

「学術情報基盤の今後の在り方について（報告）」と機関リポジトリ構築

逸村 裕

1. はじめに

2006年3月23日付けで「学術情報基盤の今後の在り方について（報告）」が公開された。これは「科学技術・学術審議会 学術分科会 研究環境基盤部会 学術情報基盤作業部会」が取りまとめたものである。

2004年11月15日に学術情報基盤作業部会が設置され、学術情報基盤としてコンピュータネットワーク、大学図書館、学術情報発信をとりあげ、それぞれにワーキンググループを置き、二年余りの審議を経て公開された。ここでは学術研究全般を支える学術情報基盤に対しての基本的な考え方や国及び大学が考慮すべきことを検討した。

学術情報基盤全体について、(1)学術研究の国際競争力を維持向上させる持続可能な仕組みの構築、(2)利活用の促進、(3)我が国の学術情報流通機能の戦略的強化、(4)国立大学の法人化・大学財政の緊縮化への対応を主な問題点として掲げた。

作業部会及びワーキンググループにはコンピュータネットワーク、大学図書館、学術情報流通関係者が加わった。名古屋大学からは阿草清滋大学院情報科学研究科教授、伊藤義人大学院工学研究科教授・附属図書館長、そして逸村が名を連ねている。

この報告書では大学図書館の基本的な役割として以下の二点を掲げている。

- a. 教育研究支援
- b. 電子情報と紙媒体を有機的に結びつけたハイブリッドライブラリーであること

これらを基として、2000年以降の電子ジャーナルの急速な普及、年間120万件を超える文献複写・現物相互貸借活動を支える国立情報学研究所の目録所在情報サービス等を踏まえ、厳しい財政下にある学術情報基盤においてなすべきことを網羅的に提示した。

報告の本文及び議事録については文部科学省のホームページに掲載されている¹⁾。

2. 機関リポジトリ (Institutional Repository) とは

機関リポジトリ、学術機関リポジトリあるいは学術情報機関リポジトリ等とも呼ばれる。大学や研究所等の学術機関で生産された学術的コンテンツ（メタデータを付した論文・報告書・教材・アーカイブ画像・データ等）を電子的に大学等の管理責任の下に蓄積保存し、無料でアクセスできるようにし、永続的に運用を行うインターネット上のシステムを意味するものである。機関リポジトリを構成する機能には(1)投稿受理機能、(2)コンテンツ

Contents

「学術情報基盤の今後の在り方について（報告）」と機関リポジトリ構築…	1	2006年秋季特別展（地域貢献特別支援事業成果報告）のご案内 ……	8
時間とともに移り変わる情報をとらえる ……	3	彙報 ……	8
図書館におけるユーザーの情報探索活動の理解を目指して ……	5		

管理機能、(3)検索機能、(4)累積機能、(5)相互運用性機能が挙げられる。

機関リポジトリが従来の電子図書館と異なるのは、国際標準形式メタデータに対応するとともに、標準的データ交換プロトコルである OAI-PMH (Open Archives Initiative Metadata Harvesting Protocol) のシステムを実装することにより世界規模でのメタデータ交換を可能とし、Google Scholar 等のサーチエンジンとの協業も企図していることである。また、学術論文を無償でインターネット上に公開しようとするオープンアクセス運動の一翼を担い、学術雑誌投稿論文を積極的に収集対象としていることもあげられる。この詳細については別稿を参照されたい²⁾。

2006年8月時点で日本国内には約20の機関リポジトリが存在する。

北米では、123の図書館が加盟する ARL (Association of Research Libraries: 研究図書館協会) において2006年1月時点での調査がある。そこでは回答のあった87館のうち、37館 (43%) がすでに機関リポジトリを構築しており、31館 (35%) が2007年中の公開を企図している³⁾。その他、海外の機関については Registry of Open Access Repositories (ROAR)⁴⁾ 等で調べることができる。

