

2 . 4

電子図書館（デジタル・ライブラリ）に関する動向 調査報告書

目 次

1	はじめに	2
1.1	背景	2
1.2	図書館利用の現状	2
2	電子図書館サービス利用上の問題点	8
3	電子図書館の目的	13
4	電子図書館に関する要素技術	14
4.1	書誌情報関連技術	14
4.2	メタデータ関連技術	20
4.3	情報発信技術の研究	24
5	電子図書館関連技術の動向	27
5.1	Dublin Core およびメタデータに関する動向	27
5.2	Z39.50 に関する動向	28
5.3	Semantic Web に関する動向	28
5.4	情報フィルタリングに関する動向	29
5.5	Personalization および情報発信に関する動向	31
6	電子図書館システム研究の波及効果	32
7	国家事業としてのデジタル・ライブラリ	33
8	日本における電子図書館の現状	35
9	海外におけるデジタル・ライブラリの現状	36
9.1	大学・非営利団体におけるデジタル・ライブラリ・サービス	36
9.2	商用デジタル・ライブラリ・サービス	37
9.3	今後のデジタル・ライブラリのあり方に関する動き	39

1 はじめに

1.1 背景

図書館の起源は古く紀元前にさかのぼるが、当初は文書の保存や特別な利用目的のために建設されたものであったといわれている。一般の市民が図書館を公共のものとして利用できるようになったのは、米国で1848年、英国で1850年に公共図書館法の成立後であり、以後約150年を経過している。

わが国においては奈良時代に朝廷や貴族の有する文庫があったが、一般市民が利用できる図書館は1950年の図書館法により、公立図書館の無料公開原則が定められた以後である。ただし、図書閲覧が開架式となったのは日本図書館協会の「中小都市における公共図書館の運営」(1963年)後であり、まだ実質的に40年程度の歴史しかなく、図書館を有効に利用する手法については欧米に比し相当遅れている状況である。

情報化の進展と共に、図書館の形態が次第に変化してきており、特に最近の数年は、インターネットにホームページを掲載する図書館も徐々に増えてきた。なかには図書館間の連携により、個々の図書館の蔵書範囲を超えた情報を提供する図書館も出てきている。流れとして専門化向けの特殊用途から一般向けに開放され、現在は図書館情報の共有化の時代に入りつつあるといえる。インターネットの普及とデジタル技術の進展は図書館の電子化を促し、情報の発信基地としての新しい役割も期待されている。しかし、従来の図書館の持つ数千年の歴史に比べれば、電子図書館(デジタル・ライブラリ)の研究は世界的に見ても高々40年程度の歴史しかなく、必ずしも電子図書館のあるべき姿についてのコンセンサスができている状況にはない。

1.2 図書館利用の現状

現在わが国では、都道府県・市区町村・私立および国公私立大学・短大・高専図書館あわせて現在4279館の図書館が存在しており、約5億4000万点の蔵書を有し、年間約4億8000万点が貸し出されている(「日本の図書館」(2000年版))。

図書館の利用統計を公表している例は非常に少なく、実態のすべては把握しがたい状況にあるが、公表されている例では下記の通りである。

東京都立図書館の場合中央・日比谷・多摩の3図書館で計約250万点の蔵書を有し、年間約80万人の入館者を数える。横浜市立図書館は中央・戸塚・保

土ヶ谷など 18 館からなっており、340 万点あまりの図書資料を有する。平成 12 年度の貸出冊数は一般書 808 万点、児童書 276 万点で計 10,84 万点となり、市民一人当たりの貸出数は 3 冊強である。

<市区立図書館の例>

(注)横浜、東京都中央区は新聞・雑誌・AV 含まず。鎌倉は含雑誌。

	蔵書数(冊数)	個人向け貸出数	貸出者数等
横浜市立図書館	3,403,335 冊	10,846,181 冊	貸出数 3.155 冊/市民一人当たり
東京都中央区立図書館	527,825 冊	795,909 冊	個人貸出者数 341,993 人
鎌倉市立図書館	460,753 冊	1,093,176 冊(*)	貸出者数 404,332 人

(*) 個人・団体区分不明

<都道府県立図書館の例>

(注)新聞・雑誌・AV は含まない。

		蔵書数(書籍)	個人向け貸出数	入館者数・貸出者数等
東京都立図書館	中央	1,630,818 冊	館内閲覧のみ	入館者数 379,592 人
	日比谷	283,891 冊	87,543 冊(*)	305,341 人(*)
	多摩	558,773 冊	館内閲覧のみ	86,993 人
	合計	2,473,482 冊	約 230,000 冊(**)	約 1,380,000 人(**)
大阪府立図書館		1,429,395 冊	1,112,894 冊	入館者数 1,017,710 人
神奈川県立図書館		868,342 冊	95,874 冊	入館者数 455,028 人

(*)日比谷図書館は耐震工事のため 117 日間のみ開館。(**)合計欄は年換算値で算出。

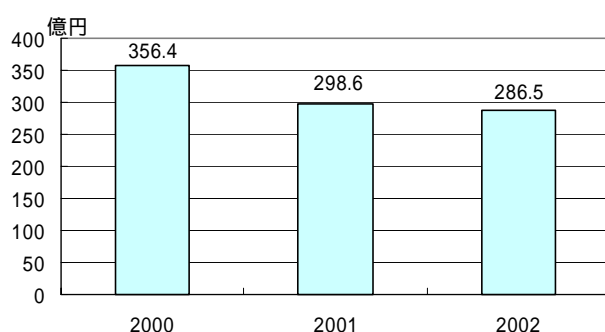
全体的に見ると、公共図書館は有効に利用されているといえるが、利用される蔵書に偏りもあり、それによる問題も生じている。昨年横浜市立図書館は小説「ハリー・ポッターと賢者の石」を 18 ヶ所の図書館で計 150 冊を購入し、貸出予約者が 2,500 人に達した。図書館側としては利用者ニーズに対応したということであるが、作家・出版社側から見ると、2,500 冊売れたはずのところは 150 冊しか売れなかったことになる。折からの出版業界の不況もあって、このような状況が作家・出版社の問題視するところとなっている。出版社は図書館貸出の有料化を主張しており、日本文芸家協会も昨年文化庁に対して「公共貸与権（公貸権）」導入を要望した。横浜市の場合は図書館の規模が大きいだけに特異な例と考えられ、出版社側も横浜市立図書館だけを問題にしているわけではない。しかし、一般的には話題作や人気作家の作品への需要は高く、公共図書館としてはその立場上その要望に応えざるをえないという事情もある。

図書館のあり方という点では、このような話題作や娯楽作でない蔵書がど

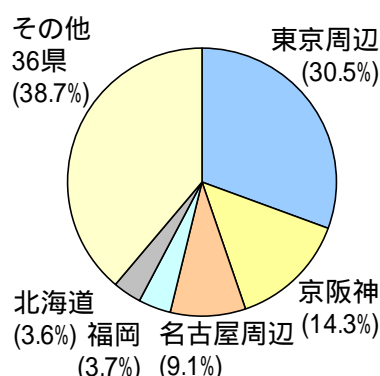
れだけ利用されているか、あるいは利用しやすくなっているかということが課題であるが、この点についてはあまり議論されていない。

公共図書館においては資料購入費の減少という深刻な問題がある。下記は全国47都道府県の公立図書館における年間資料購入費の推移と分布を示している。資料購入費は過去3年間確実に減少しており、2000年から2002年までの間に約20%も減少した。

47 都道府県の公立図書館の資料購入費推移



資料購入費の分布 (2002 年)



(注) 東京周辺：東京・神奈川・埼玉・千葉・茨城 京阪神：大阪・兵庫・京都・滋賀 名古屋周辺：愛知・静岡

資料購入費は大都市周辺に集中しており、東京・大阪・名古屋・福岡・札幌周辺の11都道府県の1都道府県あたりの資料購入費は平均13.5億円であるが、それ以外の36県の平均は3.3億円弱と1/4に過ぎない。さらに2001年から2002年にかけての購入費減額率は11都道府県の-3.6%に対し、36県は-4.8%と地方の方が大きい。

このような背景があるために、書店で簡単に購入できるか否かを問わず、利用者ニーズの高い書籍を所蔵することで図書館の存続を維持している面は否定できない。資料購入費の減少は図書館にとって死活問題であるが、当分状況の好転は望めない。図書館間の蔵書の相互利用など、資料購入費減少の影響を最小限にする必要がある。

さらに、公共図書館においては図書館員の削減も始まっている。専任の司書・司書補の数は7,848名(1999年)から7,641名(2000年)に減少するなど、図書館全体としては専任者を削減する傾向にあり、今後もさらにこの傾向は拡大すると予想される。現在、図書館員の業務ではレファレンスサービスの対応は大きいウェートを占めており、レファレンスサービスのほとんどを職員が口頭や電話で対応している。下記は、東京都立図書館、大阪府立図書館のレ

ファランスサービスの実態である。

図書館におけるレファランスサービスの実態（2002.4～2003.3）

	東京都立図書館	大阪府立図書館
口頭による問い合わせ、調査依頼	104,819	172,964
電話による問い合わせ、調査依頼	50,846	61,186
FAXによる問い合わせ、調査依頼	2,491	
文書による問い合わせ、調査依頼	899	1,731
E-mailによる問い合わせ、調査依頼	992	

それぞれの図書館において、口頭および電話による問い合わせ、調査依頼に対する年間対応人員を試算すると下記の通りとなる。

問い合わせ対応人員試算

	問合せ件数	対応時間(5分/件)	就業時間	対応人員
都立図書館	155,665 件	12.972 時間	1,900Hr/人・年 (推定)	6.8 人
府立図書館	234,150 件	19,512 時間		10.3 人

(注) 職員対応時間 5 分/1 件、職員年間就業時間 1,900 時間は試算のための推定値

両図書館ではそれぞれ 6.8 人、10.3 人が出勤時から退勤時までフルに問い合わせ、調査依頼に対応していることになり、相当な負荷がかかっていることになる。利用者が自分の見たい蔵書を探すのは大きい図書館になればなるほど大変な労苦を伴い、司書などの専門家の助けを借りることが多い。専任図書館員の削減に対応するためには一般の利用者が容易に資料・蔵書を検索技できる術を開発し、対応することが必要となっている。

このような状況は地方の方が厳しい状況にあり、それぞれ地域の事情に応じた独自の方法で図書館の利用拡大を試行しているのが実情である。2001 年に開館した仙台市の「仙台メディアテーク」は美術館風の建物に映像音響ライブラリーやスタジオも設置し、地元の商店街と共催してジャズフェスティバルを開催するなど来館者数拡大の努力を行っている。青森市民図書館や兵庫県西宮市の図書館のように商店街や生協などの毎日の生活の場と一体化を目指した公共図書館もあり、これらは年間来館者数がいずれも 100 万人前後と大都市の図書館並みの規模となっている。ただし、このような例は旧来の図書館の老朽化

