

**市民はコンテンツにたどりつけるのか？ ー検索技術・システムー**

デジタル・ネットワーク社会における出版物の利活用の推進に関する懇談会  
技術に関するワーキングチーム  
(第3回)

2010年4月27日  
常世田 良  
社団法人日本図書館協会

## 市民はコンテンツにたどりつけるのか？ —検索技術・システム—

---

### 1. 市民は何を求めているか？

- 公共図書館における利用状況  
大部分はベストセラーではなく、膨大な種々雑多（超多品種超少量）な資料  
→ 「超ロングテール」
- 図書館で貸しているのは「小説ばかり」という誤解  
70%弱は文学以外の多様な専門書、実用書、統計書、写真集、マニュアル本など  
予め書名が明らかなものはごく一部 → テーマで探す利用者が多い
- 図書館で貸しているのは「新刊書ばかり」という誤解  
80%以上は刊行後2年以上経たもの  
購入予算が削減され続けているためにこの傾向はさらに強まっている  
書店の店頭には既に存在しない（返本制度） → 知りようがない
- \* きわめて多様なニーズに対応する検索手段が必要とされる
- \* 検索の対象は当該図書館の所蔵資料ではなく、世の中に存在するコンテンツ全て
- \* 専門職（司書）の重要性 → 個別利用者のニーズに対応した検索

### 2. デジタル出版における「検索」の重要性

- 超多品種超少量で、かつ膨大なコンテンツの提供の可能性
- 効率的に検索する必要性が増大
- \* 「書名」に対するキーワード検索のみでは不十分  
書名の中にキーワードが存在しない場合が少なくない  
サブプライム → 「金融危機」  
アイスクリーム → 「ジェラート」  
誘拐 → 「かどわかし」  
盲腸 → 「虫垂炎」
- \* 現状の Web 検索では「関連付け」の取組み不十分？  
適切なソーラスを利用していない？  
数量処理では、ヒット率の偏在化を招くおそれ  
売れるものはさらに売れ、売れないものはさらに売れなくなる
- \* 全文検索では膨大にヒットし、さらに効率性が低下する可能性も

### 3. 検索システムにおける「非統制語（自然語）系」「統制語系」

#### ・ 非統制語系検索

検索システムに対して、検索システムに特化した方法ではなく、話し言葉のような、人間的に自然な文章を入力し、必要な検索結果を得るもの

例：一般的な Web ブラウザーの検索エンジン

#### 利点

特別なシソーラスの作成、維持管理が不要

一次データそのものを検索対象とすることができる

利用者も特別のスキルを必要としない

#### 欠点

検索対象が膨大な場合、ヒット数も膨大となる

キーワードが対応しなければヒットしない

そのために「ヒモ付け」等の技術開発が必要となる

それでも人間からみて不自然な検索結果が少なくない

#### ・ 統制語系検索

データベースの検索する際に、類義語・同義語、語形変化、言葉の階層などを示した検索用の限定（統制）されたキーワードを用いるもの

索引・検索の際に使う言葉の標準化を目的としている

#### 特徴

ディスクリプタ（優先語）と、それに対する非ディスクリプタ（参照語）からなる

ディスクリプタは統制語であり、検索の際にはディスクリプタを用いる

ディスクリプタの上下関係が階層として示される

#### 利点

同義語や、語形変化を集めてあるので、検索モレが生じにくい

下位概念の言葉も同時に検索できる。

優先語は文献の主たる概念について、索引者によって付与されているので、適合率が高い論理演算（ブール演算）を用いることが可能で、検索の効率化を図れる

#### 欠点

シソーラスの維持管理、拡張のため専門的作業が必要となり委員会組織が組織される

対象となるコンテンツに対して、専門化が限定（統制）されたキーワードを付与する膨大な事務量が発生する

利用者がシソーラス（統制語）の体系を理解しなければならない

- \* 非統制語系検索への極端な偏り → 非統制語系と統制語系の多様な組合せ必要  
非統制語系検索の裏で統制語系を走らせる

#### 4. 目録

図書館の目録には、資料と利用者をつなぎ付ける機能・役割がある。所蔵資料が多い場合、あるいは、ほとんどの資料が書庫に保管されている場合、資料本体にかえて、その「身代わり」である目録を利用せざるを得ない

