

「パイロット電子図書館システム事業」の概要

堀口 光

藤原 達也

通商産業省機械情報産業局情報処理システム開発課

100 千代田区霞が関1-3-1

Tel : 03(3580)3922 Fax : 03(3501)6631 E-mail : LDF03444@niftyserve.or.jp

概要

通商産業省では、情報処理振興事業協会において、国立国会図書館、文部省、自治省、科学技術庁との協力の下で、「パイロット電子図書館システム事業」を実施することとしており、この紹介を行う。

ここでは、文献等を電子化した上でデータベース化し、ネットワークにより内外に提供するパイロット事業を実施することによって、ニーズ面、技術面での課題の抽出と解決を図るとともに、より高度な検索や文書処理等を可能とする先進的技術開発を行う。

Outline of the Pilot Digital Library System Operations

Hikaru Horiguchi

Tatsuya Fujiwara

MITI, Machinery and Information Industries Bureau,

Information Systems Development Division

(Ministry of International Trade and Industry)

1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo, 100 Japan

Tel : 03(3580)3922 Fax : 03(3501)6631 E-mail : LDF03444@niftyserve.or.jp

ABSTRACT

The Ministry of International Trade and Industry (MITI) will carry out " Pilot Digital Library System Operations " within the Information Technology Promotion Agency, Japan (IPA) in cooperation with the National Diet Library, the Ministry of Education, Science and Culture, the Ministry of Home Affairs, and the Science and Technology Agency.

The overview of this operations, which compile literature into databases after rendering them into digital form and then to supply them through networks, will be realized to identify and solve demand-related and technical problems. And advanced technical development which allows more advanced retrieval and document processing will also be implemented.

1. 展望

世界には、膨大かつ多岐にわたる知的財産が存在しているが、これらの資産の多くは、従来各地の図書館に書物などの形で収集・蓄積されているため、人類の共通財産として、世界の人々がこれを迅速に活用することはできなかった。

しかしながら、近時の情報処理技術、ネットワーク技術の進展により、次のような機能を有する電子図書館が構築され、世界に分散しているあらゆる知的資産を個人の知識ベースとして活用できる環境を創り出すことが技術的には可能となりつつある。

- ・世界中の図書館へのネットワークによるアクセス

情報（文献、絵画、映像、音声など）が電子的に保存され、利用者は、ネットワークにより、自宅、学校などから、世界中の図書館に蓄積されている情報にアクセスできるようになる。

- ・高度な検索機能

あいまいな指示でも検索できたり、求める文献等がどこの図書館に所在するかが分からなくても検索できるような高度な検索手法によって、利用者が容易に情報を得られるようになる。

- ・個人の幅広い創作物の蓄積と一般への提供

経済的な理由などにより、従来出版できなかったような個人の創作物を蓄積し、広く一般に提供することができるようになる。

2. 現状と具体的な施策の方向

2. 1 図書館ネットワークの構築

2. 1. 1 現状

電子図書館において、ネットワークを通じて文献等の検索を行うためには、まず各図書館の蔵書の書誌・所在情報の流通が図られている必要がある。米国においては、OCLC (Online Computer Library Center)、RLIN (Research Libraries Information Network)、WLN (Western Library Network) 等のネットワークが構築されており、ほとんど全ての公立図書館、大学図書館が接続され、各図書館の書誌・所在情報の検索が可能となっている。我が国においては、学術文献の全国総合目録の構築と相互貸借の効率化を目的として、学術情報センターに大学図書館を中心に300以上の図書館が接続され、書誌・所在情報の流通が図られており、各地の公共図書館を結ぶネットワークは、一部の地域においては構築されているが、全国的なネットワークは未だ構築されていない。この要因としては、図書館におけるコンピュータ導入の遅れや、各図書館が作成する書誌情報の統一化が図られておらず書誌・所在情報の流通のための基盤が形成されていないことなどが挙げられる。(米国においては、US MARCが統一的に用いられているが、我が国においては、JAPAN MARC、民間MARC等が存在し、各図書館では互い共通利用が図れない形で書誌情報データベースが構築されている。)

2. 1. 2 具体的な施策—総合目録ネットワーク化の推進

通商産業省としては、21世紀に本格的な電子図書館を建設するとの国立国会図書館の構想を踏まえ、また、上記の電子図書館の機能を想定しつつ、情報処理振興事業協会において、国立国会図書館、文部省、自治省、科学技術庁との協力の下で、「パイロット電子図書館システム事業」を実施する。その一環として、全国の図書館がネットワークにより接続され、オンラインによるアクセスが可能となるシステムを構築するための第一段階として、書誌・所在情報の相互流通を図るための事業を実施する。

