



電子図書館の正しい概念を持つ

長尾 真

京都大学

電子図書館が関心を集めているが、その構築には種々の困難を伴っている。それらの課題を指摘し、解決法について提案をする。

電子図書館に何を入れるか

電子図書館はこれまでの図書館の単なる電子化というレベルで考えてはならない。これまでの図書館の持つ情報を電子化するのは当然として、さらに電子図書館ならではのことを実現しなければならない。図書館は何十年、何百年という長期にわたって人類の知的資産を蓄積して利用に供するところであるから、かなり保守的な考え方で長期的展望にたつて、電子図書館の構築を考えざるを得ない。

図書館は書籍と雑誌を中心として収集を行っているが、他にも地図等の図面、楽譜、音楽や映像のテープ、その他いわば何でも収集の対象としているといつてよい。

電子図書館においても、それは当然であるが、書籍と雑誌・資料が最も大切な対象となることは間違いない。したがって、電子図書館化の第一歩は、図書カードの電子化である。雑誌については目次、すなわち各巻号に含まれる論文のタイトルと著者名のデータを電子化することである。これはOPAC (On-line Public Access Catalogue) として、通常の図書館においてもオンラインで公開することが前提となる。しかし、これだけのことすら、日本の図書館で完全になされているところはほとんどないというのが現状である。

書物の解体と自由な再構築

電子図書館に書物を記憶すべきことは当然である。しかしそこには電子図書館でなければならない工夫をすべきである。従来の図書館では蓄積と取り出しの単位は冊子であった。電子図書館においては、1冊の書物の任意の部分を取り出せるので、それを実現すべきである。目次をたよりに第1章第j節だけを取り出したり、第n頁だけ、第m図だけ、あるいは、あるキーワードを含む頁をすべて取り出すといったことが自由にできるように、構造化して記憶しておく必要がある。

電子図書館に記憶されている書籍、雑誌その他をハイパーテキスト的に関係させて検索出力できる機能を持たせ、利用者は自分の考えにしたがって関連情報を再構築することが自由にできる必要がある。これは書物の解体である。著者の考え方で書物を押しつけられるのではなく、利用者の立場から必要な情報を利用できるという、従来とはまったく逆の書物に対する考え方が電子図書館によって可能となる。

強力な検索機能

電子図書館の1つの特徴は、種々の観点からかなり自由な検索ができることであろう。OPACはもちろんのこと、目次をデータベース化し、目次の持つ階層性(章・節・項などの)をうまく利用した階層構造検索(詳しくは、長尾 真: 電子図書館, 岩波書店,

1994を参照)や、キーワードによる本の索引からの検索、全文検索、ハイパーテキスト構造を電子図書館に作っておき、これをたどることによる検索など、種々の検索機能を用意することが必要である。特にソースを必要に応じて利用できることと、日本語の世界だけでなく、検索要求は英語など他言語にも自動的に翻訳されて、電子図書館における英語やその他の言語の世界についても日本語の世界と同じように検索できることが必要である。

電子読書機能

従来の図書館は本や雑誌を貸し出したら、それがどのように利用されるかに関心を持つことはなかった。電子図書館においては読書端末のソフトウェア機能としてどのようなものを用意するかが大切であって、これは利用者ではなく電子図書館側の責任である。

たとえば、表示文字の大きさを変えられる、縦横どちらの表示もできる、音声合成ソフトを動かして自動朗読を聞くことができる、必要な部分を機械翻訳システムで翻訳して読める、メモを自由に記入できる、カット・アンド・ペーストができる、辞書を簡単に引ける、関連する資料を同時に見ながら読める、などといった読書に伴って必要となる機能をいろいろと持たせることが必要である。

解決すべき課題

最も大切なことは著作権のある出版

物をどのようにして電子図書館に入れて利用者に提供するかという問題であろう。これは、利用した量に応じた料金を支払うことによってしか解決の道はないだろう。電子図書館の端末装置のあるところまで来て読み、コピーを取らない場合には無料とし、コンピュータに取り込んで読むとき、またこれを再利用する時などには適当な料金を払うのである。

過去の出版物はOCRで文字コード化して電子図書館に入れねばならないが、これからの出版物は出版社と交渉して電子版（フロッピーなど）を入手

するようにし、イメージ入力などはないようにすることが適切な態度であろう。そのためのコストが膨大であるのに比べて、入力された情報の利用効率はよくないからである。それだけの資金があつたらOPACの完全なものを作ったり、目次の入力などに使った方がはるかに有効である。

他にも論ずべきことは多いが、紙数が制限されているので、その他の問題については小著電子図書館、その他を参照していただきたい。

(1998.12.21)

