

基調講演 講師紹介

横山伊徳

所属

東京大学史料編纂所情報処理主幹

1984年から『維新史料綱要』データベースの開発に着手.

1997年から2年間情報処理副主幹.

SHIPS for Internet を加藤友康主幹（当時，現在所長）の指揮下にて開発する.

2001年4月から現職.

最近の悩みは，心ときめくPCやソフトに出会わないこと，
ネットワーク性善説信奉者だが，偽悪者として振舞わねば
ならない場面がふえたこと.

歴史情報の共有化 その現状と課題

横山伊徳(東京大学史料編纂所)

はじめに

学術研究に関する蓄積された英知を学術情報というのにならって、私は歴史研究に関する英知を歴史情報と呼びたいと思う。歴史学(あるいは古文献学)の研究者は、古い文献の読解を続けその成果を蓄積してきたばかりか、研究対象である史料そのものを目録や史料集として蓄積してきた。こうした研究情報や史料情報は、出版物として形をなしたものを含め、有形無形の膨大な歴史情報を形成している。

一部の歴史学(史料学)研究者たちは、すでにこれらの歴史情報を広く多面的に活用し、もって歴史学研究の推進を図りたいと考えてきた。技術の進歩を積極的に活かし、歴史学における史料研究の過程を分析し、この過程に適合的な電子技術を用い、歴史情報研究を推し進めてきている。この研究の結果、今日では、私の仕事場である東京大学史料編纂所をはじめ、日本史研究のための歴史情報は相当程度インターネットを通じて利用可能となり、歴史学研究に多くの貢献をおこなっている。今回はこうした現状に至る経緯を整理し、今後の歴史情報研究の展望を試みたいと考える。

起点

日本において文献や資料を電子計算機を用いて整理・蓄積し、検索に役立てるという試み(ここでは歴史史料における数値データに着目し、これを電子計算機の数値処理機能を利用して処理を行なうという研究は一応除外する)は意外と古い。坂野正高『現代外交の分析』(東大出版会、1971年)は、つぎのように紹介する。

外務省がIR(情報検索)システムに手をつけたのは、在外公館からの電信や省内の各種情報ファイルなど、主要な原資料いっさいをコンピュータで集中管理するという事だった。三十八(1963)年に中型電子計算機……によるウォール・ナット方式の実験に着手した。この方式は、マイクロフィルムとコンピュータを組み合わせたもので、文献や資料はマイクロ化しておきコンピュータで求める資料のフィルム番号をはじき出すシステムである。このためあらかじめ資料に対して、内容を的確に表わす、と思われるキーワードを索引としてふっておき、フィルム番号といっしょにコンピューターに記憶させておく。……

外務省に遅れること20年、初めて漢字の扱える8ビットパーソナル・コンピュータを手にして、電子計算機に漢字が入力できることを確認したとき、思い描いたが理想のシステム・モデルは、まさにこうしたIR(information retrieval)システムであった。

従来カードを用いてきた索引作成などの作業を、電子計算機システムに置き換えることを目指した。カード・システムを電子計算機に置き換えるという発想とそのシステムは、歴史学研究者にもわかりやすく、急速に普及した。たとえば、史料編纂所は影写本という原史料にできるだけ忠実な写本を所蔵しており、日本古代・中世史研究の基本史料としてその中核的史料群となっている。この史料の一点目録は、全体が膨大なだけにとても出版できないと予想されるものであったが、電子計算機と表形式を用いたRDB(relational database)を用いることによりいつでも必要な目録だけを検索して抽出できるというようなシステムが構想された。

歴史情報システムの隘路とハイパーテキスト(リンク)

いわばカード置き換え型の歴史情報処理のシステムは、いくつかの隘路を抱えていることが数年ほどして理解されてきた。これは一言で言えば、歴史史料が大量でかつ雑多である、ということに起因する。たとえばある史料で有効な項目設計も、別の史料で有効である保証はない。そのため、ひところ

は目録（テーブル）の項目（フィールド）設計ばかりが議論されていた。その一方で多くの研究者が多少の無理は承知で特定のフィールドにデータを押し込むということを余儀なくされていた。

もうひとつの隘路は、データ作成の見通しとの関係である。数千件のデータなら個人の作業でも数日で終了しある完結したデータを提供できるが、万単位十万単位になると事実上未完成の作業となってしまう。すると、データをもれなく検索するというデータベースとしての意味が半減してしまう。すると、大量で雑多であることを前提にシステムを考え直す、ということが必要となった。

