠レコードを確定する方策を発見できず、本システムでは扱えなかった。

5.4.4 団体名件名標目レコード

団体名件名標目レコードの候補は 1,398 件であるが、図 16 でもわかるように、最初から「件名の種類コード」で分類できたものは少なく、大半は普通件名の BSH4mr ノーヒット分を手作業で分類したものである。そこでまず、今回とった分類基準について、注意すべきところをあげておく。

- 国名、都道府県、市町村などの名称は、地名でもあり、またその法域の統治機関に対する団体著者標目としても使われる。本研究では、団体名には入れず、地名に分類した。地域名など他の地理的名称とあわせて構造化するためである。

- その他、地名に分類したものに、江戸時代の藩名や港湾・駅・城などの施設名、建造物名がある。
- 社寺名もどちらともとれるが、「件名の種類コード」がB(団体名)になっているものもかなりあり、団体名に統一した。

典拠レコード作成作業は、5.4.2の個人名と同じく NACSIS-CAT 著者名典拠ファイルを使用して行なった。著者名典拠ファイルの統一標目形にヒットしたものが374件、「から見よ参照」にヒットしたものが86件で、計460件(32.9%)を確定したことになる。BSH4mr 中にある団体名件名17件と合算して重複をまとめると、464件の典拠レコードとなった。

ところで、NACSIS-CATの著者名典拠レコードは、日本名はNCR1987年版、外国名はAACR2に従い作成することになっている[6]。そのため個人名や団体名を著者標目と一致させた形で件名標目としようとするれば、外国人名、外国団体名は原則的に原綴形となる。ところが本研究の範囲とした限りのCAT書誌レコード付与標目では、個人名は原綴形が多数派であるが、外国団体名はほとんど日本語形となっていた。ヒット率の低い一因にもなっている。「から見よ参照」の日本語形にヒットしたものは、統一標目形である原綴形で置き換えられたことになる。なお、BSH4mr中には「国際連合」「国際通貨基金」など日本語形の外国団体名が若干存在しており、これらはBSH標目形を優先した。不均衡は否めず、件名標目表側でも付与する側でも、扱いを検討する余地がある。

5.4.5 統一書名件名標目レコード

統一書名件名標目レコードの候補は、映画や音楽の作品名も含めて847件である。

本システムでは統一書名件名の典拠レコードは作成できなかった。NACSIS-CATには統一書名典拠ファイルがあるが、当面の運用が無著者名古典、聖典、音楽作品に限られているので、それほどヒットすると思えなかったためである。

ただし、BSH4mr中には「聖書」など23件の書名件名を含んでいるので、これらは統一書名件名標目の典拠レコードとして納めた。

5.4.6 地名件名標目レコード

地名は、単独でも件名標目となりうるが、「経済政策 中国」のように細目(地理区分)として、また「中国 農業」のように細目(主題区分、一般細目)をつけた形で使用されることが多い。表15をみると、細目形式のほうがずっと多いことがわかる。このように普通件名と併用されることが多く、ま

たそれ自体階層構造的性をもっている、といった点から、地名件名は他の固有
名件名とは異なった性格をもっている。

細目使用も含めて 1,649 件が、地名件名標目レコードの候補である。5.4.4
で述べたように、国名、都道府県、市町村などの名称を含んだ数字である。

件名標目間の関係構造を利用した検索システムを考えたとき、地名件名が
適切に位置付けられているかどうかは性能に大きく影響するが、他の固有な
と同様 BSH4mr ではほとんど管理されていない²⁵。また NACSIS-CAT 著者
名典拠レコードを使用しても一部の種類しかヒットしないし、相互の関係構
造をつける助けにはならない。

そこでまず、1,649 件のレコードに対して次の 2 処理を行なう。

- 地名件名を種類別に分類してコードをふる。あまりにも多様な性格の
ものが混在しているのを整理し、また普通件名と固有名間の関係構造
をつける材料とする。
- 『日本十進分類法新訂 9 版』([20] 以下、NDC9) の分類記号 (地理区
分) を与える。NDC9 に準拠して地名件名の階層構造を作成する材料と
する。

地名件名の分類方式に標準的なものがあるのか、といった調査は行なえな
かったので、分類項目は任意に設定した。コード付与は「市」「村」「遺跡」
「川」といった末尾の文字を手がかりにある程度自動的に行ない、残りは手
作業である。

種類別の件数等は表 16 の通りとなる (ただし、最初の 1,649 件ではなく、
最終的な 1,575 レコードでの数字)。地名といっても実に多様なものがあるこ
とがわかる。国名・都市名などの行政地名は全体の 3 分の 2 ほどにすぎず、
自然地名や歴史的な地名も無視しえない数がある。

次に NDC9 分類記号の付与である。NDC9 では補助表として、485 項目よ
りなる「地理区分」と、40 項目よりなる「海洋区分」をもっており、それぞ
れ階層化されているので、分類記号を付与できれば、NDC9 準拠の地名件名
階層系を構築できる。記号付与作業には、NDC9 の機械可読版である NDC・
MRDF9²⁶ ([21] 以下、MRDF9) を利用した。手順は次の通りである。

- MRDF9 の関連索引ファイルから、行き先記号が地理区分、海洋区分の
ものを抜き出す (1,577 件)。
- 地名件名と関連索引をマッチングして、完全一致するものに記号付与
(257 件)。

²⁵ 地名も含むが数が少ない。また、BSH には標目表本体とは別に「国名標目表」を収録して
いるが、標目表外のデータなのでまだ DB 化していない

²⁶ これも、日本図書館協会より研究使用の許可をいただいた

表 16: 地名件名の種類別統計

種類	件数	具体例
国名	143	日本、ロシア、南アフリカ (国名)
旧国名	6	ドイツ (西)、ソビエト連邦、チェコスロヴァキア
州、省など	34	山東省、カンザス州、スコットランド、チェチェン、内モンゴル自治区
外国都市名	104	ソウル、上海、ミュンヘン、アトランタ
都道府県	47	北海道、東京都、兵庫県
市町村、郡、特別区	650	神戸市、池田町 (北海道)、小川村 (長野県上水内郡)、郡上郡 (岐阜県)、東京都杉並区
国より大きな地域名	37	アジア、オセアニア、ヨーロッパ (東部)、バルカン諸国、アマゾン、イスラム圏、北極地方
一国内の地域名	59	中国 (南部)、シベリア、プロバンス、東チモール、東北地方、南九州地方、武蔵野地方、伊豆地方
都市内の地名	25	浅草、祇園、山谷 (東京都)、大阪市平野区
山岳、山脈	71	大雪山、朝日岳 (山形県)、日本アルプス、鈴鹿山脈、エベレスト、カラコルム山脈
河川	54	利根川、鴨川 (京都府)、黄河、メコン川
島	89	礼文島、伊豆七島、竹島 (島根県)、トカラ列島、スマトラ、ガラパゴス諸島、グリーンランド
湖沼	13	琵琶湖、雨竜沼、バイカル湖
海洋、内海、湾	18	太平洋、カリブ海、瀬戸内海、津軽海峡、東京湾、有明海
その他の自然地名	32	知床半島、草津温泉、黒部峡谷、天神崎、和白干潟、ゴビ砂漠
歴史的遺跡	24	吉野ヶ里遺跡、高松塚古墳、陶器古窯址群、万里の長城、アンコ - ル・ワット、莫高窟
藩名	36	仙台藩、土佐藩、名古屋藩
歴史上の国名、都市名	22	エジプト (古代)、高句麗、江戸、長安、平安京、ボンベイ
その他の歴史的地名	20	東海道、奥州道中、シルクロード、三河国、オリエント
鉄道、運河、港湾、駅、空港	20	高山本線、パナマ運河、小樽運河、横浜港、新宿駅、成田空港
鉱山	14	土呂久鉱山、北炭夕張新鉱、院内銀山
各種の建築物	40	江戸城、ル - プル王宮、エッフェル塔、自由の女神、朝香宮邸、ニコライ堂、瀬戸大橋
その他人工的施設	10	新宿御苑、セントラル・パーク、千里ニュー・タウン、東海自然歩道、渡良瀬遊水池
不明など	7	
総合計	1,575	