な情報は、たとえばインターネット検索エンジンの検索ディレクトリのように、目的とするコンテンツへのより効率的なアクセスや、検索結果の柔軟なナビゲーションを可能にする。ひよつとすると、司書に代わってこのような情報を専門に提供する職種が誕生するかもしれない。

今年からスタートする米国のDLI (Digital Library Initiative) の第2フェーズ⁴⁾でも、医療分野のエキスパートの情報検索過程を他者が参考にできる技術や、参考司書 (reference librarian) の機能を実現するソフトウェア・エージェントの技術を含んだプロジェクトが採択されている。電子図書館の蓄積する知的遺産を有効に利用する技術や事例が、データの格納や検索のための基盤技術をリードする状態に入りつつあるといえよう。

最後に著作権のある出版物、あるいは電子図書館の商業的な側面であるが、これについては切実に解決策を待ち望んでいる。せつかくの電子図書館なのに、施設に接続された端末からしか特定の出版物が閲覧できないのでは技術的な利点が生かせない。たとえば、電子図書館が本来サービスすべきユーザには電子的な利用カードを発行し、年間のユーザ当たりの有料コンテンツの利用上限額を設定しつつ、予算の範囲で有料コンテンツの利用料をアクセス回数による従量制で支払うといったモデルは実現できないだろうか。もちろん制限のないコンテンツには誰でもアクセスできるようにしておくこと、電子透かしのように不正コピーを検出する手段を実現すること、個人の認証、デジタル・コンテンツを紙のコンテンツよりも安価で魅力的なものにするといったことも必要であろう。ACMは会員向けに同学会の過去8年間の20万ページ以上の出版物を自由にダウンロードできるサービスを年間\$86で提供しはじめた。このようなビジネスモデルを順次確立して堅実な電子図書館の運営につなげるべきであろう。

参考文献

- 1) 長尾 真: 電子図書館, 岩波科学ライブラリー-15 (1994).
- 2) 杉本重雄: デジタル図書館, コンピュータソフトウェア (16,1), pp.57-63 (1999).

誰のための電子図書館か

武田 浩一

日本アイ・ビー・エム (株)

インターネットを利用して3億ページともいわれる世界中のWWWページにアクセスできる今日では、一般ユーザが電子図書館という概念^{1), 2)}を理解することは容易であろう。すでに存在する、このような膨大なコンテンツの多くは限られた期間や限られた読者にとってのみ意味があるものや、果てはまったく信憑性のない情報であったりするかもしれない。ただ、長尾先生のおっしゃるような知的遺産としてのコンテンツ以外にも、ニュースや書評、トレンドを反映した評論や解説といったダイナミックな情報は我々の日常生活に欠かせないものであり、両者が別々にアクセスされるのではなく、関連付けられて利用できることが重要であろう。このような可用性の立場から考えると、小中学生にとっての電子図書館、高齢者にとっての電子図書館、研究者にとっての電子図書館といった、コンテンツとサービスの両面からより多様で実用的な電子図書館の概念が考えられるのではないだろうか。OPACのようなデータ共有のための基本的な情報が不足しているのは深刻な問題であるが、それらの技術的な要素

を包括したサービスの次元での議論も大いに期待したい。

書物が解体され、より高度な検索や新たな情報の創造に利用できるという考え方は、電子図書館のポテンシャルを強く感じさせる。仮に人間の認知能力や創作能力が電子図書館を利用することによって飛躍的に向上するとしたら、電子図書館の情報インフラとしての社会的/文化的な貢献は計り知れない。ただし、書物の解体による情報の粒度の変化は、検索結果の増加や不均質化、検索式の複雑化や類似度判定の精度劣化につながる可能性がある。このようなユーザと検索したいコンテンツとの「インピーダンス不整合」ともいえる問題は、従来の情報検索やデータベースの分野だけでなく、情報可視化や認知科学の成果も積極的に取り込まなければ解決しないであろう。ハイパーリンクによる情報の関連付けの機能は、XLink³⁾のようなより強力な記述形式によって一層高められ、メタデータ (データに関するデータ) とともに、コンテンツの分類、統合、再利用といったより高度な目的に利用されるようになるであろう。このような豊富

3) W3C Working Draft on XML Linking Language.
http://www.w3.org/TR/WD-xlink

4) Digital Libraries Initiative Phase 2,
http://www.dli2.nsf.gov
(1999.1.29)