- 全国レベルでの公共図書館の書誌・所在情報共通フォーマットの策定。
- このフォーマットに基づき書誌・所在情報をデータベース化した上で、どの図書館にどのような情報が存在するのかを一元的に検索できる環境を実現するためのパイロット事業を実施する。

2. 2 電子図書館システムの構築

2. 2. 1 現状

電子図書館の実現のためには、図書館のネットワーク化とあわせて、文献等の蓄積や検索などに関する技術開発や実証実験を通じて、蓄積、検索機能の高度化を図っていくことが重要である。現在、米国においては、電子図書館実現のための実験として、カリフォルニア大学の9キャンパス（バークレー、UCLA、UCSF等）で電子的に蓄積された文献の内容をネットワークを通じて流通させる実験を行っているほか、コーネル大学を中心とするCOREプロジェクトやカーネギーメロン大学のマーキュリープロジェクト等において検索手法などに関する研究が行われている。

我が国においては、学術情報センターでデータベースの構築を含む電子図書館の実証的研究開発が行われている。また、国立国会図書館等においても電子図書館の機能に関する基礎的な調査研究が行われているところであるが、電子図書館システムの構築に向け、今後一層の研究を行う必要がある。

2. 2. 2 具体的施策—パイロット電子図書館システムの構築

「パイロット電子図書館システム事業」の一環として、将来の電子図書館のモデルとなりうるパイロット電子図書館システムを構築し、以下の事業を行う。

(1) 文献等の電子的蓄積とネットワークによるアクセスのためのパイロット事業

文献等を電子化した上でデータベース化し、ネットワークにより内外に提供するパイロット事業を実施し、ニーズ面、技術面での課題の抽出と解決を図る。

(2) 電子図書館システムのデータ利用・検索・提供系の研究

既存の技術を活用して、検索、閲覧を容易かつ効果的に行うためのユーザインタフェースの研究や実証実験を実施するとともに、多種多量情報の蓄積・管理技術、類似語による検索や意味的検索等の高度検索・利用技術、分散情報の統合利用技術の開発等を行い、電子図書館システムの一層の利便性向上に資するものとする。

電子図書館システムの構築に活用しうる技術開発課題

No. 1

研究テーマ及び概要	技術開発課題	達成時期			備考
		短	中	長	
<p>1 様々な形態の情報を容易に検索・利用 (マルチメディア・データベースの構築)</p> <p>世界中に様々な形態で存在する知的財産(文庫、映像、音声等)をメディアの違いを認識することなく検索利用に利用するため、様々な形態の情報を (マルチメディア情報) を包括的なメディア変換等を伴わずに容易かつ統一的に活用できる環境を提供。</p>	<p>(1)マルチメディア情報管理技術 ・データ変換形式、管理システム (オブジェクト型データベースによる対応が進行中)</p> <p>(2)マルチメディア情報検索技術 ・マルチメディア内容のインデクシング (画像認識、自然言語処理)</p> <p>(3)大容量データ蓄積技術 ・ディスクアレイ、光ディスク記憶装置 (信頼性・負荷低減の向上)</p> <p>(4)データ圧縮技術 ・テキスト、画像、音声</p> <p>(5)問合せガイダンス支援技術 ・利用者の問合せガイダンス、類似の問合せ事例を活用したリフレクション</p> <p>(6)知識の蓄積・再利用技術 ・専門家による知識のノウハウの自動蓄積・再利用</p>	○	○	○	<p>(1) Gemstone(Servio:USA) ・ Carol(Onto: USA) ・ Objectivity(Objectivity:USA) ・ ObjectStorm(Object Design, USA) ・ Veritas(Veritas Object Technology:USA) ・ UASQL(UASQL: USA)</p> <p>(3) - SCSI interface/formal規格</p> <p>(4) - JPEG, MPEG</p>
<p>2 世界中に分散する情報への迅速なアクセス (広域分散データベースの構築)</p> <p>世界中に分散する様々な形態の知的資産 (文庫、映像、音声等) を時間的、空間的制約を越えて、自宅など任意の場所から統合的に利用できる環境を提供。</p>	<p>(1)分散データベース統合管理技術 ・分散環境下においてダイナミックに情報内容が変化するデータベース群を統合的に管理</p> <p>(2)データベース・ナビゲーションシステム ・多数の分散するデータベースの中から置かれた検索条件に最も適したデータベースを指示するナビゲーションシステムの開発</p> <p>(3)機械翻訳 (精度向上、高速化)</p>	○	○	○	<p>(1) - Oracle(Oracle:USA) - Sybase(Sybase:USA) - Esri(ESRI:USA)</p> <p>(2) - www/lotus</p> <p>(3) - 機械翻訳</p>
<p>3 必要とする情報を容易かつ的確に入手 (検索機能の高度化)</p> <p>膨大な情報の中から自己の必要とする情報を容易かつ的確に入手できる環境を整えるため、利用者の意図を理解し、あるいは指示でも的確に必要な情報を提供する機械等、高度な検索・利用環境を提供。</p>	<p>(1)高度全文検索 ・キーワードによる高度な全文検索(既に製品化済み、高速化が課題)</p> <p>(2)関数検索、意味検索等高度検索 ・シソーラス構造を用いたキーワードの展開 ・自然言語処理に基づく意味検索・検索 (原文翻訳、意図解釈、質問文解釈技術) ・文書内容解析技術 (文書中の関連の自動抽出、シソーラスの自動生成等)</p> <p>(3)居住検索 ・検索条件や検索の意図を指定してその条件に合致する情報を検索 (画像認識)</p> <p>(4)音声検索 (音声認識等)</p> <p>(5)検索エージェント ・複数のデータベースの中から利用者が必要とする情報を探し出す仕組み (ウェブ、Teletext)</p> <p>(6)知識データベース ・知識というデータに意味付けされた情報を効果よく蓄納し、人間の持つ知識を高度化し自由な検索・検索を実現</p> <p>(7)3次元ヒューマンインターフェイス ・3次元ウォークスル・技術、3次元ハイパーメディア技術</p>	○	○	○	<p>(5) - Teletext(General Magic)</p> <p>(6) - 第5世代コンピュータ</p> <p>(7) - VRmaker(Stereo Graphics:USA) - World Tool Kit(World:USA)</p>