- さらに最後の「市」をとってマッチするもの(地名件名はほとんど市となっているが、相関索引は「市」のない形となっている)や、「町(県)」のように都道府県を確定できるものに記号付与(607件)。
- 残り約半数は、手作業で分類する。ただし、作業の負荷もあり、歴史的地名と建造物等は付与できなかった。最終的に付与したのは、行政地名、地域名、自然地名、あわせて1,389件である。
- この時点では複数の地名件名に同一の分類記号がつくことが多い。例えば日本の場合、NDCでは一部を除いて都道府県までしか記号が展開されていないので、「137」(神奈川県)という記号は、件名「神奈川県」の他に「横浜市」「三浦半島」等にも付与される。このとき「神奈川県」と「横浜市」以下を区別しておかないと、正しい階層関係が表現できない。そこで、「神奈川県」のように分類記号の指し示す範囲と一致するレコードには「NDC一致フラグ」をたてた。全体の17.5%にあたる276件にフラグがたっている。

これらの作業の過程で表現の僅かな違いによる重複などもみつきり、最終的には地名件名標目レコード1,575件が確定された。

なお、BSHが「国名標目表」を規定していることもあり、外国地名もすべて日本語形で作成したが、NACSIS-CATにおいては(行政地名の場合)、団体名(5.4.4)と同様にAACR2形の著者名典拠レコードが作成され、両者に食い違いが生じる。本研究では著者名典拠は扱わないが、典拠管理システム全体を考えたときには検討の要がある。

5.4.7 細目つき件名標目レコード

細目つき件名標目レコードの候補は9,343件であるが、5.3で述べたように、主標目の種類にかかわらず、細目形式のものはすべて含んでいる。

本システムでは細目つき標目レコードと、主標目レコード・細目レコードの間にリンクをはるという手法をとる。そのため、主標目・細目レコードがともに確定されているもののみを、細目つき件名典拠レコードとして作成した。処理の手順を図17に示す。

まず、細目レコードについてふれておく。BSH4版における細目には、次の6種がある。

一般細目 基本的にすべての標目のもとでも使えるもの。「歴史」「辞典」「研究法」など。約25種。

分野ごとの共通細目 分野を限定して使用できるもの。例えば、「美術・文化財共通細目」として「材料」「鑑定」などが、「古典名共通細目」として「諸本・諸版」「研究」などがある。約65種。

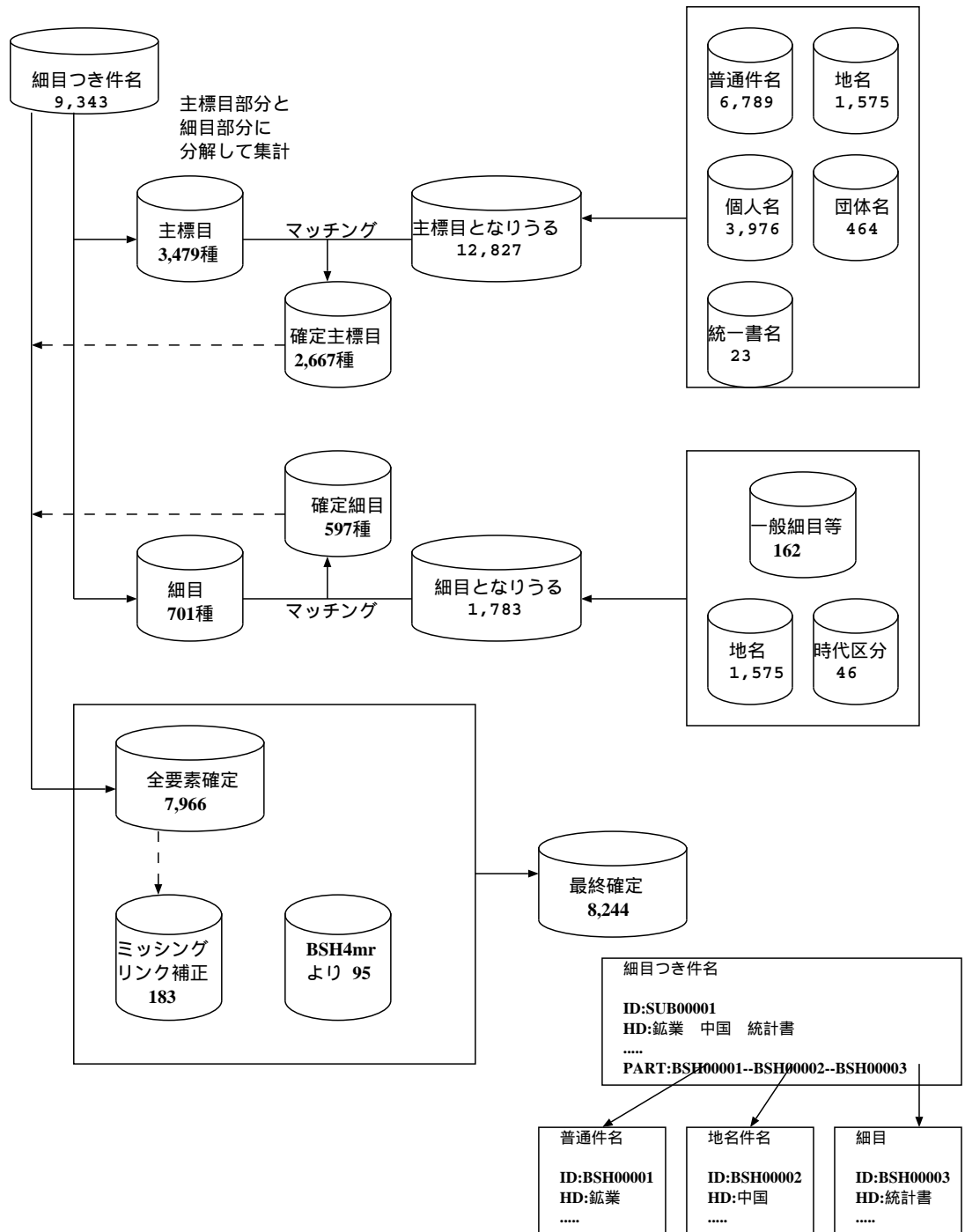


図 17: 細目つき件名標目レコード確定作業

言語細目 各言語名のもとで使えるもの。「会話」「語源」「外来語」など。
約 55 種。

地名のもとでの主題区分 地名を主標目として使用する主題細目。特に地名が重要視される主題について、地名を先行させて集中するためにある。「対外関係」「農業」など。約 20 種

地理区分 「地名のもとでの主題区分」に該当するもの以外は、すべての主題標目で地名を細目として使用できる。細目レコードとしては管理されていない。

時代区分 歴史を表す標目のもとで時代を限定するために使用する。細目レコードとしては管理されていない。

この他に「特殊細目」があるが、これは表中にあげた特定の標目のみに使用されるものなので、「細目つき」の範疇に含めず全体を主標目として扱う。

このうち、「地名のもとでの主題区分」までの 4 種は、BSH4mr 上で「細目レコード」として管理されている。ただ、本研究では「歴史」「農業」のように同じ形の普通件名標目が存在するときには、統合して普通件名標目のレコード ID を使用することとした。普通件名としての「農業」から「アフリカ 農業」等へリンクをはりやすくするためである。残念ながら普通件名では形がことなるもの(例えば細目として「対外関係」があるが、普通件名として対応するものは「外交」「国際政治」等である)は、統合するわけにもいかず、関係づけができなかった。

次に、地理区分は BSH4mr では管理していないが、地名件名標目 1,575 件がそのまま使える。

最後に時代区分であるが、従来から BSH ではこれを独立の項目として標目表にあげてこなかった。「日本 歴史」などのページをひらけば「日本 歴史 江戸時代」等が並ぶので、これを参照して「日本 経済 歴史 江戸時代」といった標目を作成していたわけである。機械処理のためには独立して作成しておく必要があるので、BSH4mr 中で使用されている細目データを集計して、そこから一般細目等をのぞき、残った中から手作業で特殊細目を除去して、46 件の時代細目レコードを作成した。

