

# データ検索からみた学生の自立性

井上 弘行

倉敷芸術科学大学図書館

(2002年9月30日 受理)

## はじめに

図書館にいとデータ検索は、日常茶飯事である。業務としてみれば、N I I (National Institute of Informatics:国立情報学研究所)の提供するDB検索などを駆使して、多様な利用者サービスに込えている。大学図書館では、学術研究・教育に関する資料情報を提供することが大半である。

このとき、教員ならびに院生はその道として、リソースへのアプローチには慣れたものであろう。学部の学生においては、どんな状態が待ち受けているであろうか。学習上、教員が指示したものが、すんなり見つければよい。往々にして、なかなかそうは行かない。そうしたとき、いわゆる「見つからない」現象につき当たる。「はい、ありませんでした」ではお粗末であろう。館員は、諦める前に相談を受ければ、最大限の対応をする。しかし、お呼びの掛からないことには、対処申し上げようがない。

かと言って、レファレンスの申し入れにおいて、必ずしも短時間で解決するものばかりでもない。そういう時は、検索のための猶予も欲しい。が、大抵の場合、「もう、いいです」と辞退を受ける。これには、一種の侘びしさが漂う。

どうしてそういうことになるのか。本論は、データ検索からみた学生の自立性について考察してみたい。

## 1. なぜ自立性か

図書館で学生が資料を借りるとき、カウンターに来てから学生証(図書利用カード)をおもむろに探し始める。ひどいケースでは、探しても探しても見つからず、結局のところやっとの思いで見つけた資料も借りずじまいということもないではない。

図書館に詰めていると、例えばこういうレファレンスを受ける。

「先生から『〇〇を調べて来い』と言われたのですが、どうしたらよいですか」

館員が一番泣かせられるタイプである。

質問が抽象的で具体性がなく、こちらから絞り込もうとして探りを入れようとすると、タジログ。これが求められる自立性への第1の局面である。

こんな場合、調べること自体が何であるのかが分かっていない。よって、こちらが一歩近づこうとすると、質問者も一歩下がる。結果は、堂々巡りに陥ってしまい、「それなら、もうい

いです」となってしまう。

この状態から抜け出させてやるのが、館員の教員に代わる支援であり、また成果が得られたときには、学生の自立心が芽生えてくることに繋がる瞬間であろう。

第1の局面は、調査の目的物を明確にすることであった。第2の局面は、情報は何処にあるのか。ここでの前提は図書館である。それもDBを基本とする電子情報の探索である。いまさら説明するまでもなく、DBは、一定の規則に準じて構築されている。この規則がユーザに見えて来ないのである。これを確実なものにするのが、第2の局面である。

第1に関しては、学習の途上で明確にするしかない。第2においては、高度な技法が駆使されればされるほど、利用者（学生）はその技術を磨かねばならない。それが、ここにいう自立性への課題である。

## 2. 諦めの実態

図書館の運営を預かっていると、大変、残念な事象を知ってしまう。新学期の始まりと同時に、新入生に対しての資料の貸出に応ずる「図書利用カード」を用意して待つ。大半は、新1年生の前期、後期へと時の経過と共に利用されて行き、さらに、2年、3年と学年が進むにつれ未使用の図書利用カードは減って行く。しかし、卒業時期が来ても一度も利用された形跡のないことを物語るケースも無いではない。

これは、ここでは例外扱いとさせて貰う。

問題は、図書館に利用に来て、「資料が無かった」という場合、利用者はどうしているかということである。調査によると、「諦めた」というのが、約3割もいる。あるいは、「他の図書館へ行った」とか「インターネットで調べた」とかというもあり、それは、それで自立の証であり、望ましいことである。

探しても資料が無かったなら、10人に3人は、諦めると表明しているのである。本人の熱意、性格、その日の体調など色々原因はあろう。これについては、Carol C. KUHLETHAUも、経験的に、“... but I suspect the number is substantial.”と述べている<sup>1)</sup>。

ここで、number は、断念する人の数を指している。

学生は、図書館に来て初めのうちは、どうしてよいのか分からず困惑し、そのことについても KUHLETHAU は、次のように説明している。

They were confused and disoriented, often expressing annoyance at the assignment, the library, and themselves.

