

# 千葉市図書館 OPAC についての報告

2013 年 12 月 21 日

道廣勇司<sup>1)</sup>

## 1. はじめに

多くの公共図書館と同様に、千葉市図書館も収蔵資料の OPAC\*がウェブで公開されており、インターネットに接続できれば館外からでも 24 時間資料の検索ができる。

この千葉市図書館 OPAC の動作を調べたので報告する。目的は二つある。一つは千葉市図書館 OPAC の改善のための参考に供すること。もう一つは他の公共図書館の利用者が、それぞれの図書館の OPAC の動作を検証するための参考に供することである。

右肩に「\*」マークの付いた用語は、第 9 節で簡単な説明を行った。

## 2. 検索機能

検索機能こそは OPAC の要である。ある目的を持って検索を行ったとき、検索漏れ\*も検索ノイズ\*も共に少ないことが望ましい。

### 2.1 異体字対応

**【問題点】** 異体字で検索できない場合がある。

漢字は「島・嶋・嶌」のように異体字が多い。異体字は、形が異なるが意味と音が一致しており、検索においては同一視することが望ましい。このような同一視を検索の「異体字対応」と呼ぶことにする。

異体字対応が無いか不十分だと、大量の検索漏れが発生することがある。

千葉市図書館 OPAC は異体字対応しているが不十分であることを次節以降で具体的に述べる。

#### 2.1.1 鷗と鷗

森オウガイのオウはカモメという字であるが、「鷗」とも「鷗」とも書かれる。森自身が両方の字を自署しており、どちらも正しいと言えるが、書籍ではふつう「鷗外」と印刷されている。

しかし「鷗」の字は JIS コード (JIS X 0208\*) に含まれなかったため、昔のコンピューターでは扱えなかった。世田谷区立図書館のように「鷗外」でないとヒット\*しない OPAC もある。「鷗」は JIS X 0213\*や Unicode\*には含まれる。

千葉市図書館 OPAC は世田谷区と逆に、「鷗外」でなければヒットしない。「オウガイ」で変換して「鷗外」しか出

てこない日本語入力ソフトもあることを考えると、これは望ましくない。

#### 2.1.2 間と聞

内田ヒヤッケンのケンは、印刷物ではふつう「聞」のように門構えの中が「月」になっている。「聞」は「聞」の異体字である。JIS X 0208 には含まれないが、CP932\*には含まれる。

千葉市図書館 OPAC では「百間」でなければヒットしない。「百間」のほうが見慣れた字体であり、現在では「ヒヤッケン」で変換してこちらを出す日本語入力ソフトがあることを考えると、これは望ましくない。

オウガイとヒヤッケンの例から分かることは、千葉市図書館 OPAC では、データが「鷗外」「百間」で統一されていること、「鷗・鷗」「間・聞」は異体字として扱われていないということである。

#### 2.1.3 瘦と瘦

「瘦」は「瘦」の旧字で、JIS X 0208 には含まれないが、人名用漢字であり、第三水準漢字として JIS X 0213 に含まれる（「瘦」のほうは JIS X 0208 にもある）。

千葉市図書館 OPAC の書誌情報では、どちらの字もデータ中に存在する。しかし異体字としては扱われないため、「書名」に「瘦身」と入れれば 4 件、「瘦身」と入れれば 27 件と、異なる資料がヒットする。

「鷗外」「百間」の例では検索結果がゼロなのでおかしいと気づくが、瘦身/瘦身はどちらでも何かしら出てくるため、他方の可能性に気づきにくいという落とし穴がある。

#### 2.1.4 その他の異体字

「高・高」「崎・崎」は異体字として扱われない。

よって、「高田」と「高田」では異なる検索結果となる。「山崎」と「山崎」も同様。

ちなみに「崎」「高」は JIS X 0208 内だが、「崎」「高」は JIS X 0208 外で、CP932 に含まれる。

#### 2.1.5 異体字として扱われる文字

最初に述べたように、適切に異体字として扱われる文字もある。

「龍・竜」「斎・齋」「島・嶋・嶌」「崎・寄」「渡・邊・邊」「船・舩」「鉄・鐵・鐵」「秋・穉」「国・國」「亞・亜」といった組合せは異体字として扱われる。これらは全て JIS X 0208 に含まれる字である。

よって、「芥川龍之介」「芥川竜之介」や「与謝野鉄幹」「與謝野鐵幹」はそれぞれ同じ検索結果となる。

<sup>1)</sup> 千葉市在住。連絡先は <http://moji.gr.jp/mich/>

### 2.1.6 まとめ

以上のことから、千葉市図書館 OPAC の検索機能は異体字に対応しているが、異体字辞書<sup>2)</sup>は JIS X 0208 の範囲内に収まる小規模なものなのではないかと推測される。せっかくの異体字対応がこれでは勿体ない。異体字辞書を拡充すれば検索漏れがぐっと減らせるだろう。

いずれにせよ、図書館の蔵書検索システムで「森嶋外」「内田百閒」の検索結果がゼロでいいはずがない。

## 2.2 件名標目による検索

OPAC では、各資料にその主題を表す件名標目\*が付与されており、これを検索条件に含めて検索することができる。具体的には、「詳細検索」で「件名」のところ是件名標目を入れる。

たとえば、ヨーロッパの歴史を主題とする書籍には「西洋史」という件名標目が付与されている。これを知っていれば、「書名」に「宮廷」を入れ、「件名」に「西洋史」を入れて検索することで、中国史における宮廷を主題とした資料を排除した検索が可能になると期待できる。