電子コンテンツ形成と活用の実践を

安達 淳

学術情報センター研究開発部

電子図書館に入るもの

電子図書館の議論では、対象として想定するコンテンツの範囲を確認しておかないと、往々にして議論が噛み合わなくなる。長尾先生の問題提起では、従来から本や雑誌のような形で形成されてきた情報の電子化、利用、蓄積、保存といったことを論点としていると受け止めた。一方、まったく新たに電子化情報を作るところから考えると、違ったアプローチも出てくる。私がここで申し上げる意見も、基本的には後者の見方を併せ持っている。

図書館の現状

出版はグーテンベルク以来の長い歴史を持ち、情報の組織化として社会に根付いている。そして出版資料は、図書館という組織とそこで働く図書館員という paraprofessional に支えられて社会システムとして機能してきた。しかし数十年後にはこれらの機能がほとんどすべてコンピュータに置き換えられると確実に予想できる。したがって、情報工学の研究がこのような変貌にどのように寄与し、積極的に働きかけていけるかが今の課題といえる。図書目録のようなメタデータの作成に関しては、我が国の図書館では随分進んできているが、雑誌記事の目次にまでは電子化が及んでいないのが現状のようだ。学術雑誌などで組織的に行われている程度である。また、図書館で電子化されている情報を一般市民がインターネットから引き出せるところまで至っている場合は少なく、長尾先生の危機意識には同感せざるを得ない。

しかし、実際のところは出版量は増

加しているのであるから、図書館にとって業務量は増えており、電子情報の作成がなかなか追いつけないでいる。

図書館にとって、過去からの膨大な資料を継承しつつ、新たに情報技術のスキルを装備し、積極的に電子化の方向へ進みたいのは当然であろうが、なかなか追いつけずもどかしい状態なのである。

入れるものをどのように選ぶか

さて、過去の蓄積も含めて、今後電子化すべき資料の対象は多く、どのように電子化していけばいいのか、方向性が見えないことがフラストレーションの種になっている。ただ、ここで注意すべき点は、「情報」の量は増えているものの、「知識」の量が果たして増えているのかという点である。図書館では「コレクション」という言葉を使う。ある観点から資料を選択して、コレクションとして利用者に提供するということである。電子的資料についても同様なことがいえる。図書館には、対象とする利用者およびそのコミュニティに対して十全な情報の集合を提供維持する役割を担わなければならない。

そこで、たとえば、大学図書館では、先生や学生とが密接に協力して電子的コレクションを整備していく必要があるのではないか。このためには図書館という組織とその働き方が従来以上にオープンになるよう積極的な変革を進めていく必要があると思う。

いろいろな困難

電子化する対象を絞ったとしても、著作権などの旧来からの制度的なしがらみのために、現状では電子コンテン

ツの入手と提供がきわめて困難である。これを技術発展の流れに合った新しい仕組みに変えていくには、出版ビジネス自体の変貌も必要で、長い時間を要すると思われる。そこで、長尾先生のご指摘にもあるように、慣習上、著作権の制約から逃れられるギリギリのところ、表題や目次までの電子化に努力を注ぐべし、という考え方も出てくる。

長尾先生は、文書を画像として電子化することにやや否定的な見解をお持ちのようであるが、利用者にとっては紙でも画像でも文字でもオリジナルな情報を入力できることが重要なのである。最小限それを確保するための最も経済的な方法として、画像入力も過渡期においてはそれなりの有用性はあると考える。特に古い資料や多言語資料の電子化の場合、このことがいえるのではないかと。

コンテンツ作成の組織化

従来本のような形をしていた情報を電子化することにより、その利用方法がどのように変わるかについては、大いに研究する必要がある。情報の可視化や大量情報の高速ブラウジングなど、研究すべきテーマは数多い。長尾先生が Ariadone で先駆的に実現してみせられた電子読書機能や本の「解体」については、その後類似の機能が百科事典やコンピュータソフトウェアのマニュアルを対象として一部ではあるが提供されるようになってきている。この種のコンテンツでは、付加価値の分だけさらなる投資が必要である。そのためこのような方向に向かうよう積極策を考えなければならない。

1つには、まとまった量のコンテンツを作り、これをもとに活発な研究を行うことである。具体的な進め方の例として、特定分野の教科書、専門書、規格書などの相当量を、著作権者の了解を得た上で、公的な研究助成を受けて電子化するのである。成果のコンテンツは自由な利用を許し、そこから発生するさまざまな問題を総合的に検討するというようなアプローチが考えられる。

このような実践を行い評価結果をフィードバックすることによって、これ

から新たに作ろうとする情報のとるべき形式などで共通の認識と方向性が得られる。そうしないと、電子化といえ、PDFにすればいいのだという安易な考え方に落ち着いてしまうことにもなる。狭い専門分野であっても、電子コンテンツの利用がクリティカルマスを越えると、情報の電子化へ向かう雪崩現象が起きる可能性が期待できるので、組織的なコンテンツの電子化という試みを、並行していくつかの対象分野で行うことが必要ではないだろうか。