以上により、細目となりうるレコード 1,783 件(大半は地名であるが)の集合ができた。一方主標目となりうるのは、5.4.6 までに確定してきた全標目レコード 12,827 件である。これらと、細目つき件名を分解して得た主標目データ 3,479 件、細目データ 701 件をマッチングすることにより、主標目として 2,667 件、細目として 597 件を確定することができた。

なお、図 17 では省略したが、主標目部分については、普通件名で行なったように、別途 BSH4mr の参照語レコードとのマッチングを行なった。60 件

ほどが参照語にヒットしている。また、細目に対する参照語レコードというものはないが、BSH3 版からの改訂作業で一般細目「統計」を「統計書」と改めるなど若干の修正が行なわれているので、3 版での形式のものもヒットするように調整した。上記の確定数にはこれらも含まれている。

ここでもとの細目つき件名に戻り、主標目と細目の全要素が確定できているものを抜き出す。85.3%にあたる 7,966 件で全要素が確定できており、5.4.6 までと比較すると大変確定率が高いといえる。

確定したレコードのデータ内容に関しては次の 2 点の処理を行なった。

- 主標目レコード、細目レコードへのリンクデータとして「PART:BSH300055300-BSH300669900-PRD000000013」のように各要素の ID を連ねた「分解要素 ID」フィールドを持たせる。
- 標目形を、各要素の標目形を合成したものに置き換えた。例えば、主標目が BSH4mr では参照語になっていた場合には、参照先の正しい標目に置き換えた。

ところで、本システムでは「和歌山県」の下位に「和歌山県 農業」、そのまた下位に「和歌山県 農業 統計書」、というように表示する仕様としている。そのため、「和歌山県 農業 統計書」があるときには、ミッシングリンクが生じないように付与実績がなくとも「和歌山県 農業」のレコードを作成しておく必要がある。このような例を調査したところ 183 件みづかり、機械的にレコード作成を行なった。確定したレコードに、このミッシングリンク補正用レコード、さらに BSH4mr 中にあるが付与実績のない細目つき件名をも加えて、最終的には 8,244 件の細目つき件名標目レコードが確定された。

なお、以上の処理においては、主標目や細目があらかじめ確定されているかどうかを調べているわけであるが、その組合せ規則等はチェックできていない。実際に、「農業」が「地名のもとの主題区分」であるにもかかわらず、「農業 日本」という逆転した形のものも存在しており、確定されてしまった。極端に言えば「日本 フランス」でも確定されてしまう。典拠管理という側面からは、より細かなチェックロジックを確立する必要がある。

5.5 典拠ファイル中のその他のレコード

以上で各種の件名標目レコード作成が終了したわけであるが、検索システム上の必要から、典拠ファイルには標目レコード以外にもいくつかの種類のデータを収録した。

5.5.1 細目レコード

細目レコードについては、細目つき件名とともに 5.4.7 で説明した。検索上も例えば細目「対外関係」のついたものを一覧するためには上位に「対外関係」というレコードがあったほうがよいので、一般細目等のうち普通件名としては使われない形のもの 115 件と時代区分細目 46 件をおさめた。

5.5.2 参照語レコード

BSH4mr 中にある参照語 2,810 件をおさめた。個人名件名標目レコード等では「から見よ参照」を標目レコード中に残しており、参照語レコードは作成していない。

5.5.3 補助地名レコード

5.4.6 で地名件名標目への NDC9 付与について述べたが、これをもとに階層構造を作ろうとすると、ちょうど細目つき件名でも生じたようなミッシングリンクが発生することがある。これに加えてアクセスポイントを増やそうという目的もあり、記号の指し示す範囲と一致する件名レコード（「NDC 一致フラグ」がたっているもの）のない分類記号について、分類項目見出しを使用した「補助地名レコード」225 件を作成した。「道西：留萌支庁 [天塩国]」のように件名標目としての典拠形とはいいがたい形が多いので、地名件名標目レコードとしては数えない。

5.5.4 主題+地名共出現レコード

これは、検索システムを設計する過程で、これまでの BSH の規定によっておこる不都合を多少とも解消するために作成したものである。

BSH4 版ではすべての標目を地理区分することが可能になるが、現行 3 版では地理区分の可能な標目は表中で指示されたものに限られていた。そのため現行 3 版を忠実に守ったデータでは「北海道の総合開発」という主題に対して「総合開発」と「北海道」かもしくは「総合開発」と「北海道-経済」の 2 件名を与えている。これでは、「北海道 総合開発」からは件名データをヒットさせることができないし、「北海道」という件名の関連語に「北海道 経済」は提示できるのに「北海道 総合開発」は示せない。関連標目を幅広く提示するという検索形態では悪影響が大きいので、複数の件名標目を付与された書誌レコードを調査し、地名件名と普通件名の共出現を調べて「総合開発+北海道」という形式の「共出現レコード」を作成した。

なお、ここで筆者は細目形式の事前結合方式が、事後結合方式より優れていると主張しているわけではない。主題と地名が分けて付与されているという前提なら、また違った検索システムで関連主題の提示ができるであろう。両

表 17: BSH4mr 中の階層関係

標目総数	7,450
BT をもつもの	6,896
NT をもつもの	1,839
BT, NT 総数	各 7,909
BT を持ち NT を持たないもの (最上位語)	314
BT, NT ともに持たないもの (孤立語)	238
	普通件名に限ると 6

方のやり方がデータベース上で混在しているのでは設計に支障があるといっているのである。その点で、すべての標目に地理区分を許すという BSH4 版の改訂は、事前・事後の優劣とは別に、大変意義のあることと考えている。

5.6 標目間の関係構造データ

典拠ファイル中のレコード作成について述べてきたが、本研究の目標は標目間の関係構造を生かした検索システムであるから、以下に関係構造データの作成について説明する。もちろん、一つ一つの標目を吟味することは不可能なので、BSH4mr 中の情報やここまで作成したデータをもとにどれだけ自動処理できるか、という点が中心となる。

5.6.1 BSH4mr の階層構造

2.3 で述べたように、BSH4 版では 3 版までの連結参照 (をも見よ参照) が階層関係 (BT, NT)、連想関係 (RT) に整理される。中でも階層関係が重要であり、改訂作業においても特に時間をかけて整備されている。

BSH4mr 中の標目レコードには、BT(上位語)、NT(下位語) フィールドが設けられ、それぞれ相手先のレコード ID と標目形が埋め込まれている。常に機械的チェックをかけているので、BT, NT の相互対応は完全に保証されている。

ID(レコード ID): BSH300020600
 HD(標目形): 遺跡
 BT(上位語): 考古学 BSH300204100
 NT(下位語): 貝塚 BSH300068100
 NT(下位語): 経塚 BSH300138000
 NT(下位語): 巨石記念物 BS8900043300
 (以下、省略)

NT はもちろんのこと、BT も 1 レコードに複数存在することがある。例えば、「植物生態学」は「植物学」と「生態学」を BT としている。

図 17 に BSH4mr 中の階層関係に関する諸元をまとめた。

標目には若干の固有名称や細目つき件名も含まれるが、普通件名に限ると「自然」「論文集」など孤立語はごくわずかである。

この BT, NT はすべて典拠ファイル中の標目レコードに持たせ、検索システムで提示することとした。また、各標目間の BT, NT 関係から「階層構造表」を導くことができる。

BSH300204100
BSH300204100/BSH300020600
BSH300204100/BSH300020600/BSH300068100
BSH300204100/BSH300020600/BSH300138000
BSH300204100/BSH300020600/BS8900043300
BSH300204100/BSH300020600/BSH300238500
BSH300204100/BSH300020600/BSH300238500/BS8900113900
BSH300204100/BSH300020600/BSH300245600
BSH300204100/BSH300020600/BS8900136800
BSH300204100/BSH300020600/TSU300000192
BSH300204100/BSH300020600/TSU300000556
BSH300204100/BSH300075900
BSH300204100/BSH300182700
(以下、省略)