や都市計画にそって図書館の増改築あるいは移設が可能となったという幸運な面もあり、必ずしも全国的に実施できるわけではない。

図書館の蔵書を特化することで利用価値を高める動きもある。浦安市立図書館ではビジネスマンや技術者の利用拡大を図るため、専門書や判例集を充実させ、ビジネス支援を前面に出している。また、神奈川県立川崎図書館では5年前から一般図書の所蔵を廃止し、ビジネス情報専門の蔵書に特化しており、同館の収集する企業関連情報論文誌は国立国会図書館より多い。これらの動きは、ベンチャー企業や中小企業の多い大都市周辺で見られる特徴である。米国においては起業に関して図書館が非常に大きい役割を果たしているという事情もあり、ビジネス支援図書館推進協議会（2000年発足）においても、企業経営支援の立場からの図書館利用を検討中である。

次に、大学図書館の現状と問題点を見てみよう。まず所蔵する蔵書数は下表の通り、一般の公共図書館に比し、図書館数は少ないものの、蔵書数はほぼ匹敵している。

わが国の大学図書館数および蔵書数

		図書館数	蔵書数(千冊)
大学図書館	国立大学	302	85,283
	公立大学	101	15,635
	私立大学	792	141,967
	計	1,195	242,813
短期大学図書館		384	19,682
高専図書館		61	4,444
計		1,540	266,939
(参考) 都道府県・市区・町村立図書館		2,609	284,462

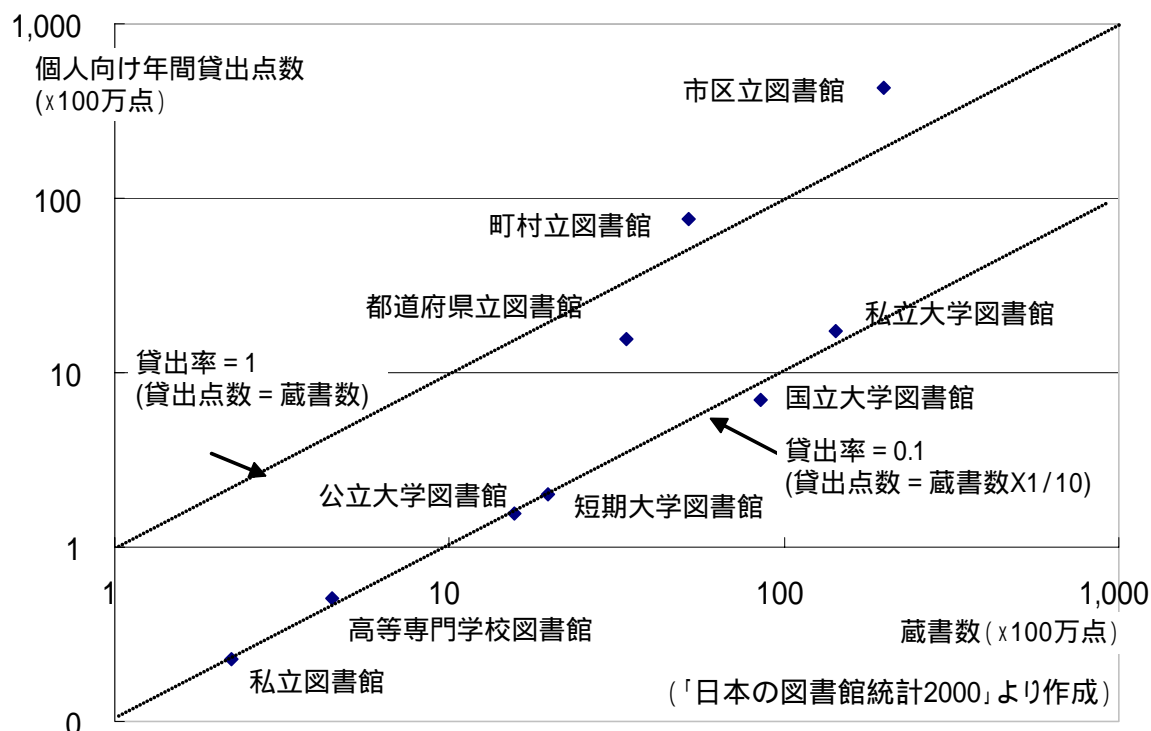
(出典：「日本の図書館- 2000年版」)

ただし、利用状況から言えば必ずしも活発とはいえない状況にある。大学図書館は必ずしも学内関係者の利用に制限されているわけではないが、手続き上の問題などもあり、実態としては学外特に一般市民が利用しているケースはほとんどないというのが実態である。

次図は、図書館における総括的な利用実態として、公共図書館、大学図書館等における蔵書数と個人向け年間貸出数を示したものである。本図では、図

書館の蔵書数に対する、年間貸出数を蔵書の貸出率として、公共図書館、大学図書館のそれぞれを評価している。貸出率は、蔵書のすべてに対し、蔵書 1 冊が年に何回貸し出されているかの平均値である。

図書館における蔵書の個人向け貸出状況



市区立図書館、町村立図書館および都道府県立図書館の貸出率はそれぞれ 2.2、1.5 および 0.5 となる。すなわち、図書館の所蔵する全資料に対し、市区立図書館では蔵書 1 冊が年平均 2 回以上、町村立図書館では 1 回以上、都道府県立図書館では蔵書の半分が年間 1 回貸し出されていることになる。

これに対して、大学図書館は大学の種類、規模に関係なくほぼ一律に蔵書 10 点に対して年間 1 点の貸出である。大学図書館の性格から、貸出対象とならない貴重な古典文書・学術文書も多いため、一般の公共図書館と同列には議論できないものの、この状況は問題がある。

今後の、大学図書館は学術情報発信拠点としての役割を期待されているが、実態としてはなかなか進んでいない。「学術情報の流通基盤の充実について」(科学技術・学術審議会研究計画・評価分科会答申 2002 年 3 月)では、科学技術基本計画において重点分野の一つとされるライフサイエンス関連における電子ジ

ジャーナルの体系的購読においてすら“諸外国の大学等における電子化された学術情報の収集・利用実績と比較して、大幅に立ち遅れている”という指摘がなされ、大学図書館の電子図書館的機能として“大学等からの情報発信機能の整備に関して、総合的な企画・立案を行なう機能および発信される情報のポータル機能”を担うことが要求されている。大学図書館における電子媒体による情報発信は下表の通りで、まだ緒についたばかりである。現在の大多数の大学図書館はまだ、所蔵資料・蔵書のデジタル化に追われており他のことには手が回らないという状況である。

電子媒体による主な情報発信部署

()内は回答大学数(回答大学 99 校中)

発信情報	最多担当部署	第2位担当部署
学位論文(抄録)	図書館(13)	学部・研究科(8)
学位論文(本文)	図書館(6)	学部・研究科(2)
紀要、学術報告(研究報告)等	学部・研究科(5)	図書館(2)
科研費等の研究報告書等	図書館(4)	学部・研究科(4)

(出典:「学術情報の流通基盤の充実について」付属資料)

なお、参考までに、紙媒体による情報発信状況を示すと下表の通りである。現時点ではまだ、学部・研究科による紙媒体の情報発信が主となっている。

(参考)紙媒体による情報発信状況

()内は回答大学数(回答大学 99 校中)

発信情報	最多担当部署	第2位担当部署
学位論文(抄録)	学部・研究科(38)	広報(20)
学位論文(本文)	学部・研究科(49)	図書館(13)
紀要、学術報告(研究報告)等	学部・研究科(63)	図書館(32)
科研費等の研究報告書等	研究協力課(51)	学部・研究科(27)

(出典:「学術情報の流通基盤の充実について」付属資料)

2. 電子図書館サービス利用上の問題点

欧米のデジタル・ライブラリ商用サービスは、母体となる企業が長い歴史

を持ち、それぞれ独自のコンテンツを有しているため、それぞれ特徴あるサービスを提供している。また、商用サービスであることから検索画面についても操作性の面で利用しやすいようにさまざまな配慮がある。これらは、デジタル・ライブラリ提供サービスとして価値のあるモデルであるが、各社の方針に基づいて集積されたコンテンツを、各社がターゲットとするユーザに対し、資金と人手をかけることでさらに利用拡大を図ることを目的とするものである。従って、基本的にはそれぞれの企業のビジネスに限定された範囲で最適化を図ったモデルである。

現在世界的に推進されているデジタル・ライブラリの研究は、大学や国立図書館などがリーダーシップをとって公共性の高い次世代デジタル・ライブラリのあり方とそれに関連する技術開発を行っている。デジタル・ライブラリ・サービス提供各社はその結果を踏まえて各社のコンテンツを効率よく提供するための有効性の検証とユーザへ利便性の高いサービス提供を行うという立場で協調と役割分担がなされている。ここでいう公共性とは情報の利用者と提供者が共に科学技術発展に貢献し、その恩恵を等しく享受するというものであって、娯楽や個人の余暇を利用した一般教養としての自己啓発とは異なるものである。

下記は大学で広く利用されている NACSIS Webcat により、蔵書の検索を行なうときに入力する指定項目である。利用者はこれによって、自分の要求する蔵書の所蔵の状況を調べることができる。単に自分の目的とする図書館に所蔵されているかどうかだけでなく、全国のどの図書館に所蔵されているかも調べられることも可能である。

NACSIS Webcat による蔵書検索時の指定項目

- | | | |
|-----------|------|-------------------|
| 1. タイトル | ---- | 書名、シリーズ名など |
| 2. 著者名 | ---- | 著者、編者、翻訳者など |
| 3. 出版者 | ---- | 出版者、発行者（出版社、団体）など |
| 4. 出版年 | ---- | 出版年（西暦） |
| 5. 標準番号 | ---- | ISBN または ISSN |
| 6. フリーワード | ---- | タイトル、著者名、件名、分類など |
| 7. フルタイトル | ---- | サブタイトルを除くタイトル全体 |

各大学図書館、公共図書館で構築するオンライン蔵書検索（OPAC：Online Public Access Catalog）も同様の項目を指定して蔵書検索が可能である。しかし、OPAC は基本的に各図書館ごとにシステム開発しているため、そのインタ

フェースには統一性がない。指定項目を個別に見ると蔵書のタイトルと著者名などの共通項を除けばシステムごとに異なっている。東京都立図書館の場合などは「タイトル、巻次、著者名、出版者、出版年、ファイル種別、ISBN、ISSN、所蔵館、ソートキー、ソート順」の指定項目がある。入力時に半角カナを許容するかどうかまで、システムによって異なるため、大学図書館などではシステムごとに入力説明書を掲載しているが、自治体の図書館では説明のないものも多い。

機械は基本的に自然言語を理解できないので、検索語が人名か事象名などの判断が難しく、また人名であっても著者名であるのか、その人名を冠した書名であるかなどの判断ができないために、細かく指定する必要がある。しかし、現状では、各図書館のデータベースやシステムがおのこの独自に構築されていることが、利用者インタフェース不統一の要因となっている。

大学や研究機関などで所蔵する資料は一般書籍から論文、研究報告、ジャーナル、画像資料その他あらゆる形態の資料が数 100 万点にもおよび、データベースにもさまざまな形態がある。現時点では、それらが一元的に管理されているわけではなく、OPAC でうまく検索されるのは条件が整った場合に限られている。