彼らに探索への自信をつけてやりたいが、日常的に経験することを、さらに KUHLETHAU は、的確に触れている。

...a dip in confidence is commonly experienced once an individual has initiated a search and begins to encounter conflicting and inconsistent information.

一般的に自信の喪失を経験する、という。課題解決への途上で、必要な情報は何かということが、明確に表現できないで躊躇していることも否めないであろう。

しかし、そこには、探し方の問題も潜んでいるのではないか。以下に、その辺の事情に迫ってみたい。

### 3. 来館者の性質

大学図書館の場合、大半の利用者は学生と教員である。最近では、大学の施設も社会に開放される傾向にあり、一般の利用者も若干はある。

ここで学生としての図書館利用が、他の利用者と比較してどういう位置関係にあるかを確認しておく。表1は、利用者をカテゴリとしてみたときの特性をまとめたものである。

この表で、学生の場合における特色をみると、「特定物でものは既知」としている。こういう形で来館すれば、館員も容易に対応できる。授業をしっかりと聴いていなかったのか、授業の進め方がどうだったのかは推し量る由もないが、先のような質問が飛び込んで来るからには、館側としては平素から検索技術を磨いても、なお、追いつかない。

表1 利用者別資料検索の特性

カ テ ゴ リ	理 由 付 け	特 色	難 度
I 学生	・自由検索 ・授業で指示されたもの ・学習	不規則 特定物でものは既知 範囲をやや限定	－ ＋ ＋
II 院生	・自由検索 ・研究 ・学修	研究と同化 教員指導と自立的働きかけ 目的が明確	－ ＋＋ ＋
III 教員	・自由検索 ・研究検索	研究と趣味が重畳 ものの既知性は不詳	＋ ＋＋＋
IV 一般利用者	・自由検索	不特定対象	－
V サービス提供者	・自由対象 ・要求に応える	研修を積む 道具が揃っている	＋ ＋＋

難度：容易（－），難しい（＋），やや難しい（＋＋），非常に難しい（＋＋＋）

これは単に、DB検索の技量を言っているのではない。いわゆる高い会話術が求められる。今様に言えば、コミュニケーション能力である。その点においては、一般利用者も似た面を持っている。

ついでに触れておくと、質の面では、教員の館利用、特にデータ検索では神経を使う。高度な専門用語の上に、きわめて数少ないデータヒットを狙うには、それだけのための専門家 (searcher) が活動しているくらいであるから、通常の館員が対応するには、余りあることも確かである。

#### 4. データは何処にあるか

次にデータは何処にあるか、という視点で資料を探す方法を挙げる。実際にモノが何処にあるかを示すものであり、必ずしも利用している自館にあるとは限らない。

①館内を自由に探す …………… 目標のものは、有っても無くても差し支えないことが前提。

一般的には所蔵が期待されている。仮に見つからなくても、あるのかも知れない。

探し方が不十分ということが考えられる。

②OPAC<sup>注1</sup>を活用 …………… ①ほどではないが、同じ理由は残存する。

③友人に聴く …………… 慰めにはなっても、根本解決には至らない。

④館員に尋ねる …………… かなりの確からしさと結論づけられる。

館には所蔵していなくても出版物であれば、入手の方策は種々ある。

館外所蔵の場合、手続きが必要で若干の面倒が伴う。これを越えるモチベーションが要求される。

⑤インターネットで探す …… 館のHPから入る場合は、ほぼ②と同等である。

仮に見つかりと、④以上に、その後の手続きに面倒を伴うこともある。

⑥その他の手段 …………… テレビ、電話、PDA<sup>注2</sup>などがある。

これらは、いずれもメディア上を流れる情報であり、固定利用には向かない。

注1 OPAC (Online Public Access Catalog:オンライン閲覧目録)