目録は資料の「身代わり」である以上、その特徴を正確かつ簡潔に示す必要がある。さらにさまざまなポイントから検索可能なものでなければならない

目録を作成する手法には、大きく分けて記述目録法と主題目録法のふたつがある

資料のタイトル、著者、出版社、大きさなど外見上の特徴を記録する記述目録法に対して、主題目録法は、資料の内容上の特徴を、すなわち、どのようなテーマについて書かれているのかを表現する

特に、テーマ・内容からの検索には、件名標目と分類が非常に有効である

書名にはその図書のテーマを示すことばが現れることが多いが、その現れ方はさまざまであり、明確な形で現れない場合も少なくない。そのような場合でも、件名や分類のしくみを利用すれば、図書のテーマ、内容からの検索を簡便かつ効率よく行うことができる

- ・ 書名目録

- 典型的な「非統制語系」目録

- 「書名」「シリーズ名」「各巻名」「別書名」などを分かち書きし

- それぞれを50音順にインデックスを作成

- 単に与えられた書名をデータ化すればよい → 事務量少

- ・ 分類目録、件名目録（シソーラス）

- 典型的な「統制語系」目録

- 森羅万象を十進法の数字によって体系化する

- 対象コンテンツに対して人為的に分類行為を行う必要がある → 事務量大

- 体系を維持するために委員会組織が常にメンテナンスする必要がある。

- ・ 「日本十進分類（NDC）」日本図書館協会

- ・ 「基本件名表目標（BSH）」日本図書館協会

#### 件名標目数

| 項目     | 語数    |
|--------|-------|
| 標目数    | 7847  |
| 参照語数   | 2873  |
| 説明つき参照 | 93    |
| 細目     | 169   |
| 総計     | 10982 |

## 5. MARC (Machine Readable Cataloging)

- ・ コンピュータ化された目録
  - ・ コンピュータによる目録や書誌情報の流通を容易にする
  - ・ データベースによる多面的な検索が可能
  - ・ 標準化、高い信頼性、詳細なデータ
- 
- ・ JAPAN/MARC : 国立国会図書館
  - ・ TRC MARC : (株)図書館流通センター
  - ・ NS-MARC : (株)日販図書館サービス
  - ・ OPL MARC : (株)大阪屋
  - ・ 標準マーク : 書籍データセンター
  - ・ ゼロマーク : (株)トーハン
  - ・ 銘柄マーク : (株)日販
  - ・ 日書連 MARC : 日本書店商業組合連合会など
  - ・ WorldCat : 米国 OCLC
  - ・ MARC21 : 米国議会図書館、カナダ国立図書館・文書館
- 
- ・ JAPAN/MARC (日本全国書誌) の例

|          |                            |
|----------|----------------------------|
| 請求記号     | UL5-E1                     |
| タイトル     | 講座図書館の理論と実際. 第5巻           |
| タイトルよみ   | コウザ トショカン ノ リロン ト ジッサイ. 5  |
| 出版地      | 東京                         |
| 出版者      | 雄山閣出版                      |
| 出版年      | 1991.4                     |
| 形態       | 259p ; 21cm                |
| 各巻タイトル   | 情報検索 / 細野公男編               |
| 各巻タイトルよみ | ジョウホウ ケンサク                 |
| 注記       | 監修:岩猿敏生ほか                  |
| ISBN     | 4-639-01006-0              |
| ISBN     | 4-639-00968-2 (set)        |
| 入手条件・定価  | 2600 円                     |
| 全国書誌番号   | 91037678                   |
| 個人著者標目   | 細野, 公男 (1940-)    ホソノ, キミオ |
| 普通件名     | 情報検索    ジョウホウケンサク          |
|          | →: 上位語: 情報処理    ジョウホウショリ   |
|          | →: 上位語: ドキュメンテーション         |
|          | →: 下位語: データマイニング           |
|          | →: 下位語: 検索エンジン    ケンサクエンジン |
|          | →: 関連語: データベース             |