検索エンジンによっては「主題」によって検索できる機能もある。しかし、主題として与えられたキーワードを機械的に一致検索するだけなので、正しく検索することもあれば、文章の流れでたまたま一致したパターンが出現すれば検索リストに刑されてしまうこともある。そのパターンだけ取り出すと日本語として何の意味のないものであっても検索エンジンはその判断ができない。例えば、「たいし」は「聖徳太子の業績」「少年よ、大志を抱け」「検索したい資料（しりょう）の探し方」「熱く焼いた石（いし）を使った料理法」「海外旅行にも行きたいし、国内旅行もしたい時の知恵」などが検索対象と見なされる。さらに、これらの精度も検索エンジンによって差異がある。

最近開設したデジタル・ライブラリにおいて、“多くの資料がデジタル化され、インターネットで検索できる。1400 年代の文献もある”という趣旨の新聞報道があったので、「1400 年代の文献」をキーワードとして検索したところ、ヒットした第 1 画面として下記の結果を得た。

キーワード「1400 年代の文献」による検索結果

項番	書名	発行年	情報	内容
1	チョーサー:その時代・文学・言語	1982		

2	「バースの女房」をめぐって：フォ-サ-カンタベリ物語	1985
3	戦後作文教育文献解題：昭和 20 年代・30 年代	1999
4	科学者伝記小事典：科学の基礎を築いた人々	2000
5	流通研究の現状と課題	1995
6	VOA 基本 1400 語文例集	1992
7	高校生、英語初級者のための英語基本熟語 1400 語	1992
8	80 年代における産業組織問題・文献調査報告書	1979
9	おかやま文化山脈：戦後の軌跡	1995
10	ヒューマニズムの芸術	1987

(注 1) 情報：所蔵図書館情報あり 内容：内容・目次あり

(注 2) 著者名、出版社名等省略、その他画面表現については紙面の関係で変更した

結果的には、検索エンジンが「1400 年代」を意味のある一つのブロックとして検索しなかったために、受験参考書や関係の不明な蔵書がヒットされる結果となった。自然言語はもともと人間同士が理解しあうための言語であるため、機械が正確に理解するには高度な技術を必要とする。特に日本語は英語などと異なり単語ごとに区切る習慣がないため、単語を抽出するために形態素解析という技術が使われる。形態素解析技術は現在の水準で約 96～98%の精度といわれており、要素技術としてはかなりの水準にあると考えてよい。ただし、検索後の理解という意味では、異なった技術が必要である。「情報技術」を一つのブロックと扱う場合と「情報」「技術」と分ける場合は検索結果が異なってしまうので、索引語を抽出するためのインデキシング技術や単語辞書を利用した自然文解析技術などの研究が行なわれている。「1400 年代」の場合は「年代」という語の性格に関するルールの問題である。

情報検索技術の最終的な目標は、利用者の要求に合致した検索結果を提供するかに尽きる。この技術は、一般に情報フィルタリング技術と呼ばれている。情報のフィルタリング・サービスには大きく 2 通りの手法がある。一つは Yahoo に代表されるように、サービス提供者がすべてのインターネットのページを収集して利用者に分かりやすいように分類表示するディレクトリ・サービスである。これは、基本的には人手で収集、分類・登録するもので目的とする情報が登録されていれば無駄なく、非常に効率よく目的の情報にたどり着けるし、情報のフィルタリングも人手で行うため、有害情報などの掲載・非掲載なども利用者の感覚に合致した基準で判断できるという長所がある。反面、人手によるだけにインターネット上のすべてのページを網羅することは難しい。そこで、最近では他の手段である検索ロボットを併用して使うことが多い。

検索ロボットを利用するシステムでは、クローラーあるいはスパイダーなどと呼ばれるソフトウェアでページの収集を行う。ページの収集はページのリンク関係をもとに機械的に行われ、ページの分類や利用者からの要求であるキーワード（単語）との一致比較もすべて自動的に行われる。一致比較は利用者が入力したキーワードに対して対象ページの全文検索を行うのが原則であるがどこまで一致検索するか、リンクをどこまで辿ってページ収集するかは検索エンジンの性能に依存する。世の中のすべての検索エンジンが、すべてのリンクのすべてのページを網羅して一致比較するわけではない。どこまで実施するかによって検索対象のヒット数が変わってくる。また一致比較についても種々の手法がある。最近では、不必要な情報いわゆるノイズやジャンクが検索画面に表示されることが問題となっており、重要な研究課題となっている。

検索ロボットによるページの自動収集と全文検索では Google が最も規模が大きく、その収集ページも 30 億ページを超えている。Google は独自の PageRank 技術により、検索されたページのうち多くのページからリンクされているページは価値が高いと判断をしている。これらは、利用者にも受け入れられ、ある種の検索には非常に有効な手法であるが、基本的には提供サイドの論理によって情報の価値が決められている。

Web における一般の商用検索サービスでは、入力された検索語により、対象ページの全文と比較しその検索語が存在するかどうかで検索が行なわれる。従って、入力された検索語がタイトルであっても、そのタイトルのテキストを検索するのではなく、その検索語を使って記述しているページが検索の対象となる。最近では性能が非常に向上したため、かなり精度の良い検索も行なわれているが、検索対象がある程度の情報量を有する Web のページであり、また出力リストには、各社の基準によるランク付けが行なわれるので、書誌情報をもとに行なわれる蔵書検索も含めて万能というわけではない。

情報検索においては、検索語に適合する情報を提供するほかに、個々の利用者の特性に適合する情報を提供する考え方もある。この分野の技術は “ Personalized Information Filtering ” または “ Personalization ” と呼ばれている。この技術で最も知られているのは協調フィルタリング (Collaborative Filtering) 技術である。この技術は 1992 年に Xerox PARC から発表され、現在ではオンライン書店の Amazon.com が採用していることで有名である。協調フィルタリング技術は、Web にアクセスする履歴の経緯を蓄積し、類似アクセスする利用者の履歴経緯と合わせて動向分析し、類似傾向を持つ利用者は類似

情報への欲求を持つという考え方である。この技術の利点は、利用者自身が意識していない自分の傾向まで把握できるので、商品マーケティングの立場から見ると利用価値の高いものである。Amazon.com の場合、昨年 11 月 1 日から 12 月 23 日までの 2 ヶ月弱で全世界 220 カ国以上から 5,600 万件の注文を処理している。ピーク時の 12 月 9 日（月）、11 月 11 日（水）ではそれぞれ 1 日に 1,700 万件、1,600 万件の発注があった。このように、莫大な数のユーザが利用する分野では、利用者特性分析の母集団が大きいと、協調フィルタリングもかなりの精度が期待できる。

本調査のテーマであるデジタル・ライブラリでは主目的が資料・蔵書の検索であり、上記のような膨大な規模の消費者を対象にした商品情報検索とは環境条件が非常に異なる。協調フィルタリングは、利用者の履歴蓄積が少なかったり、あるいは検索対象の知名度がなくアクセスが少ない場合にはそれほどの精度が期待できないので、電子図書館サービスに適用するにはさほど大規模でない利用者を前提とした技術開発が必要となる。

3. 電子図書館の目的

わが国の公共図書館は一般利用者への市民サービスという性格が強いため、利用者ニーズの高い書籍、具体的には新刊書や雑誌、漫画など話題性のあるものや日常生活に身近な書籍のサービスが中心にならざるを得ない面がある。市民の教養を高めることに貢献するのは公共図書館の重要な役割のひとつであるが、最近では、ビジネスや産業に直結する高度な技術に関する書籍・資料への関心も次第に表面化しつつあり、地域経済や産業の発展に貢献することも今後の公共図書館の役割として期待されている。その意味では、現在の公共図書館はそのような期待に応じ得る蔵書も十分ではなく、大学図書館と連携することを含め、従来の枠を超えた機能を持つ図書館のあり方を検討すべき時期にきている。

科学技術・学術審議会研究計画・評価分科会の答申「学術情報の流通基盤の整備について」においても、“図書館資料の共同利用”として、“資料の学外者、他大学への貸出”などを積極的に進め、学術情報発信機能の強化として、“一次情報を登録及び更新し、目録所在情報、メタデータなどの二次情報を体系的に電子化して学内外からの利用に供する”ことや、構築された“データベースが公開され、広く利用される”ことの必要性が指摘されている。

大学図書館は、高度に専門的な蔵書を豊富に有し有益な蔵書が多いだけに、学内だけでなく、学外からの利用も含めて図書館の利用を積極化する仕組みを開発する必要がある。

広い意味でのデジタル・ライブラリ関連研究としては、わが国ではいままですでデジタル・アーカイブの研究を中心に進められてきた。デジタル・アーカイブは歴史的資料や古書・貴重書など保存を主要な目的とし、技術的には電子図書館と共通するものが多いが、図書館の蔵書は日常的に利用されることを目的としているため、異なった視点での研究が必要である。

学術雑誌などをデジタル化してインターネット経由で購読できるものは電子ジャーナルと呼ばれ、欧米ではかなり普及している。これは、研究開発に利用したり、技術調査などに非常に有効なものであるが、資料・書籍の利用時には実物を手にしなければならないこともあれば、資料・書籍に関連する情報だけで足りることも多い。従って、自分の必要な資料がデジタル化資料あるいは非デジタル化資料を含めて、どこにどのような形で存在するかを容易にかつ迅速に検索することができなければならない。

4 電子図書館に関する要素技術

4.1 書誌情報関連技術

電子図書館における情報検索は書誌情報による資料・蔵書検索が基本である。例えば印刷媒体による蔵書を有する従来の図書館において、数 10 万冊あるいは数 100 万冊の蔵書が所蔵されていても、分類に従って整然と配置されているために利用者が容易に閲覧蔵書を探し出せることを見ても書誌情報（メタデータ）の重要性がわかる。Google の全文検索の動きは従来の図書館にたとえて言うなら、30 億冊の蔵書が開架式書棚に無造作に並べられ、蔵書要求があるたびに片端から順に要求蔵書があるかどうかを探す動作をしているのと同じで、効率という点では必ずしも良いとはいえない。

日本では書誌情報は科学技術情報流通技術基準（SIST: Standards for Information of Science and Technology）によって科学技術論文などの書誌情報が決められつつある。図書館情報システム向けに米国標準の MARC (Machine Readable Cataloging) があり、わが国では国立情報学研究所が採用しているが、基本的には国家規模レベルの大規模システムを前提とした規格である。

一般の図書館でも利用できる書誌情報のメタデータ記述としては、Dublin Core がある。海外ではかなり利用されているが、わが国ではまだほとんど利用されていない。

Dublin Core は DCMI (Dublin Core Metadata Initiative) によって制定されたメタデータ記述の標準仕様である。国際標準ではないが、英、豪、加、

デンマーク、フィンランド、アイルランド、ニュージーランドの政府が標準として採用しており、今後国際標準になる可能性は高い。DCMI は 1994 年に WWWC2 でネットワーク上の情報資源を記述する共通のメタデータ標準が議論されたのを機に組織された。Dublin Core は翌 1995 年に米オハイオ州ダブリンで OCLC/NCSA Metadata Workshop が開催されネットワーク上の情報資源に関する最低限必要な項目として提案されたものである。仕様は Dublin Core Metadata Element Set, Version 1.1 として纏められ、下記の 15 項目 (Element set : 基本要素) からなっている。

