

# 問題群データベースのコンセプトと応用

平田 周

## Question-Clusters-Based Database Its concept and applications

Schu Hirata

### Abstract

Burst of information in quantity has brought a paradox that people have more difficulties to find out necessary information. For the solution, many studies on the information retrieval have been conducted and not a few tools for searching have been developed. But many have experiences that they could hardly pin-point particular information they required. This paper, however, does not propose a new retrieval technology. We have observed that most of non-specialists, particularly Japanese are not well-aware of their problems they want to solve, more precisely, they can not fully define what they want to know. It will be useful if databases are provided with a structure quite different from the existing ones searched by use of key words. We called the new type database as "question-clusters-based database".

### 1. はじめに

情報量は指数関数的に増加を続けている。情報量の増大は、かえって必要とする情報を見出すことを困難にするというパラドックスを感じる。そのために、コンピュータのための検索エンジンなどが開発されているが、そのほとんどはキーワードの入力によって必要な文献や記事を引き出すという方法である。しかし、キーワードをいくつかクロスさせて選べる情報の内容が求めていたものと大きく違うという失望は誰しもが経験するところである。その原因については、いろいろな研究もなされている [文献1]。

しかし、専門家は別にして、一般人がデータベース検索で求めるピンポイントな情報が得られないのは、検索技術のほかに、当人がほんとうに自分が知りたいことの定義ができていないことにも一因があるのではないかと考えた。これはあたかも、調べる課題が十分明確にされていない状態で大きな図書館に入って本を探すのに似る。経験によれば、とりわけ日本人は、このような問題定義が下手であり、それがデータベースの利用を阻む検索技術以前の問題として存在するのではないかと考えるのである。

### 2. 日本人と問題定義

日本人は、専門家をのぞき、一般に自分の抱えている問題について、明確な定義をしていないのが普通であると感じる。的確な問題定義なくして、情報検索により自分が必要とする情報が得られないということに異議はないであろう。しからば、なぜ日本人が一般に問題定義をしないうまま問題意識を抱えているのか。それにはいろいろな説明が可能である

うが、国民の意識や知的レベルの水準がかなり普遍的で、ことさら自分の問題を明確に定義せずとも普段の生活、あるいはビジネス活動においても困らないことにあるのではない。いうなれば、他人が抱く問題は多くの場合自分の問題でもある。また、不用意に自分が直面している問題を他人に語ることは、自分自身の弱点を見せることにもなる。そのような日常生活は、次第に自分にとっての問題を定義できるほどには明確に意識しないという態度を植えつけていく。“問題定義”という言葉自体がすでに日本語的感覚のものではない。

中小企業事業団は、「スーパーネット」という名称（元スマイルス）のデータベースを開発・運用している。しかし、キーワードを入力して欲しい情報を検索させるこのデータベースに対する評判はきわめて悪く、肝心の中堅・中小企業関係者がほとんど使っていないのが実状である。それには検索端末の利用環境の問題やデータベースにのせられているデータや情報にも問題は多い。しかし、実際に使用される状況を見て、とても実用的でないことを実感する。

その根本にあるのは、すでに述べたように、これら利用者が、自分自身が知りたいことの問題意識が曖昧であることに原因があるように思われる。専門家であれば、このような問題はあまりなく、表示された情報の利用価値についても即座に評価できる。それは、情報を検索する前提に、当然調べることの目的が明確にされているからにほかならない。

### 3. 検索された情報の意味づけ

緑川信之（図書館情報大学）は、同じ情報を見てもその人の知識状況によって反応が異なるということを指摘し、そこから発展させて、人々が情報検索で得た情報は、その人が次の発想や新たな情報を求める動機づけとなる「情報触媒」であるという興味深い論を述べている [文献2]。たとえば、富士山の高さを検索で求め、3,776メートルであることを知って満足するのは、その高さを知る必要性が別の次元にあるとする。それは、原稿を書いてその数字が必要になった場合もあり、アジアで最も高い山を知ることが課題なのかもしれない。最初に得た「富士山は3,776メートル」という情報は、それが一つの“触媒効果”として次の求める情報へ到達する役割をもっているのだと考えるのである。

このような形式での情報検索は、コンピュータによる場合、ハイパーテキスト型の検索はきわめて有効である。しかし、一般に、情報を検索しながら思考を続けるという環境は普通のビジネス関係者などにはあまりない。自分が抱く問題に対し直接的に的確な解決を与えてくれる情報を求めるものである。

一つの情報を見て、人はさまざまに反応し、それはその人の検索前の状態において持っている知識レベルにかかわるとすることは正しいが、それは欲しい情報を得るための方法を説明するものではない。もし、検索された情報がたとえ有用なものであったとしても、それを読む人の知識レベルによって、価値のあるものにもなれば、ならないものでもありうるとなると、データベースの構造や検索方法とは異なる領域の問題であろう。

また、例に挙げた「富士山の高さ」といった問題であれば、得られる情報は完結型であり、満足は100%達成される。しかし、多くの問題は、完結型のものではなく、価値的にも価値有り無しの二極的なものではなく、傾斜的価値を有する。たとえば、「2000年の経済

情勢はどうか」というような問題の場合、読み手によってさらに内容が分析され、評価されることになる。

だが、多くの非専門家的なユーザーの場合、得られるものが、自分の求めている解答にできるかぎり至近距離にあることを望む。ここに、データベース利用に対する不満感の源があり、わが国でデータベース利用が活発化しない原因があるのではなからうか。