件名標目による検索が重要な理由はいくつかあるが、内容が書名にストレートに反映されていない書籍<sup>3)</sup>が多いことも挙げられる。

多くの公共図書館は日本図書館協会件名標目委員会が編集した『基本件名標目表第4版』に基づいた件名標目を採用している。千葉市図書館も然り。

### 2.2.1 件名標目を知る手段

【問題点】どんな件名標目があるか分からない。

千葉市図書館 OPAC には件名標目表そのものを検索する機能が無い。それゆえ、ヨーロッパの歴史について書かれている本を探したくても、件名に「西洋史」と入れればいいとは分からない。件名に「ヨーロッパ史」と入れても何もヒットしないのである。

実は裏技がある。筆者があるときレファレンスカウンターで国語辞典について相談した際、対応してくださった方は国語辞典の件名標目が分からないため、具体的な国語辞典名をまず検索してその書誌情報から件名標目「日本語一辞典」を見つけ、調査を進めてくださった。あとで分かったが、「ヘルプ」の「※検索のヒント」にもこのやり方がちゃんと載っていた。

とても有効なテクニックだが、一種の「バッドノウハウ<sup>4)</sup>」と言えなくもない。これをもって「OPAC は奥が深い」と感心しては改善が進まない。

それに、よく似た件名標目はいくつもある。例えばある書誌には「民主主義」という件名標目が付いているとして

も、別の書誌には「民主政治」が付いているかもしれない。件名標目自体の検索機能があるなら、「民主」で検索することで、これらのほかに「社会民主主義」や「民主社会主義」などの件名標目があることが知れるだろう。

### 2.2.2 中間一致問題

【問題点】件名の中間一致による検索ノイズが発生する

たとえば、ニンジンについて書かれた本を探そうと「件名」に「人参」と入れて検索すると、人参の本のほかに、外国人参政権や婦人参政権について書かれた本が大量にヒットする。

また、国語について書かれた本を探すため「国語」と入力すると、中国語についての本までヒットする。

件名による検索（のデフォルト）が完全一致でなく中間一致\*だからである。このような仕様になっているのは、どんな件名標目があるかを知らない利用者の要求にできる限り応えようとした、つまり検索ノイズを増やしてでも検索漏れを減らそうとしたからだろうか。

しかし、検索ノイズが半端無い。件名標目を特定できるようにし、完全一致を基本とすべきではないか。

### 2.2.3 仮名表記問題

【問題点】件名標目が仮名の場合にうまく検索できないことがある。

実は件名検索では「前方一致\*」「完全一致\*」の条件で検索ができる。よって「国語」で前方一致検索させれば「中国語」の件名標目を排除することはできる。

しかし、「人参」で前方一致の検索を行うと何もヒットしない。なぜなら人参という件名標目は「にんじん（人参）」という形で格納されているからだ。『基本件名標目表第4版』では人参の件名標目は仮名表記の「にんじん」である。推測だが、このままでは「人参」でヒットしなくなるため、後ろに「(人参)」を追加した形でデータベースに格納されているのだろう。しかし現行システムでは「にんじん（人参）」というデータを単一の文字列として検索を行うらしく、中間一致でなければ「人参」でヒットしないのである。

いまの場合、「人参」は「にんじん」の別表記なのだから、両者に対する前方一致検索の結果を合わせたものを表示するようにならなければならない。

### 2.2.4 曖昧検索問題

【問題点】過剰な曖昧検索\*機能により多量の検索ノイズが発生する。

件名検索機能は、件名標目の読みでも検索できる。そこ

<sup>2)</sup> システム内に、どの字とどの字を異体字とみなすかという一覧表データがあるはずで、それを仮に異体字辞書と呼ぶことにする。

<sup>3)</sup> 例：『菊と刀』（日本文化）、『神は老獺にして』（アインシュタインの伝記）、『沈黙の春』（環境問題）、『モスクワの森』（物理の問題集）、『文字の食卓』（書体デザイン）

<sup>4)</sup> 「計算機を使っていると、何でこんなことを覚えなさいといけないのだろうか、とストレスを感じつつも、それを覚えなさいとソフトウェアを使いこなすことができないためにしつこく覚えなければならない、といった類いのノウハウは多い。そうした雑多なノウハウのことを、本来は知りたくもないノウハウという意味で、私はバッドノウハウと呼んでいる。」（高林哲「バッドノウハウと『奥が深い症候群』」より）

で、「外国人参政権」などを排除しつつ「人参」の本を探そうとして「にんじん」と入力して読み検索すると<sup>5)</sup>、今度は「妊娠」について書かれた本までヒットする。なぜか。

これは、仮名文字列を入力した場合に濁点の有無を無視する曖昧検索機能が働くからだろう。このような曖昧検索は、ときとして有効に働く（用語集「曖昧検索」参照）。

しかし、件名標目でこのような曖昧検索はやり過ぎではないか。一つ一つの件名標目には多数の書誌が対応するため、件名標目の検索ノイズが1件増えただけで、書誌の検索ノイズは数十、数百と増えることになる。

さて、「ペルシャ」で件名検索すると、個人件名<sup>6)</sup>「Turgot Anne Robert Jacques」を持つ書籍がヒットする。これは恐らく「Robert Jacques」に「ロベールジャック」と読みが付与されており<sup>7)</sup>、その「ペールジャ」の部分に一致したためだろう。濁点・半濁点に加え、長音符（ー）をも無視する曖昧検索ならば「ペルシャ」は「ペールジャ」に合致する。長音符を無視する曖昧検索は広く行われている（用語集「曖昧検索」参照）。しかし、「ペルシャ」が「Turgot Anne Robert Jacques」に化けるとは、まるで判じ物である。