Dublin Core Metadata Element Set

項番	項目	内容
1	Title	情報資源の名称 (書名、論文名など)
2	Creator	情報資源の内容責任者 (著者: 個人、機関など)
3	Subject	情報資源の内容を表すトピック (キーワードやキーフレーズ、あるいは標準分類コードで表現)
4	Description	情報資源の内容 (アブストラクトや目次など)
5	Publisher	情報資源を利用可能にした責任者 (出版社や出版責任を持つ個人、機関など)
6	Contributor	情報資源の内容に貢献した責任協力者 (個人、機関など)
7	Data	情報資源が作られた日付または現在の形になった日付
8	Type	情報資源の内容の性質 (一般的なカテゴリーや機能) やジャンル。 情報資源の内容が集成物であればその集成レベル。
9	Format	情報資源の物理的形式あるいはデジタル形式。通常情報資源のメディア・タイプやディメンジョンなどで、情報資源を表示したり動作させるときに必要なハードウェアやソフトウェアがわかるもの。ディメンジョンの例は情報資源のサイズや存続期間など。
10	Identifier	情報資源を一意的に表すもの (URI (含 URL)、DOI、ISBN など)
11	Source	情報資源があるソースの全体または部分から派生した時のソース
12	Language	情報資源の内容を表現している言語 (英語(en)、仏語(fr)など)
13	Relation	関連する情報資源
14	Coverage	情報資源の内容が有する (適用される) 範囲と視野。地理的・時間的範囲や (行政でいう) 支配権や管轄権のおよぶ範囲など。
15	Rights	情報資源や情報資源内に含まれている権利関係 (著作権、コピーライトその他) に関する記述

Source: Dublin Core Metadata Element Set, Version 1.1: Reference Description

基本要素の数が 15 では少ないという指摘もあり、15 項目の基本要素をそれ以上分けない考え方を Dublin Core Simple (DCS)、15 項目からさらに細かく分ける考え方を Dublin Core Qualifier (DCQ) という。DCQ としては DCMI 利用委員会で Element Refinement と Element Encoding Scheme の 2 案を提示しているが、研究機関によっては独自でさらに細かく分けたり、あるいは 15 項目に対して独自の拡張を行なっている例もある。Element Refinement または Element Encoding Scheme を適用しても、独自の拡張を行なっても、Dublin Core Simple を採用している図書館と交信する場合は変換を行なう必要があるため、それも考慮して拡張する必要がある。

また、Dublin Core には実装上の規約が含まれていないので、今後各図書館が Dublin Core を採用する場合には、さまざまな実装方法がとられることも考慮する必要がある。今後は XML による文字列検索メタ・データベースが増えていくと予想されるが、SGML による場合や、リレーショナル・モデルの構造を持つ場合もある。また、各図書館の情報蓄積も、Web 文書やテキスト、静止画などさまざまな形で行なわれているため、情報の構造表現もさまざまな形態がある。

電子図書館においては、1 図書館内だけでもさまざまなデータベースが存在するが、利用者は各データベースに対して、その構造を意識することなく、かつ横断的に情報検索することが必要である。これが可能になると、利用者は 1 データベースに対して検索して必要情報が得られなかった時に再度他のデータベースに同じ検索要求を出すことなく、一度の検索要求で複数のデータベースに対して、同時に検索を行なうことができるようになる。さらに、1 図書館内だけでなく、図書館間の情報検索も統一できる可能性がある。現在、図書館間の蔵書検索は国立情報化研究所の NACSIS Webcat において実現されており、NACSIS Webcat では全国 855 機関の 5,200 万図書の総合目録データベースを有している。しかし、このように大規模な蔵書目録はすべての図書館で準備することは不可能で、現在の通常の図書館では、せいぜい学内あるいは 1 自治体の管理下にある複数図書館での蔵書検索ができる程度に過ぎない。今後、各図書館がこのような負担を負うことなく図書館間の連携を可能にし、また 1 図書館内でもさまざまな構造を持つデータベースにアクセスすることができるための技術開発がなされていくと考えられる。

以上を可能にするための基本技術の一つとして Z39.50 がある。Z39.50 は ANSI (American National Standards Institute) で規格化された米国標準で、

クライアント/サーバー間の通信プロトコルを規定している。Z39.50 は基本的に英語を前提としている規格であるため、英語圏ではかなり使われているが、わが国では先進的実験例がある程度に過ぎない。Z39.50 は OSI モデルのアプリケーション層に位置づけられ、情報検索時に利用者の端末とデータベース・サーバーの間との通信規約として 1988 年にバージョン 1 が制定され、現在は 1995 年に拡張されたバージョン 3 が使われている。Z39.50 は現在でも拡張の検討がなされており、2002 年 5 月に次バージョンの Draft Standard の投票が行われた。Z39.50 は機能的には 11 の Facility からなっている。実際の検索サービスを実現するときには、これらすべての機能を用いる必要はなく、Z39.50 の開発者が実現するサービスに応じてその中から選択することができる。

Z39.50 の機能

	Facility	Service	内容
1	Initialization	Init	初期化处理
2	Search	Search	検索要求処理
3	Retrieval	Present	検索結果の返戻 (へんれい)
		Segment	イメージデータやマルチメディアデータなどサイズの大きい検索レコードに対するセグメンティングの指示
4	Result-set-delete	Delete	検索結果の一部または全体の破棄の指示
5	Browse	Scan	用語や属性リストに従った、データベースまたはデータベース間のスキミング
6	Sort	Sort	クライアントの要求に従った複数検索結果のソートとマージ
7	Access Control	Access-control	アクセス制御
8	Accounting/ Resource Control	Resource-control	リソースを制御する範囲とその処理内容
		Trigger-resource-control	検索作業の開始や停止の指示
		Resource-report	Z39.50 手順のどの段階にあるかの照会
9	Explain	<i>uses the Search of Search and Retrieval</i>	サーバーの実装 (検索可能なデータベース、アトリビュートセット、スキーマ、レコード・シンタックスなど) やサーバーの利用可能時間や特別な処理要求ができるサーバーなどの情報取得処理
10	Extended Services	Extended-services	検索結果の後日出力、周期的検索、デー

			データベース更新、印刷指示などの拡張処理
11	Termination	Close	終了処理

(注) 本表は ANSI/NISO (非公式) 資料他を基にした。内容は規格上の表現とは異なる。

現在、米国で提供されているサービスに共通に含まれる最も基本的なサービスは (1) Init (2) Search (3) Present である。

Init は、サービス要求に先立ち、クライアントからサーバーに対して、自分が何者であることを知らしめて、合意を得るためにパスワードを渡し、一連の認証手続きを行なう。Search はクライアントからサーバーに対して問い合わせを行なうサービスである。この処理では、サーバーに蓄積された Result set を含んだ形で検索を行なうことができる。Present は検索結果をクライアントに返す処理を行なう。この時、クライアントは検索の内容やフォーマットを指示したり、どの検索要求する Result set のレコードを指定することができる。また、数多くの検索結果や、サイズの大きい検索結果に備えてクライアント/サーバー間でバッファの取り決めを行なうことができる。

Z39.50 の特徴は Stateful かつ Connection-oriented という性質にある。ある検索要求によって検索されたレコードの集まりは“Result set”と呼ばれ、サーバー内に保存される。この Result set は後に続く検索要求でも利用することができる。この点は単にブラウジングを目的とする Web サーバーや SQL サーバーと基本的に異なる。また Connection はクライアントとサーバー間だけでインタラクションの定義が行なわれるので複数のサーバーと同時にセッションを張ることができる。これらの機能をソフトウェアで実現することにより、従来の検索サービスに比較して複数のデータベースやサーバーを横断する高度な検索が可能となる。

以上の基本機能をベースにして、そのほかの諸機能を利用すれば構造の異なるデータベースや図書館間の情報検索連携が可能と考えられる。ただし、Z39.50 はあくまで基本的なプロトコルを規定しているだけであるため、その特徴を十分に活用し、さまざまな電子図書館サービスを実現するためのアルゴリズムなど技術開発を行なう必要がある。

Z39.50 では、ある検索語 (Search term) が与えられた時、検索エンジンが正しくデータベースから検索するために、与えられた検索語の性質や条件などの属性を与える規定として Attribute Set がある。代表的な Attribute Set が下記の Bib-1 Attribute Set である。

Bib-1 Attribute Set

Attribute Type	内容
Use	Search のためのアクセスポイント (title, author, subject etc)
Relation	Search term と Database Index 内の value との関係 (less than, greater than, equal to, phonetically matched etc)
Truncation	Search される Index 内の value のどの部分が (the beginning of any word in the field, the end of any word in the field etc)、例えば "Smith" を Search する時、Index 内の value は Smith で始まるのか、Smith で終わるのか 等
Completeness	Search term を Index 内の特定の value に限定するかどうか
Position	Index field のどこから Search するか (at the beginning , anywhere etc)
Structure	Search されるものの構造 (a word, a phrase, a date etc)

(注1) 本表は LOC 資料(Bib-1 2001.9.9 版)他を基に作成。内容は規格上の表現とは異なる。

(注2) value は Bib-1 attribute set によって決められた数値。例えば、個人名:1、企業名:2、会議名:3、タイトル:4、--- ISBN:7、ISSN:8、--- 出版日:31 --- OCLC Number:1211 など。5,000 --- 10,000 は個別ユース(ZIG による Interoperability 保証なし)。Dublin Core 用は Title:1097、Creator:1098、Subject:1099 --- 略 --- RightsManagement:1111。

Attribute Set Bib-1 は米国議会図書館 (LOC)、OCLC、RLG などが 1980 年代半ばに作成したものを、欧州規格を取り入れて Bib-1 とし、Z39.50-1992 に規格として組み入れられた。当初は図書館用に作成されたが現在は図書館だけでなくさまざまな用途に使われている。Bib-1 は 1992 年当時明確に意味づけしたドキュメントがなかったので、実装したときに Attribute の使い方の上で混乱が生じたため、Bib-1 Semantic Document を作成されたが、これはいまだに公式ドキュメントではない。Bib-1 Attribute Set はその後も拡張されて、1995 年に現在の仕様として確定した。しかし、Bib-1 は第 1 儀的に蔵書検索を前提としているため汎用的な情報検索としては限界もあり、米国化学協会などによる科学技術関連情報検索向け STAS (The Scientific and Technical Attribute and Element Set) や、米政府刊行物の公的アクセスサービス GILS (Government Information Locator Service) などの Attribute Set も作成されている。現在の Bib-1 では Query としては Type-1 query を前提としているが Type-1 の限界もあり、ZIG (Z39.50 Implementors Group) では GILS や STAS との共通化を含めて新しい Attribute Architecture を検討している。

今後は、ZIG の動きなども見ながら Z39.50 や Bib-1 Attribute Set およびその拡張仕様を含めて電子図書館サービスへの適用可能性が検討されていく

と予想される。また、Z39.50 はクライアント/サーバー間だけのプロトコルを規定しており、クライアントと実利用者との間のユーザ・インタフェースは規定していない。従って、Z39.50 の特徴を十分生かした利用者インタフェースをあらたに開発する必要がある。