ID を標目形に置き換えると、次のようになる。

考古学
考古学/遺跡
考古学/遺跡/貝塚
考古学/遺跡/経塚
考古学/遺跡/巨石記念物
考古学/遺跡/古墳
考古学/遺跡/古墳/前方後円墳
考古学/遺跡/祭祀遺跡
考古学/遺跡/洞穴遺跡
考古学/遺跡/住居址
考古学/遺跡/チャシ
考古学/鏡
考古学/原始社会
(以下、省略)

この階層構造表から「階層パス DB」(3.3.2参照)ができ、BSH-OPAC の「階層系一覧機能」で提供されている。

なお、ここで、BT, NT 関係の設定基準について述べておく。Aitchison らの『シソーラス構築法』[1]では ISO2788 に基づき、階層関係として「属関係」「階層的部分全体関係」「事例関係」をあげている。「階層的部分全体関係」には「身体の組織と器管」「地理的位置」「学問分野あるいは研究領域」「階層的社會構造」があげられ、その他の部分全体関係は連想関係 (RT) とみなされる。IFLA「件名典拠ファイルのためのガイドライン」[8]や LC "Subject Cataloging Manual"[16]でも部分全体関係を階層関係とする範囲はともかく、属関係(「類種関係」とも)、部分全体関係、事例関係の3つを階層関係とする点では変わらない。ところが、BSH4 版の階層関係はこれと大きく異なり、本来連想関係となるものを多く含んでいる(先にあげた例では、

「考古学」と「遺跡」「原始社会」など)。必然的に、重合階層関係(複数の上位語をもつ関係)も多くなっている。厳密な基準を適用しなせば、恐らくBT, NTは半数くらいになるのではないか。こうした状態の原因には、一つには従来の「をも見よ参照」を踏襲している部分が多いこと、もう一つにはなるべく階層構造表から関連主題を一覧できるように配慮したことがある。本研究で階層関係を画面に表示してみると、様々な関係のものが交じりあって表示され、いささかわかりにくいことがあり、整理の余地があるだろう。

しかしながら、図書館目録の主題検索機能は必ずしも専門家によって使われるわけではないから、専門を特化したシソーラスに比べて、幅広く関連する主題を提示する仕組みが大変重要である。したがって、階層関係の定義を厳密にするだけでは、有効な検索システムを構築しえないと思われる。代わりに連想関係(RT)を大幅に増やすことにもなるだろうが、現状での連想関係は1対1の関係を表すだけで構造化は困難であるから、あまり多くなると始末におえない。件名標目表の構造自体に改善の余地があるのか、件名ですべてをカバーすることをやめて分類による検索との組合せなどを考えた方がよいのか、十分に検討されるべきである。

5.6.2 地名件名間の階層構造

5.4.6で、地名件名標目レコード作成にあたって、階層化のためにNDC9分類記号を付与したことを述べた(行政地名、地域名、自然地名、計1,389件)。また、5.5.3では、分類項目見出しから「補助地名レコード」(225件)を作成したことを述べた。この両データとNDC・MRDF9ファイル[21]から、地名件名の階層づけを行なう。

まず、NDC・MRDFから、NDC9記号(地理区分、海洋区分)に上位記号、下位記号を付したデータを作成する。NDCは十進記号であるから普通は数字の桁を操作すれば上位、下位がわかるが、時に記号の階層性が崩れる場合がある(印刷体ではインデクションで示す)。このときMRDF9のレコード上では「縮約記号」(-171山陰地方-172鳥取県の場合の-172)、「不均衡記号」(-348チェコ-3483スロバキアの場合の-3483)というコードが付されているので、これを考慮に入れて正しい階層関係をつける。なお、NDC9より導入された「中間見出し」(“-345/-349中欧諸国”)のように一定の範囲を示して階層構造を明示する方式)には対処できなかったもので、完全な階層でないところがある。

次に、分類記号の指し示す範囲と一致する件名標目レコード(「NDC一致フラグ」がたっているものと、「補助地名レコード」)を抜き出して、NDCに沿って上位語・下位語をつける。NDC9記号ごとの上位記号・下位記号が作成済なので簡単である。

最後に、「NDC一致フラグ」のたっていないレコードを処理する。これ

は同一 NDC9 記号をもつ、一致フラグのたったレコードを上位語とすればよい。

これにより、地名件名間の階層関係が完成した。上位・下位がはっきりすれば、BSH4mr 中の BT, NT と同様に階層構造表を作成できる。

日本
日本/北海道
日本/北海道/道北 (中略)
日本/東北地方
日本/東北地方/青森県
日本/東北地方/岩手県
日本/東北地方/宮城県
日本/東北地方/秋田県
日本/東北地方/山形県
日本/東北地方/福島県
日本/関東地方
(以下、省略)

地理区分においては「-1 日本」のように 1 桁の記号には上位記号が設定されないから、これらが最上位語となる。すなわち「日本」「アジア」「ヨーロッパ」「アフリカ」「北アメリカ」「南アメリカ」「オセアニア」の 7 系ができる。海洋区分は別に「-1 太平洋」から「-8 南極海」までが設定されているが、これは 8 系に分かれる必要はないので、「海洋(地名)」というレコードを別に作成して、その下の系にまとめる。

以上はあくまで NDC9 に準拠した階層構造であり、例えば日本国内はほとんど府県レベルまでの展開なので、その下に多くの地名がランダムに並んでしまうことになるが、仕方がない。また、「-36 スペイン」には「南欧には、この記号を使用する」という注記があるが、この場合「ヨーロッパ(南部)」に「-36」を与えると、上記ロジックから「スペイン」の下位語とみなされてしまう(NDC 一致フラグは「スペイン」にたつため)。こうした逆転現象は 10 箇所近くでおこるが、特に対処はできなかった。

5.6.3 固有名と普通件名との事例関係

「事例関係」とは「普通名詞で表現される事物と事象の一般的なカテゴリーと、そのカテゴリーにおける個々の事例間の関係」²⁷であり、一般的に属関係(類種関係)、部分全体関係とともに階層関係とみなされる。例えば「山岳」と「大雪山」「六甲山」といった個々の山名との関係である。

BSH3 版では(2.3参照)、さまざまな種類の個別名称のグループが「例示件名」「省略件名」とされ、一部の代表的なものだけを標目表にあげていた。この場合、連結参照として、「言語: 個々の言語名(例:日本語)」といったものがつけられる。個々の図書館では、表中にない言語を扱った図書があれ

²⁷[1] pp. 48

ば新たに件名標目とするが、利用者への参照は「： 個々の言語名をも見なさい」でとどめるのが一般的運用である。

BSH4 版では、言語名、病気名、動植物名等の個別名称について、文献的根拠のあるものはなるべく標目表中に掲げ(十分には採録しきれてはいないが、それぞれかなりの数にのぼる)、一般名称との間に階層関係を設定することとした。よって、「言語」の下位語として、採録した言語名が列挙されることになる。

しかし、個人名、団体名、書名、地名については、BSH4 版でも採録数は少なく、また一部をのぞいて階層構造に組み込んではいない。したがって、普通件名から個々の固有名件名への参照指示は、以下に例を示す「参照的注記」によってなされることになる(従来の連結参照は BT, NT, RT 構造になったが、特定の一標目レコードを指し示さない参照は NT とはできないため、「参照的注記」(仮称)とする)。

標目: 彫刻家
参照的注記: 個々の彫刻家名(例:ロダン)も標目となる。

標目: 地方議会
参照的注記: 個々の地方議会名(例:神戸市議会)も標目となる。

標目: 説話文学
参照的注記: 個々の説話文学作品名(例:今昔物語)も標目となる。

標目: 遺跡
参照的注記: 個々の遺跡名(例:登呂遺跡)も標目となる。

BSH4mr 中では、個人名約 45 件、団体名約 15 件、書名約 10 件、地名約 20 件を数えるが、必要なものすべてに参照的注記がついているとは限らず、実際には固有名を事例関係としてもつべき普通件名はもっと多いと思われる。