3. 国立情報学研究所 (NII) 「次世代学術コンテンツ構築事業」

平成18年度 NII 次世代学術コンテンツ構築事業公募事業が行われ、2006年7月その結果が公表された。

平成17年度は委託事業として19大学を対象に行われた。この成果は平成17年度委託業務成果報告発表会として報告され、内容は公開されている⁵⁾。

18年度は公募形式で行われた。最終的には77大学 (国立64、私立13) が応募し、57大学 (国立47、私立10) が採択された⁶⁾。

この次世代学術コンテンツ構築事業は NII が進める「CSI (Cyber Science Infrastructure)」の一環として位置づけられ、学術コンテンツの確保・発信を目的としている (図参照)。これは国公立大学が機関リポジトリ構築を行うに当たっての支援を行うもので、構築事業と新規研究開発事業の二点が含まれている。

18年度において注目されるのは公募の条件として領域を二つに分けたことである。「事業領域1」は従来通りの機関リポジトリ構築・運用であり、「事業領域2」は先端的研究開発事業として、機関リポジトリ発展のための研究開発・評価事業等を対象とした。

これまで NII が行う事業においては NII が発案して各大学がそれに対応するという経緯の事業が中心であった。今回の次世代学術コンテンツ構築事業は、公募形

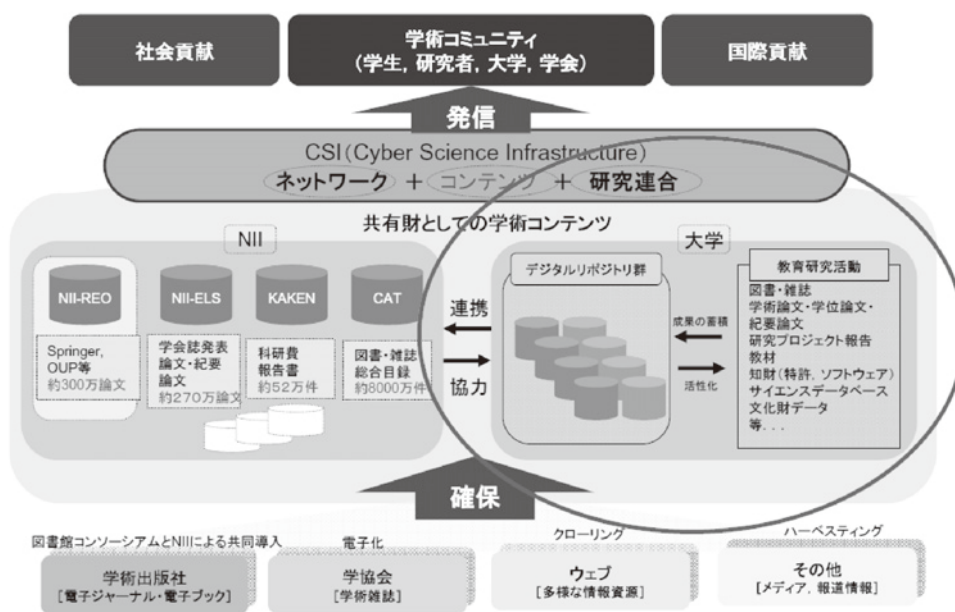


図1. 次世代学術コンテンツ基盤 (<http://www.nii.ac.jp/irp/info/2006/jigyogaiyo0712.pdf>)

式という NII にとって新しい試みであり、今後の大学と NII の関係を考えると興味深い事象であった。

またいくつかの私立大学においては「機関リポジトリ」を構築するが NII からの資金は不要であるとの見解が示されている。機関リポジトリは本来、大学個々が構築するものであり、その姿勢は評価できよう。

今後は機関リポジトリの量的及び質的充実、各機関リポジトリの評価、さらに一般への広報周知が問題となろう。さらに考えられることとしては、大学の認証評価において機関リポジトリの位置づけが明確化できるかどうかが挙げられよう。

参考文献

(1) http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/06041015/020.pdf

(2) 逸村裕. 機関リポジトリ. 名古屋大学情報連携基盤センターニュース. Vol.5, No.2.2006.p.101-105.

http://www2.itc.nagoya-u.ac.jp/pub/pdf/pdf/vol05_02/101_105campus02.pdf

(3) SPEC Kit 292, Institutional Repositories. University of Houston Libraries Institutional Repositories Task Force. 2006. 176p.

