

研究班報告 3 Media Studies Group

学術文献の電子化を巡って

岩橋 俊哉

I. 電子図書館と電子化された文献の特徴

電子図書館のビジョンは Ted Nelson の Xanadu Project に遡ることができる (URL=<http://www.sfc.keio.ac.jp/~ted/>)。彼は、ネット上での引用の形式ハイパーリンクを提唱し、全世界の図書がひとつの図書館の中で互いに引用され合う仮想の電子図書館という壮大なモデルを描いてみせた。その中には既に著作権を守る方法や、それに関連した利用料金等の徴収方法までが含まれている。そして今や、彼のビジョンは、現在インターネット上で徐々に実現されつつある。

それでは、電子図書館の前提となる電子化された情報はどのような特徴を持っているのか。まず最初に電子化された文書等の情報の特徴を書籍と比較してみることにしたい。

特長

1. 保管場所がわずかで済むこと。

書籍の形態ならば、大型の本数十冊で構成されるような百科辞典ほどの情報量であっても、それが電子化された場合にはコンパクトディスク数枚に収まるほどになる。

2. コンピュータが使えることで情報の検索性が高くなること。

特定の語句や、いい回しなどを調べるのが非常に容易になる。コンピュータを用いた検索では、文書内のあらゆる文字を検索の対象にできる。

3. 出版や公開に伴う配布の手間の低減。

電子化された情報は、インターネットなどの電子的通信手段によって、配布可能なことから、配布の時間も費用もほとんどかからなくなる。

問題点

1. 文書の電子化には手間がかかること。

紙に書かれた情報の電子化は自動的にはできない。この作業は、校正など手作業を含んだ手間のかかる作業である。

2. 電子化された文書を閲覧するにはコンピューターが必須であること。

データ本体は、場所をとらないが、記録されたメディアのみではその内容を閲覧するこ

とはできない。この点は書籍に比べて閲覧の手軽さがだいぶ劣る。

3. データがすべて読めなくなってしまうようなデータ消失が生じやすいこと。

これもデータの記録方式による問題だが、一部でも欠けると、データ全体の繋がりがとぎれてしまい、極端な場合、メディア内の情報がすべて読めなくなってしまうようなことが起きる可能性が高い。

4. 文字情報が文字コードとしては電子化できない資料があること。

これはコンピュータ上で現在用いられる文字コードは、すべての文字を網羅していないことによる。文字コードとして変換できない場合には、文字情報を画像データとして蓄積することで代替できるが、画像になった文字は、文字としては検索ができないという問題が残る。この問題に対処するため、世界中のさまざまな文字データを無償で配布している「今昔文字鏡」の活動 (<http://www.mojikyo.gr.jp/html/index.html>)、その文字をすべて用いることができる OS である BTRON 超漢字 (URL=<http://www.tron.org/>) などもあるが、今後の普及が待たれるところである。

5. 著作物の原本を守る方法が確立されていないこと。

これはもっとも重要な問題のひとつであろう。デジタルデータの性質上、完全なコピーが作れることから、データそのものからはどれがオリジナルでどれが複製かを判別することは原理上はできないため、現在の所確立された方法はない。

このような電子化文書が収められた電子図書館は、ネットワーク上に作られた仮想的な図書館である。我々はそこに外向くことなく、コンピューター上で、電子化された書籍を閲覧したり、借り出したり (=複製を作る) することができる。しかも、電子図書館では同時に複数の人が同じ文書を閲覧できる。これは特に希少本などに関して重要な点であろう。

実際に実現されている例として、ネットライブラリー (URL=<http://www.netlibrary.com/>) と青空文庫 (URL=<http://www.aozora.gr.jp/>) のふたつを挙げておきたい。これらのサイトにアクセスすることで、我々はそこに収められた文書をコンピューターの画面上で、あるいは印刷して紙の上で読むことができる。

電子化された文書が揃っているならば、それを収蔵する電子図書館のシステムは、実は容易に構築することができる。建物についてはまったく不要で、文書を収める記録装置があれば良い。インターネットのサイトを立ち上げて、ホームページを作り、そこからこれらの文書を閲覧できるように設定するだけなのである。

II. 文献を電子化して公開するために必要な作業

すでに書籍の形態で存在するものを電子化する場合には、1) スキャナーで文献を画像情報に変換し、画像ファイルを作成し、2) OCR (光学式文字読取装置) で文字情報を文字コードに変換した後、3) 何らかの書式によってレイアウトを整える、という手順になる。この場合、紙に記録されたレイアウトができるだけ再現できる書式が好ましい。この作業で一番の問題は、手順2である。ここで文字の校正が必要になるのだが、この部分は現在の機器の性能では、100%の精度はないので人が最終的にはチェックしなくてはならない。