注2 PDA (Personal Digital Assistant:携帯用情報機器)

自明なことではあるが、情報を求める際、ここに示した種々のいずれに頼るにしても、強い意志力が大きな比重を占めてくる。学生の場合、案外と教員の近くに答えはあるにも、質問が出ないのは畏れであり、現代風に変形した敬意であろうか。

#### 5. コミュニケーション能力

学生にとって、一番良いのは、教師から提示された資料が自学の利用館にあることである。彼らの自立心を剥ぐ元凶は、折角に閲覧を期待して入館しても、目的のものが得られなかったときである。この延長上に、その所在が研究室というケースがある。学生は、直接、研究室へは行けない。

ここで、学生の自立への道は二分される。

①仕方なく諦める

②館員に相談する———教員の了解が得られる  
———非承諾

教員が出張とか何とかで連絡がとれない、あるいは非承諾となれば、学生は勉学の意欲を失う。自立への支援という観点では、サーキュレーションを期待するケースである。

館員は、本人の期待する所に向けて可能な限りアシストする。

しかし、館の能力から限界もある。それよりも、何としても答えに近づこうとして努力の姿勢を見せることの方が大事である。逆にそういう場から逃れようとする態度や素振りを浴びせられることもある。データを求めるには、時間も要るし、話し込みも必要である。この質問の問答に耐えきれなくなって諦めてしまう学生がいる。

専門的に解決の糸口に迫るには、バーンスタインの言う精密コードが欠かせない。すると、質問者は、館員からのリアクションに追従困難となるのである。このとき、無かったら他館へ行くというケースもあるが、案外安易に諦めてしまうことも多い。

この諦めるというケースは、本人の熱意の程度もあるだろうが、教員が強く求めないのだろうかという問題も含んでいるのではないだろうか。

深田は、「…自分が生まれ育った文化から自らを切り離して行動することはできない。」と指摘している<sup>5)</sup>。資料の探索には、決して少なくない時間が取られる。その一瞬、彼らは文化を育てているのであり、体験をすることがコミュニケーションの能力向上に直結するのである。

## 6. DB検索

情報を蓄えているものをDB (Data Base) という。紙に印刷された辞書類は、記録情報ではあっても、普通、DBとは呼ばない。電子化された場合に、DBという呼称が浮上してくる。問題は、このDBの中から如何に効率よく目的とする情報を引き出すか、これである。

図1は、その評価として説明されるものである。このとき、データベースシステムの良否を測る尺度として次のものが用いられる。

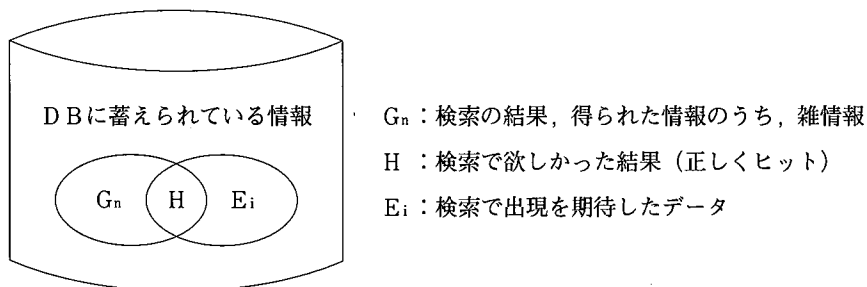


図1 検索DB情報

検索をして出てきた情報 =  $G_n + H$  ……得られた情報には、雑情報 ( $G_n$ ) と本当に求  
めていた使える情報 ( $H$ ) とが混じる

検索をして出て欲しい情報 =  $H + E_i$  ……出現を期待しても出て来ないものがある ( $E_i$ )

