

システム間連結のための著者名典拠ディレクトリシステムの開発

Development of the Name Authority Resolution System

1) 名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University

2) 株式会社ナレッジサイエンス
Knowledge Science Co., Ltd.

3) 名古屋大学附属図書館
Library, Nagoya University

棚橋 是之¹⁾ 山本 哲也¹⁾ 梶田 健史²⁾ 次良 丸章³⁾
TANAHASHI, Koreyuki YAMAMOTO, Tetsuya KAJITA, Kenji JIROMARU, Akira

Abstract

Author name is an important access point for academic information. But academic information systems do not always control the author name. Above all, the format of author names varies from system to system. We must resolve this problem to develop an institutional repository (NAGOYA Repository). Our plan is to develop a system of resolving author name in order to connect systems with one another. This resolution system links existing systems via author name with minimum customization. This paper describes our approach, outline of the system and the remaining issues.

Keywords: Institutional repository (学術機関リポジトリ), Name variations (著者名の揺らぎ), Name authority (著者名典拠), Web service (WEB サービス), DSpace (DSpace), CSI : Cyber Science Infrastructure (CSI : 最先端学術情報基盤)

1. はじめに

本を読んで、その著者について知りたくなることや、その著者の別の本を読んでみたくすることは誰にでもあることだろう。学術情報の探索においても、重要な著者を見つけたら、その著作をたどることは、効率よく情報を見つけるための常套手段である。

インターネットの時代になり、我々は未知の著

者と出会う機会が多くなったのではないだろうか。検索エンジンで出会う有用な論文の著者は、必ずしも自分の専門領域の著者ばかりではない。そんなとき、その著者についての正確な情報を参照したり、その著者の別の著作を読んだりすることは、出会った情報をより深く正確に理解するための手助けとなるだろう。

本稿では、学術機関リポジトリ（後述）の構築

をきっかけとして我々が着手した、学内研究者の「著者名」というアクセスポイントの整備について、その概要を報告する。

1.1 学術機関リポジトリの構築

学術機関リポジトリとは、インターネットを通じて各機関における教育・研究成果を広く一般に公開する仕組みである。リポジトリには研究成果の本文（通常、PDF ファイル）と、それに加えてメタデータ（タイトル、作者、掲載先等のデータ）が付与され登録される。登録された研究成果は機関の資産として保管される一方、インターネット上でメタデータが広く流通することにより、新しい読者を獲得するとされている。

わが国での学術機関リポジトリ構築は、国立情報学研究所の最先端学術情報基盤事業（CSI：Cyber Science Infrastructure 事業）のひとつに位置づけられたことにより、大きく前進した。2008年1月現在、77の学術機関リポジトリが稼動している¹⁾。名古屋大学でも2005年度からCSI事業の委託を受け、構築を進めることができた。リポジトリのシステムには、オープンソースであるDSpaceを採用し、NAGOYA Repositoryとして2006年2月に運用を開始した。初期データとしては紀要データ、博士論文等約3,000件を収録した。その初期データ構築の過程で、次節に述べる著者名の問題が顕在化し、その解決のため「著者名典拠ディレクトリシステムの構築」を構想したものである。幸い、2006年度からCSI事業の領域2（先端的な研究開発を行う事業領域）に採択され、開発を進めることができた。また2007年度は「多様な情報資源の統合・提示」というテーマのもと、他大学とも連携しながら開発を進めている。

1.2 著者名表記及びシステム間連携における問題

学術雑誌掲載論文や学会発表等をDSpaceに登録する過程で、氏名の書き方、とくに外国語表記については、論文によって表記方法が異なることが問題となった。これは、発表媒体の編集規約等による避けがたい問題である。たとえば「山本 国男」の場合、「山本 國男」、「yamamoto, kunio」のほか、「YAMAMOTO, Kunio」、「Yamamoto, K.」

などの表記が混在する可能性がある。メタデータは論文からの転記を原則とするので、同一人物の氏名表記にバリエーションが生じることになる（ただし、大文字・小文字の異同についての検索はどちらからでもできる）。特に、イニシャルによる表記の場合、著者名が識別できないケースが多く発生してしまう。

そのためNAGOYA Repositoryにおいては、メタデータの登録にあたり日本語表記の著者名を補完的に追記する運用をしているが、著者名を統一的にコントロールする作業は非常に労力がかかる。また、DSpace内で統一できたとしても、他のシステムとの連携を考えた場合、システム間で著者名の表記が異なるという問題がやはり生じてしまう。