さて、現状では登呂遺跡に関する図書には「登呂遺跡」という標目のみが付与されており、また「登呂遺跡」と「遺跡」という 2 件名典拠レコード間には明示的なつながりはない。これでは「遺跡」を検索した利用者に、「遺跡」とその下位語(「住居遺跡」など)を付与された図書は提示できるが、個々の遺跡名を付与された図書は提示できない。「個々の遺跡名も標目となる」といわれても、全ての遺跡名を想起して検索することは不可能だから、結局遺跡に関する図書を通覧することはかなわない。このような状態を解消するには次の 2 つの方法が考えられる。

- 固有名と普通件名の両方を付与することにする。例えば「登呂遺跡」とあわせて「遺跡 静岡県」を付与しておけば、「遺跡」からも探せるし、もちろん個々の遺跡名でもアクセスできる。件名付与規定による対処である。

- 普通件名標目レコードと固有名件名標目レコード間の事例関係を何らかの形で明示的につける。件名典拠ファイル側での対処である。

本研究では後者の方法を試みる。どちらの方法も、「遺跡」からのアクセスに対して関連資料を集中して提示する役割は果たせるが、利用者に典拠ファイル中に存在する「個々の遺跡名」のリストを提供できる点で、件名典拠ファイル中に明示的に情報をもつ方式のほうが優れていると考えるからである。

とはいえ、この関係づけは機械的に行なうのは不可能で、手作業に頼らざるを得ない。本研究ではそのような時間をかけられなかったので、先に地名件名に付した種類別コードを利用して、モデルを作成するにとどまった。実装できたのは、「河川」「遺跡」「鉱山」など13の普通件名と、対応する地名件名合わせて696レコードとの関係にすぎない。

具体的には、地名件名レコード側には種類コードに対応して「一般件名ID」を持たせ、普通件名レコード側には「事例関係ID」として地名件名レコードのIDリストを持たせる。

標目: 石狩川
一般件名ID: BSH300086600 (「河川」のID)

標目: 河川
事例関係ID: PLC000000195; PLC000000201; PLC000000218; ...
(「釧路川」「十勝川」「石狩川」...のID)

なお、事例関係は階層関係の一種ではあるが、件名典拠ファイル上で関係構造を持たせる場合には、普通件名間のBT、NT関係とは区別した方がよいと思われる。階層構造表などを提示する場合に、例えば「住居遺跡」「古墳」といった普通件名の下位語に、多数の固有名が混在すると、甚だわかりにくくなる。関係の種類が異なるということを明示的にしておいた方がよい。

5.6.4 細目つき件名の関係構造

5.4.7で述べたように、典拠ファイル中には8,244件の細目つき件名標目レコードが作成され、これらは主標目、細目レコードへのリンクデータとして、「分解要素ID」フィールドを持っている。

標目: 音楽 歴史 現代
分解要素ID: BSH300055300 BSH300669900 PRD000000013
(「音楽」「歴史」「現代」の各ID)

この「分解要素ID」を手がかりに、細目つき件名の関係構造データを機械的に作成することができる。

まず、それぞれの細目つき件名について、「 」という標目を「 」という細目で展開したもの、という情報を作成する(や は実際には

IDである)。「音楽 歴史」なら「音楽」を一般細目「歴史」を用いて展開したもの、であり、「音楽 歴史 現代」の場合は「音楽 歴史」を時代区分「現代」で展開したものと最後の細目のみで考える。これを全ての細目つき件名に施した後、転置して「」で集積すると、標目「」を展開している細目つき件名の集合が得られる。このとき、検索システム上の便宜を考えて、「」の種類によって「主題展開(一般細目など)²⁸」「地名展開(地理区分)」「時代展開(時代区分)」に分けておく。一方で、今度は全データを「」で集積すると、細目「」を使用して展開している細目つき件名の集合が得られる。

結果として次のような関係データが作成される。

標目形: 音楽
主題展開 ID: BSH300055400;BSH300055700;BSH300056200...
(「音楽 演奏」「音楽 作曲」「音楽 歴史」...の各 ID)
地名展開 ID: BSH300056500;BSH300056600;BSH300056700...
(「音楽 イタリア」「音楽 東洋」「音楽 日本」...の各 ID)

標目形: 音楽 歴史
時代展開 ID: BS8900196200
(「音楽 歴史 現代」の ID)

標目形(細目): [演奏]
細目使用 ID: BSH300055400...
(「音楽 演奏」...の各 ID)

標目形: イタリア
主題展開 ID: SUBGEN:SUB000000290;SUB000000291...
(「イタリア 教育」「イタリア 経済」...の各 ID)
細目使用 ID: BSH300056500;SUB000001547...
(「音楽 イタリア」「演劇 イタリア」...の各 ID)

通常の普通件名標目レコード(それに個人名、団体名なども)では「主題展開」、「地名展開」、「時代展開」が現れる可能性がある。ただし、時代展開は歴史に関する標目に限られるので、他の二つよりはずっと例が少ない。5.5.1で作成した細目レコード(細目としてしか使われないもの)には「細目使用」が現れる。地名件名標目レコードは主標目としても細目としても使用されるので、「主題展開」「細目使用」の双方が現れる可能性がある。また、一般細目等のうち同形の普通件名があるものは統合しているので、一部の普通件名標目レコードにも「主題展開」「細目使用」の両方が現れる。例えば、「年鑑」では「年鑑 書誌」(年鑑類の書誌)は主題展開であり、「新聞 年鑑」(新聞に関する年鑑)は細目使用である。最後に細目つき件名標目レコードでは、さらに細目展開されるので、「主題展開」「地名展開」「時代展開」が現れる可能性がある。

²⁸細目には形式区分も多いので「主題展開」の語にはやや違和感があるが、主題と形式を分ける術もないので、この呼称を使う。

ところで、5.5.4では、BSH3 版までの地理区分の制約規則による不具合を解消するため、「総合開発+北海道」のような「主題+地名共出現レコード」を作成したことを述べた。これは地理区分による展開を補うためのものであるから、細目つき件名と同じように扱う。すなわち「 + 」を主題と地名 にすべて分解しておいて、それぞれの主題、それぞれの地名ごとにまとめて「共出現 ID」を各普通件名標目レコード、地名件名標目レコードに格納しておく。

さて、ここまでの関係データは、結局のところ細目つき件名を分解した各要素と関連づけるということである。ところが、分解した各要素である普通件名や地名件名にはそれぞれ階層構造がある。例えば、「札幌市 歴史」は「札幌市」の「主題展開」であるから、「札幌市」の関連標目としてリンクされる。一方で「札幌市」は「日本/北海道/札幌市」という階層構造の中にあるので、「北海道」と「札幌市」はリンクされており、「北海道」「札幌市」「札幌市 歴史」という上下関係をたどることができる。しかし、問題はここに「北海道 歴史」という標目があった場合の「札幌市 歴史」との関係である。本システムでは、「北海道」に対するデフォルトのヒット書誌表示には「札幌市」を、ひいては「札幌市 歴史」をも含むようにしている。であるなら、「北海道 歴史」が「札幌市 歴史」を含まないと理屈にあわない。同様に、「日本 農業」を考えると、やはり「北海道 農業」が必要である。また「農業」の下位語には「農村」があるので「農村 日本」(「農業」は「地名のもとの主題区分」なので「日本」が先行するが、「農村」では地理区分をする)も必要である。さらに考えれば、「農村 北海道」にも何らかの経路が必要である(「農村 日本」を介してもよい)。

これとは別に、「日本 歴史」は「日本 歴史」へのリンクを持っているが、一方で「日本 経済 歴史」へのリンクはない。これも上位・下位に疑似した関係と考え、リンクをはることとする。

