ディジタル図書館に関する最近の話題から

杉本重雄 図書館情報大学

ディジタル図書館に関する様々な活動が活発に進められている。本稿ではディジタル図書館に関する最近の話題として、National Digital Library Federation (NDLF)、Dublin Core、NSF/NASA/ARPA によるDigital Library Initiative (DLI) を取り上げ、簡単に述べる。

キーワード:

メタデータ、Dublin Core、Digital Library Initiative、National Digital Library Federation (NDLF)

Recent Topics on Digital Libraries

Shigeo Sugimoto
(University of Library and Information Science)

There are many R&D projects on the digital library and related activities. This paper briefly describes recent activities of the National Digital Library Federation (NDLF), Dublin Core, and the Digital Library Initiative by NSF/NASA/ARPA.

Keywords:

Metadata, Dublin Core, Digital Library Initiative, National Digital Library Federation (NDLF)

はじめに

ディジタル図書館 (Digital Library) に関する様々な活動が活発に進められている。アメリカでは議会図書館や大学図書館を中心として進められている資料のディジタル化を中心としたプロジェクト、NSF/NASA/ARPAの大規模な研究助成 Digital Library Initiative[1] による研究プロジェクトが活発に進められている。ヨーロッパではイギリスでの eLib[2] による研究助成や各国の国立図書館を中心とする資料のディジタル化プログラム、また我国ではこれまでに進められてきたプロジェクトに加えて、日本情報処理開発協会(JIPDEC)による次世代電子図書館システム研究開発プロジェクト、国立大学図書館での電子図書館機能の推進等、新しい活動が始められている。ディジタル図書館に関する国際会議も活発に開かれており、ADL'97(1997.5、主催:米国議会図書館、米国医学図書館、IEEE、NASA ほか)、ACM DL'97(1997.7、主催:ACM)、1st European DL Conference(1997.9、主催:ERCIM(European Research Consortium on Informatics and Mathematics))、ISDL'97(1997.11、主催:図書館情報大学)など数多く催されている、

本稿では、National Digital Library Federation(NDLF)、Dublin Core、DLI について簡単に紹介する。

- ◆ NDLF[3] は議会図書館や大規模な大学図書館を中心として 1995 年に結成されたもので、ディジタル 図書館に関する様々な話題を検討し、ディジタル情報の共同利用を進める上で重要な役割をになって いくと考えられる。
- ディジタル図書館の観点からメタデータは情報の組織化と検索のために非常に重要な役割をになっている。Dublin Core[4] はネットワーク上の情報資源のメタデータの記述形式として提案され、現在定義が進められている。
- 1994 年秋から 6 大学 (UCB, Stanford, UCSB, U. Michigan, UIUC, CMU) によって進められてきた DLI は 1998 年秋には終了する。それに続いて新しい研究助成 (DLI II) が計画されている。

National Digital Library Federation

NDLF は議会図書館や有力な大学図書館などの研究図書館 (注) が集まって 1995 年 5 月に構成された、ディジタルコレクションの形成と有効利用等、学術情報のための DL に関わる基本的な問題に取り組んでいる。特に、NDLF では大きなコストをかけて作成するディジタル資料の共有 (共同利用・相互利用) や大規模な情報資源の有効利用に焦点を合わせている。NDLF では参加組織による Policy Board をもち、議論を進めている。また、NDLF ではタスクフォースを設け計画の基本プランを作ることにした。このタスクフォースは Commission on Preservation and Access (CPA[5]) をコーディネータとし、各参加組織のスタッフから上級メンバーが参加することになった。このタスクフォースは 1996 年 8 月にレポートを出し、その中で、現時点での重要項目として以下の 3 点を中心に進めていくことを推奨している。

- ディジタル情報の発見 (Discovery) と検索 (Retrieval)
- 知的財産権と経済モデル
- ディジタル情報のアーカイブ

本年5月に Washington DC の議会図書館で開催されたディジタル図書館に関する国際会議 ADL'97では NDLF の活動の中心的役割を果たしている CPA の Deanna Marcum の招待講演があり、NDLF の活動に ついて述べていた。NDLF のメンバーは基幹的な図書館ないしは大規模な有力大学の図書館である。資料のディジタル化はコストがかかるため大規模な組織でなければなかなか取り組むことができない。同じく

ADL'97 の招待講演のひとつであった Smith College 図書館の Sarah Pritchard による講演では、大学の規模 (すなわち図書館の規模) が小さいので資料のディジタル化の計画は特別には持っていないこと、また NDLF のような活動によってディジタル資料へのアクセス性が高まることを期待していることを述べていた。