→: 関連語: セマンティックウェブ  
→: 関連語: 文献調査 || ブンケンチョウサ

|          |                |
|----------|----------------|
| NDLC     | UL5            |
| NDLC     | UL41           |
| NDC(8)   | 010.8          |
| 本文の言語コード | jpn: 日本語       |
| 発行形態コード  | 0101: 図書       |
| 出版国コード   | JP: 日本国        |
| 西暦年      | 1991           |
| 校了日      | 19950627       |
| 最終更新     | 20020624105800 |
| 書誌 ID    | 000002103043   |

#### 電子出版物 (パッケージ系)

- (1) 大宅壮一文庫創刊号コレクション [電子資料]  
: 日本の雑誌. 大正編 3 / 大宅壮一文庫編著
  - (2) 画像データ
  - (3) 東京 : 紀伊國屋書店, 2003.9
  - (4) CD-ROM2 枚 ; 12cm + 解説書(94p ; 26cm)
  - (5) システム要件: CPU i486DX2 66MHz 相当以上(Pentium 推奨). + OS Windows 95(SP1)/98/Me/NT 4.0(SP4)/2000(SP1)/XP 日本語版. + メモリ 32MB 以上. + ハードディスク 20MB 以上の空き容量. + CD-ROM ドライブ 2 倍速以上. + ディスプレイ 表示色 256 色以上(フルカラー推奨), 解像度 800×600 以上
  - (6) ISBN 4-314-90027-X : 88000 円
  - (7) NDC(9): 051(
  - (8) JP: 20447840
- 
- (1) タイトルと責任表示に関する事項
  - (2) 電子資料の特性に関する事項
  - (3) 出版・頒布等に関する事項
  - (4) 形態に関する事項
  - (5) 注記に関する事項
  - (6) 標準番号、入手条件に関する事項
  - (7) 分類 (日本十進分類法 NDC9 版)
  - (8) 日本全国書誌番号 (JP 番号)

## 6. メタデータ

メタデータとは、もともとは、データに関する（構造化された）データという意味だが、特に、ネットワーク情報資源の発信、検索、アクセス等のために作成されるデータという意味でつかわれることが多い。したがってプロトコルなどの意味で使われることも多い。ダブリンコア基本記述要素（Dublin Core Metadata Element Set）がメタデータの国際標準となっている。これは、インターネット上の資料のための書誌的事項のようなもので、核となるデータ要素によって構成されている。

メタデータは、従来の目録規則の書誌的事項より簡便で、一次情報の作成者・出版者によるメタデータ作成が可能であり、また、膨大な量のネットワーク情報資源への対応することができる。

ネットワーク情報資源特有の要素、たとえば、リンク情報、権利関係なども含む。

目録の分野では、従来の目録を統合する目録、つまり目録の目録を意味する場合が多い。

## 7. 雑誌記事索引

- ・ 非統制語系

- 「国立国会図書館雑誌記事索引」

- \* 非統制系であるためキーワードの選択がヒット率に影響する

- ・ 統制語系

- 「JOIS」JST、「医中誌データベース」医学中央雑誌刊行会

- \* 同義語、類義語も関連付けられているので検索精度が高いが  
関連付けのための膨大な事務量が発生する
    - \* キーワードの選択に専門性が必要とされる

- ・ 記事単位のデジタル販売の可能性

- 書籍の検索よりも精緻な検索システムが必要

- 全文検索では、書籍よりも効率が悪くなる可能性が高い

## 8. MLA 連携

- ・ 図書館、博物館、美術館、公文書館など知的資源所蔵機関の連携

- 第一段階 目録の連携＝標準化

- ・ 「メタ」メタデータ作成の必要性

## 9. 新しいビジネスモデル

- ・ 「品切れ」コンテンツで再販困難なもの  
→ デジタル化し公費負担により図書館から提供
- \* 米国では10年以上前から、韓国でも10年近く前からは図書館（公費負担）からのe-BOOK、電子ジャーナル、商用データベース提供は当り前の状況となっている
- \* 内容を判別できないものにエンドユーザーは直接コスト負担できない！
- \* 携帯電話キャリアなど Web 関連企業による MARC データ（書名以外）の利用開始