4.2 メタデータ関連技術

検索情報適応技術としては、Semantic Web 技術の適用は重要課題である。Semantic Web は Tim Berners-Lee 氏が提唱している技術で、同氏はすでに 1994 年 9 月に CERN で開催された第 1 回 WWW 国際会議の基調講演で “ Web Semantics ” について触れているなど Web の開始当初から Web における Semantics を課題と考えていたようである。1998 年のいくつかのスピーチでは “ Metadata ” は “ Machine understandable information ” と明確に位置づけており、この頃から Semantic Web を提唱し始めたといわれている。なお、1996 年の “ The World Wide Web: Past, Present and Future ” では、子供たちを Web の有害情報から守ることの重要性と、その対応としての W3C の Platform for Internet Content Selection (PICS) initiative に触れており、それが可能になるのは、PICS ラベルが HTML と異なり、Machine-readable であるからだと言っている。また、サーチエンジンが巨大なインデックスの中から高速に情報検索することへの有能と同時に多くのジャンクを生じさせる無能にも触れ、その無能の原因は単に語彙だけをみてサーチし、ドキュメントの品質に対するコンセプトがないからだとも指摘している。“ An early but long term goal of the web development ” は人々の知識や協働作業を正しく反映して、それによって人々の協働の分析や作業を支援することであり、その第一歩が “ a machine-readable form with defined semantics ” を導入し、Machine や Filter software が読めるようにすることであると述べている。

現在、W3C では主要課題として Semantic Web に取り組んでおり、W3C の Director である Tim Berners-Lee 氏は最近の各地のスピーチでも必ず Semantic Web を強調している。

海外などの研究機関では、研究結果などを広く Web で紹介することなどが日常的に行なわれており、最近日本でも大学などを中心に研究結果や注目される技術の紹介や研究者のプレゼンテーションを Web に掲載したり、海外規格の翻訳を私的に掲載するなどの傾向がでてきている。このような利用者の特性を有効に活用することができれば、検索精度が非常に向上することが期待できる。

検索対象の資料・蔵書書誌情報として Dublin Core を適用する場合は、基

本的には 15 の基本要素で表すことになる。検索要求時に書名などが明確にわかっていれば検索対象にたどり着けるが、書名が分からないことも多いし、特定分野に関連した資料・著作を検索したい場合もあるので、Dublin Core を採用すればすべてが解決する訳ではない。このような場合に、多くの利用者が対象物に対してさまざまなコメントを付けることができ、さらにそのコメントを資料・蔵書検索に利用できれば、利用者の要求に対する関連付けが可能になり、適合度も精度の高いものになる。本研究では、以上の協調動作を可能にする技術の一つとして Semantic Web 技術の適用を検討する。

Semantic Web 技術ではメタデータを記述・交換するためのフレームワークとして RDF (Resource Description Framework) が決められており、その基本的なデータ・モデルは Resource、Property、Statement からなっている。Resource は URI(Uniform Resource Identifier)を持つものならば何でも良い。HTML で書かれた Web ページの全体やその一部であってもよいし、Web サイトの集合体でも良い。印刷書籍のように直接 Web 経由でアクセスできなくても良い。Property は Resource を表現するための特徴や属性、関係を記述するもので、Resource の Value や Type あるいは他の Property との関係を定義するものである。Property は RDF Schema 仕様として別途定められる。Resource が Property と、その Property の示す関係における Resource の Value を伴っているとき、この全体を RDF Statement という。この時の Statement の 3 つの要素すなわち Resource、Property および Value をそれぞれ Subject (主語)、Predicate (述語) および Object (目的語) という。Statement の Object (すなわち、Property value) は URI で定義された Resource でも良いし、文字列や XML で定義されたデータタイプのような literal でも良い (ただし、XML の markup による literal は RDF processor では単に literal とみなしている)。

例えば、

Ora Lassila is the creator of the resource <http://www.w3.org/Home/Lassila>.

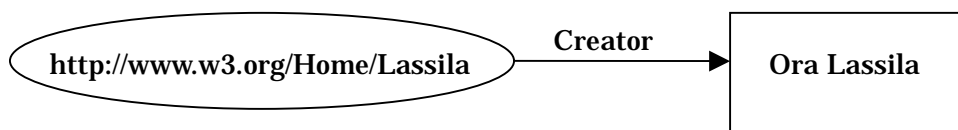
の場合は、下記の通りとなる。

Subject (主題 Resource)	http://www.w3.org/Home/Lassila
Predicate (述語 Property)	Creator
Object (目的語 literal)	“ Ora Lassila ”

これを、RDF/XML 形式では下記のように表現できる。

```
<rdf: RDF>
  <rdf: Description about=" http://www.w3.org/Home/Lassila" >
    <s: Creator> Ora Lassila </s: Creator>
  </rdf: Description>
</rdf: RDF>
```

また、この RDF Statement を Diagram で表現すると、下記のようになる。

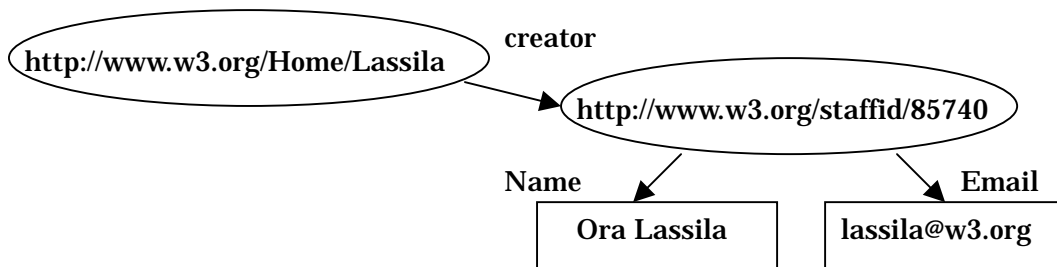


RDF モデルにおいては、目的語が他の Statement の主語になることができるし、主語が同時に 2 つの目的語を持つことができる。従って、Statement を組み合わせることで複雑な表現が可能である。例えば、下記 2 つの文章を一連の RDF モデルで示すことも可能である。

The individual referred to by employee id 85740 is named Ora Lassile and has the email address lassila@w3.org.

The resource <http://www.w3.org/Home/Lassila> was created by this individual.

これに対する RDF モデルは下記となる。



(注)上記の例文、RDF/XML 形式、RDF モデルダイアグラムはすべて “ Resource Description Framework (RDF) Model and Syntax Specification ” (W3C) による。

RDF モデルでは、基本となる単純な RDF モデルを組み合わせることで複雑な記述が可能となり、XML を使って表現することができる。重要な点は、RDF Schema を定義することで、だれでも自由にアプリケーションを構築でき、これらの Statement をコンピュータが理解できることである。

現在、何の条件も与えずに Google の日本語ページで“ RDF ”で検索してみると、18,900 件がヒットされるが、トップ 10 件のうち 2 件は RDF (Refuse Derived Fuel:ごみ固形燃料) である。この 2 件はいわゆるノイズである。逆に、燃料関係者から見ると 8 件がノイズである。もし、このとき検索エンジンが、その利用者が Semantic Web の研究者であることを予め知っていれば、ごみ固形燃料がヒットされることはない。利用者の経歴や実績、あるいは現在関心を持っているテーマなどは RDF モデルを使って登録しておき、検索エンジンがそれを理解したうえで検索を実行すれば検索精度は非常に向上すると考えられる。また、W3C の Semantic Web Activity における RDF Schema についての新しい動きを察知したいときなども、エージェントを利用することで、その情報を自動的に収集することなども可能である。また、研究対象としての Semantic Web については最近多くの研究者が Web にさまざまな意見、コメントなどを掲載しているので、それらを検索エンジンが理解できれば、研究者が Semantic Web のある特定分野についての情報を入手したいときでも有効に利用できる。

Semantic Web はいくつかツールなども提供されているが、まだ研究途上の技術であり、特にキラーアプリケーションが見つからないということが研究者共通の重大事である。従って、その関心の高さにも関わらず Semantic Web そのものの評価が定着していないのが実情である。本研究の対象である Dublin Core も Semantic Web の基盤技術として主に電子政府関係で各国の導入が進みつつあるが、RDF モデルはより広い分野での応用が期待されており、さまざまな立場からの研究が急がれている。

Semantic Web は当初から Information Filtering とのかかわりが深く、今回の研究開発でより進んだ電子図書館サービスを実現できる可能性は高い。

本研究開発では、電子図書館における検索情報適応技術への Semantic Web 技術の適用可能性を検討し、検索情報適応技術に最適な RDF モデルや RDF Schema、また、多くの人々が情報を共有しエージェントを利用して自動的に相互運用するためのオントロジー言語などの研究を実施する。

なお、最近の商品マーケティングでは、Web 分析ツールも試みられている。

これは、利用者の Web に対するアクセスの行動を分析するもので、ログやアクセスパスを分析するなどいくつかの提供されている。まだ市場に出始めた段階でそれほど実績もなく、分析の精度や規模、性能といった面からまだ改善の途上にあるが、本研究開発ではこのような動きも考慮して研究を進める。

4.3 情報発信技術

情報発信技術の一つとして CDF が研究されている。CDF は物理的な Web サイトのディレクトリ構造とは無関係にコンテンツ情報を構造化して提供する試みであり、W3C で標準化を検討している。

CDF (Channel Definition Format) は文字通り Channel を定義するためのフォーマットである。ここで、Channel とは利用者に情報を提供する窓口となる Web サイト (あるいは Web ページ) を示している。CDF の Channel は、TV のチャンネルにおいてさまざまな番組が編成されて提供されているように、さまざまな Web の情報を編成して提供できるようにすることが目的である。CDF は、Microsoft が原案を作り 1997 年 3 月に W3C に提案した。急速に膨張する Web ページの世界では Web publisher や content author が自分たちの提供するものをユーザに見てもらうことが非常に難しくなりつつある。そこで、見やすい形でしかも active に、ユーザに対してアピールする技術の基盤として考えられたものである。従って、Web publisher や vendor などの間ではかなり広い支持を得ている。

CDF では、Channel を構成する要素として、下記の 7 つの構成要素 (major element) を規定している。各 element は XML で記述されるので、CDF は XML のアプリケーションの一つである。

CDF Channel の Element

Element	内容
Channel	Channel 自身の定義 (Channel の Title, Abstract, URI など) CDF ファイルの階層構造 (Subchannel) を持った時は Top-level.
Item	Channel のコンテンツ (Item を表す Title, Abstract, URI など)
UserSchedule	利用者への情報配信スケジュール (monthly/weekly/daily)
Schedule	情報提供者側の情報有効期間や更新スケジュール (任意)
Logo	情報提供者側のロゴ (Channel や Channel 内の Item に対応)

Tracking	Channel に対する利用者の tracking
CategoryDef	“ News ” “ Sport ” などのカテゴリ

(注) 内容は W3C、Microsoft 資料他から抜粋して作成した。規格上の表現とは異なる。

CDF の Channel は CDF ファイルによって表現された Web サイトということができる。CDF ファイルは多くの Web ページの階層を定義付ける。従って、電子図書館のようにさまざまなデータベースからなる集合に対しては、サービス提供を行なう際に、それらのうちから利用者の関心あるページを集めて一つの Channel を構成することが可能となる。CDF は本来ニュースや天気予報など頻繁に内容が変わるものに対して、利用者が毎回各 Web サイトにアクセスしなくても変更内容を自動的に利用者のデスクトップに表示したりサーバーに蓄積することができるようにするためのフォーマットであるが、単なるフォーマットの規定であるため、CDF を利用した新たなアプリケーションを開発することで利用者に新しいサービスを提供することができる。