なお、Executive summary は以下で入手できる。

<http://www.arl.org/spec/SPEC292web.pdf>

(4) <http://archives.eprints.org/>

(5) <http://www.nii.ac.jp/irp/info/2005.html>

(6) <http://www.nii.ac.jp/irp/rfp/2006/>

(研究開発室研究協力者・筑波大学図書館情報メディア研究科／図書館情報学)

時間とともに移り変わる情報をとらえる

石川 佳治

*

まずは自己紹介から始めたい。私はこの4月に筑波大学から転任し、名古屋大学情報連携基盤センターの学術情報開発研究部門教授に着任した。兼任という形で名古屋大学附属図書館研究開発室の室員も担当することとなった。

筑波大学の前は、奈良先端科学技術大学院大学に勤務していた。その間、同大学では、国内に先駆けて電子図書館がオープンした。できて間もない大学であったため、最初から電子図書館として設立されたことが特徴であった。私自身はその立ち上げに直接関わったわけではなく、あくまでもユーザとしての立場ではあったが、電子図書館の時代の始まりを体感することができた。その後筑波大学では、すぐそばの図書館情報大学の関連する研究者とも何かと行き来があった。平成16年4月に筑波大学と図書館情報大学が統合した後では、さらに身近な存在となった。名古屋大学でも図書館関係の仕事に携わるということで、何かしら縁を感じる。

* *

私のそもそもの専門分野は情報科学で、より具体的に

いえば、データベースを中心に、データマイニング、情報検索、ウェブに関係する研究などを行っている。いずれにしても何かしらデータを対象としてきたのだが、大学の卒研でデータベースの研究にはじめて触れた頃（80年末）

を考えると隔世の感がある。その頃はコンピュータの価格も高く、能力もきわめて低かったことから、少量のデータ（その当時は大量と思っていたのだが）をきちんと管理する、というところにデータベース研究の力点が置かれていたように思う。以下では多少古い話も含め、データベース研究の現在までの流れを簡単に振り返ってみたい。

現在でもデータの管理は研究開発の重要なトピックではあるのだが、ハードウェアの高速化・大容量化と低価格化がもたらした影響は非常に大きい。以前であれば、本当に必要なデータのみをコンピュータで管理し、無駄なデータはできるだけ保持しない、というのが基本的戦



略として有効であったのだが、現在では、無駄かどうかを判断するコスト（つまり手間やお金）の方が、データを蓄積するコストよりもむしろ大きくなっている。特にハードディスクの大容量化・低価格化によるビットあたりの単価の減少の効果が大きい。今日ではデータをとりあえず消さずにとっておくということが多くの場合リーズナブルなのである。

このようなデータの蓄積コストの低下は、ビジネスや科学技術分野におけるデータ管理にも大きな影響を与えている。たとえばコンビニの売り上げデータ、企業の営業・取引におけるデータ、オフィスの文書データ、システムへのアクセス履歴、実験・観測データなど、さまざまなデータがストレージに蓄積されて何かしら有効な利用を待っている。一昔前にはテラバイト（1TB = 1,024GB）のデータという途方もない量、というイメージであったが、現在ではもう特別なものではない。すでに世の中はペタバイト（1PB = 1,024TB）の時代に入っている。

そのような流れの中で90年代に入ってきて現れてきた技術の一つが、OLAP（On-Line Analytical Processing）やデータウェアハウス（data warehouse）である。通常のデータベースが、たとえば企業の現在の取引データを蓄積し、リアルタイムに更新や問合せを処理するのに対し、OLAPやデータウェアハウスでは、分析業務に特化したデータの管理を行う。しばしば前者は基幹系、後者は分析系と呼ばれている。このような動きが出てきた背景には、ハードウェアの性能向上と低価格化があり、得られた余力を分析処理にあてることが可能になったという理由がある。

もう一つ、90年代前半から起こった新たな研究分野がデータマイニング（data mining）である。マイニングとは「発掘」の意味であり、知識発掘（knowledge discovery）と呼ばれることもある。大量のデータの中から役に立つ知識を発見しようというのがその目的で、ビジネス面への応用（例：商品の売れ筋の予測）が盛んになされているが、科学技術分野のデータからの知識抽出など、ビジネス以外の分野でも多くの試みがなされている。前述のOLAPやデータウェアハウスと同様、データマイニングも低価格かつ高性能のコンピュータの普及や大容量のストレージの普及を背景として発生してきた技術であり、そのままでは無駄になってしまうデータか