なお新規に作成する場合には、最初にワードプロセッサで作成してあると、容易にテキストファイルあるいは特定のファイル書式に変換できるので、特に校正の手間が省ける。

次に、ファイル書式 (=フォーマット) の選択肢としては、以下のものが挙げられる。

a. テキストファイル

あらゆるコンピューターで扱える、もっとも汎用性が高い書式であるが、この書式は文字情報のみで、レイアウト情報がまったく含まれないという問題がある。なお文字の検索は、レイアウトとは関係なく可能なので、ワードプロセッサ等で読み込むことなどで、まったく遜色なく行える。

b. HTML(Hyper Text Markup Language) Tim Berners-Leeが、ブラウザーのアイディ

アと共に、インターネット上で機種を越えて文書を共有するための書式としてまとめたデータ書式。現在インターネットで標準の書式となっているので、ブラウザー以外には特別なアプリケーションは不要で、汎用性も高い。また見出しなどを指定するためのタグも含めてファイル全体がテキストファイルであることから、テキスト本文のみに関してならばブラウザー以外のアプリケーションでも読むことができる。ただしレイアウト関連の規定は非常に少ないため、閲覧の条件によっては見た目が変わってくるという問題がある。

c. PDF(Portable Document Format)書式

Adobe社 (URL=<http://www.adobe.com/jp/products/acrobat/adobe.pdf.html>) が開発、配布している書式。電子的に文書を配布するというを前提に設計されたデータ書式である。コンピュータの機種やOSに限らず同じレイアウトで文書を見ることができ、さらに画面でも紙でも同等なレイアウトで表示ができることを目標としている。また、文書の配布を容易にするために、文書を作成するアプリケーション Acrobat と、作成されたデータを閲覧・印刷するためのアプリケーション Acrobat Reader を分けて、Readerの方は、非常にゆるい条件で無償配布している。Acrobatの方は有償ではあるが、個人の手が届く価格に設定されている。

その他にHTMLがもとにした仕様で、1986年に国際規格となったSGML (Standard Generalized Markup Language) や、Donald Knuthによって1970年頃から開発された印刷システムであるTeX、Leslie LamportによるTeXのマクロパッケージLaTeX、エキスパンドブック (Expanded Book) 書式 (URL=<http://www.voyager.com/jp/eb.html>) などの選択肢もあるが、一般的なユーザーがそれを用いるという点で今一步の所にある。

なおワードプロセッサの文書ファイルは、作成や編集の容易さなどの長所もあるが、その書式が公開されていないものが多いため、データを閲覧できるソフトウェアが限定されること、閲覧や印刷だけの場合でも、原則としてそのアプリケーションのパッケージを購入する必要があることなどの問題があり、電子化のための書式としては不適だと考えられる。

これらをまとめると、書式が公開され普及しているとはいえ一企業の開発した書式であるため、将来に渡っても確実に標準の書式であり続けるかどうかは確実ではないが、現時点では、配布書式としては、PDFを選択するのが望ましいと思われる。

Ⅲ. まとめと結論

電子化された文書は、特にコストに関して、データの保管、配布などの面で、紙の文書に比べて非常に有利である。特に学術文献については、対象となる読者の数が限られることから、配布コストを下げるためには有効な方

法だと考えられる。ただし、データの性質上、著作権保護の問題、記録書式の問題などに注意を払う必要があるが。

文書を電子化する書式としては、現時点では、PDFとHTMLというふたつの書式で作成し、保管、配布するのが現実的な選択肢だろう。理由は、PDFに関しては、現在、もっとも理想的な書式であるが、今後とも標準であり続けるかどうかということについてやや疑問があり、その問題を回避するために今後も生き残る可能性がもっとも高いHTML版も作成しておくことが良いと考えられるからである。

研究班報告 4 Global Studies Group

エリツィン時代の終焉

— 1998年8月金融危機 —

内田 健二

今年度、わが研究班は班員それぞれが鋭意研究を進めながらも、班としては統一テーマのもとに集団的な研究に取り組むことができなかった。そこで、私が1999年11月17日に本研究所の定例研究会で行った報告の一部を以下に要約することで、班としての報告の責めを塞ぎたいと思う。なお、これは私が本大学の海外研修制度により1998年5月から1年間モスクワに滞在した際の見聞をまとめたものであり、エリツィン辞任など、最新の動向には触れていないことをお断りしておきたい。

はじめに：1998年金融危機

ロシア政府と中央銀行は1998年8月17日、ルーブリ切り下げ、民間対外債務の3ヵ月間モラトリアム、短期国債の償還繰り延べを内容とする共同声明を発表した。当初、多くの国民はこの声明がもつ意味を計りかねていたようであったが、その後、銀行預金者の預金保障措置が発表されるにおよび、人々は生活防衛に向けて行動を開始した。両替所には市民がドルを買い求めて長い行列をつくり、ルーブリは政府の予測を越えて急激に下落した。

しかも8月末、穀物収穫量が大幅に落ち込むとの見通しが発表された。市民は食料品の買いだめに走り、麦粉や食用油など多くの食料品が、一時的ではあれ店頭から姿を消した。

消費物資の豊富なモスクワは、物資の流出を阻止するための検問を開始した。クラスノヤルスクでは一種の配給制が試みられた。自由市場経済への移行をめざした諸政策の失敗が、如実に示された瞬間であった。

この金融危機は第一に、エリツィン時代の終焉を意味した。エリツィン自らルーブリ切り下げの可能性を否定したわずか4日後に切り下げが実施されたことは、彼がもはやリーダーシップを喪失したことを物語る。

そもそもエリツィンのリーダーシップは投機的リーダーシップの性格を強くもつ。それは「敵」との闘争を通じて初めて人々の支持を獲得する。しかしそれは、新たな政策の立案と実施に関してはオポチュニズムに陥りがちになる。ソ連時代、共産党の指導性は人事(ノメンクラトゥーラ制度)によって担保されていたが、彼はこの伝統を色濃く引き継いでいる。チェルノミルディン以下、ステパーシンに至る4人の首相の突然の解任は、命令と首のすげかえが彼の指導の中心的方法であることを改めて示した。こうしたリーダーシップこそが政治の攪乱要因となった。

金融危機の第二の意味は、IMF路線の破綻が明らかになったことである。金融危機を経て、西側世界はロシア経済とどのように向き合うかという論点が改めて浮上した。これ