その他の情報発信関連技術としては、OSD (Open Software Description Format) の適用可能性について研究する。OSD は Microsoft と Marimba が 1997 年 8 月に W3C に共同提案したソフトウェア配布に関するデータフォーマット規定である。OSD は XML で記述されるので XML アプリケーションの一つである。OSD で規定されているは Element 下記の通り。

OSD Element

<Major Element>

Element	内容
SOFTPKG	ソフトウェア・パッケージ全体の定義 (名称、格納 URL など)
IMPLEMENTATION	SOFTPKG の Child element で、ソフトウェア・パッケージの実装指定 (Browser, OS など)
DEPENDENCY	IMPLEMENTATION または SFTPKG の Child element で、ソフトウェアが正しく配布されたり、正しく動作するために必要な関連ソフトウェア・コンポーネントとの関係の定義

<Minor Element>

Element	内容
TITLE	ソフトウェアパッケージの名称 (Title Bar に表示される文字列)
ABSTRACT	IMPLEMENTATION や SFTPKG の Child element になり、ソフトウェアの実装やソフトウェア自身の説明の要約を示す文字列

LICENCE	ソフトウェアのライセンス情報や著作権情報が取得できる場所 (URL)
CODEBASE	IMPLEMENTATION の Child element で、インストールされるファイルや配布ユニットのファイル名称や格納場所 (URL) を規定する
OS	ソフトウェアがサポートしている OS
OSVERSION	ソフトウェアがサポートしている OS のバージョン
PROCESSOR	ソフトウェアが動作するプロセッサ
LANGUAGE	ソフトウェアを国際的に配布する時に必要な言語 (英語: "en" など)
VM	ソフトウェアの実行に必要な仮想マシン
MEMOSIZE	ソフトウェア実行時に必要なメモリ量
DISKSIZE	ソフトウェア配布時に必要なディスクスペースの量
IMPLTYPE	インプリメンテーションのタイプ

W3C のテクニカル・ノート (TN) では、OSD の最終目標について下記のような主旨で記述している。

「ソフトウェアの配布に関する "Pull"-based software distribution は必要なソフトウェアを探し、ダウンロードし、アップデートするという User Action を必要とするが、HTML4.0 の<OBJECT>tag や HTML3.2 の<APPLET>tag を OSD リソースのポインティングに使うことで OSD-aware user agent が必要なソフトウェアを (もちろん、ユーザのインフォームド・コンセントを得た上で) 自動的にダウンロードし、アップデートすることができる。ソフトウェアの配布に関してユーザ主体であることは "Pull" であり、自動配布は "Push" である。OSD の最終目標は、Pull であるか Push であるかに関わらず、ソフトウェア・パッケージとソフトウェア・パッケージ間の関連を記述できる XML ベースの Vocabulary を提供することである。CDF File と OSD-based software package を組み合わせることで True-push を実現することができる。」

CDF や OSD のフォーマットを活用することで、電子図書館サービスの利用者端末にも検索情報に対応したソフトウェアを配布することができるので、電子図書館サービスの利便性の向上が期待できる。ただし、CDF も OSD も共にフォーマットの定義であるため、これらを有効に使うアプリケーションの開発が必要である。例えば、Microsoft 社ではこれらを適用して、ニュースやスポーツ記事の自動配信やソフトウェアのアップデートできる技術 (Active Channel Technology) を開発し、Internet Explorer に実装している。

Personalization 対応として、利用者ごとに異なる情報からなるページは動的コンテンツといわれ、従来アクセスがあるたびにサーバーがページ生成を行っていたが、最近はレスポンスをあげるためにページのうち静的な部分と変

化する部分を分けてそれぞれ別のキャッシングを行なう方法がとられ始めている。いくつかのソフトウェア製品がサーバーに搭載されているが、まだ動作環境やキャッシュ対象がそれぞれ独自で、電子図書館向けにそのまま適用できる段階にはない。

5 電子図書館関連技術の動向

5.1 Dublin Core およびメタデータに関する動向

米国においては 1994 年に NSF/NASA/ARPA による Digital Library Initiative フェーズ 1 (DLI1) (現在はフェーズ 2 段階にある。<http://www.dli2.nsf.gov/>) が開始され、新しい情報技術開発を行っている。このプロジェクトにおいてメタデータの扱いは主要な研究課題となっている。イリノイ大学、ミシガン大学、スタンフォード大学、カリフォルニア大学サンタバーバラ校などがメタデータデータを用いた記事検索やあるいはメタデータそのものを研究課題として研究を進めている。

EU においては、1999 年より、Information Societies Technologies 分野の研究としてデジタル・ライブラリへの研究助成が始まった。また EU は米 NSF と共同研究も開始した。EU は ERCIM (European Research Consortium on Informatics and Mathematics) の Digital Library Working Group (DELOS, <http://www.iei.pi.cnr.it/DELOS/>)、NSF 側はミシガン大 (School of information, http://www.si.umich.edu/UMDL/EU_Grant/home.htm) が全体纏めで、両者で (1) Interoperability of Digital Library Systems (2) Metadata (3) Intellectual Property Rights and Economic Issues (4) Resource Indexing and Discovery in a Globally Distributed Digital Library (5) Multilingual Information Access の 5 共同 WG を構成している。ここでもメタデータは中心的な課題である。

日本では図書館情報大学で 1994 年 8 月以来デジタル図書館ワークショップを継続的に開催しており、2002 年 3 月で第 22 回を数えている。その間、海外の研究者による招待講演や、タイのアジア大学や奈良先端技術大学院大学における共同ワークショップの開催など行っている。また同大学では 1995 年、1997 年、1999 年にデジタル図書館の国際会議 ISDL も開催した。最近ではメタデータに関する議論などを積極的に展開している。

実際に Dublin Core を実験した例には、国文学研究資料館の「国文学研究資料館コラボレーションシステム」がある。これは国文学研究資料館が、東京大学史料編纂所、国際日本文化研究センター、歴史民族博物館、大阪市立大学メディアセンターなどと共同で行なっている歴史系データベースの作成に関する

る相談会の活動の一環として実施されている。また、図書館情報大学では、経済学関連のデータベースである WoPEc、および米国の政府刊行物を対象にした WAGLIS のメタデータと Dublin Core の対応付けなどの研究を実施している。特に自然科学系を対象にした Dublin Core 研究は見当たらない。

5.2 Z39.50 に関する動向

Z39.50 は ZIG (Z39.50 Implementors Group) で継続的な研究が行なわれている。ZIG は非公式な会合であるが、Z39.50 の公式な管理機構である Z39.50 Maintenance Agency (米国議会図書館の 1 機構) と密接な連携の元で、毎年オープンな国際会合を持つなど活発な動きをしている。

なお、Z39.50 は、米国 ANSI (American National Standards Institute) の下部機構として、情報、ドキュメンテーション、図書館などを専門的にあつかう規格制定組織 NISO (National Information Standards Organization) が制定した規格である。米国図書館協会、国立図書館 (議会図書館など) や米国科学会、主要企業などが投票して規格を制定し、ANSI が認証している。ここで制定される規格は分類番号「Z39」となる。

日本では国文学研究資料館の「国文学研究資料館コラボレーションシステム」において Z39.50 を採用し、Dublin Core との連携も含めて研究している。大学では、OPAC で Z39.50 を適用しているサイトがあるが、海外大学との連携において、大学名を指定した上で英文資料を検索する機能に限られる。日本ではまだ、Z39.50 は先進的な研究機関が採用を試みている段階であり、その機能を十分活用しているとは言えない段階である。

5.3 Semantic Web に関する動向

米国では DAML (DARPA Agent Markup Language) の研究プロジェクトが 2000 年 8 月にスタートしており、DARPA から 7,000 万ドル (5 年間の総額) が拠出される。EU においては、すでに第 5 次 RTD (Research and Technological Development) (1998 ~ 2002 年) の Framework Program (FP5)/User friendly information society (IST Program) で推進されていたが、2002 年 8 月に第 6 次 RTD がスタートし Semantic Web 研究のために Semantic Web Advanced Development (SWAD)-Europe プロジェクトが開始された。W3C では 2001 年 2 月に Metadata Activity Group を Semantic Activity Group と改称し、同年 8 月には Web Ontology Working Group (WOWG) を立ち上げている。W3C の

Director である Tim Berners-Lee 氏は、2003 年 1 月の米 SIA (Software and Information Industry Association) のミーティングで、現在の W3C のホットトピックは Voice, Multimodal, new devices Web Services Semantic Web として、その中でも特に Semantic Web を強調している。現在の W3C は主に Semantic Web を中心に活動しているといっている。

日本では、昨年人工知能学会に“セマンティックウェブとオントロジー研究会”が設立され、11月6日に第1回研究会が開催された。研究会委員は8大学(は大阪大学、静岡大学、京都大学、法政大学、大阪市立大学、名古屋大学、筑波大学、前橋工科大学)、2研究機関(国立情報学研究所、産業技術総合研究所)および3企業(NTTコミュニケーション科学基礎研究所、米国富士通研究所、日本ユニシス)からなっている。他には慶応大学がW3Cのアジア拠点でもあることから、INTAPと共催で2001年11月にはW3Cの公式セミナーであるW3C Dayを開催したり、2002年10月にはSemantic Conferenceを開催するなどSemantic Webの普及促進に努めている。INTAPは電子政府応用の面から調査研究を行なっている。

5.4 情報フィルタリングに関する動向

検索ロボットを用いる商用サービスとしては、Googleが世界最大の検索エンジンとして有名である。

あるページの重要度を評価する場合、そのページが他のページからリンクされている数が多いほどページの価値が高いと判断することができる。この手法そのものは、論文の評価などで引用回数が多いほど論文の価値が高いとする手法と同じでGoogle以前から存在する手法であるが、インターネットでは、多くのダミー・サイトを作り、そこからリンクすることであたかも多くのサイトからリンクされているかのように見せかけることが簡単にできる(このようなページをScam Web Pageと言う)。そこで、GoogleではこのScam Pageを排除するために、リンクしているサイトがどれだけ他のサイトにリンクしているかを考慮し、他のサイトに多くのリンクを張っているサイトからリンクされたページの方が重要度が高いと判断する技術(PageRank)を開発・採用している。この技術によれば、ダミー・サイトの価値はほとんどなくなるので、Scam Web Pageの重要度評価は非常に低くなる。利用者にとっても、ヒットしたランク上位のページは多くのリンクを持つので関連情報が多く入手できるという利点もある。Googleは1998年9月にLinux PCサーバー30台、収集ページ5,000万ページでスタートしたが、現在はサーバー10,000台以上、収集ページ30億ペー

ジ以上を有し、毎日 1 億 5000 万件以上の検索要求を処理している。

キーワード検索型検索エンジンの検索効率と正確さの向上のための手法として、SEO（検索エンジン最適化 Search Engine Optimization）が注目されている。これは各ページがロボットにより検索されやすく、また検索されたときにランクの上位に設定されやすいようにページを作成する手法である。多くの検索ロボットに収集されやすいようにキーワードを文章の上位に配置したり、テキスト中のキーワードの出現頻度調整、索引付けを誤りなくするために無駄なタグの省略などのほか、ページデザインやレイアウトに至るまで、検索ロボットの特性を熟知し、それに適合するようにページを作成する。本来は検索ノイズをなくし、検索効率を上げるための手法であるが、一部業者による悪用などにより重要度の低いサイトが上位ランクでヒットされることもある。検索エンジンの特性を熟知して悪用する手法（Search Engine Spam）も日々進化しており、検索エンジン開発者との競争状態にある。