ら、何かしら有益な情報を取り出そうという考え方に基づいている。

このような流れは図書館の世界とも無縁でない。電子図書館にそもそも現実味が出てきたのは90年代からの電子文書技術の普及が一つの要因であるが、それを支えるコンピュータ技術や情報インフラの普及などなくしては、その実現は困難であったろう。

* * *

実はこれまでは前置きである。大量のデータが容易に蓄積でき効果的に利用できるようになったということは、会計の用語でいうところの「ストック」が増えた、ということに例えることができるだろう。それでは「フロー」は何に対応するかというと、ウェブ上に存在する情報や、ネットワークを介して流通するさまざまなデータであろう。ウェブ上に存在する多種多様な情報を例にとると、ある時点において存在していたページが次に訪問するとなくなっていたり、その内容が大幅に変わっていたりすることはしばしば見られる。また、ウェブ上のニュースや最近盛んなブログやソーシャルネットワークなども、時々刻々と情報に変化するという点では共通している。つまり、今日では、蓄積されている情報だけでなく、蓄積されないままに消えていく情報も大量に存在しているということである。

このような流れを受けて、最近のデータベースの研究分野では、ネットワーク上を流通する大量のデータを扱うための研究が盛んになっている。そのような研究の領域として、ストリーム処理がある。一言でストリーム（stream）といってもいろいろあるのだが、典型的な例としては多数のセンサによる環境のモニタリングがあげられる。次々と到着するセンサからの観測情報（例：温度など）をリアルタイムに集積し、異常があれば即座に反応したり、情報を継続的に集約したりする。先に述べたような蓄積することを前提としたデータとは異なり、後々まで保持すること自体にはあまり意味はなく、むしろその場その場で瞬時に応答できるか、また、大量の情報から不要なものを適切に廃棄できるかなどが焦点となる。

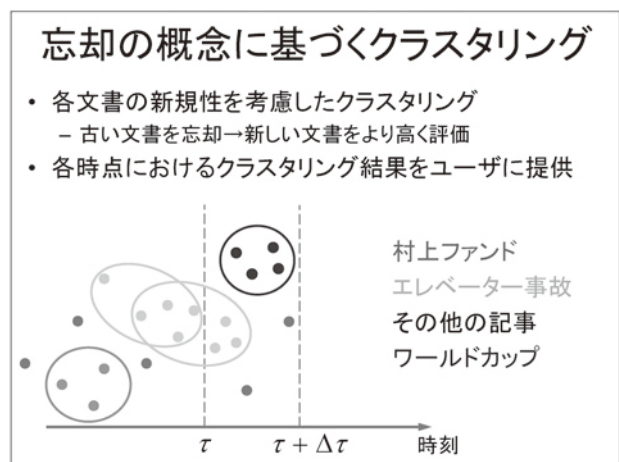
ストリーム処理の別の例としては、かなり異なる立場であるが、情報配信も挙げられる。情報配信にもいろいろなものがあるが、たとえばウェブ上のニュースの配信を考えよう。従来のニュース配信では、単にウェブページが定期的書き換えられ、ユーザが好きなタイミング

でサイトを訪問する形であった。しかし最近では、RSSなどを用いて、サイトの最新情報を配信するサービスが各社により提供されている。RSS自体はそもそもサイトの更新情報の記述のためのXMLに基づく規格であるが、この機能をうまく用いることで、ニュースサイトやブログなどから最新の記事をかき集めてくるプログラムが容易に作成できるようになった。ユーザの立場からは、あたかも情報が次々と送りつけられてくるように見えるため、これは一種のストリーム型の情報であると捉えることができる。

* * * *

いろいろと書いてきたが、考えてみると、ストックとフローに分けた場合にフローの方に分類される、時間とともに変化して変わりゆく情報をどのようにとらえ、有効活用していくかということが私の研究の基本的なテーマの一つであるように思える。これに関連してしばらく前から研究を行っているのが「忘却の概念に基づく文書クラスタリング」である(図参照)。対象として、配信される大量のニュース記事を適切にグループ化し、ホットな情報を中心に提供するような応用を想定している。