以上から、筆者は件名検索において現行のような曖昧検索は適切でないと考える。件名標目がはっきりと分かるなら件名検索は「完全一致」条件でいいわけで、問題の所在はどんな件名標目があるかを調べる手段が無いということにある<sup>8)</sup>。

### 2.2.5 隠しデータ？

件名に「ヨーロッパの歴史」と入れて検索すると、41件のヒットがある。『基本件名標目表第4版』には「ヨーロッパの歴史」という件名標目は無いはずなので、何らかの工夫がなされていると思われるが、よく分からなかった。

実際、ヒットした書誌情報の件名の項目を見ても、「ヨーロッパの歴史」というものは無い。

ただ、41件のほとんどは件名標目に「西洋史」とか「フランス—歴史」などを含む、ヨーロッパの歴史について書かれたものだった。

しかし、中には『学研まんが世界の歴史6 激動の東アジアと唐の皇帝玄宗』のように明らかにヨーロッパの歴史が主題ではないものも含まれている。表示される書誌情報のどこにもヨーロッパとの関連を窺わせるものは無い。これが検索アルゴリズムの何らかの欠陥によるのか、単なるデータ（隠しデータ？）の誤りなのかは判断がつかなかった。

次に件名に「鷗外 漱石 芥川 堀辰雄」と入れてみる

と、唯一『日本の名作文学案内』がヒットする。これの件名は単に「小説(日本)」となっていて、件名標目としてはどこにも作家名が無い。件名標目の隠しデータがあるか、あるいは「書誌詳細」において件名標目が全ては表示できていないとしか思えない。

## 2.3 AND/OR 検索

【問題点】AND 検索と OR 検索を組み合わせたとき、検索条件が曖昧。

詳細検索では、「検索条件1」から「検索条件5」までの入力欄がある。これらの中でAND 検索、OR 検索、さらにはNOT 条件を用いた組合せ検索ができる。



AND, OR, NOT 条件は、五つの検索条件のちょうど間にある四つのドロップダウン<sup>9)</sup>で選べるようになっている(図)。

検索条件1~5を  $C_1 \sim C_5$  と書くことにしよう。

$C_1 \text{ AND } C_2 \text{ AND } C_3$  や  $C_1 \text{ OR } C_2 \text{ OR } C_3$  のように、すべてがAND かOR の場合は曖昧性は生じない。

また、 $C_1 \text{ NOT } C_2$  は、論理的には  $C_1 \text{ AND } (\text{NOT } C_2)$  の意味であり、これも曖昧性は無い。

しかし、AND とOR の混じった  $C_1 \text{ AND } C_2 \text{ OR } C_3$  は以下の二通りに解釈しうる。

- $(C_1 \text{ AND } C_2) \text{ OR } C_3$
- $C_1 \text{ AND } (C_2 \text{ OR } C_3)$

両者が等価でないことは論理学の教えるところだ。

このどちらであるかは検索画面にもヘルプにも書かれていない。

また、 $C_1 \text{ NOT } C_2 \text{ NOT } C_3$  についても、

- $(C_1 \text{ AND } (\text{NOT } C_2)) \text{ AND } (\text{NOT } C_3)$
- $C_1 \text{ AND } (\text{NOT } (C_2 \text{ AND } (\text{NOT } C_3)))$

の二通りに解釈し得る。

では、 $(C_1 \text{ OR } C_2) \text{ AND } (C_3 \text{ OR } C_4)$  という検索条件はどのように表せばいいのだろうか。実はどのように順序を工夫

<sup>5)</sup> 既に見たように人参の件名標目は「にんじん (人参)」となっているので、「にんじん」で検索して人参の本が出てくるのは読み検索の結果とも言い切れないが、ここでは利用者の意図を記述した。

<sup>6)</sup> 「個人件名」とは何かという説明は見当たらなかったが、「ヘルプ」の「※検索のヒント」に言及があった。「その人物に関する評伝などは、この項目で探ることができ」とある。

<sup>7)</sup> 件名標目の読みデータは検索結果に表示されないため、これは推測に過ぎない。

<sup>8)</sup> もちろん冊子版の『基本件名標目表第4版』は各館に備えられている。しかし、中央館の1冊を除いて禁帯出であるし、館内OPAC端末の横に置かれているわけでもない。

<sup>9)</sup> この図にあるような複数の選択肢から一つを選ばせるユーザーインターフェース部品。「ヘルプ」では「プルダウンメニュー」と表記されている。

しても、表すことはできない<sup>10)</sup>。

## 2.4 字数制限

### 2.4.1 あるはずの書誌が出ない

【問題点】1文字の検索文字列でうまく検索できない。

鄧小平について調べたいとする。「鄧」の字は日本では鄧小平本人かせいぜい一族の姓としてしか使わないので、これ1字で検索しても検索ノイズはほぼ無いと考えられる。この1字で検索すれば、もしかすると「鄧語録」といったような「小平」の付かない書誌が見つかるかもしれない。

しかし、「鄧」を「書名」「著者名」「件名」のどれに入れても「該当する書誌がありません。」と表示され、何もヒットしない。実際には『現代アジアの肖像4 鄧小平』とか鄧小平著『鄧小平は語る』といった書誌があるにもかかわらず。一体なぜか。

実はこれらの検索項目では、検索文字列が1文字だけのとき、完全一致条件でしか検索できない仕様になっている。よく見ると、入力欄に1字だけが入っているとき、検索の種類はデフォルトの「この言葉を含む」から「この言葉と一致する」に強制的に変えられ、他の選択肢が選べないようグレイアウト（操作できないようにすること）される。

この言葉と一致する ⇅

つまり、「書名」に「鄧」と入れて検索すれば書名が『鄧』である書誌しかヒットしないのである。実際にはそのような書誌は無いので「該当する書誌がありません。」ということになる。