結論を先取りすれば、ハイパーテキストという考え方の登場である。すでに情報科学の方々には了解されているであろうが、マウスの発明者として有名エンゲルバートは、多数の頭脳労働者の間でオンラインの情報を共有する「コミュニティ」という概念と、その上で情報を交換する仕組み（構造化ハイパーテキスト）を60年代に考案した。この後に「ハイパーテキスト」の命名者ネルソンが登場し、彼の考えたXanaduというシステムは、「世界中のあらゆる文献、図表、データからなる世界的規模のネットワークで、何億というユーザが同時に利用することを目的としている」システムで、「分散化されたネットワークを稼働するためのソフトウェア」により「すべてのサーバに保存されたドキュメントやファイルの情報が、いろいろな種類のリンクを通じて、各種のサーバ上にあるデータと結合できる」という大構想であった。こうして、正規化されたデータではなくてもテキスト記述を可能にすること、テキストとテキストの関係をリンクで明示できること、そしてテキスト記述への検索システムという手法をとること、などが隘路を切り開く突破口となった。

一方、90年ころからインターネットが実験的ネットワークの役割を終え、本格的なワイド・エリア・ネットワークとして利用されるようになった。インターネットとハイパーテキストとは本来別概念であるが、インターネットはハイパーテキストという仕組みを利用する（完全であるかどうか別にして）ことになった。インターネットを基幹とするシステムは、歴史情報に適合的な仕掛けであることが了解されることとなった。このことへの理解を進めたのは、重点領域研究『沖繩の歴史情報研究』（代表：岩崎宏之）の存在が大きい。

インターネット技術の展開＝歴史情報のマルチメディア化

インターネットを前提とすることで、生み出されてきた新しい歴史情報のリソースは、画像情報である。特に、全体としてネットワーク共有を最大限に活かすためには、大量のあるまともをもった画像群を一挙に（あるいは一貫した手法で継続的に）デジタル化し、格納＝共有するためのサーバを用意し、画像サーバとのリンク情報を付与することにより、テキスト検索の結果を踏まえ関連する画像閲覧を行なうことが必要である。つまり、史料の構造を反映する画像情報群のメタ・データの管理として、テキストによる歴史情報の管理を活用しよう、というわけである。

こうした画像サーバの現実的な有効性としては、史料保存と利用の両立があげられる。現在ではこうした用途のためにマイクロフィルムが積極的に利用されているが、このメディアは、画像の検索がしにくいというデメリットを抱えている。しかし、画像サーバの形をとれば、画像の管理データに画像のアノテーションをつけることで、検索の効率は一挙に向上するし、負担も軽くてすむ。

拡散するコミュニティ・協働するコミュニティ

歴史情報は物理的にはTCP/IPとハイパーリンクによって結びついている。こうして結びついた歴史情報は、その作成主体のコミュニティにおいてインターネット技術によって各人から利用できる。ここで着目しておきたいのは、このコミュニティは、それを実現したインターネットの技術によって、外部のネットワーク（とそれにつながるPC）とつながっていることである。日本の前近代史研究者集団というコミュニティ（が仮に存在するとして）へと拡散することができる。ひとつのコミュニティ内のデータベースによって管理される歴史情報だけではなく、コミュニティを超えて誰もが、リソース・ロケーションを示せばそのまま歴史情報へリンクを形成できるという点こそ注目すべきである。

このように一つの歴史情報のリソースがさまざまに形を変え、あるいはいくつか絡まりあって別のリソースを形成するようになる。これは、言い方をかえれば、研究者（集団）、図書館・文書館、大学

などの各コミュニティが標準的なしくみで立ち上がって、コミュニティとコミュニティが拡散しながら融合していくことを意味する。

ネットワークとして存在し、非求心的な拡散構造をもつインターコミュニティな社会が生まれつつあるもう一つの動きは、「全体として有機的に関連した情報を構築」する動きである。現実にはリソースリストと（自動巡回）検索エンジンが存在する。それ以外に、ILS 所在情報サービスがあるといわれている。歴史的に見ると、80年代アメリカの地質調査所 USGS が取り組んだ地球環境変化研究計画 Global Change Research Program がそのプロトタイプを作ったという。すなわち、地球環境変化の調査のために必要とされる情報資源は、地球観測衛星から、自然史博物館、文書館、図書館、岩石園から昆虫コレクションまでひろがっている。この研究計画では、制限のない調査、自由な（無料の）情報流通、という原則を確立しながら、それらリソースのありかを示す Locator 機能が考案された。これらの考え方やプログラムが、90年代前半に行政管理庁や NARA(国立公文書館)などで受容されて、90年代後半にはガーバメンタル（アメリカ政府に特化した場合、汎用の場合はグローバルという場合もあるもよう）・インフォメーション・ロケータ・サービス GILS として研究開発と法制化が進んだ。

この考え方は、歴史学の分野にも応用できよう。日本（世界）中にある日本史史料、文書、関連情報などのリソース・ロケータ（ヒストリカル・インフォメーション・ロケータ・サービス/HILS）を構築しようという試みがあり得る。GILS は、先ずコミュニティのもつリソースがネットワークへリンクされてくること自体を促すこと、そして、デジタル化されていないリソースのデジタル化に際して、そのコミュニティのプロフェッショナルたちの智慧を活かすこと、をガイドラインとして掲げています。つまり、融合するインターコミュニティの形成とリソース・ロケータの構築とは車の両輪なのです。