この手法の特徴は、文書を手にしたその時点では文書の価値が最も高く、時間が経つにつれその価値が薄れていく(つまり忘却される)という仮定をいれて文書のクラスタリング(グループ化)を行う点にある。古い文書を積極的に忘れることで、逆に新しいトピックを中心に文書群をグループ化して、ホットな情報をユーザに提供できると考えている。



この研究のアイデアを考えた後でよくよく調べてみると、図書館学の一分野である、引用分析(citation analysis)の領域で古くから研究されている老化(obsolescence)の概念(雑誌記事が出た後、引用数が次第に減っていく現象)との関連が深いことがわかった。ここでも図書館学との関わりを感じている。

今後の電子図書館は、ネットワークを介した情報の統合やウェブとの連携など、より流動的な情報への対応が求められるだろう。何らかの形で本研究が貢献できないかと考える次第である。

(研究開発室兼任室員・情報連携基盤センター／データベース・データ工学)

図書館におけるユーザーの情報探索活動の理解を目指して

寺井 仁

1. はじめに

人間の認知プロセスを明らかにする際の最大の問題点は、対象そのものを直接観察することができないという点にある。認知心理学では、このようなブラックボックスである人間の心を、実証的アプローチと情報处理的アプローチと呼ばれる2つのアプローチを用いて明らかにしようと試みる。実証的アプローチでは、被験者を対象とした心理実験を通して、問題解決時の行動や内省報告

から認知プロセスを推定する。情報処理アプローチでは、実証的アプローチによって得られたデータをもとに認知プロセスのモデルを仮定し、計算機シミュレーションを通してモデルの妥当性を検証する。これら2つのア



アプローチを通して、モデルの妥当性を高めてゆくのが認知心理学における研究手法の大きな特徴となっている。

研究開発室に赴任する以前は、名古屋大学情報科学研究科において、以上のような認知心理学的手法を用いて主に実験室という環境の中で、人間の問題解決プロセス、特に「ひらめき」と呼ばれる「洞察 (insight)」のメカニズムについて検討を行ってきた。(詳細に関しては、寺井・三輪・古賀、2005; Terai & Miwa, 2006を参照)。実験室という閉ざされた環境と図書館という公共の開かれた環境という違いはあるが、ユーザーとしての人間が主体であるという意味において、これまでの研究の延長として図書館に貢献できる研究を行っていきけるのではないかと考えている。

2. 問題解決と情報探索

実験室から日常に視点を移してみると、我々の日常もまた問題解決の連続である。問題解決は初期状態（現在自分が置かれている状態）から、様々な制約の下で状態を変化させ、目標状態へといたるプロセスである。日常における問題の多くは、目標が漠然としたものであり、何を行えばよいのかが明確には定まっていない「よく定義されていない問題 (ill-defined problem)」である。このようなよく定義されていない問題を「よく定義された問題 (well-defined problem)」とするために、情報ニーズが生まれ、情報探索が行われる。ここで重要な点は、情報ニーズを満たす情報が見つかったとしても、そこで情報探索が終わるわけではない点である。情報探索の結果、取得された情報が更なる情報ニーズを生み出すとともに、漠然とした目標状態にも影響を与える (図1)。以下では、「ある講義で、各自の視点で『認知科学について』のレポートの提出を求められた学生」を例に考えてみたい。認知科学に関する知識をほとんど持っていない学生の場合、(1)まず、与えられた課題に含まれる「認知科学」というキーワードで情報探索を開始すると予想される。(2)そして、「認知科学」についての情報探索の過程で、「認知科学」に関連した「実験心理学」や「人工知能」についての情報ニーズが生まれ、(3)その結果、不明確であった目標状態がより明確に、例えば、『認知科学について』のレポートが、『情報処理システムとしてみた人の認知過程』のように変化するかもしれない。そして、このような目標状態の変化は更なる情報ニーズ

