

ハイパーテキストによる歴史資料の情報化

横山 伊徳：東京大学史料編纂所

1. はじめに 標準化とエンドユーザ指向

歴史資料を情報化する、もっと単純に言えば(歴史)資料をコンピュータで処理するという作業は、日本においても決してまあたらしいものではない。それは既に1960年代から行なわれてきた。

しかし、1996年における「歴史資料の情報化」は、60年代のそれとは質的に異なる。つまり、単にファイル容量が増えたとか、検索スピードが実用的になった、という量的な問題に還元できるものではない。

私は、この差異の基本にあるものは、一つは標準化であり、もう一つはエンドユーザ(分散)指向であると思う。この二つの傾向が、ネットワークにおいて、また、テキスト処理においてほぼ同時に現れたことにより、90年代後半において、歴史資料の情報化は新しい段階に入ったのである。

2. インターネット

1) ネットワーク接続の標準化 TCP/IP

私は、コンピュータサイエンスの課程教育はほとんど受けてきていない。つねにOJTで切り抜けてきた。このような私にとって80年代は、コンピュータとコンピュータをつなぐネットワークのことは、ほとんどメーカー任せの状態であった。それだけメーカーの独自性がつよくあらわれる世界がネットワークの世界であった。

私がネットワークを意識し始めたのは、90年代初頭に勤務先にTCP/IPによるネットワークが敷設され、とにかくMSDOSマシンとこのネットワークをつながなければならない、という羽目に陥ってからである。その時は、大学だからTCP/IPというぐらいの認識で、また、当時の議論でもTCP/IPはネットワーク標準になるという見通しには慎重論が多かったように思う。

TCP/IP成立の経緯についての議論はひとまず措くとして、TCP/IPがいったん標準として確立してしまうと、はじめに採用されたUNIXは勿論のこと、PCや最近ではメインフレームもTCP/IPへの対応を遂げてしまった。つまり、PCであれ、ワークステーションであれ、メインフレームであれ、ネットワークに関係するときは、同じ理屈に従うことになった。スピードや機能では他の上位のコンピュータに劣るかもしれないが、ネットワークへの関与という点では、PCで十分対応できるようになったわけである。しかも、だれしものが実感しているように、スピードや機能においてもPCの飛躍的な進歩によりある程度カバーされるようになってきた。

2) 情報発信単位としてのPC Windows95/MacによるTCP/IP標準サポート

とはいえ、PCからのTCP/IP利用は、到底素人のなせる世界ではなかった。極端な情報不足、不親切な設定プログラムなどなど。恐らく、90年代初頭は、OSやマシンのメーカーも、TCP/IPを一部のマ

ニア的研究者の tool ぐらいにしか位置づけていなかったのではなからうか？

しかし、現在では主要な PC 用の OS が TCP/IP を簡単に利用できるように変化した。PC のプリンタ設定程度の手軽さで TCP/IP を設定できるようになった。この結果、ごく普通のユーザの PC から実現することとなった。

個人による publish の可能性

80 年代にも PC から外部のネットワークを利用する方法はなかったわけではない。いわゆるパソコン通信である。しかし、これはどこかに大きなコンピュータを置き、そこに PC から情報を入手するという、どちらかといえば、今までのマスコミュニケーションと共通する情報の流れを形成するものである。

しかし、この新しいネットワークは(原理的には)PCからの情報発信を可能にした。おそらく近代におけるマスコミの自由の確立過程には、個人や小集団から公衆へ publish する回路(ピラ・チラシなど)を抑圧する過程を伴っていたが、今度は少々様子が異なってくる。

例えばこうである。私は現在主として 19 世紀オランダ語を利用した沖縄研究を進めている。ここで蓄積されたオランダ語史料は、将来この分野を研究する場合、くり返し利用される可能性が高い。日本語資料ならば、何らかの形で(日本で)出版するという(奇特的な)出版社も現われよう。しかし、オランダ語ではその可能性は絶無である。それならば、インターネットでこれらのオランダ語史料を public にすれば、実質的に出版と同じ効果が得られるわけである。

3. ハイパーテキスト

最近の特徴がハードウェアにおける標準化とエンドユーザ指向だとすれば、ソフトウェアにおけるこれら二つの動向の(同時的な)現象が、ハイパーテキスト HTML という考え方である。