もう一つ例を挙げる。書名に「雁」と入れて検索すると、森鷗外の『雁<sup>がん</sup>』はヒットするが、宮沢賢治の『雁の童子<sup>かり</sup>』はヒットしない。

理屈が分かってみれば、この動作には一貫性があり、説明可能という点で合理性もある。また、「ヘルプ」にも中間一致検索の字数制限のことはちゃんと記載されている<sup>11)</sup>。

### 2.4.2 検索条件の強制変更気づきにくい

しかし、「雁」で検索して「雁の童子」がヒットしない理由にいったいどれだけの人が気づくだろう？ 筆者はこの現象を見つけてから理由を見出すまで数週間を要した（その間、日常的にOPACを利用していた）。

分かりにくさの原因の一つは検索結果の表示の仕方にある。「鄧」の例では、ヒットが無い。この場合、JavaScriptによって出されるアラートには「該当する書誌がありません。」と表示されるのみであり、検索が完全一致条件に強制

的に変えられたことがどこにも表示されていない。グレイアウトされたドロップダウン「この言葉に一致する」を見て気づけ、というのは無理がある。

「雁」の例では、ヒットがあるので画面が切り替わり、どのような検索条件で検索を行ったかが一応表示される。この表示の問題点は3.1節に譲るが、ここでは、よほどよく見ている利用者でなければ完全一致条件に変えられたことに気づかないだろうことだけ指摘しておく。

### 2.4.3 字数制限は不可避か

改めて言うまでもなく日本語の語彙には漢字1字で表される言葉が大量にある。「鯛」や「恋」や「癌」についての資料を探すにはどうすればいいのか<sup>12)</sup>。

しかし、そもそも1字の検索文字列を認めないのはなぜだろう。ありそうな理由として、1字で検索するとヒット数が多くなりすぎてシステムに過度の負荷をかけるから、ということが考えられる。しかし、「鄧」1字でヒットするはずの件数は数えるほどしかないのに対し、2字でも「全項目<sup>13)</sup>」に「から」と入れてヒットする件数は20万件を超えている。現行の字数制限では、システムの負荷を一定以下に抑え込む効果は無く、ただ統計的に負荷を軽減するだけと言えるかもしれない。

既に見たように、字数制限は利便性を著しく落とす。字数の制限は無くし、ヒット数の上限を設けることで負荷を抑えるという仕様にはできないだろうか。実際、1000件ヒットしてもそれを全部見たいということはまず無いだろう。最初の1000件のみ表示することにすればいい。検索結果がそんなに多い場合は絞り込みを検討するところだ。

### 2.4.4 検索条件の強制変更が全く表示されない例

最後に「鄧」で何もヒットしない問題をもう一度取り上げる。今度は「書名」に「鄧★」と入れて検索してみる。「★」でなく「。」とか「@」とかスペースでも構わない。

この場合、入力欄には2字入れているわけなので、検索の種類はドロップダウンが勝手に「この言葉と一致する」に変えられたりはしない。ところが、「★」のような記号は検索において無視される仕様になっている。検索システムは受け取った「鄧★」をまず「鄧」に変え、字数が1字であることから完全一致条件で検索してしまう。よってヒット数はゼロである。ヒットが無い場合は画面遷移が行われず、ただJavaScriptのアラートで「該当する書誌がありません。」が表示される。このケースでは画面のどこを見ても検索の種類が完全一致に変えられたことがまったく分からない。

この件について、「[鄧★]で検索する人などいない。意味の無い例だ。」と思われるかもしれないが、そうではない。「★」を使ったのは説明の都合であり、現実にはスペー

<sup>10)</sup> 筆者はそのような検索ができなければならないとは考えていない。OR条件は検索を二回に分けることでこれに代えることができる。検索の自由度を求めたらずらにユーザーインターフェースを複雑化することは望ましくない。

<sup>11)</sup> ヘルプの「2-2(2). 検索処理における有効文字数」参照。

<sup>12)</sup> これらの例はいずれも読み検索では大量の検索ノイズが発生するものである。逆に「柊」であれば読みの「ひいらぎ」で検索してもほとんど検索ノイズは無いだろう。

<sup>13)</sup> 「全項目」での検索は「書名」「著者名」「出版者」「件名」「内容注記」のOR検索である。

スが問題を引き起こすことが多いだろう。ウェブを閲覧していて気になった言葉を OPAC に入れて検索してみるとはよくあることだが、ブラウザの画面からコピー&ペーストすると、しばしば該当文字列の前後にスペースが付いてくることがある。「鄧」1字をコピーしたつもりが、見えないスペースのためになぜヒットしないのか分からない、という事態になりかねないのである。

## 2.5 全項目での AND 検索

**【問題点】** 全項目検索でスペース区切りの AND 検索が使えない。

著者に「芥川龍之介」と「夏目漱石」の両方を含む資料を検索したいとする。

著者名に「芥川龍之介 夏目漱石」と入力して検索すると 45 件ヒットする。スペースで区切って複数のキーワードを入れると AND 検索になるという仕様である。ここまでは正常。

次に、著者名でなくとも、書名や注記などここに含まれていればよい、という条件で検索すべく、「全項目」に「芥川龍之介 夏目漱石」と入力して検索すると何もヒットしない。本来なら、探索する項目を増やしたのだからヒット件数も増えることが期待されるが、全く逆の結果になった。原因は分からない。

全項目検索で AND 検索を行いたければ、詳細検索において、検索条件 1、検索条件 2 それぞれを「全項目」にして、それぞれに検索語を入れなければならないようである。