本を越えて

いままでの議論は、本のような形式の情報を念頭において行ったが、これから新しく作られる情報では、動画や音声を含む構造化された情報にシフトしていくのは必然で、そのような情報の可能性に期待するところが大ではあるが、そのためにも、ある程度の量の素材データを作って実証的に研究開発を進める必要がある。このような情報の取り扱いに際しても、言語的に表現

された情報が中核となるわけで、長尾先生の主張される電子読書機能のさらなる発展がキーとなることには変わりはない。著作権の処理を含めた社会システムにまで問題を広げた場合、既得権益の再配分も含み、容易なことでは解決のつかない問題であるといわざるを得ない。私見としては、コンテンツ作りの実践とその活用に対する一層の研究を通して、新しい素材とそれを活用する環境を作ることから着手するのが、長尾先生の危機意識に応える一番近い道程ではないかと考えている。たとえば、次世代を担う子供の教育に活用する教科書や参考資料の電子図書館であれば、相当な公的投資も社会的に受け入れられるのではないかと。また、情報処理学会員の立場からは、情報技術の電子図書館を作るという共同研究プロジェクトを行って、コンテンツ作り、HCI、分散ネットワークなど、多分野の研究を並行して推進するような方策があつてよいと考える。

(1999.1.26)

の通り、利用料金をとらざるを得ないでしょうが、本や雑誌などの種類、その古さなどによって料金はそれぞれ異なつて設定され得ることを考えて、自動料金徴収のシステムを作り、ある程度の規模で実験することによって、出版社や著者が十分収入が得られるということを示し、出版社や著者が電子図書館に積極的に本や雑誌を提供するようになる状況を作ることが必要でしょう。

著作権の切れたものについては、世界中のボランティアに手分けして重複せずに入力してもらおう運動を国立国会図書館かどこかがリーダーシップを発揮して行うことが必要だと思います。入力状況の把握とその管理運営は各国の中心的機関が行う必要があります。

電子図書館のもう1つの重要な仕事は、ネットワーク上の情報を選別して永久保存することでしょう。ほつておけば大切な情報でも電子形態の情報は消滅してしまう可能性が高いので、これは急を要します。その時電子図書館として注意すべきことは、その情報の価値と信憑性を保証することにあります。こうすることによって、電子図書館に頼めば信頼できる情報が得られるという安心感を社会全体に対して与えることが大切であります。これは大変難しい仕事ですが、やらねば人類の知的資源の多くは失われるでしょう。

なお、書物の解体をした時の情報の粒度についてコメントをいただきましたが、うまい階層的な書物の解体、索引による解体、頁単位による解体などを融合的に提供するとともに、これらの解体に関する構造情報は本文データの上に乗っているという構造形態をとりますから、本文データの粒度は利用者の目的に応じて大きくも小さくもなるものであることを申し添えます。

いずれにしても、こういったことの具体的な建設的作業は、私のような年令の者には不可能で、これからの時代を背負って行かれる皆様方に期待をしなければなりません。皆様のご活躍を期待しております。

(1999.2.3)

再コメント

長尾 真

京都大学

お2人の方々から大変建設的なご意見をいただきありがとうございました。参考になるとともに、いろいろと考えさせられる内容でありました。まず、十分な量の情報を持った電子図書館ができあがるまでには長期間を要するわけですから、その間に解決すべき重要な課題について研究開発を行っておく必要があるだろうと思います。それらのご指摘のあった通り、メタデータの作成とレファレンスサービスの自動化であると思います。これらは言語情報処理のみならず、知識工学からも、また人工知能の立場からも非常に興味のある課題であります。両者とも、しっかりしたシソーラスと図書館分類 (6

～7桁の詳しいもの) 情報が利用できることが必要です。自動レファレンスサービスは特に知識と文脈を考慮した対話システムの構築という現在最も多くの関心を集めている問題でもあります。この場合、知識の中には電子図書館の持っている参考図書、収納している本や雑誌の目次、索引、本文テキストといったものと、そのハイパーテキスト構造が必要です。このような内容は一挙には作れませんから、安達先生のご提案のように、ある分野をとりあげて、その分野についてだけでもできるだけ完全に近い情報を入れてテストをすることが望まれます。

電子図書館の場合、武田様のご指摘

議論の続きは次をご覧ください
<http://www.ipsj.or.jp/magazine/interestsay.html>