1)ハイパーテキストの定義

コンピュータに保存されているテキストのある部分を、他の部分とリンクする機能
非シーケンシャルなライティング

ハイパーテキストは、原理としてはやさしい。私達がよく見慣れたところでは、いわゆる「見よ」(see, Cf.)に当たる。通常の文献では、「見よ」という指示があったとしても、それを実際見るには、別のページをめくったり、別の文献を探しに図書館に出かけなければ、「見る」ことはできない。これをその場で直ちに「見る」ようにしたものがハイパーテキストである、と考えられる。

search から navigate へ

このような関連付けが実現したシステムを想定すると、それは、従来のコンピュータによるデータ処理とはかなり異なったシステムとなる。ここでいう従来型の処理とは、知りたい言葉を指定して、コンピュータに蓄積された多くの情報から情報を絞り込んで、必要な情報に到達するという過程をさしている。

このような絞り込んでいく検索方法の最大の特徴は、一つ一つのレコードは独立している、という点である。例えば、検索の結果二つのデータがヒットしたとしよう。しかし、この二つのデータがどう関連するのか、それを見つけたすのは、データを分析するがわにかかっている。

ある情報に到達するのに、このように絞り込んでいく方法と、逆に芋蔓式にあちこちたどっていく内に、必要な情報に到達するという方法がある。ハイパーテキストはどちらかといえばこの後者の方

法を実現したものである。だから、情報は拡散していくのであり、目的の情報に到達できないまま「迷子」になる可能性をはらんでいる（故に「(ホームページに)戻る」というボタンが不可欠)

2)ファイル) /リンク/ボタン

以上から了解されるように、ハイパーテキストにとって、一つ一つのデータ(ファイル)と、そのデータ同士を関係付けるリンクと、そのリンクを表示するボタンの三者が重要な役割を果たしている。

具体的に述べれば、データファイルは、文章テキストそのものである。最近では文章だけでなく、画像や音声、動画などが取り扱えるようになってきている。リンクはソフトウェアによって実現されるが、リンクによって二つのデータファイルが結び付けられることになる。ボタンは、そのリンクを表示するものであり、リンクの種類によってボタンも変化する。

3)HTMLによるハイパーテキストの標準化

インターネットはその通信方法を標準化することによって広く普及した。その結果、インターネットメールやテルネット、FTPなどのネットワーク利用方法も、多くは標準化したもの、共通した仕組みのものとなった。

インターネットで利用されるハイパーテキストも、標準化されている。この結果、送り出すがわのコンピュータやソフトが何であれ、また受取る側のコンピュータやソフトが何であっても、ネットワークをいったりきたりするデータの形式は同じ、ということになった。

そして、更に特筆すべきなのは、そのデータの形式が、特別なソフトウェアなしで記述できることである。この形式を取極めたものがHTMLである。これによって、いざとなれば、OSだけでハイパーテキストを記述することができるようになったのである。実際は便利なソフトが出まわっているが、それらを使っても、ファイルを保存する際には、非常に単純なプレーンテキストとして保存される。

4)ハイパーテキストの可能性

史料集

問題は、歴史学にとってどのようなデータが、ハイパーテキストにふさわしいか?ということを見極めることである。史料集における傍注や参照注のようなものは、ハイパーテキストになじみやすいことは了解されよう。また、史料集の目次も、目次自体がハイパーテキストになっていけば、目次から本文へという流れは、ハイパーリンク以外の何者でもない。しかし、例えば、史料集において、ある言葉から必要な史料を探し出すような「索引」的な機能は、今までのsearch型検索が必要であろう。だが、全体が完結していて索引切出しも終わっているような史料集では、逆にその索引をハイパーテキスト化すれば、索引から本文へという流れも、ハイパーリンクである。

構造記述型文書目録

私が史料集と共にハイパーテキストにむいていると考えているのが、文書目録である。従来文書目録こそRDBの独壇場と考えられてきたが、私はこの考えに最近では直ちには賛成しがたいものを感じている。その最大の理由は、文書目録の記述方法の深化である。詳しく述べる機会は別に譲りたいが、文書は一つ一つが単独で孤立して存在することはない。ある文書は別の文書とともにこよりに綴じられていたり、また、同じ性格の帳簿が何年分も経年的に残っていたりする。それらを一点の文書または帳簿を一データとして独立して取り扱うのは、たとえデータベースでも正しくないのではなからうか?。そのようなデータ相互の関連を含んだ目録記述(これを仮に構造記述型文書目録と呼んでおく)が、現在国際的に標準化されようとしている。この目録標準化には相当程度ハイパーテキストの

考え方が反映していると思われる。

4 . ハイパー『琉球外国関係文書』

1) 史料の選定

2) プレーンテキストの入力

Anonymous FTP との共存

3) ファイルの分割とリンクの設定

原秩序重視

論理的再構築

4) FTP 転送

このレポートは、1996年12月9日に開催された総括班研究会(会場：富士フイルム本社ホール)での報告を再録した。4 . の部分については「WWW サーバーによる『琉球外国関係文書』の公開」(『人文学と情報処理』15号、1997年)を参照のこと。