なお、キーワードによる検索エンジンのほか最近には以下に示すようにキーセンテンス（文章）入力による情報検索エンジンも使われ始めている。

サンフランシスコ SOMA の H5 Technologies は、収集した情報の分類に際し、テキストに記載された主要キーワードを抽出し、AI 手法を用いた H5SMF（H5 Subject Matter Framework）と称する自動分類技術を用いて索引付けを行う。利用者が検索のために文章を入力すると、システムが文章の中のいくつかのキーワードをピックアップして表示する。利用者はその中から自分の必要とするキーワードをクリックすると関連情報が検索されて出力される。性能は、1 分当たり 5,000 ページの速さで文書分類と検索が同時に行なわれる。ただし、この技術は特許出願中ということもあって詳細は公表されていない。

Yahoo や Google と異なり、インターネット上で検索サービスの提供は行わず、企業や政府・官公庁向けに検索ソフトのパッケージ販売や検索ソフトを軸にしたシステム構築サービスを行なっているため、今後もアルゴリズムが公表される可能性は薄い。H5 Technologies は 1999 年創設のベンチャー企業であるが 2002 年夏には米海軍情報局に H5SMF 応用システムを納入している。

Autonomy Inc.も文章入力による情報検索を行なっている。同社の検索技術の根拠は Thomas Bayes と Claude Shannon の理論に基づいた、独自のパターンマッチング技術（Non-linear adaptive digital signal processing）である。同社のソフトウェアは、ある特別な主題やコンセプトに関連して単語や用語がテ

キスト内に通常どう現れるかを認識しており、コンピュータが、あるパターンが他のパターンに対してどれだけ優位かを基にして、そのドキュメントがある主題に対して何%の関連を持っているかを判断できる。こうして、ドキュメントのエッセンスを抜き出して、キーコンセプトの“Signature”としてコード化しているのが、オペレーションが自動化できると述べている。

Bayes は 18 世紀の英国の数学者で、死後出版された“*Philosophical Transactions of the Royal Society of London*” (1763)に確率に彼の研究主題であった多変数と 1 変数が他の変数に与え影響の範囲の関係についての業績が記されている。Bayse 理論では、結果の分布 (a posteriori probability : 事後確率) は各変数による結果の分布 (a priori probability : 事前確率) により計算可能で、今日の典型的応用は、クエリおよびエージェント・プロフィールに対する関連ドキュメントの判断基準を与えるものである。Autonomy 社は、検索ロボットによって収集された結果がどの程度検索要求に合致したものであるかを判断する基準としてこの理論を適用した。

Shannon は“*Mathematical Theory of Communication*” (1949)により、コミュニケーションにおける情報の価値に対して定量的評価を導入して情報理論の基礎を築いたことで知られている。コミュニケーションにおいてその単位 (語または語句) が互いに独立であるときに、その単位の発生確率が分かれば、その単位によって伝達される平均の情報量 (エントロピー) が定義できる。自然言語の会話はもともと冗長度が高いので、ニュース情報などは、この理論を応用することで、容易に元のテキストの最重要コンセプトを抽出したり、自動要約を作成することができる。

Autonomy 社は、1996 年創設のサンフランシスコのベンチャー企業であるが、顧客は GM、エリクソン、独バンク、サーベース、ルーセントなど 575 社、OEM 先 60 社、提携先アクセンチュア、コンパック、IBM など 130 社を有する。

5.5 Personalization および情報発信に関する動向

Personalization 技術の代表的なものは Collaborative Filtering (協調フィルタリング) と Content-based Filtering である。Collaborative Filtering は米 Net Perceptions Inc が実用化し、Amazon.com が採用したほか、日本ではアスクルが採用するなどの動きはあるが、日本での普及は芳しい状況とは言えず、2002 年には Net Perceptions 社は日本法人を閉鎖している。Content-based Filtering は予め、利用者が登録したプロフィール (嗜好情報) とコンテンツとの情報をマッチングした上で、利用者に情報提供するものである。Collaborative Filtering も Content-based Filtering もそれぞれ長所、短所があり、両者を組み

合わせた技術などの開発も行なわれている。

インターネットのサービスとして、利用者の嗜好に合わせた情報を提供するサービスは 1996 年に Yahoo が “ My Yahoo ” としてサービス開始した。My Yahoo では利用者の嗜好に合わせた個人専用の Web ページを構成できる。

最近日本では GPS を使った情報フィルタリング技術や情報家電向けを意識した研究が行なわれている。最寄の病院やショッピングセンターを探すときなど、利用者の位置する地理的条件を加味して情報提供することでより利用者にとって適合した情報が提供できる。また、情報家電向けでは好みの番組の自動録画などの応用も考えられている。

情報発信技術としては、米 PointCast Network 社が 1996 年に Push サービスを開始し、一時は 120 万人以上の会員を集めたが、2000 年 5 月には単独サービスを停止した。日本では、昨年（2002 年）6～11 月にかけてエキサイトや BIGLOBE など 4 社が相次いで Push 型配信サービスを開始した。まだ提供コンテンツも少なく、市場による評価が出るまでにはしばらく時間を要する。

CDF、OSD については、提案元である米 Microsoft 社が Active Channel Technology と称して、CDF の利用技術を開発している。Active Channel Technology では、Offline Browsing, Site Crawling, Webcasting の機能を実現している。Offline Browsing は予めキャッシュ・メモリなどに Web ページを収集しておくもので、利用者はオフラインで Web ページを見ることができる。接続料金の節約や、大容量ファイルなどはダウンロード時間の短縮になる。Site Crawling は自動的に収集したあるページのリンクを見て、リンク先のページを自動的に収集する。Webcasting は TV やラジオと同様に、決まったスケジュールに従って、Web Publisher が利用者に自動的にコンテンツをブロードキャストする。Microsoft 社では、Web publisher に対して、積極的に CDF file 作成を呼びかけている。

6 電子図書館システム研究の波及効果

Dublin Core はまだ国際標準ではないとはいえ、すでに世界 30 カ国（25 言語）でコンセンサスを得られた規格である。また、各国政府は下記の通り Dublin Core をベースとした規準をおのこの制定しており、社会的基盤技術としてのメタデータ研究が促進されている。

- ・米国 GILS (Government Information Locator Service)
- ・英国 e-GMS (e-Government Metadata Standard)

- ・デンマーク OIO-metadata (Offentlig Information Online Metadata)
- ・アイルランド IPSMS (Irish Public Service Metadata Standard)
- ・フィンランド Finnish Dublin Core Extension
- ・オーストラリア AGLS (Australian Government Locator Service)
- ・ニュージーランド NZGLS (New Zealand Government Locator Service)

また、Dublin Core を含めメタデータの可能性として、e - learning、e-business、knowledge management な社会のさまざまな分野で新しい Web の可能性を提供するものと期待されている。しかし、少なくとも日本においては、現実にくつつかのツールなどがすでに提供されているにもかかわらず、その本格的な利用に至っていない。この原因は、効果的な利用方法が確立していないためといわれている。電子図書館システムの研究開発はまさにその利用技術を中心課題とするものであり、実システムとしての検証を行なうことで日本のメタデータの研究と実用化に向けての大きな進展をもたらす効果がある。

7 国家的事業としてのデジタル・ライブラリ

現在、デジタルライブラリの研究が世界的に行なわれているのは 1970 年代初頭に米国議会図書館が所蔵する資料を書誌情報による検索を計画したことや NSF、DARPA のプロモーションが契機となっており、デジタル・ライブラリは常に米国の国策とのかかわりで発展してきたといえる。

図書館の情報化・ネットワーク化は 1960 年代に発足した OCLC (Ohio Colleges Library Center: 当時名称) に見られるように特定の大学間という限定された範囲でその萌芽が見られるが、現在の形で電子図書館の研究が本格的に始まったのは NSF/NASA/ARPA がスポンサーとなって DLI1 (Digital Library Initiative フェーズ 1) をスタートさせた 1994 年ころからである。発足以来 DLI は常に電子図書館の技術発展のリーダーシップを発揮してきたし、現在もフェーズ 2 として新たな技術開発に挑戦している。

米国では、NSF の資金を中心に DARPA , LOC (Library of Congress) 、NASA、NLW (National Library of Medicine)、NEH (National Endowment for the Humanities)、NARA (National Archives and Records Administration)、SI (Smithsonian Institution)、IMLS (Institute of Museum and Library Services)などがスポンサーとなり DLI にもとで 現在下記 15 研究プロジェクトが進行している。

現在進行中の Digital Library Project

	研究機関	研究期間	予算総額
1	UC Berkeley	99. 4. 1 - 04. 3.31	\$5,000,000
2	UCLA	00. 9. 1 - 03. 9.30	\$650,000
3	UC Santa Barbara	99. 9. 1 - 04. 8.31	\$5,400,000
4	CMU	99. 5. 1 - 03. 4.30	\$4,000,000
5	Columbia Univ.	99. 9. 1 - 04. 8.31	\$5,002,375
6	Cornell Univ.	99. 5. 1 - 03. 4.30	\$2,268,608
7	Univ. of Hawaii	00.10. 1 - 03. 9.20	\$349,619
8	Univ. of Illinois	00. 9. 1 - 03. 7.31	\$360,555
9	Indiana Univ.	00.10. 1 - 05. 9.30	\$3,056,913
10	Univ. of Massachusetts	00.10. 1 - 03. 8.31	\$450,000
11	Michigan State Univ.	99. 9. 1 - 04. 8.31	\$3,599,989
12	Univ. of S. Carolina	99. 4. 1 - 03. 3.31	\$1,199,215
13	Stanford Univ.	99. 4. 1 - 04. 3.31	\$4,297,585
14	Stanford Univ.	00.10. 1 - 03. 8.31	\$528,896
15	Tufts Univ.	99. 6.15 - 04. 5.31	\$2,758,400
予算総額			\$38,922,155

(注) 表は DLI2 の下で推進中のプロジェクト。EU 等との共同プロジェクトは除く

なお、過去には 1999 年 4 月から 2002 年 12 月末までに 20 プロジェクトが推進・完了し、投入された総研究費用は \$10,953,259 となっている。完了したプロジェクトと現在進行中のプロジェクトを単純に比較すると、1 プロジェクトあたりの研究費用は約 55 万ドルから約 260 万ドルへと大規模化していることから見て、米国が本研究分野を重視していることがわかる。

なお、EU においてもデジタル・ライブラリは第 5 次 (1998~2002 年)、第 6 次 (2002~2006 年) の RTD (Research and Technological Development) のプロジェクトで推進されている(4.10.1 Dublin Core およびメタデータに関する状況、4.10.3 Semantic Web に関する状況 参照)。

日本における企業横断的電子図書館プロジェクトとしては、1996 (H8) 年から実証実験として下記が推進された。大学においては、現在各大学で電子図書館プロジェクトが推進されている。

日本で実施された電子図書館プロジェクト

プロジェクト名称等	目的	推進元	参加企業・団体
次世代電子図書館システム研究開発プロジェクト (H8~H11 年度)	次世代電子図書館の実証実験	(財)日本情報処理開発協会	日立、富士通、日本電気、日本 IBM、東芝、三菱、沖、日本ユニシス、リコー