## 2.6 曖昧検索

**【問題点】** 曖昧検索が期待通りに働かないことがある。

件名検索において濁点・半濁点や長音符を無視するといった曖昧検索が働くことを述べた。この機能は「件名」だけでなく「書名」「著者名」などでも働く。

ここでは、曖昧検索が期待通りに働かない例を挙げる。濁点の有無には頓着しないのだから、書名に「びぶんしょ」「ひぶんしょ」「ひぶんじょ」のどれを入れても同じものがヒットするはずで、実際そうなる。ところが、書名に「ものづくり」と入れると、「ものづくり」とは結果が異なるのだ。「ものづくり」では 21 件なのに対し、「ものづくり」「ものづくり」「ものづくり」「ものづくり」のいずれも 571 件ヒットする。この結果を見れば「づ」は「つ」とは対応せず、「ず」「す」と対応しているように見える。「づ」を「ず」に変換したうえで濁点を取り除いているのかもしれない。

## 2.7 全角／半角問題

**【問題点】** 出版年月などに全角数字が使えない。

詳細検索のオプションの「出版／発行年月」に全角数字を入力すると、

出版／発行年に使用できる文字は半角整数のみです。

などというエラーが出る。

まず「半角整数」という用語がおかしい。ここでは「半角数字」と表示すべきだろう。

しかし問題なのは全角数字を受け付けないことだ。

すべての利用者が文字の全角／半角を意識して入力し分けられるわけではない。また、自分が入力した数字が全角か半角かを目で見て確認するのは実は容易ではない<sup>14)</sup>。一方、全角で入力された数字を半角に変換することは、コンピューターにはたやすいことである。であれば、変換はシステムが自動的に行うべきである。利用者に半角入力を強いるべきでない。

一方、書名のほうは英数字を全角にしても半角にしても同じ検索結果となる。この統一性の無さが利用者の混乱を招く<sup>15)</sup>。

全角文字を受け付けない検索項目は他に「分類記号」「ISBN」「タイトルコード」がある。

## 3. 検索結果一覧

この節では、検索結果の一覧の表示について述べる。

### 3.1 検索条件の表示

#### 3.1.1 検索条件表示の分かりにくさ

**【問題点】** 検索条件の表示が分かりにくい。

たとえば「書名」に「いた」を入れて検索すると、検索条件の説明として、

「書名=いたこの言葉を含む' & 所蔵館=全て」の検索結果です。

と表示されるが、表現として雑ではないだろうか。

「いたこの言葉を含む」は恐山のイタコの言葉ではなく、キーワード「いた」を「この言葉を含む」という一致条件で、という指定を表している。それならば、たとえば

書名 = 「いた」を含む

<sup>14)</sup> 「全角」という呼び方は歴史的なものであって、現在では全角文字も全角で表示されるとは限らない。たとえば、Windows に搭載されている「MS P 明朝」「MS P ゴシック」や Mac OS に搭載されている「Osaka」は全角英数字も全角ではなくプロポーショナル（可変幅；文字ごとに固有の幅を持つ）であり、文字を扱うプロでも見分けにくい。Osaka は半角数字より全角数字のほうが幅が狭めにデザインされており、たいへん紛らわしい。千葉市図書館 OPAC の表示フォントには Osaka が指定されているため、Mac ユーザーはこのようなフォントで半角入力を強要されるのである。なお、Windows については、同様の問題がある「MS P ゴシック」で表示させたいようだが、実際には「MP P ゴシック」が指定されている。管見の限りこのようなフォントは存在しない。

<sup>15)</sup> この不統一は、〈実装上の都合〉によるのかもしれない。つまり、データベースの設計上、出版年・月は**整数型**の項目なので、全角数字を入れると数値とみなしてくれず検索が働かないが、書名は**文字列型**の項目であり、データベースエンジンが自動的に全角／半角の違いを捨象して検索してくれる、というわけである（あくまで推測だが十分あり得る）。いずれにせよ、入力を受け持つプログラムが入力値を半角化してデータベースエンジンに渡せばすむ話である。

とでも表示するほうがいいのではないか。

「所蔵館=全て」で、「全て」の右側にシングルクォートが抜けているのも気になる。

複数項目で OR 条件を使って検索するともっと分かりにくくなる。たとえば「書名」に「△△」を含むか「▲▲」を含むという検索を行うと、以下のように表示される。

```
「書名='△△この言葉を含むまたは (OR条件) '&
書名='▲▲この言葉を含む' & 所蔵館='全て」の検索結果です。
```

要するに①検索文字列、②一致の種類（中間一致、先頭一致、完全一致）、③AND/OR の三つを単純に横につなげただけなのである。

決定的におかしいのは「または (OR 条件)」の入っている位置だ。「または」は書名に関する二つの検索条件の間にあるべきなのに、一方の検索条件の表示に含まれている。では検索条件間の「&」は何を意味するのだろうか？

本来、上記の検索条件は

```
((書名に「△△」を含む) OR (書名に「▲▲」を含む)) AND (所蔵館=全て)
```

という内容であるが、現状の表示ではそれがまったく分からなくなっている。

### 3.1.2 文字化け

書名に「r★」と入れて検索してみる。既に述べたように「★」は無視されるため、「r」だけによる検索となる。この場合、字数制限により、強制的に完全一致条件となる。書名が『R』の書誌と、内容タイトルが「R」のものを含む書誌が合計7件ヒットする。