以上のリンクをはるには、次の操作を行なう。

- 細目つき件名の各要素を、5.6.1で述べた階層構造表データで置換する。実際にはIDであるが、模式的に示すと「遺跡 札幌市」なら「考古学/遺跡 日本/北海道/札幌市」となる。これを各要素(「考古学」から「札幌市」まで)から検索できるようDB化する。
- もとの細目つき件名集合に戻る。そして「考古学 日本」ならば上のDBから「考古学」と「日本」を両方含むものを選ぶ。「考古学 北海道」「遺跡 日本」「遺跡 札幌市」などが自然に選ばれる。この場合はありえないが、「考古学 日本」があれば、それも選ばれるはずである。「考古学 日本」も選ばれるが、必要ない(既に「展開」として処理済)ので除く。

- 「日本 歴史」に対する「日本 経済 歴史」のように、もとの要素（「日本」と「経済」）が全部残って別の要素が加わった形の場合は、細目展開に準じる形と考え、「準細目展開 ID」という関係データとする。
- その他のものは BT, NT 関係に準じる形と考え、「準下位語 ID」という関係データとする。

この方式には一つの欠陥がある。それは、「辞典 歴史」(辞典の歴史)の「準下位語」に「経済 歴史 辞典」などが入ってしまうことである。分解した各要素の順序を考慮して調整する必要があり、なお改善の余地があるが、本研究では対処できなかった。

5.6.5 連想関係

5.6.4までに述べてきた関係構造は、すべて階層関係といえる。本研究では特に標目間の階層構造を生かした検索システムを目指すので、典拠ファイル作成にあたっては階層関係の整備に労力を集中した。そのため、他の関係構造についてはあまり追究できていないが、実装した部分、できなかった部分について簡略に述べる。

連想関係 (RT 関係) とは、階層関係にはないが関連する関係で、通常 2 標目レコード間で相互的に参照が示される。

BSH3 版では階層関係との間に区別がなく、ともに連結参照（「をも見よ」参照）とされてきたが、BSH4 では BT, NT とは区別して関係づけることとなった。とはいえ、階層関係に比べると整備がゆきとどかず、数も 279 組 (相互性は保証されている) にとどまる。なお、RT が少ない一因には、5.6.1 で述べたように、BSH4 では一般には連想関係とみなされる関係の多くが階層関係として処理されているという事情がある。

BSH4mr 中の RT(連想語) フィールドは、BT, NT と同様、相手先のレコード ID と標目形が埋め込まれている。

HD(標目形): 衛生	HD(標目形): 医学
RT(連想語): 医学 BSH300015600	RT(連想語): 衛生 BSH300038900

RT 関係については、BSH4mr 中のもの以外には、全く追加していない。

5.6.6 年代的関係

IFLA の件名ガイドライン [8] では、「階層関係」「連想的関係」とともに「年代的関係」をあげ、「団体や地理的名称のようなカテゴリには不可欠である」としている。本研究では、5.6.2で地名件名間の階層構造を、5.6.3では地名件名と普通件名との事例関係を設定したが、実は地名件名に関しては、この年代的関係、年代的変遷が大変重要である。ただ、これは単純な新旧の

関係だけではなく分離や統合が頻繁にあり、さらには同一の名称でも時代によって指し示す範囲が異なるということも多く、本研究では全く手がつけられなかった。階層関係と年代的关系を統合した関係構造をどのように構築するのは、大きな問題である。

なお、団体名典拠レコード間に年代的关系がある場合、もととなったNACSIS-CAT 著者名典拠レコードのSAF(からも見よ参照)で相互に関係づけられているが、これも特に利用できなかった。

5.6.7 直接参照関係

5.5.2で述べたように、BSH4mr 中にある参照語 2,810 件を参照語レコードとした。直接参照関係は、USE(を見よ参照)、UF(を見よ参照あり)によって相互に示されている。

HD(参照形): 碁	HD(標目形): 囲碁
USE(を見よ): 囲碁 BSH300017500	UF: 碁 BSH300192400

もっとも、検索システム上では USE のみを使用しており、UF は本システムではなくてもよい。なお、BSH4mr では、参照先が複数になることもある。例えば「油」は、「石油」と「油脂」の2つを参照先とする。

個人名件名レコード等では「から見よ参照」を標目レコード中に残しており、参照語レコードは作成していない。

5.6.8 参照的注記(仮称)と説明つき参照

BSH4mr 中の注記は、標目使用の範囲を示すものが多いが、それとは別に5.6.3で述べた「参照的注記」がある。複数の標目に対して「をも見よ参照」を出すものである。「をも見よ参照」の対象には下位語と連想語が考えられるが、下位概念への参照が多い。

参照的注記は次の2パターンのもものがほとんどを占める。

- 一般的名称から「個々の 名」を参照するもの。「遺跡」に対して、「個々の遺跡名(例:登呂遺跡)も標目となる」など。
- ある細目を付した形への参照。「商業」に対して、「各国・各地方名のもとの主題区分 商業をも見よ」など。

印刷体の標目表ではこれしか仕方がないが、OPAC では可能なら関係する標目レコードの一覧を出した方がよい。ここにあげた例は、「遺跡」は事例関係として、「商業」の場合は細目つき件名と細目レコードとの関係として、解決した。事例関係のほうは、ごく一部でモデル的に示したにすぎないが、固有名群への参照すべてに応用することは可能である。

しかし、対処できてない例もある。

- 「外国語」に対して、「個々の外国語名 (例:英語, フランス語) も標目となる」のような例。「個々の外国語名」が BSH4mr 中に存在していないわけではない。階層構造として、「言語」の下位語に各言語名を位置づけているので、「外国語」からはこのような参照になっている。
- 「外交」に対して、各国名もとの主題区分 対外関係 (例:アメリカ合衆国 対外関係) をも見よ」のような例。このように主標目としての形と、細目としての形が異なっていると、5.6.4の手順では関係づけられない。

このあたりは不十分になっており、改善の余地がある。

また、複数の標目に対して直接参照 (を見よ参照) を出すなどのために BSH4mr では「説明つき参照」レコードが設けられているが、本研究時点では整備が進んでいないこともあり、実装できなかった。

5.7 最終的な件名典拠ファイル

以上で、本研究における件名典拠ファイルが完成した。

既に、3章の表 6に、最終的な典拠ファイルのレコード件数を示し、本章はじめの表 13に、典拠レコードのフィールド一覧を示した。

5.3で付与標目データを分類した時点での標目候補は 27,337 件であるが、そのうち、19,010 件 (69.5%) が確定された (付与されていない標目もあるので、典拠レコード中の標目総数は 21,071 件)。これを、表 14に示した「CAT 書誌レコード中の BSH 標目」までさかのぼって算出すると、付与された延べ標目数では 215,854/247,390 で 87.3%を今回の典拠ファイルはカバーしている。また、リンクしている書誌レコード数 (複数標目がある場合は 1 件でもリンクされているもの) では 164,305/179,893 で 91.3%のカバー率である。

なお、検索システムはこの件名典拠ファイルをまず検索する仕様なので、この時点でカバーされなかった書誌レコード、付与標目は除去した。従って検索システム BSH-OPAC は 164,305 件の書誌レコードに付与された 215,854 件の標目を対象とすることとなった。

6 まとめ

以上、件名標目間の関係構造を利用した主題検索支援システムについて述べてきた。

3章で機能を説明した検索システム BSH-OPAC について、4章で評価実験を行なった。4.3でまとめたように、限られた範囲の実験ではあるが、「与えられた主題検索質問に対して、タイトルキーワード検索と比べて優位性のある結果が得られるか」「実装した様々な機能が利用者に理解され、使いこなされるか」の2点で、概ね肯定的な結果が得られた。しかし、常に BSH-OPAC がうまく機能するとは限らず、有効な検索ができないケースもかなり存在することも明らかになった。

5章では、BSH-OPAC の機能を実現するために必要な、件名典拠ファイルの構築作業について述べた。特に5.6で行なった典拠レコード間の各種関係構造データ作成は、BSH-OPAC 検索機能の根幹を左右するものである。