を生み、目標状態と情報ニーズは相互に影響を与えながら変化してゆくと考えられる。

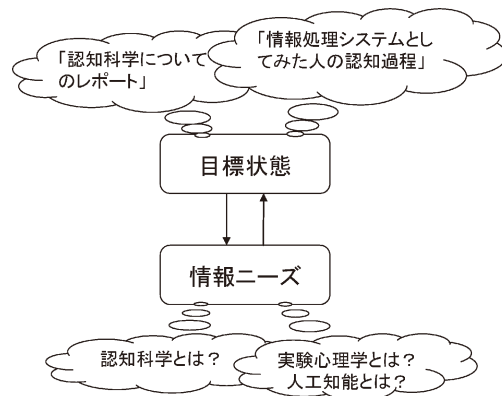


図1. 相互に影響を及ぼしあう目標状態と情報ニーズ (Bandura, 1986; 三輪, 2000).

以上のように、情報探索とは、単に目的の情報を探し出し、そこで終了するような単純なプロセスではなく、問題解決そのものを変化させながらダイナミックに進む現象であると予想される。そして、このようなプロセスを通して、当初は意図しないが、「興味深い情報との出会い (情報遭遇: Information Encountering)」が、目標状態や情報ニーズを大きく変化させ、「思いがけない発見 (serendipity)」を導くと考えられている (Erdelez, 2004; Foster & Ford, 2003)。

3. 図書館におけるユーザーの情報探索活動

図書館は、日常において情報ニーズを満たすために情報探索を行う場所のひとつであり、様々な情報やサービスが提供されている。

図書館が提供する情報源は大きく2つに分けることができる。その一つが、従来の高度に構造化された物理的情報源 (書籍) であり、他方が、電子的情報源である。現在の図書館を取り巻く環境は電子的情報源の急速な普及により大きな影響を受け変化してきている。その結果、図書館が保有する情報は、伝統的な紙媒体資料だけでなく、一部は電子媒体の資料に置き換わり、また検索等の利用形態にも変化が生まれてきた。情報の構造化の程度、そこに含まれる内容、アクセスの容易性など質的に異なる複数の情報源は、ユーザーの情報探索プロセスにも大きな影響を与えていると考えられる。つまり、情報ニーズを満たすために、従来の書籍情報だけでなく、OPACによるインデックス検索や Web での情報検索が行われ、内

容や質が異なる情報が獲得される。そして、このような異なる情報源に対する情報探索の結果が、情報ニーズや目標状態に異なる影響を与えるものと考えられる（図2）。

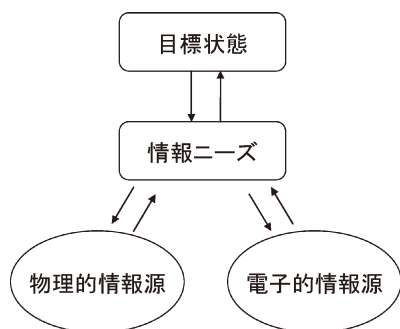


図2. 情報探索と情報源

附属図書館研究開発室では、全学における教育研究の支援機能の高度化を測るため、電子的情報源・物理的情報源を統合したハイブリッド・ライブラリーの実現に向けた研究開発が進められており、私はユーザーの視点からこの問題に取り組んでいきたいと考えている。情報源のハイブリッド化においては、情報源の質の違いがユーザーの問題解決に与える影響を考慮し、何を電子媒体に置き換え、何を紙媒体のまま残すのか、またユーザーの認知特性を考慮した上でどのようなサポートが可能なのかについての議論が不可欠であると思われる。実際に、多くの先行研究において、電子媒体と紙媒体の資料の利用形態に差が認められることが報告されており（e.g., Christianson & Aucoin, 2005; Liu, 2006）、また、このような差異は、たとえ含まれる情報が等価であるような場合（e.g., 書籍と電子Book）であっても同様である。

では、どのような情報源のあり方が情報探索を行うユーザーにとってより望ましく、また、どのようなサポートが可能なのか。それを明らかにするには、(1)ユーザーが実際にハイブリッドな情報源を如何に活用しているのかを明らかにし（実証的アプローチによる事例研究）、(2)その上で、ユーザーが直面している問題の質（よく定義された問題／よく定義されていない問題）と、ユーザーが探索する情報源の質（書籍、電子Book、OPAC、Web等）がどのような影響を及ぼしあっているのかを実験的に検討し（実証的アプローチによる実験研究）、(3)そして、その結果から問題解決と情報探索の認知モデルを導くこと（情報処理アプローチによるモデル研究）が必要となる。