それはいいのだが、検索条件の表示がなぜか

```
「書名='Rこの言葉と一致する' & 所蔵館='全て」の検索結果です。
```

のように「R」であるべきところがキリル文字の「Я」と表示されてしまっている。他のアルファベットではこのような現象は見あたらなかった。

また、「★」を付けずに「r」だけで検索した場合は、検索結果は同じだが、検索条件の表示で「Я」に化けたりはしなかった。

なお、書名に「Яeal」と入力すれば書名に「real」を含むものが検索できるので、「R」と「Я」は便宜的に異体字のように扱われている<sup>16)</sup>ののだろう。確かにラテン文字の「A」、ギリシア文字の「Α」、キリル文字の「А」は見た目では区別が付かず、これらを異体字のように扱うことは妥

当と言える。しかし、「R」と「Я」はたまたま鏡像のような形をしているが、紛れる恐れはまず無いし、音価もまるで違う<sup>17)</sup>。これを異体字のように扱うのはやり過ぎではないだろうか。

## 4. 検索結果書誌詳細

書誌詳細の表示におかしな点はいくつかある。

### 4.1 ページ数表示

まず、ページ数のところが「ページ数 (枚数)」となっているが、これは分かりにくい。「(枚数)」は何を意味するのだろうか。ヘルプにも特に説明は無い。

また、ページ数の表示にも問題がある。例えばタイトルコード 1000800976961 の『やさしい Java 活用編 第4版』を見てみよう。ページ数の表示は「11,487p」となっている。

これは前付<sup>18)</sup>が11ページで本文が487ページという意味であろうが、カンマが位取り(3桁区切り)のカンマにしか見えないので、1万ページ超の大著かと思ってしまう<sup>19)</sup>。

ここは例えば「11p+487p」とでも表記すべきではないだろうか。「p」より「ページ」のほうが分かりやすいかもしれない。

### 4.2 ISBN 表示

「ISBN/レーベル番号」という項目がある。

適当に検索して表示させると、例えば「4-7973-7477-3」というように表示される。

一見、現行規格の13桁ISBNではなく旧タイプの10桁ISBNのように見えるが、実はそうではない。10桁ISBNとすればチェックディジット<sup>20)</sup>が合わない。この書籍の正しいISBNは「978-4-7973-7477-3」であって、画面に表示されているのは実は先頭の「978-」を省いた部分なのである。このようなISBNの表示の仕方は認められていない。誤りである。

## 5. ヘルプ

OPACのヘルプを読む利用者は実際には少ないかもしれない。しかし、辞典の凡例と同じで、きちんとした記述が求められるはずである。

<sup>16)</sup> つまり、検索上、同じ文字として扱われている。

<sup>17)</sup> キリル文字のЯはいわゆる母音字で、「ヤ」の音を表す。

<sup>18)</sup> 目次や前書など、本文より前の部分をまとめて前付(まえづけ)という。

<sup>19)</sup> 筆者は長らく、単純にデータが誤っているのだらうと思いついてきた。しかし、あまりにも数が多いうえ、いずれも5桁であるのが不自然に思われた。そこで自分の持っている本が書誌データでどう表示されているかを調べ、前付と本文のページ数をカンマで区切っていると分かった。

<sup>20)</sup> 10桁ISBNも13桁ISBNも、最後の桁は他の桁の値から計算によって決定される。これにより、どこかの桁を間違えると、計算によって異状が分かるようになっている。誤りのチェックのための桁(digit)であるためこの桁をチェックディジットと呼ぶ。

## 5.1 不親切な説明

### 5.1.1 NDC

「検索条件の表」において、「分類記号」の説明として

NDC分類による検索ができます。

とあるが、NDCが何なのか、どこにも説明がない。検索画面にもどこにも「NDC」という表記は無いのだから、ここでは

日本十進分類法の分類記号による検索ができます。

とでも書くべきではないか。

### 5.1.2 単語単位

「(2). 検索処理における有効文字数」において、「検索対象」として「文字列」と「単語単位」が挙げられているが、「単語単位」の説明がどこにもない。どこかで指定するものなのかどうかも分からない。

### 5.1.3 スペース区切りによる AND 検索

入力欄に「火星 探査」と複数のキーワードをスペース区切りで入れればそれらの AND 検索になる。効率の良い検索の要とも言える重要な機能なのに、このことが一切書かれていない。

なお、書くのであれば区切りのスペースは全角でも半角でもよいことを記述すると思う。スペース区切りで AND 検索になるソフトウェアはあまたあるが、海外製のもので半角スペースでないと区切りにならないものが稀にあるからだ。

### 5.1.4 読み検索のための文字種

「1-1.書誌検索 検索条件入力画面」には以下のような記述がある。

ヨミで検索する場合は、すべてカタカナで入力してください。(書名についてはひらがなでも検索できます。)

素直に読めば、「書名の場合は平仮名でも片仮名でも読み検索ができるが、他の項目では片仮名でなければ読み検索ができない」ように受け取れる。

類似の記述は詳細検索の画面にもある：

表記が分からないときは、すべてカタカナによる読みで入力してください。書名の読みはひらがなでも検索することができます。

しかし、「著者名」「出版者」「件名」にいろいろな言葉を片仮名と平仮名で入れてみたが、仮名の種類ではヒット件数は変わらなかった。ヘルプの文言の私の解釈が間違っ

ているのか、ヘルプが間違っているのかは分からなかった。

## 5.2 細かい文言の間違い

検索項目について、以下のように書かれている。

書名、著者名、出版社、件名、内容注記、分類記号、ISBNから検索できます。

しかし、検索画面では「出版社」ではなく「出版者」である<sup>21)</sup>。

実際にはこれらに加え、「タイトルコード」でも検索ができるが、そのことが書かれていない。「2-4.絞込み項目の指定」でもタイトルコードが抜けている。

また、「1-1.書誌検索 検索条件入力画面」には以下のように書かれているが、「検索条件3」は「検索条件5」の誤りだろう。

「検索条件3」に分類記号やISBNなどを入力するときは半角で入力してください。

## 6. データの不備

ここでは、システムの不具合でなく、データがおかしいものを少しばかり取り上げる。

### 6.1 特殊なラテン文字

【問題点】字上符や斜線などのついたラテン文字がデータとして入っていないことがある。

『Arne』というタイトルの書籍があるが、この著者名が「Bj=rnstjerne Bj=rnson」となっている。「=」は印刷業界で「ゲタ」と呼ばれる記号で、元々は活字が無かった場合に臨時に入れておくものであった<sup>22)</sup>。正しい著者名は「Bjørnstjerne Bjørnson」であり、「ø」の字が入力できなかったか、文字化けを起こしたかの理由で、「=」にされてしまったものであろう。