## 10. 専門職（司書）の役割

- ・ 自動化、AI の限界  
素人のスキルでは、高度な検索は不可能
- ・ 専門職が自動システムを駆使するメリット  
利用者の潜在的ニーズ、理解レベル、発展の可能性・方向など勘案した検索が可能
- ・ 米国では、商用データベースが図書館を介して自宅で利用可能となっているが  
専門職の高度な検索スキルを利用するために、むしろ来館者が増加している
- \* 効率性、最適解の担保

## 資料1

### 十進分類法

- 0 総記（図書館、図書、百科事典、一般論文集、逐次刊行物、団体、ジャーナリズム）
- 1 哲学（哲学、心理学、倫理学、宗教）
- 2 歴史（歴史、伝記、地理）
- 3 社会科学（政治、法律、経済、統計、社会、教育、風俗習慣、国防）
- 4 自然科学（数学、理学、医学）
- 5 技術（工学、工業、家政学）
- 6 産業（農林水産業、商業、運輸、通信）
- 7 芸術（美術、音楽、演劇、スポーツ、諸芸、娯楽）
- 8 言語
- 9 文学

### 3 社会科学

#### 300 社会科学

- 301 理論、方法論
- 302 政治・経済・社会・文化事情
- 303 参考図書（レファレンスブック）
- 304 論文集、評論集、講演集
- 305 逐次刊行物
- 306 団体
- 307 研究法、指導法、社会科学教育
- 308 叢書、全集、選集
- 309 社会思想

#### 310 政治

- 311 政治学、政治思想
  - 311.1 政治哲学
  - 311.13 政治社会学
  - 311.14 政治心理学
  - 311.15 政治道德、政治倫理
  - 311.16 政治学方法論
  - 311.19 計量政治学
  - 311.2 政治学史、政治思想史
  - 311.3 \*\*\*
  - 311.4 \*\*\*
- 312 政治史・事情
- 313 国家の形態、政治体制
- 314 議会
- 315 政党、政治結社
- 316 国家と個人・宗教・民族
- 317 行政
- 318 地方自治、地方行政
- 319 外交、国際問題

#### 320 法律

- 330 経済
- 340 財政
- 350 統計
- 360 社会
- 370 教育
- 380 風俗習慣、民俗学、民族学
- 390 国防、軍事

## 資料 2

### 医中誌データベースの検索例

検索語 → 「花粉症」

| 同義語     | 見出し語       |
|---------|------------|
| 結膜花粉症   | 結膜炎-アレルギー性 |
| 花粉症     | 花粉症        |
| スギ花粉症   | スギ花粉症      |
| 杉花粉症    | スギ花粉症      |
| ヒノキ花粉症  | ヒノキ花粉症     |
| 花粉症-ヒノキ | ヒノキ花粉症     |

### シソーラスの内部構造

#### 気道疾患

鼻疾患

鼻炎

鼻アレルギー

花粉症

#### 気道疾患

気道過敏症

鼻アレルギー

花粉症

#### 耳鼻咽喉疾患

鼻疾患

鼻炎

鼻アレルギー

花粉症

#### 免疫系疾患

アレルギー

過敏症-即時型

気道過敏症

鼻アレルギー

花粉症

#### ヒットした記事の例

鼻アレルギーの手術的治療 アレルギー性鼻炎にたいする Nd:YAG レーザー治療効果の検討

Author : 伊藤博隆(名古屋市立大学 医学部耳鼻咽喉科), 馬場駿吉, 高木一平, 鈴木元彦, 間宮紳一郎, 北尾俊二, 杉山和子

Source : 頭頸部外科(1349-581X)6 巻 2 号 Page93-97(1996.10)

論文種類 : 原著論文

Abstract : アレルギー性鼻炎にたいして ND:YAG レーザー治療を通年性、季節性のものに行い、よい効果を得たので報告する。通年性では術後 2 週目より以後には有効率 86.4%となり、季節性では飛散季節前治療で、併用薬使用群で飛散最盛期では 44.4%に症状抑制効果があった。また、併用薬を使用しなかった症例は最盛期で 85.7%に抑制し、総数では 10 症例(62.5%)に有効であった。この結果により通年性ばかりでなく花粉症にたいしても効果のあることが判明した。この治療法で特に副作用を認めなかった。(著者抄録)