注:NDLF のメンバー

- (1) Commission on Preservation and Access (CPA)
- (2) Columbia University
- (3) Cornell University
- (4) Emory University
- (5) Harvard University
- (6) Library of Congress
- (7) National Archives and Records Administration (NARA)
- (8) New York Public Library
- (9) Pennsylvania State University
- (10) Princeton University
- (11) Stanford University
- (12) University of California, Berkeley
- (13) University of Michigan
- (14) University of Southern California
- (15) University of Tennessee
- (16) Yale University

メタデータ

メタデータ、すなわちデータに関するデータは大規模な情報資源の中から所望の情報 (データ) を見つけ出すために必要不可欠のものである。たとえば、従来の図書館情報システムでは MARC に代表される目録情報が電子化され、広く利用されてきた。図書館資料に関するデータである目録は代表的なメタデータである。また、索引、辞書やシソーラスもメタデータと考える。さらに書評のように他の文書に関する文書もある種のメタデータと考えられる。

従来のデータベースは単体で使われることが一般的であった。従って、利用者は特定のデータベースの利用方法を知っていればよく、また利用するデータベースのメタデータの特性を知っていれば十分であった。ところが、インターネットを介して多くの情報資源(データベース、コレクション、単体の文書など)が接続されるようになって、複数の情報資源にアクセスすることが一般的になり、意識しない間に複数の情報資源を利用していることすらある。一般にメタデータは情報資源毎に与えられるので、複数の情報資源にアクセスできるようにするにはメタデータ間の相互利用性(Interoperability)を向上する必要がある。さらにデータそのものだけではなく、情報資源に関する情報(たとえばデータベースに関するメタデータ)を用意することも必要である。

メタデータは巨大な情報資源(インターネット、およびそこにつながる電子図書館)から情報を見つけだすためのキーとなる。そのためメタデータに関する研究は活発である。たとえば、IEEE ではメタデータに

関する国際会議(第1回1996.4)を開催した。大規模な情報空間の中から所望の情報を見つけ出す情報発見 (Information Discovery) がキーワードの一つとなっている NSF/NASA/ARPA の DLI(Digital Library Initiative) でもメタデータが特に注目されている。たとえば、ミシガン大学ではエージェント技術を用いてメタデータによる利用者に応じた検索システムの構築を進めている。カリフォルニア大学サンタバーバラ校では地理情報に関するシソーラス、地名事典等に関して研究を進めている。また、スタンフォード大学では種類の異なる情報資源(データベース)に対する一様なアクセスを実現するため情報資源や利用条件に合わせたメタデータの変換等を行なうシステムの研究を進めている。

yahoo に代表されるインターネット上のディレクトリサービスの場合、人手による分類とメタデータ作成が行なわれている。この場合、データを作成する対象範囲を広くするには多人数が必要であり、また文書の出版からメタデータの作成までのタイムラグが長くなりがちである。さらに、適切な管理メカニズムを持たないと消滅した文書に関するメタデータの削除が行なわれず、存在しない文書を紹介することになる。一方、AltaVista や lycos、opentext などの文書検索サービスの場合、ロボットや Crawler と呼ばれるソフトウェアで機械的に文書を集めてデータベースを構成している。これらは機械的に多数の文書を集めることができるのでカバーする範囲は広がり、タイムラグもなくなるが、意味的な自動分類が困難なので検索ノイズが多くなる。検索ノイズを減らすには適切なメタデータを文書に埋め込んでおき、ロボットが集めた文書から自動的にメタデータを抽出してメタデータのデータベースを構成することが有用であると考えられる。以下に示す Dublin Core Metadata Element Set (通称 Dublin Core) はこうした考え方の下に提案されたものである。

OCLC と NCSA はインターネット上の文書オブジェクト (DLO) のためのメタデータを検討するためのワークショップを 1995 年 3 月に Ohio 州 Dublin で開催し、そこから Dublin Core の提案がなされた。その後、米国、ヨーロッパ、オーストラリアを中心として Dublin Core に関する実験や評価が進められた。1996 年 4 月には、Dublin Core をベースとしてメタデータを議論するためのワークショップがイギリスのWarwick で開催され、以下のような成果を残した。

- Dublin Core の構文を表す SGML の DTD の提案。
- 既存の HTML 構文の中への Dublin Core のためのタグの組み込み方法の提案。
- Warwick Framework と呼ぶメタデータの流通のためのコンテナ構造 (container architecture) の提案
- Warwick Framework の MIME, SGML, CORBA 上での記述方法の提案
- ▶ドキュメントの著者やコレクションの管理者のためのメタデータ作成方法に関する提案

Warwick Framework では 13 項目が提案されたが、その後イメージデータに関するメタデータワークショップ (1996.9) 等を経て現在以下の 15 項目になっている。