ところで、2.1において、統制索引語を用いた主題検索に関わる問題は、「検索システムの問題」「索引付与作業の問題」「索引ツールそのもの問題」に大別できると述べた。本研究で追求してきた件名検索においては、索引付与作業の問題とは「件名典拠管理の問題」であり、索引ツールそのもの問題とは「件名標目表の問題」である。

最後にこの3点について、若干のまとめと展望を行なっておきたい。

検索システム

本研究における評価実験では、大学図書館でなら一般利用者の使用にも堪えるのではないかと、いう結果が出た。もちろん、時間的制約や筆者の力量不足、さらには CGI インターフェースの限界もあり、一見してわかりやすいシステムになったとはいえない。システム環境にもよるが、より直観的に理解できるインターフェースを作ることは可能であろう。どうしても多くの情報を画面に詰め込むことになるので、マルチウインドウでそれぞれ色を変えるなどして各種情報の区別が容易にできるようにした方がよい。また、3.1で「遺跡」を中心とする関係構造図を示したが、そもそも関係構造を文字だけで表現しようというのは無理があり、より視覚化された情報提示が必要である。

また、BSH-OPAC は実験用なので「件名からしか検索できない」システムである。実用的には当然タイトルキーワード検索など、他の検索機能とも融合させなくてはならない。これも分かりやすくインターフェースを設計するのは、それほど簡単ではないように思われる。

ところで、件名標目、分類記号、フリーキーワードといった各種の主題検

索方法の優劣に関する議論は、「一長一短」という結論に落ちつくことが多いようである。例えば、分類は体系的な関連主題の検索にすぐれ、件名は特定主題の検索にすぐれている、とよくいわれる。オンライン環境下では複数のツールを提供することも可能であるが、利用者にとっては主題検索の方式がいくつもあると言われても、その都度どれを使えばよいのかわからず、かえって煩雑に感じるだけであろう。諸方式の長所を生かして融合したシステムがどのようにすれば可能か、また本当に可能性があるのか、現時点の筆者には何ともいえないが、今後探求される必要があると感じる。

件名典拠管理

BSH-OPAC は、5章で作成した典拠管理ファイル抜きには存在しえない。その点で、まず典拠管理ファイルの構築と維持を図書館システム中に組み込まなくてはならない。

いうまでもないが、典拠管理ファイルは書誌レコード集合とのリンクを保って存在すべきもので、日々の目録作成とともにメンテナンスされる必要がある。本研究では付与標目の集合からまとめて件名典拠管理ファイルを作成した手順を説明したが、これは本来の道ではなく、付与されていた標目種類数の3割ほどが無視される結果となった。

NACISIS-CAT には著者名と統一書名の典拠ファイルがあるので、大学図書館においては典拠管理作業も少しは身近なものになってきた。しかし、本研究で構築したような件名典拠ファイルを作成しようとする、件名典拠レコードと書誌レコードのリンク管理は著者名典拠等とそうかわらないにしても、典拠レコード間の関係構造がきわめて複雑である。

普通件名間だけを考えても、BT, NT 関係は相互参照ではあるが二方向で意味が異なるので、「をも見よ参照」では処理しきれない。また、細目つき件名については、主標目、細目に分けた各部分とのリンクが必要である。その他にも普通件名と固有名の事例関係など、様々な関係構造がある。典拠ファイル上でどのように表現していくかというレコードフォーマットにも整理の余地があるし、実際に典拠管理を行なうには支援システムが必要であろう。BSH のシソーラススタイル化を行なった北、芝 [10] は件名標目表の維持作業に関して「シソーラス・エディタ」の必要性を指摘しているが、典拠管理作業においても何らかの支援システムが欠かせないと思われる。

典拠管理作業の支援システムは、件名付与作業の支援システムにもなる。関係構造データを含めた典拠ファイルの情報が適切に表示できれば、正しい件名標目を付与する助けになるであろう。件名付与率をあげることが件名検索を有効にする条件であることはいうまでもない。

各図書館で全く独自に件名を付与し、典拠管理作業を行なっていくこと

は、現在の諸情勢では不可能に近いと思われる。各種 MARC はそれぞれ件名典拠ファイルを持ち、管理を行なっているため、MARC を利用している公共図書館等では、MARC 作成元から件名典拠データを提供してもらおう、という方向が現実的であろう。MARC の書誌レコードと件名典拠レコードがきちんとリンクされているなら、件名標目の変更などの場合もデータのやりとりで対応が可能である。

大学図書館等では、NACSIS-CAT の書誌レコードによって自館の OPAC を構築するしかない。共同分担目録システムである NACSIS-CAT では、十分な典拠管理を行なうことは MARC のような集中管理よりもずっと困難である。著者名典拠のように、典拠レコードの作成と書誌とのリンクをオプションの作業にして強制しないとしても、実際の運営は相当に難しい。仕組みが複雑であるうえに、独自に件名規程として定めなくてはいけない部分が非常に大きい。担当者による付与件名のゆれは、著者名などとは全く桁違いであろうから、どこでどう調整するのかという問題もある。

しかしながら、わが国の図書館目録が主題アクセスにおいて著しく貧弱であるのは確かであるし、主題典拠管理の問題を将来的に避けて通ることはできないのではないだろうか。NDC9、BSH4 など主題関係ツールの機械可読化が進んできた現在、そろそろ検討の必要があると思われる。

OPAC 環境下の件名標目表

評価実験で明らかになったことの一つは、検索質問に意味内容の一致する件名標目があるかないかで、BSH-OPAC の有効性が大きく左右されるということであった。「長期波動」という主題に対して、付与されている件名が「景気」や「世界経済」では、有効な検索はできない。

件名標目表は十分な特定性を持たなければならないと思われる。従来、あまりに特定の件名標目を作ることは小件名の乱立につながり、目録をかえって使いにくくするといわれてきた。確かにカード目録の環境では、いくら参照があるとはいっても、一つのことを調べるのに、いくつものカードボックスを相手にしなくてはならなくなって不便である。しかし、OPAC では本研究のように下位項目を包含した検索を提供することで、広く網をかける検索が可能である。となれば、特定主題を検索する利用者のために、少なくとも文献的根拠があるならば、特定の件名標目をどんどん採用すべきと考える。

また、OPAC ログからの検索質問で件名標目にヒットしなかったものは、見方を変えれば件名標目ないしは参照語の候補になる。佐川 [22] は OPAC ログにおける主題からの検索語を調べて、ノーヒットの語を必要に応じて件名辞書に「別名」として追加する、という実践を報告している。こうした手法は、件名標目表の維持編集においても有効である。これまでの件名標目選定

は概ね図書の出版状況に応じてなされてきたわけであるが、利用者の検索行動にも目を配ることが、今後必要になってくると思われる。

件名典拠ファイル構築や検索システム作成の過程で、BSH4mr データに起因する問題で処理に無用の複雑さが生まれた部分もいくつかあった。大きなものを一つあげれば、地名と主題の列挙順序が地名先行と主題先行の2種類あるなど細目つき件名の作成規則が複雑で、標目間の関係づけに苦労したほか、規則上ありえない標目を排除するアルゴリズムも作成できなかった。

もう一つ、わが国にはBSHとNDLSHという2つの件名標目表があり、NACSIS-CATでは混在使用されているという現状がある。2.4で述べたように、最近はBSHの割合が高くなっているが、全体で見ればまだNDLSHが付与された書誌レコードの方がずっと多い。

統合化の試みも既にあるし[10]、BSH4改訂にあたってもNDLSHを意識した作業がなされているが、もし明示的に相互リンクできれば非常に有効である。NDLSHはBSHに比べると語数が豊富で、国立国会図書館の蔵書とリンクしている性格上、確実な文献的根拠に基づいてメンテナンスされているという長所がある。一方で、今のところ標目間の関係構造がほとんどつけられていない(「をも見よ参照」がほとんどない)という短所もあり。両者を統合した検索システムができれば、単に検索対象書誌レコードが増えるというだけでなく、双方の長所を生かせる可能性がある。統合検索のための相互リンクは典拠管理ファイルのレベルで行なうこともできるが、やはり件名標目表維持機関のレベルで何らかの配慮があったほうが望ましいであろう。

