

コンピュータ利用の可能性と問題点

長 沢 洋

近年、歴史資料保存利用機関でのコンピュータ導入が増えつつあり、史料の整理と利用にコンピュータを活用する際の諸問題についても、実践例に基づいた議論がなされるに至っている。当館においても現在コンピュータの導入を検討しつつあるが、実現にはまだしばらくの時間が必要な状況にある。

コンピュータの導入についての利点は多々あるが、この紙面では検索について（特に、不特定の利用者のための検索手段構築に問題を絞って）考えるところを述べてみたい。

文書館が、一般利用者にとって利用価値のあるデータベース・検索体系を作ろうとした場合、困難な問題は二つあると考える。

ひとつは、検索されるであろう事柄が時代と分野を問わず、しかも検索の対象となる文書群が中世から近代・現代にいたる様々な性格のものであるという点である。地域を対象にした文書館の場合、これはごくあたりまえのことではある。

ごく大まかに言うならば、従来のコンピュータ利用は、比較的

コンピュータ利用の可能性と問題点（長沢）

均質な資料や情報をデータベース化する試みを中心として進められてきた。そこでは意識するとしないとにかかわらず、検索者が検索対象に一定の予備知識を持っていることが前提とされていたように思われる。

たとえば、ある史料をテキストとしてその全文を入力してデータベースとするような場合、あるいは同一範疇に属する多量の情報をデータベースとする場合などにおいては、その史料を利用することの意味や、データベースを検索することのメリットを、ある程度利用者自身が理解していることが必要条件とならざるをえない。言い換えれば、その史料・データベースから自分の知りたい情報が得られる人（あるいは、得られるかもしれないと考える人）にしか利用価値がないのである。

あまりにも当然のことだと言われるかもしれない。しかし、データベースの利用についての知識を持っている人はそう多くはなく、文書館の利用者の中には、史料操作に未経験なこともあって、史料に基づいて歴史に関する情報を得ようと考えても、どの史料を見ればよいのか分からない人たちも多い。

また、史料の利用についてその要求も多様化している。利用者が文書館に求めるのは、地域の歴史と史料に関するさまざまな情報であり、それはあらゆる分野・時代にわたらざるをえない。今のところ、これに応えようとするには、文書館職員の個人的な知

識と史料研究によるほかはないが、そのような方法・手段の限界性は明らかであり、いずれ文書館収蔵史料の全体にわたる多角的・総合的な検索手段の構築が求められねばならないであろう。

現時点において、かかる検索手段構築の目論見が、いささか非現実的なものに見えるであろうことは否定しない。史料の検索にコンピュータを導入する試みの順序としては、右述のように、特定のテキストを入力したり、個別の文書群や書誌情報などをデータベース化したりするところから始めるのが穏当であることも理解できる。当館でも、現実的にはそのようなところから始めなければならないであろう。

しかし、収蔵史料の全体を対象にした検索が可能か否かは、我々が日常の業務の中で確かに感じている問題である（他館でも同じではないだろうか）。当館では、多様な史料の目録群とでも言うべき二冊の複製資料目録を作成したが、これは所蔵者の所在地域別に構成されており、索引は所蔵者名索引があるのみというものである。したがって、この目録を活用できるのは、「自分はそのような史料を見ればよいのか」が分かっている利用者だけであって、館の職員でさえも、これをもとにして自在に史料を探せるだけの経験を持っているとは限らないのである。

この問題は、結局、検索語彙（キーワード）の問題に行き着かざるをえないだろう。右述のような検索手段を構築しようとする

ならば、あらゆる検索主題に対応できるように検索語彙を付けるか、あるいは検索者を有効な検索語彙に導くような仕掛けを作らなければならない。しかし、資料に与えられる情報のうちでも、キーワードというのは整理者の恣意に最も左右されやすい部分である。整理者各自が詳細に付与しようとすればするほど、結果として、利用に耐えない無秩序な検索体系がで上がる可能性が大きい。膨大な検索語彙に体系的な秩序を与える作業はやはり不可欠であろう（ただし、これが検索語彙そのものの統一化・規格化に直ちに結びつくかどうかは、また別問題であると考えている）。いずれにせよ、試行錯誤のための相当な時間と手間が必要なことである。

困難な問題の二つめは、文書群の階層構造に関わる問題である。周知のように、文書整理と目録作成の理論として、近年特に、組織体の構造に基づいた文書群の階層構造の再構成という考えが主張されている。検索手段の構築にあたって、この考えを生かすことは自と求められる。これはもちろん文書群自体が階層構造を持つからであるが、それと同時に、検索する側の求める情報にも深浅の階層とでも言うべきものがあるからである。

たとえば、文書の解題集を通覧するだけでこと足る場合、シリーズレベルに付与された情報を検索するのが一番合理的である

場合、最小単位の文書一点一点を対象に検索しなければならない場合など、さまざまである（実際は、これらを総合した多角的な検索がもっとも現実的であろう）。どのような場合でも常に文書一点ごとに検索され、その結果も一点ずつの文書リストが出力されるのでは、かえって不便である。特に、入力されたデータの数が多ければ多いほど、この不便さは増してゆかざるをえない。

たとえ、検索という局面に限定するにせよ、文書群の構造を捨象して一点ごとの文書を並列に扱うならば、個々の文書の位置付けは不明確になり、必要な情報にたどり着くことは結果として困難になるものと思われる。複数の文書群にわたる検索であれば、特に、この難点は避けがたい。

とはいえ、時代も性格も異なる種々様々な、しかも各々独自の階層構造を持つ文書群を、ひとまとめにしてデータベース化し、多角的な検索手段を作り上げるのはきわめて難事である。当館の場合、図書館等との複合施設であるために既設の汎用コンピュータを導入する可能性もあるが、拙速を避け、可能な限り多様な文書群を整理しながら、データベース化のための仮説を練っていかなければならない状況にある。

（ながさわひろし 研究員）