- (1) タイトル (Title): オブジェクトの名前
- (2) 作者または著者 (Author or Creator): 情報資源の内容に関して責任を持つ人(複数可)
- (3) 主題およびキーワード (Subject and Keywords): 情報資源に述べられたトピック
- (4) 記述 (Description): アブストラクトやイメージデータの説明など内容に関する記述
- (5) 出版者 (Publisher): 情報資源を流通の形態にしたエージェント
- (6) 他の関与者 (Other Contributors): 編集者や翻訳者等文書の内容の作成に関わった人
- (7) 日付 (Date): 出版の日付
- (8) 情報資源タイプ (Resource Type): 小説、詩、辞書といった情報資源のジャンル

- (9) 形式 (Form): Postscript ファイルや Windows 実行形式といった、情報資源の物理的な形式
- (10) 情報資源識別子 (Resource Identifier): 情報資源を一意に識別するための番号あるいは名前
- (11) ソース (Source): 情報資源オブジェクトの出所となった情報資源 (印刷物あるいはディジタルデータ)
- (12) 言語 (Language): 情報資源の内容を記述している言語
- (13) 関係 (Relation): 他の情報資源オブジェクトとの関連づけ
- (14) カバレッジ (Coverage): 地理的場所や時間的な内容に関する情報資源の特性
- (15) 権利管理 (Rights Management): 著作権記述などの権利に関する記述や利用条件に関する記述へのリンク (URL もしくは何らかの URI)

Dublin Core は基本項目だけを決め、これ以外にドキュメントの性質や提供者の環境に応じて他の項目を加えることを認めている。

Dublin Core では各エレメントをフリーテキストで表すことが許されている。これはメタデータに関する専門知識を持たない著者であっても容易に記述できるという点からは重要な点である。一方、フリーテキストだけで記述すると、検索ノイズが多くなることは否めない。そのため、各エレメントにサブエレメントを設け、どのような規準にしたがって記述しているのか (Scheme)、記述データはどのような形式なのか (Type) を表すことにしている。また、サブエレメントの種類を表す記述子を qualifier と呼んでいる [6]。下にタイトル・エレメントを例として示す。

Element: Title (TITLE)

qualifier とその値:

Scheme

Internal 他の記述方法とは無関係に決められた記述方式 (フリーテキスト)。 デフォールト値

AACR2 AACR2 に従って決められた記述方式に従う。

Type

Main 主タイトル。デフォールト値。
Long 長形式タイトル(Long Form)
Short 短形式タイトル(Short Form)
Subtitle サブタイトル
PartTitle タイトルの一部
Spine 背表紙のタイトル
Translated 翻訳されたタイトル

1997年3月にオーストラリア・キャンベラのオーストラリア国立図書館で開かれた第4回ワークショップ (DC-4) では上述の qualifier の記述方法に関する議論が進められた (Canberra Qualifier)[7]。 Dublin Core メタデータは任意の電子文書を対象とはしているが、WWW 文書への適用を主要な目的のひとつとしているので、HTML の文法と qualifier の記述の整合性等が議論された。 Dublin Core は構造の単純さを特徴とする。一方、あまりに構造を単純化しすぎると記述内容の意味的な解析に対する要求が大きくなりすぎ、効率的な情報検索が難しくなり、相互利用性が低下するという危険をはらんでいる。 qualifier はあまり Dublin Core 自体の定義が複雑にはならないように配慮しながら、従来から利用されてきたメタデータで意味が明確に定義されている記述方式や語彙の利用を進められるようにしている。

これまで Dublin Core に関する研究プロジェクトはアメリカ、イギリス、オーストラリア等英語圏が中心であったが、ヨーロッパ各国でもプロジェクトが開始されつつある。そのため、メタデータが表す内容が何語で書かれているかといった多言語による記述に関しても重要な話題として認められている。なお、DC-4の参加者はアメリカ、イギリス、オーストラリア、ニュージーランドの英語圏以外にヨーロッパ(フランス、ドイツ、北欧諸国)、タイ、日本からであった。また、第5回のワークショップは本年10月にフィンランドのヘルシンキでの開催が計画されている。