より根本的に、BSH4mrにおける関係構造そのものの妥当性や、細目つき件名方式を維持することの是非、などの問題や、OPAC時代にあるべき件名標目表の構造については、現時点で整理して説明できる状態にない。こうした点は今後の課題としたい。

謝辞

本研究の遂行にあたっては、多くの方々のお世話になりました。

まず、指導教官の宮澤彰先生には、多くのアドバイスを頂きました。特に、問題意識の拡散しがちな私に対して、いつも方向性を整理して頂きました。また、データ処理作業において、先生の手になるコマンドツール D-Series のおかげで、大量データの複雑な処理を効率的に行なうことができました。

教育研修部の諸先生方、濱田先生、橋爪先生、神門先生、後藤田先生には、着手発表会や最終発表会の席上で、様々なアドバイスを頂きました。また、橋爪先生、後藤田先生には、UNIX に関する実習でもお世話になりました。

本研究は、BSH4 版の編集用機械可読データなしには成立しません。日本図書館協会件名標目委員会ならびに協会事務局には、刊行前の微妙な時期にもかかわらず、データの研究使用をご許可いただきました。また、NDC・MRDF についても研究使用させていただきました。

筆者は 1995 年から日本図書館協会件名標目委員として、BSH4 版の改訂作業に携わっておりますが、委員長の石塚先生をはじめとする委員会メンバーには、いつも多くのことを教えて頂いています。また特に、吉田暁史先生、北克一先生には、本研究の内容にも直接ご教示を頂きました。

5 カ月間という長期のセミナーに対し、下村事務部長をはじめとする神戸大学附属図書館の皆様には、快く研修参加をお認め頂きました。とりわけ勤務掛である情報管理課雑誌掛の、前田掛長をはじめとする皆様には、業務多忙の折りから大変なご迷惑をおかけしましたが、温かく見守って頂きました。

さらに、年度末の多忙な時期にもかかわらず、神戸大学人文・社会科学系図書館、自然科学系図書館の多くの方々には、手間と時間のかかる評価実験被験者をお引き受け頂きました。

最後になってしまいましたが、塚田課長をはじめとする学術情報センター研修課の皆様には、様々なお世話を頂きました。とりわけ、熊淵掛長には、研究遂行から生活全般にいたるまで、いろいろな御配慮を頂きました。

このような多くの方々のお力添えなくしては、本研究を遂行することはできませんでした。深く感謝を申し上げます。

参考文献

- [1] Aitchinson, Jean & Gilchrist, Alan. 内藤衛亮 [ほか] 訳 『 シソーラス構築法』 丸善, 1989. 195 p.
- [2] [American Library Association] Subcommittee on the Display of Subject Headings in Subject Indexes in Online Public Access Catalogs. *Headings for tomorrow : public access display of subject headings*. Chicago, American Library Association, 1992. 51 p.
- [3] Bates, Marcia J. “Subject access in online catalogs : a design model” *Journal of the American Society for Information Science*, 37(6): pp. 357-376 1986
- [4] Bates, Marcia J. “Rethinking subject cataloging in the online environment” *Library resources & technical services*, 33(4): pp. 400-412 1989
- [5] Drabenstott, Karen Markey & Vizine-Goetz, Diane. *Using subject headings for online retrieval*. San Diego, Academic Press, 1994. 365 p.
- [6] 学術情報センター編 『 目録情報の基準 第3版』 学術情報センター, 1997. 75 p.
- [7] 芳賀奈央子, 松井幸子 「 NACSIS-CAT の和図書書誌ファイルにおける主題情報の現状調査」 『 研究成果流通システムの研究開発 : 科学研究費総合研究 (A)(課題番号 06302076) 研究成果報告書 平成7年度報告』 学術情報センター, 1996: pp. 101-115
- [8] 国際図書館連盟書誌調整局分類・索引セクション 「 件名典拠ファイルのためのガイドライン」 作業委員会 永田治樹 [ほか] 訳 『 件名典拠記入および参照記入のためのガイドライン』 日本図書館協会, 1995(『 現代の図書館』 特別号). 56 p.
原著の書誌事項は次の通りである。Working Group on “Guidelines for Subject Authority Files” of the Section on Classification and Indexing of the IFLA Division of Bibliographic Control. *Guidelines for Subject Authority and Reference Entries*. Munchen, K.G. Saur, 1993. 62 p.

- [9] 北克一, 芝勝徳「基本件名標目表第3版(BSH3)のシソーラス構造化の試み - 外形式を中心として - 」『整理技術研究』31: pp. 53-75 1992
- [10] 北克一, 芝勝徳「我が国における主題索引ツール統合の試み」『TP&D フォーラムシリーズ』3: pp. 6-31 1994
- [11] 北克一, 芝勝徳「OPACにおける主題検索と典拠システム: 主題検索システム・プロトタイプと利用者 OPAC」『現代の図書館』32(2): pp. 115-122 1994
- [12] 国立国会図書館図書部編『国立国会図書館件名標目表 第5版』国立国会図書館, 1991. 2冊
- [13] 主題情報の組織化とユーザー・インターフェイス研究班「利用者の検索行動と主題情報—国立国会図書館におけるOPACモニター調査を中心に」『図書館研究シリーズ』32: pp. 181-269 1995
- [14] Larson, Ray R. “Between Scylla and Charybdis: subject searching in the online catalog” *Advances in librarianship*, 15: pp. 175-236 1991
- [15] Ludy, Lorene E. “OSU Libraries’ use of Library of Congress Subject Authorities File” *Information technology and libraries*, 4: pp. 155-160 1985
- [16] Office of Subject Cataloging Policy, Library of Congress. *Subject cataloging manual: subject headings. 4th ed.* Washington, D.C., Cataloging Distribution Service, Library of Congress, 1991. v. (loose-leaf)
- [17] 日本図書館協会件名標目委員会編『基本件名標目表 第3版』日本図書館協会, 1983. 2冊
- [18] 日本図書館協会件名標目委員会「件名標目表改訂の基本方針」『図書館雑誌』89(3): pp. 186-189 1995
- [19] 日本図書館協会件名標目委員会「基本件名標目表第4版機械可読フォーマット(案)」『現代の図書館』33(1): pp. 55-67 1995
- [20] もり・きよし原編 日本図書館協会分類委員会改訂『日本十進分類法 新訂9版』日本図書館協会, 1995. 2冊

- [21] 日本図書館協会分類委員会編 『NDC・MRDF9 マニュアル:日本十進分類法 (NDC) 新訂9 版機械可読ファイル・マニュアル』日本図書館協会, 1996. 23 p.
- [22] 佐川祐子 「公立図書館における OPAC の導入 -東京都杉並区立中央図書館の利用記録の分析を通して」『現代の図書館』28(2):pp. 66-75 1990
- [23] 斎藤泰則 「情報探索と質問定式化 - オンライン目録探索を中心に - 」『Library and information science』30: pp. 77-92 1992
- [24] 千賀正之 [ほか] 「件名作業の現状と今後」『図書館研究シリーズ』27: pp. 75-196 1987
- [25] 豊田邦雄 「基本件名標目表のシソーラス化を考える」『図書館界』39(6): pp. 221-228 1987
- [26] 上田修一 「OPAC における主題探索の問題」『TP&D フォーラムシリーズ』2: pp. 2-11 1993
- [27] 吉田憲一 「主題検索と OPAC : 司書課程の学生への目録利用調査から」『図書館学会年報』40(2): pp. 71-84 1994
- [28] 吉田暁史 「オンライン時代の件名標目表 -基礎的考察」『図書館界』40(2): pp. 98-102 1988
- [29] 吉田暁史 「目録と主題アクセスの今日の問題」『現代の図書館』32(2): pp. 83-89 1994
- [30] 渡邊隆弘 「NDC8 を利用した OPAC の主題検索支援システム構築の試み」『図書館界』46(4): pp. 164-176 1994