しかし、øはUnicodeに含まれ(U+00F8)、特別高いスキルが無くても簡単に入力できる文字である。ノーベル文学賞を受賞したノルウェーの作家、Bjørnsonの著作物を検索したくても現状ではできない。

同様に、『Bon app=tit bien s=r』という書名の書籍があるが、正しくは『Bon appétit bien sûr』であろう。また、この著者の一人が「Joel Robuchon」とあるのは、正しくは「Joël Robuchon」だろう。この著者は、他の書誌の著者名では「Jo=l Robuchon」となっている。

このような例は枚挙にいとまがない。このぶんでは、たとえラテン文字で書かれていても英語以外の言語はほとんどろくに検索ができないということになる。

ゲタが本番データに含まれていること自体、本来はあつ

<sup>21)</sup> 出版をするのは出版社(もっぱら出版事業を行う会社)とは限らない。一般企業や学協会、個人なども出版を行う。そこで「出版者」という用語がしばしば用いられる。

<sup>22)</sup> ゲタ専用の活字があったわけではない。何か適当な活字を逆さまに入れると活字のお尻(溝がある)にインキが着いてこのような形に印刷される。

てはならないことである。ただ、データの修正には多大な労力がかかるから、簡単に直せないことは理解できる<sup>23)</sup>。

## 6.2 JIS 外の漢字

**【問題点】**システム上扱える漢字が使われていない場合がある。

JIS コード (JIS X 0208 ; 第一・第二水準) に含まれない漢字で、システム上扱えるにも関わらず、他の表記に置き換えられてしまっているものがある。

### 6.2.1 𪗇

SMAP のメンバー「<sup>くさなぎつよし</sup>草𪗇剛」の著書・訳書を検索しようと「著者名」に「草𪗇剛」と入力しても何もヒットしない。データが「草【ナギ】剛」として格納されているからである。「𪗇」は JIS X 0208 には含まれないが、CP932 や Unicode には含まれる。また、JIS X 0213 では第三水準に収録されており、今日では取扱の容易な字である。

### 6.2.2 圳

中国の都市「<sup>シンセン</sup>深圳」について書かれた書籍を調べるのは困難である。書名や内容細目表<sup>24)</sup>では「深【セン】」と表記されているため、「深圳」ではヒットしない<sup>25)</sup>。「圳」も JIS X 0208 には含まれないが、CP932 や Unicode には含まれる。また、JIS X 0213 では第三水準に収録されており、取扱は容易である。

「草𪗇剛」の場合は「くさなぎつよし」と読み検索すれば済むのに対し、「しんせん」では「新鮮」「新選」「人選」「親善」などがヒットしてしまうため、「中国」などと AND 検索を行ってもうまく絞り込むことができない。

## 7. テスト用 (?) サイト

以下の URL のページは一体何であろうか？

<https://www.library.city.chiba.jp/licsxp-opac/WMnuTop.html>

千葉市図書館 OPAC と同じサーバー上にあり、一見本物に見えるが、そうではない。「武蔵野市内在住の方は」との文言が見えるので、武蔵野市立図書館用に作られたページがテストか何かのために一時的に千葉市図書館のサーバーに置かれ、放置されているのではないか。

実際、[武蔵野市立図書館のウェブサイト](#) は OPAC の URL 構成などが千葉市図書館と酷似しており、同一メーカーによるものと見られる。

本番サーバーにテスト用ページを一時的に置くことはよ

くあることだが、アクセス制限<sup>26)</sup>も無いまま (おそらく) 何年も放置されているのは如何なものだろう。

幸いというべきか、このページは少なくとも Google にはインデックス<sup>27)</sup>されていないようである。よって、Google 検索でここにたどり着くことは (恐らく) 無い。しかし、このページの HTML を見ると、head 要素に

```
<meta name="robots" content="INDEX">
```

と記述されていて、もし検索エンジンのボットがここをクロール<sup>28)</sup>したらインデックスされる可能性が高い。このようなページが検索エンジンの検索結果に表示されれば利用者の混乱を招くことは確実である。

テストページなのであれば、ここの記述は、INDEX でなく NOINDEX (検索エンジンのボットに対し、インデックスせぬことを指示) としておけばよかったと思う。

## 8. おわりに

自分が利用している千葉市図書館 OPAC の機能や仕様、表示などを調べてみたところ、問題が少なくないことが分かった。

しかし、実は千葉市図書館の OPAC が特別ダメなわけではなく、多くの公共図書館がこれと大して変わらないレベルであるようだ<sup>29)</sup>。図書館システムのメーカーは数が少ないので、そもそも OPAC の種類自体が少ないようだが。

千葉市図書館の開館時間は中央館を除いて 9:00~17:15 であり、会社勤めでは平日はまったく利用することができない。そこで、必要な資料は OPAC で予約し、休日に受け取るという利用者も多いと思われる。筆者もその一人だが、休日でも書棚をじっくり眺める余裕のある日はほとんど無い。