NSF/NASA/ARPA Digital Library Initiative

NSF/NASA/ARPA による共同助成プログラム Digital Library Initiative(DLI) は 1994 年から 4 年計画で 始められ、すでに3年が過ぎようとしている。助成を受けた下記の6大学(Carnegie-Mellon University, University of Michigan, University of Illinois at Urbana- Champaign, University of California at Berkeley, University of California at Santa Barbara, Stanford University) に与えられた助成総額 2300 万ドルであり、 これに加えて多くの企業が Partner として寄与している。DLI は図書館と計算機科学の分野の研究者が協力 し、さらに出版社や政府機関などからのコンテンツを対象とする電子図書館の testbed (実験台)作りのプ ロジェクトとして脚光を浴びた。DLIでは約半年に1度の割合で全プロジェクトが集まる会議を開催してお り、進展具合を公開する場であるとともに、意見と技術の交流の場となっている。(たとえば、1996年5月ミ シガン大学、同年12月スタンフォード大学、1997年6月カーネギーメロン大学)また、イリノイ大学のグ ループが中心となって6プロジェクトの情報を公開している(http://dli.grainger.uiuc.edu/national.htm)。 NSF 他は 1998 年から次のディジタル図書館研究助成計画(通称 DLI II)を、これまでよりも規模を大き くして進める予定である。DLI II の計画を議論するためのワークショップが本年3月にニューメキシコ州 サンタフェで開催された。ワークショップは、非常に大規模で分散した情報環境の中で知的作業を進めて いくための基盤環境 (Distributed Knowledge Work Environment) に関する研究にはどのような点が重要 であるかといった点から次の研究助成プログラムのプランを考えるという目的で開かれ、ミシガン大学の Daniel Atkins がホストし、NSF の Y.T.Chien を始めとする計画を策定する側からの出席者、現在の DLI の研究代表者、利用者評価、アーカイブ、美術館等の専門家による presentation と議論が行われた。そこ での議論から、現在の DLI のような大規模プロジェクトだけのものとは異なり小規模なものから大規模な ものまで組み合わせること、またより実世界指向のプロジェクトとすることなどの方向付け進められた。 このワークショップの報告書はまもなく WWW 上に公開される予定である。

おわりに

DLI のように研究指向のものから図書館自体が進めている資料のディジタル化を中心とする開発指向のものまでいろいろなディジタル図書館プロジェクトが現在も非常に活発に進められている。NDLF もあげている「情報発見(Information Discovery)」、「アーカイブ」、「知的財産権と経済モデル」は、ディジタル図書館プロジェクトに関する共通のキーワードであると思われる。知的財産権の重要性に関しては言うまでもないであろう。非常に大規模な情報空間を構成することになるディジタル図書館にとって情報検索や情報発見が重要であることは明らかである。情報発見という観点から、ディジタル図書館間の相互利用性を高め、非常に大規模で多様な情報資源の中から効率よく情報を探し出すためのメタデータの重要性は容易に理解できる。また、アーカイブの観点からは、どのような資料を電子化して保存すべきかということに加えて、電子化された資料、あるいは電子的に作成された資料は従来の印刷体の資料とは異なった意味で保存の難しさを含んでいる[8]。

また、ディジタル図書館の相互利用性や国際的な情報基盤上での情報発見等の研究を進めるには国際間での共同研究の場を設けることが重要になってゆく。NSF は ERCIM と Digital Library に関する共同研究

を始めており、現在 Metadata、Interoperability、Optimized Search and Retrieval、Intellectual Property and Economy Model の 4 テーマでの共同研究を進めようとしている [9]。 さらに、Multi-linguality を新たに加えてアメリカ-ヨーロッパの研究者による共同研究を進めようとしている。一方、Dublin Core のグループでも多言語の扱いの需要性を認識しており、DC-4 においても多言語に関するディスカッションが行われた [10]。こうした多言語情報の処理や相互利用性等の研究はこれからますます重要性を増すと思える。

参考文献

- [1] NSF/ARPA/NASA Digital Libraries Initiative (national synchronization home page), http://dli.grainger.uiuc.edu/national.htm
- [2] eLib Home page, http://www.ukoln.ac.uk/elib/
- [3] National Digital Library Federation (NDLF), http://lcweb.loc.gov/loc/ndlf/
- [4] The Dublin Core Metadata Element Set Home Page, http://www.oclc.org:5046/research/dublin_core/
- [5] Commission on Preservation and Access, http://www-cpa.stanford.edu/cpa/index.shtml
- [6] Dublin Core Metadata Element Set: Reference Description, http://purl.org/metadata/dublin_core_elements
- [7] Weibel, S., Iannella, R., Cathro, W., The 4th Dublin Core Metadata Workshop Report, D-Lib Magazine, June 1997, http://www.dlib.org/dlib/june97/metadata/06weibel.html
- [8] Preserving Digital Information: Final Report and Recommendations, http://www.rlg.org/ArchTF/
- [9] EU-NSF Supported Programme for Planning Meetings Between US DLI Projects, NCSTRL and ERCIM, http://www.area.pi.cnr.it/ErcimDL/EU-NFS.html
- [10] Baker, T., Metadata Semantics Shared Across Languages: Dublin Cores in languages other than English, http://www.cs.ait.ac.th/~tbaker/Cores.html