No. 2

研究テーマ及び概要	技術開発課題	達成時期			備考
		短	中	長	
<p>4 個人の知的活動を支援する知識ベースとして活用 (知識ベースの構築支援)</p> <p>個人の創造的活動を支援するための、文書基約検索(検索の柔軟な要約を支援する機能)や文書作成機能(目的とする文書の作成を支援する機能)を整備し、個人の知識ベース構築を支援する環境を提供する。</p>	<p>(1)文書管理情報の自動抽出 ・文書から文書管理情報(目次、タイトル、著者名等管理情報)を自動的に抽出</p> <p>(2)キーワード自動抽出 ・文書から検索項目となるキーワードを自動的に抽出</p> <p>(3)文書基約検索 ・あらかじめ利用者が指定した情報(5W1H等)を自動的に抽出し、一定の形式で提示するシステム</p> <p>(4)ハイパーメディア編集 ・各情報(テキスト、映像、音声等)ごとにポインタを設定、リンクをばるることによって、関連した情報群を1つ情報塊として蓄積・提示</p> <p>(5)自動リンク機能 ・関連する情報の(半)自動的リンク付け</p>	○	○	○	<p>(4) - XyberCard(Apple:USA) - NoteCard(Xerox:USA)</p>
<p>5 高度な表示機能</p> <p>情報の内容を容易に把握できる表示環境を整備するため、従来の基メディアが持つ優れた機能(情報の一貫性やブラウジング等)を実現していくとともに、基メディアでは実現不可視なより高度な表示機能(拡大・縮小表示、補正等)を提供。</p>	<p>(1)高速ブラウジング技術</p> <p>(2)拡大縮小・ズーム表示機能</p> <p>(3)3次元仮想空間による表示 (バーチャル・リアリティ技術)</p>	○	○	○	<p>(3) - VRmaker(Stereo Graphics:USA) - World Tool Kit(World:USA)</p>
<p>6 個人からの情報管理 (個人情報の管理)</p> <p>個人からの情報採信コストを低減し、これまで経済的理由等により出願できなかったような個人の創作活動を、広く一般に提供できる環境を提供。</p>	<p>(1)個人情報管理 ・データベース ・ネットワーク上やデータベース上の情報に対するオリジナリティの保証 (著作権技術等) ・知的財産権対応 (著作権情報の付与)</p>	○	○	○	
<p>7 様々な形態の情報を容易に電子化 (情報入力方式の高度化)</p> <p>様々な形態の情報を電子的に扱うため、膨大な情報を効率的に電子化する環境を提供。</p>	<p>(1)OCR技術 ・既存のOCRの文字認識率の向上 ・多言語の文字、罫、グラフ、図形、映像部分の認識が可能な動的OCR</p> <p>(2)不特定読者の音声認識 (1万語レベル)</p> <p>(3)創造的文書作成技術 ・創造的文書の作成支援技術</p>	○	○	○	
<p>8 検索の意による検索作業</p> <p>個人の知的活動を個人との協調活動により一層高度化するための環境整備。</p>	<p>・ CSCW</p>	○	○	○	<p>・ Colo(Xerox:USA) - LOTUS NOTES(Lotus:USA)</p>