精度  $P^*$  (precision) は,  $P = H / (G_n + H)$  \* 適合率ともいう

出現率  $R$  (recall) は,  $R = H / (H + E_i)$

ある検索結果をこの式に当てはめてみた。その結果、学生がこのシステムを使いこなすには、  
余りに低いということが判明し、安易に諦めてしまう一因にもつながっているのでは、という  
ことが見えてきた。適切なアドバイザーが欠かせない場面と言える。

$$P = H / (G_n + H)$$

$$= 7 / 627$$

$$= 0.01$$

出現率においては、通常、 $E_i$  は利用者側からは正確なところを掴めない。レファレンスの  
応対で最も悩まされる局面である。

## 7. DB 検索の比較

データベースの使いやすさというものが、あるであろうか。ある組織に所属していると、そ  
の機構が提供するシステムに唯一、依存する傾向が出てしまうのはいたしかたないことであろ  
う。そこには、限定されたリソースにしか手が届かないというジレンマがつきまとう。その意  
味においても図書館を利用する際には、館員の助言を求めることは、多くの有効な手段となる。

インターネットがあるとは言え、ここでは、限られたシステムに依存すると何が起きるか、  
いくつかのDBを比較してみて、学生の自立性向上への影響点を指摘してみたい。

表2は、提供者別のDB検索結果である。ここでは、検索結果の効率は問わない。また、D  
Bの母数も不明である。しかし、先の測度の1つである精度においては、比較値として簡単に  
求められる。

表2 提供機関別検索結果

提 供 機 関	件数	マッチ	精度	注 釈
国立国会図書館 (Web-OPAC)	1043	186	0.93	精度は200件で推定
倉敷市 (中央館)	1773	29	0.15	精度は200件で推定
NACSIS Webcat	658	182	0.91	精度は200件で推定
大学図書館	627	7	0.01	某大学
業者提供	900	900	1.00	

与えた検索用のタームは、「マナー」で、結果は、表2の内容が得られた。最初の3機関に

対する精度は、件数が余りに多く、先頭からの200件で推定した。業者提供のシステムにおいては、いずれの資料にも「マナー」が含まれていることを、全件確認した。表2に示した値は、DBの更新など随時変動が加わり、あくまで一過性の比較測定値に過ぎない。本論での推考には、大きな影響はないとみる。

問題は、大学図書館の結果である。倉敷市の場合も似たような値を示しているが、こちらは、館種が公共図書館である。

表3は、大学図書館と業者提供の検索結果の提示された最初の5件を参考に挙げたものである。前者をA、後者をBとする。

表3 システムによる入手情報の相違

No	A: 大学図書館 (整列されて提示)	B: 業者提供
1	13歳論: 子どもと大人の「境界」はどこにあるのか	楽しい生活とマナー
2	22世紀への手紙: 生命・情報・夢	これが正しい「お祝い事の」マナーです
3	30のキーワードで <u>学ぶ</u> 現代経営	これが正しい「お見舞い・お悔やみ」のマナーです
4	70の格言から <u>学ぶ</u> 現代ストックマーケットの法則	婚約・結婚親の心得とマナー
5	Braque	先生のちょっとした気づき&マナー

この2つを比較して歴然とするのは、Bの方には、検索キーとして与えた「マナー」という語が、悉く含まれている。他方、Aには全く見あたらない。それどころか、下線で示したように「学ぶ」が見えている。Noの1,2,5には、直接タイトルには含まれていないが、編著者名とか件名とかのキーに一致すると、このような形で出てくる。

例えば、著者名に「山中智恵」というのが含まれていたとする。その読みの一部には、「マナ」が潜み、これは検索対象に選ばれてしまう。

また、タイトルの一部に「さまざまなカタチと変化」のような文字列を含むと、これも対象になってしまう。

テキスト検索では、BF (Brute-Force) アルゴリズムが原始的なものとして知られているが、そのようなことが起こるのは、「マナー」は「マナ」とみなす曖昧検索の手法をシステムが採用しているからである。それは、データを逃さないという見地からは、1つのアプローチではある。これが、果たして学生に受け入れられるルールであろうか。自立心を大事に育成して行くためには、むしろ弊害にさえる疎外性を孕んではないだろうか。