#### 4. 問題群データベースの考え方

かつてアメリカに Metamorph という変わった検索ソフトがあった。自然語で自分の問題を質問形式で記述して入力すると、データベースの中からその解答にあたる部分が表示されるものであった。このソフトを使っているとされていた Boeing 社では、着陸時に飛行機の車輪が出ないという異常状態に直面した場合、「着陸準備に入ったが車輪が出ない。どうすればよいか」と入力すれば、マニュアルの中からその解答となるべき部分がモニター画面に映し出されるというような使い方がされているという話であった。詳細にチェックしたわけではないが、質問文にあるいくつかのキーワードに加え、疑問詞、あるいは動詞を識別してデータベースの中から必要な情報を選んでくるという原理だと見受けられた。

しかし、この場合も、用いられる複数のキーワードの間の関係について記述されるという点では新さがあるものの、基本的にはキーワード検索であって、検索によって表示される情報は必ずしもユーザーを満足させるものにならない場合もあった。

通常、データベースは、ユーザーがはっきりとした問題を意識して検索に向かっていることを前提としている。しかし、問題定義が十分になされないまま、かつ表示された情報を触媒としてではなく、問題解決に役立つ完結型の情報を求めるとなれば、その利用価値としてユーザーを満足させることは難しい。

そこで考えられたのが「問題群データベース」である。ユーザーは、まず自分の知りたいと思っていること的前提となる該当項目を「問題群」の中から選ぶ。中堅・中小企業に質問して集めた経営的諸問題は 120ばかりにしかならなかった。これらは類似的なものの調整など作業段階であり、紙数の制約もあるので具体的紹介は割愛せざるをえないが、その中の 1 つを使って、問題群データベースのコンセプトを説明してみることにした。

まず、大きくカテゴリーに分けられた「問題群」の中から、自分が抱えている問題に相当するものを選択する。次いで、それに関連するサブ的な属性（例えば、この問題についてのユーザーの知識水準、日本、アメリカ、世界といった地域的な条件、発表された年代など）について該当する項目を選ぶ。その結果、必要な情報が選ばれて表示されるという構造のものである。

たとえば、「今年の経済がどうなるかを知りたい」ということが問題である場合、まず [経済] というカテゴリー中の問題群を表示させ、その中から「今年の経済がどうなるかを知りたい」という項目を探し、それに [初級] [日本] [1999年以降] と入力すれば、関連するレポートや雑誌記事、あるいは参考図書が表示される。表示されるものは、経済企画庁の見通し、銀行やシンクタンクが出している定期刊行物の中にあるこの問題に関する論文や解説、総合雑誌や経済雑誌などに掲載された記事、あるいは新聞の記事などである。日本の経済に関する OECD や経済誌の記事などは、[初級] では表示されない。多くの

文献や記事情報を表示するよりは、問題への的確な解答になるとと思われるものを少数示すほうがよい。[上級]になれば、より基礎的データや、さまざまな論評の背景になる諸情報を提示することになる。

## 5. 問題群データベースの構築

本来データベースは、無限の情報源を意識し、その中からSGMLやXMLなどで標準化された文書情報、あるいは情報に含まれる検索用キーワードによって、コンピュータが自動的に選びだしてくるということを理想としている。その意味からすれば、「問題群データベース」は、このような理想から本質的にかげ離れている。なぜなら、集められた情報は有限的なものであり、データベースに組み入れられる前に、すでに想定される問題との関連において価値評価がなされているからである。

このような前処理作業は、無限の情報を対象にはできない。このことは、知り及ばない膨大な情報から必要なものを選び出すことを求める専門家には満足できないことは当然であろう。しかし、データベースの利用者の圧倒的多数は、多すぎる検出表示を好まず、自分がさらにその中から最も価値ある情報を選び出すという作業を避けたいと思っている。

ビジネスにおいては、専門的なサーチャーを利用するというケースも少なくない。またコンピュータで同様の役割を果たすエージェントなどの考えも発展しつつある[文献3]。この問題群データベースは、いわばサーチャーの代行機能という意義を持つものとも考えることも可能であろう。問題群データベースの構築者、あるいはデータ更新を行う人は、通常のデータベースの検索を駆使して、その中から問題の解答に当たるものを選ぶことが作業として求められるからである。

まず、必要な情報を自らデータベースで選ぶという習慣があまりない中堅・中小企業経営者・幹部のためのデータベースとして、構築を試みている。さらには、初等教育から高等教育までの学生、あるいは一般人を個別的对象とする問題群データベースがつけられることの意義は大きいのではないかと思われる。

### 参考文献

- [1] 日本図書館情報学会研究委員会編「情報検索の理論と実際」論集・図書館情報学研究の歩み 第19集 石川徹也「情報検索システムに関する研究課題」日外アソシエーツ 1999 p107-125
- [2] 緑川信之著「情報検索の考え方」 図書館・情報メディア双書 6 勉誠出版 1999 p66-118
- [3] David Ellis 原著 「情報検索論」細野公男監訳 斎藤泰則他訳 丸善 1994 p101-130
- [4] M・バックランド著 「図書館・情報サービスの理論」高山正也訳 勁草書房 1990 p152-154