このような利用者にとって、OPAC で見つからない資料は無いも同然であり、OPAC が妥当な動作をし、優れた検索機能を提供してくれなければ出会えるはずの資料に出会えないのである。

なお、千葉市図書館 OPAC には、使い勝手を左右するユーザーインターフェースについての問題も大小あることがわかったが、十分な時間が取れなかったのでほとんど割愛せざるを得なかった。またの機会としたい。

## 9. 用語集

本文や脚注で詳しく述べなかつたいくつかの重要な用語

<sup>23)</sup> ちなみに日本書籍出版協会の [Books.or.jp](#) や [国立国会図書館の OPAC](#) のデータにも実は大量のゲタが含まれている。

<sup>24)</sup> 「内容細目表」もヘルプに説明が無いが、例を見る限り、例えばアンソロジーにおいて、収録作品の題名・作者等を一覧にしたものをそう呼んでいるようだ。

<sup>25)</sup> 例外は、逐次刊行物「新建築」2011年7月号および「Swimming Magazine」2011年10月号で、これらの記事名では「深圳」と表記されている。

<sup>26)</sup> パスワード認証とか、特定の IP アドレスからしか閲覧できないようにしたり、といったことを指す。

<sup>27)</sup> インデックスするとは、平たく言えば検索可能な状態にデータを用意すること。

<sup>28)</sup> ボットがクロールするとは、平たく言えばロボット (プログラム) がデータを収集に来ること。

<sup>29)</sup> アトランダムにわずか数館を簡単に調べただけなので確かなことは言えない。中には CP932 外の文字が全く扱えない OPAC もあった。

と文字コードについて簡単に説明する。

### OPAC (オーパック)

Online Public Access Catalog の頭字語で、図書館の所蔵目録をオンラインで検索できるようにしたもの。今日の多くの公共図書館では、ウェブ上で利用できる「ウェブ OPAC」が公開されている。

### JIS X 0208

文字コードの JIS (日本工業規格) はいくつかあるが、単に「JIS コード」と言えば普通はこの JIS X 0208 のことを指す。1978 年に制定され、現在までに四回改正された。

第一水準漢字 2965 字、第二水準漢字 3390 字と、仮名・英数字・記号など 256 字を含んでいる。

### CP932

JIS X 0208 の文字種を拡張する形でマイクロソフト社が定めた文字コード。「Windows-31J」とか「MS 漢字コード」などとも呼ばれる。「㊦」などの丸付き数字の類いや、「高」「崎」「鄧」「圳」といった漢字を含む。

### JIS X 0213

JIS X 0208 を拡張し、第三・第四水準漢字や多くの記号を追加したもの。

### Unicode

世界の諸言語の文字を単一の文字コードで表せるよう企図された巨大な文字コード。JIS X 0213 の漢字はすべて含む。一般に書誌情報には JIS X 0208 や CP932 に無い文字が多数含まれるので、今日の OPAC のほとんどは Unicode を採用していると思われる。

### ヒット

本報告では、検索を行った結果、検索条件に当てはまるものが見出されることを「ヒットする」と呼ぶことにする。

### 検索漏れ

検索の目的に合致しているものがヒットしないことを検索漏れと呼ぶ。英文中で「ナイフ」を検索しようとして「knife」を検索文字列としたとき、複数形の「knives」はヒットしない。これが検索漏れである。また、国語辞典を検索しようとして書名欄に「国語辞典」と入れると、『岩波国語辞典』はヒットするが、『広辞苑』はヒットしない。これも検索漏れと言っている。検索漏れの少なさを「再現率」という。

### 検索ノイズ

検索の目的に合致していないものがヒットしたとき、それを検索ノイズと呼ぶ。国語辞典を検索しようとして書名に「国語辞典」と入れると、『岩波中国語辞典』もヒットする。これが検索ノイズである。検索ノイズの少なさを「適合率」という。

### 完全一致・前方一致・中間一致

検索文字列と同じであるという検索条件を「完全一致」、検索文字列で始まるという検索条件を「前方一致」、検索文字列をどこかに含むという検索条件を「中間一致」という。千葉県図書館 OPAC では、完全一致は「この言葉と一致する」、前方一致は「この言葉で始まる」、中間一致は「この言葉を含む」と表示されている。

### 曖昧検索

ある種の違いを無視するような検索方法を曖昧検索と呼ぶ。たとえば濁点・半濁点の有無を捨象する曖昧検索では「たんす」で「ようふくだんす」もヒットする(読み検索の例)。また、小書きの仮名とそうでない仮名を同一視する曖昧検索では、「フィルム」で「フィルム」もヒットする。音引(長音符)を無視する曖昧検索では「コンピューター」で「コンピュータ」、「イスタンブール」で「イスタンブル」、「チンギス・ハン」で「チンギス・ハーン」もヒットする。曖昧検索は表記の揺れなどによる検索漏れを少なくする工夫である。しかし、「ベット」で「ベッド」が、「アピール」で「アヒル」がヒットするなど検索ノイズを増やす原因にもなる。

### 件名標目

OPAC において、各資料に付与される、資料の主題を表す文字列。予め決められた一覧の中から選ぶ。例えばカメラについての書籍には「写真機」という件名標目が付けられる。一つの資料に複数の主題があるとみなされれば複数の件名標目が付けられる。千葉県図書館 OPAC では単に「件名」と呼んでいるようだ。件名標目の付与は、内容が分かっているべきなのに、たいへんな労力がかかる(実際に付与するのは館ではなく納入業者であるらしい)。公共図書館の多くは日本図書館協会『基本件名標目表』に基づいている。国立国会図書館には独自の『国立国会図書館件名標目表』がある。