図書館で「諦める」ということは、いろいろな意味付けができるであろう。その中で、安易に諦めてしまうことのない選択肢を可能な限り提供し続けるには、検索結果Bを応答するシステムではいけないのか。強い疑念を残すところである。

事情が分かっておれば、ファセットA群を#1として、

#1 not 学 not au=まな not さまざま

でもって、より厳しい条件式を与えるのも一法である。不要情報をかなりの確度で削ぎ落とす

ことに成功する（35件迄低減）。あるいは、タイトルのみを検索キーに切り替えても改善は見込める。そういうことが学部生、しかも専門学科に所属しない一般学生が思いつくかは、甚だ怪しいものである。

ただし、そのような式を与えることは、必要なデータをも逃してしまう危険性もある。

そのようなことよりも、検索結果としてAのような状態が得られた段階で、即「無い」と判断されてしまわないだろうか。表2のような効率のよいシステムもあれば、日常、業務としている館員でさえ、あきらめのムードに襲われても不思議ではない。

## 8. 自立心を脅かす曖昧検索

曖昧検索が、なぜ自立心を脅かすか。

これまでに見てきた検索語、「マナー」は、「マナ」「マーナ」のいずれにもヒットしてしまう。そこで、DBのデータに次のような文字列（読み）が出現すると検索に選ばれる結果となる。

タイトル	「学ぶ」「さまざまな形」「眼差し」「まな弟子」「真名本」
著者名	「山中智恵」「本間長世」「山本学」
件名（内容）	「山梨県」

これらは、その読みの一部にいずれも「マナ」がある。同音異義語でその結果、表2のような耐え難い精度値となった。

これは、一律的に決めつけられるものでもない。

探索語に「ガラス」を選んでみよう。こちらでは、精度は0.84を得ている。これぐらいであると、まずまずであろう。しかし、これとて残る0.16には、例えば『はるかなるわがガラスカル』がサーチされてくるのには、滑稽でさえある。

それにしても、学生に安心して使わせるには、商用向きのように精度1.0は、この上ない理想である。何故なら、彼らの検索に向けられる時間は、教員・館員ほどにはなく極端に限られる。もとより、情報科学として検索論が、基礎講座にカリキュラムされており、こうした課題に触れる機会があれば、なお好ましいことであろう。

システムとの対話は、相手が機械とは言え、コミュニケーションをとる一種のスキルが要求される。それは、主張スキルであり、会話スキルであり、質問スキルである。特に最後の質問スキルは、システムのクエリーとしての癖をある程度予測可能とする技術である。

それには、関連語をはじめ異義語・同義語、シソーラス、さらには論理演算等に関する知識が必要なことは言うまでもない。ところが、作業の途中で脱線することもある。システムに対する知識のギャップである。しかも、自ら探し求める情報に対する整理の無さも影響する。こうした事情を KUHLTHAU は、次のように捉えている<sup>1)</sup>。



... an inability to express precisely what information is needed makes communication awkward between the user and the systems.

これは、利用者側の切実な問題であり、システム設計者、システム仲介者（館員、指導者）に強いインパクトを与える課題でもあろう。

以上に見てきたものは、学生が在籍中に修得して行くものと期待する。しかしながら、学習だけに情熱を傾けられるのは一部の学生かも知れない。残る大半の者は、ちょっとした躓きで進路を見失ってしまう危険な成長期であるということにも、留意すべきであろう。