附属図書館研究開発室が掲げる主な研究課題には、ハイブリッド・ライブラリーの研究開発および図書館情報リテラシー教育の普及と高度化が挙げられている。前者に対しては、ユーザーの情報探索活動に対する認知心理学的な観点からの分析を通して、その特性をもとにしたハイブリッドな情報環境のあり方について示唆を与えることができるものと考えている。また、後者に対しては、ユーザーが現実の環境（図書館）においてどのように情報および情報源を扱い、情報探索活動を行い、そこにどのような問題を抱えているのかを教育という形でユーザーにフィードバックすることにより貢献ができるものと考えている。

参考文献

- Christianson, M. & Aucoin, M. (2005). Electronic or print books: Which are used? *Library Collections, Acquisitions, & Technical Services*, 29, 71-81.
- Bandura, A. (1986). Self-regulation of motivation and action through internal standard and goal system. In A. P. Lawrence (Ed.), *Goal Concepts in Personality and Social Psychology*, 19-85. Hillsdale, NJ: Erlbaum.
- Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information Processing and Management*, 40, 1013-1025.
- Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, Vol.59, No.3, 321-340.
- Liu, Z. (2006). Print vs. electronic resources: A study of user perceptions, preferences, and use. *Information Processing and Management*, 42, 583-592.
- 三輪眞木子 (2003). 『情報検索のスキル - 未知の問題をどう解くか』. 中央公論社.
- 寺井仁・三輪和久・古賀一男 (2005). 仮説空間とデータ空間の探索から見た洞察問題解決過程. 『認知科学』, Vol.12, No.2, 74-88.
- Terai, H. & Miwa, K. (2006). Sudden and Gradual Processes of Insight Problem Solving: Investigation by Combination of Experiments and Simulations. *Proceedings of the 28th annual conference of the cognitive science society*, 834-839.

(研究開発室専任助手／認知科学)

2006年秋季特別展（地域貢献特別支援事業成果報告）のご案内

○特別展

「江戸時代の村と地域—美濃養老・日比家文書にみる暮らしと災害—」

9月29日（金）～10月20日（金） 10：00～17：00（土・日とも開館）

会場：名古屋大学中央図書館 4階展示室

○講演会

10月9日（月） 13：00～15：30

会場：名古屋大学中央図書館 5階多目的室

講師：秋山晶則（名古屋大学）「日比家文書にみる暮らしと災害」

水本邦彦（京都府立大学）「美濃路・伊勢道・近江道」

○展示資料解説

10月11日（水）、14日（土） 各回13：30～14：00

主催：名古屋大学附属図書館・同研究開発室

後援：愛知県、岐阜県、三重県、名古屋市、大垣市、養老町の各教育委員会

問い合わせ先：〒464-8601 名古屋市千種区不老町 名古屋大学附属図書館 情報管理課庶務掛

TEL052-789-3667

E-mail：shomu@nul.nagoya-u.ac.jp

彙報

2006年

4月10日～5月5日

春季特別展「『地獄物語』の世界—江戸時代の法と刑罰—」

4月15日 ギャラリートーク：塩村耕「ドキュメンタリー文学としての地獄物語」、神保文夫「白洲と牢屋—江戸の裁判事情—」

4月17日 第1回FM

5月22日 第2回FM

6月2日 第4回友の会トークサロン：秋山晶則「江戸の礼節」

6月19日 第3回FM

第21回オープンレクチャー：石川佳治「ホットなトピックの発見と追跡—TDTに関する研究の動向—」

7月24日 第4回FM

第22回オープンレクチャー：寺井仁「ひらめき、発見、創造：洞察プロセスに関する認知科学的検討」

LIBST Newsletter No.8

編集・発行

名古屋大学附属図書館 研究開発室

〒464-8601 名古屋市千種区不老町

TEL 052 (789) 3697、5699