BBCC と国立国会図書館の実験協力 (H8～H10年度)	電子図書館におけるデータのデジタル化作業に関する実証実験	BBCC(新世代通信網実験協議会)	BBCC、国立国会図書館
----------------------------------	------------------------------	-------------------	--------------

今後の図書館のあり方としては、2000(H12)年12月に地域電子図書館構想検討協力者会議が、主に地域の公共図書館の今後について「2005年の図書館像」として報告をまとめている。ここでは、電子化された蔵書が子供から高齢者まで幅広く、容易に利用できることで生活に密着した図書館のあり方や、大学図書館と連携することで大学図書館の書籍を借りたり、遠隔学習が可能となるなど生涯教育の場としての図書館像も示されている。大学の図書館に関しては、2002(H14)年3月に科学技術・学術審議会研究計画・評価分科会の答申「学術情報の流通基盤の充実について」に詳しくまとめられている。この報告によれば、日本の国立大学では平成7(1995)年より資料の電子化を進めているものの、蔵書目録情報の電子化は入力対象冊数約6,300万冊のうち入力済は3,300万冊余り(53%)でまだ半数が入力されていない状況にある。

昨年の国立国会図書館関西館の開設や科学技術振興事業団のデジタル・ライブラリ J-STAGE など新しい動きはあるが、欧米との格差はまだ大きく、一日も早く欧米と比肩するために各大学や図書館の枠を超えた国家的見地からの研究推進が非常に重要である。

8. 日本における電子図書館の現状

現在インターネット上で構築されている WWW によるオンライン蔵書検索(OPAC)は下記の通りである。

わが国で構築されている WWW による OPAC

(2002年12月現在)

図書館の種類	OPAC 構築数	(参考)図書館数	率(%)
大学・短大・高専図書館	177	1,540	11.5%
都道府県・市区町村立図書館	125	2,609	4.5%
その他(国立図書館、研究機関の図書館など)	51		
計	353		

OPAC : Online Public Access Catalog

米国などでは、すでにほとんどの図書館がオンライン化されており、全体

としては日本の電子図書館化は欧米に比し相当遅れている。ただし、下表の例に見るように、ホームページ検索数や簡単な蔵書検索数などはほぼ来館者数に匹敵するほどに普及してきた。

ネットワークによる図書館の利用（2001.4～2002.3）

	東京都立図書館	大阪府立図書館
ホームページ検索回数		1,014,909
インターネットからの蔵書検索件数	1,120,929	396,303
館内 OPAC からの蔵書検索件数	約 1,840,000	
(参) 来館者数	約 1,380,000	1,017,710

(注)都立図書館の OPAC 検索件数、来館者数は日比谷図書館開館日数 117 日を年換算推定

日本の大学・研究機関向けの電子図書館サービスの代表的なものは国立情報学研究所（NII：National Institute of Informatics）の提供するサービスである。電子図書館サービス NACSIS-ELS (NACSIS Electronic Library Service) は、日本の人文科学、法学、経済学、理学、工学、農学、医学の主要な学会のすべてを対象に学会が発行する学術雑誌から論文を検索できるので、多くの大学図書館のホームページからリンクされている。蔵書を検索するために必要な蔵書目録に関しては、国立情報学研究所目録所在情報サービス NACSIS-CAT (NACSIS Catalog Information Service) がある。このサービスは全国規模の総合目録データベースを形成している。2001 年 3 月末現在で参加している図書館は 855 館、対象蔵書数は 5,200 万冊である。NACSIS Webcat ではこのデータベースを Web 上で検索である。

現在、科学技術振興事業団（JST）の科学技術情報発信・流通総合システム J-STAGE（Japan Science and Technology Information Aggregation Electronic）、研究開発支援総合ディレクトリ ReaD（Directory Database of Research and Development Activities）などとの間で図書館サービスに関する分担見直しが行なわれている。

9 海外におけるデジタル・ライブラリの現状

9.1 大学等・非営利団体を中心としたデジタル・ライブラリの状況

米国の OCLC (Online Computing Library Center, Inc) は世界 86 カ国・地域の 42,489 図書館に対してサービスを提供している。対象図書館は米国だけで

も 30,000 館を越え、あらゆる種類の図書館をカバーしている。OCLC は米国オハイオ州の大学学長が集まって、図書館のサービス向上と経費節減を目的として 1967 年に設立された。当初はオハイオ州立大学の図書館内に設置され、計算機センターも同大学のリサーチセンターを利用し、州内の 54 大学を対象にした地域限定のサービスで、名称も **Ohio Colleges Library Center (OCLC)** であった。その後、1971 年にオンラインで共有できる目録システムの導入、1977 年に国際ネットワーク化し、1979 年に図書館間の貸借システムを導入した。現在に至るまで世界の 6,928 図書館との間で 1 億 2200 万冊の貸借実績がある。1981 年にオハイオ州立大から独立し、名称も現在の名称に変更して非営利のメンバー組織としてスタートし、現在に至っている。レファランス・ツールとして **FirstSearch Service** が開始されたのは 1991 年で、現在では世界の 19,246 図書館にサービスされている。本社はオハイオ州ダブリンにある。

米 **RLG (Research Library Group)** も非営利のメンバー組織である。1974 年に設立されたが、翌年コロンビア大、ハーバード大、エール大、ニューヨーク・パブリック・ライブラリが参加し、推進母体となった。会員は 161 団体(2002 年 11 月現在)で、大学メンバーは約 60 校、そのうち半数は米国外の大学である。大学以外では各国の国立図書館や美術館、博物館などであり、**RLG** としてはアーカイブなど資料の保存やデジタル情報の検索などが大きいテーマとなっている。そのための標準化や技術開発も進められており、2000 年 3 月には **OCLC** と共同でデジタル・アーカイブのインフラストラクチャ構築についての議論を開始し、2002 年 6 月には、**OCLC/RLG** の共同 **WG** の報告として、デジタル情報の保存に関するメタデータのフレームワークについての報告書を作成している。**RLG** はメンバー組織であるが、サービスは非メンバーにも開放され、技術開発に際しても非メンバーの利用を前提として検討を進めている。

そのほかにも、**CLIR (Council of Library and Information Resources)** やその関連組織である **DLF (Digital Library Federation)**、**ARL (Association of Research Libraries)**、**RAP (Regional Alliance for Preservation)**、**ULG (University Libraries Group)** などの組織がさまざまな視点から相互に連携をとりながら電子図書館のあり方や次世代電子図書館の検討などについて検討を進めている。

9.2 商用デジタル・ライブラリ・サービスの状況

LNCS (Lecture Notes in Computer Science) / LNAI (Lecture Notes in Artificial Intelligence) はハイデルベルグ (ドイツ) を本拠地とする Springer-Verlag Heidelberg の提供するサービスである。Springer-Verlag 社は 1842 年に設立されている。古くはディーゼル、ファラデー、ポアンカレなどの科学者の論文を出版したり、ゲッチンゲン大学と関連して著名な数学者・物理学者などの出版物を扱っており、非常に歴史が古く、実績のある出版社である。LNCS は 1973 年に設立され、2002 年 11 月には印刷媒体での累積発行数 2,500 巻に達した。10 万人以上の著者により、100 万ページが掲載されている。1998 年 5 月に LNCS online が導入されて以来、印刷物でもオンラインでも入手可能となっている。なお、2,500 巻のうち 1000 巻がオンラインでアクセス可能である。2002 年だけで言えば 304 巻が発行され、17,000 人の著者による 13 万ページ超の論文が掲載された。デジタル・ライブラリとしては 50,000 人超の科学者・技術者による 35,000 を超える論文が装備されている。各巻には特定分野の専門論文、複数著者によるもの、会議の論文集、各種ワークショップの資料などが含まれ、特別号には特別なチュートリアル、先進技術、最新の話題、記念論文、ACM の優秀論文などが掲載されている。

ScienceDirect はアムステルダム (オランダ) に本拠を置く Elsevier Science 社の提供するデジタル・ライブラリ・オンライン・サービスである。Elsevier は 1880 年に古典や百科事典の出版・販売を行うためにロツテルダムに設立された。創立以来 100 年以上の歴史をもつ世界最大の科学・技術、医学に関する出版社である。ScienceDirect では 1,700 以上のジャーナルがオンラインでアクセスでき、5,900 万以上のアブストラクトが検索できる。

情報提供サービス会社としては米ペンシルバニア州フィラデルフィアを本拠地とする ISI (Institute for Scientific Information) 社がある。同社は科学論文の引用調査で日本でも良く知られている。1997 年にインターネットによる情報提供サービス ISI Web of Science が開始され、翌 1998 年には全米組織である ARL (Association of Research Libraries) のメンバーの 53%、1999 年には約 70% にサービス提供するなど急速に成長している。2001 年には科学技術のトレンドを分析したり、国や研究者、企業などのランク付けを行うツール ISI Essential Science Indicators や、文献や特許などを横断的に調査できるプラットフォームとして ISI Web of Knowledge を提供するなど商用としてのサービス拡大を図っている。引用を利用することで関連情報の検索も容易になると同時に、最近では論文に対する評価基準のひとつとして引用回数も重要な位置づけにあり、今後も成長が期待できる。日本でも ISI Web of Science サービスを利

用している大学は多い。

9.3 今後のデジタル・ライブラリのあり方に関する動き

米国においてはDLI1が始まった1994年ころからデジタルライブラリの構築が盛んに行われるようになった。バージニア工科大のNDCTD (Networked Digital Library of Theses and Dissertations) のように大学間で学位論文を共有する試みや、NSFのSMETE (Science, Mathematics, Engineering and Technology Education Library) プロジェクトのように、科学技術教育のために教育資源の共有や教育のための協調を研究するなど、目的に応じた研究が進められている。

英国においては、大学における教育環境向上のための政府組織 JISC (Joint Information Systems Committee) の助成によりデジタル・ライブラリ・プロジェクト eLib (<http://www.ukoln.ac.uk/services/elib>) を推進中。ここでは、図書館における技術やサービスおよび環境などに関する研究を進めている。

デジタル・ライブラリの今後のあり方や技術を議論する国際会議としては、1994年5月に米ニュージャージー州 Newark の Rutgers 大学で開催された Digital Library Workshop を受けて IEEE-CS が ADL'9x を開催。また Texas A&M 大学が開催した Digital Library 国際会議を受けて ACM が DL'9x として開催したが、2001年に両会議が統合され、ACM/IEEE-CS 共催の JCDL (Joint Conference on Digital Library) となって継続されている。JCDL2003はテキサス州ヒューストンにおいて開催される(5/27-31)。欧州においてはEUのERCIMが国際会議 ECDL (European Conference on Research and Advanced Technology for Digital Library) を開催している。2003年度は第7回会合でノルウェーの Trondheim で開催される(8/17-22)。アジアにおいては1998年に第1回 ICADL (International Conference on Asian Digital Library) が香港で開催され、続いて台北(台湾1999年)、ソウル(韓国2000年)、バンガロール(インド2001年)、シンガポール(2002年)と毎年開催されている。その他にも、OCLC、RLGなどが独自に行っている Workshop などが頻繁に実施されている。