### おわりに

成長期にある学生が、提供されたシステムに左右される。本人たちは、このことに案外気づいていないであろう。

曖昧検索は、そうした人とシステムとの親疎性の問題でもある。この辺りを議論するには、もっと多くの事例を集めることが欠かせない。

多くのOPACが、ネット上から利用できる今日、身近なシステムが自由にアクセスできない仕組みを未だ残しているのには、実に残念である。

当拙論では、いくつかのデータを集約し、館員の眼で観た学生の図書資料へのアプローチと学生の自立性との相関を追究してみた。

実際には、検索した結果から目標の資料に辿り着き、それをさらに閲覧する作業が待っている。見つけた資料は、その先を目指して読解行（reading, understanding, and doing）へと進んで、新たな指標の生産に結びつくことを期待する。

データ検索においては、“Hang in there!”である。

### 引用・参考文献

- 1) Carol C. KUHLTHAU, The Information Search Process(ISP) : A Search for Meaning Rather than Answers; 電子ネットワーク環境下における研究者・学習者の情報利用行動に関する実証的研究, PP.1-17, 国立情報学研究所, 2001
- 2) 市村, 長谷川, 渡部, 佐藤: テキストマイニング事例紹介, 人工知能学会誌, vol16, No2, PP.192-98, 2001
- 3) 河野, 川原: Web検索におけるテキストマイニング, 人工知能学会誌, vol16, No2, PP.212-17, 2001
- 4) 小野欽司他著, 『情報学シリーズ3 情報学とは何か』, 丸善, 2000年
- 5) 深田博己著, 『コミュニケーション心理学』, P144, 北大路書房, 2000年
- 6) 図書館情報管理システム『リメディアオ 図書検索 4』, (株)リコー, 1999年
- 7) 神門典子他著, 学術情報センター編, 『全文検索 技術と応用』, 丸善, 1998年
- 8) Information retrieval: data structures and algorithms, edited by William B. Frakes, Richard Baeza-Yates, PP.219-40, Prentice Hall, 1992
- 9) 社団法人情報科学技術協会編, 『新サーチャー入門』, 日外アソシエーツ, 1992年
- 10) 高橋恒介著, 電子情報通信学会編著, 『テキスト検索プロセッサ』, PP.21-58, オービット, 1991年
- 11) 千賀正之著, 『検索入門書』, 日本図書館協会, 1991年
- 12) 斉藤孝・佐野眞・甲斐静子著, 『文献を探すための本』, 日本エディタースクール出版部, 1989年
- 13) 有川, 篠原: 文字列パターン照合アルゴリズム, コンピュータソフトウェア, Vol4, No2, PP.2-8, 1987

## Students Puzzled with Ambiguous Search Systems

Hiroyuki INOUE

*Library,*

*Kurashiki University of Science and the Arts,*

*2640 Nishinoura, Tsurajima-cho, Kurashiki-shi, Okayama 712-8505, Japan*

(Received September 30, 2002)

Today, OPAC systems are available for use at each library. They are very useful or not so, because the systems do not always indicate high precisions.

We have tried to search materials on internet DB systems. The used DB systems are as follows:

Web-OPAC(National Diet Library)	0.93
Public Library(Kurasiki-city)	0.15
NACSIS Webcat	0.91
University Library	0.01
Commerce DB	1.00

Their precisions are indicated above.

A permitted key word is “マナー.” That can be pronounced [mana:] in Japanese, and the meaning is ‘manner’ in English. Japanese is written in an ideogram. So, the “マナー” is classified into same group of “学ぶ [manabu],” “眼差し [manazashi],” and “真名本 [manabon].” They have a common pattern of ‘mana’ in the phonetic sign. If a search system works by an ambiguous search method, hit data probably include uncertain noises.

An example of the lowest precision is that in University Library. It is a problem that many students are made to use the system for their exercises. We would like to emphasize the importance of this problem. We know that they will stop searching, and simply begin to think “our searching objects are not found out.”

After they express such annoyance, they have a suspicion about the system in the end. On just this time, they are puzzled with an ambiguous search system, and the growth of their mind may be slowed down temporarily.

Librarians in general feel the mood through their references, and say, “a high precision system such as commerce DB(cf. prec.;1.00) should be installed for the students.” We support their opinion, too.