そこで我々は、著者名典拠のようなしくみをシステムの外部につくり、そのしくみの中で、システム内の著者名の揺らぎとともに、システム間の著者名の揺らぎをも吸収できないかと考えた。

2. 著者名典拠ディレクトリシステム

2.1 システムの基本概念

今回開発する著者名典拠ディレクトリシステム（以下、本システム）は、大規模なシステム改修やデータ内容の見直しをすることなく実現でき、システム内では表記のゆらぎを持った著者名をそのまま扱うことが可能なしくみである。いわゆる著者名典拠のように、異なる表記からコントロールされた表記を参照でき、さらに著作一覧のリンクをたどることができる。またDSpaceだけでなく、異なる複数のシステムを相互連結することを指向している。著者名の揺らぎを解消しながら、多様な情報源を、著者名をキーとして結びつけるWebサービスということができる。

具体的には、表記のゆらぎをもった著者名を受け渡すと、その文字列を専用の照合ルールに照らして著者を同定し、必要なリンク先へ適切なパラメータをつけてリダイレクトするというものである。これに類似する既存のしくみとしては、論文名や雑誌名を同定してリンク先を示すリンクリゾルバがある。この動作を研究し、基本的な仕様の設計を行った。名前を受け取り、求めるサービスへのリダイレクトを行うイメージを図2-1に示す。

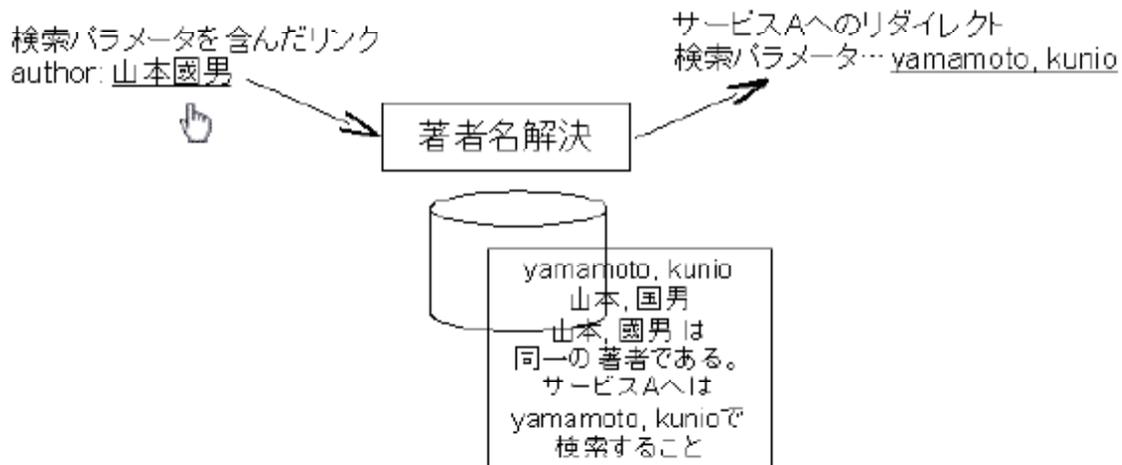


図 2-1 基本概念

つまり、リンクリゾルバになぞらえれば、著者名解決システムと呼ぶことができる。この「解決」に当たっては、著者名同定とリダイレクトに必要な情報をそなえたデータベース（著者名ナレッジベース）が必要となる。

2.2 これまでの取り組み

平成 18 年度は、仕様の検討を繰り返した後、システムのプロトタイプを製作し、初期データの整備を行った。

平成 19 年度は、以下の開発を行った。また開発したシステムを他機関でも利用可能なようにオープンソースとして公開していくことにしている。

(1) 典拠データベース管理画面の追加

リダイレクト用データ、また著者同定用のデータをより簡易に登録・編集するためのインターフェースを開発した。

(2) Web サービス化

WebAPI により、任意の外部システム（DSpace、研究者プロフィール等）から著者名の同定と標準形の取得ができるようにした。

(3) リダイレクト可能なサービスの追加

(4) 著者名検索のアルゴリズムの整備

3. システムの構築

平成 18 年度に構築したプロトタイプ・システムについて概要を述べる。

3.1 システム構成

本システムのソフトウェア構成を表 3-1 に示す。

システムの構築にあたっては、すべてオープンソースのプロダクトを使用した。

Web Application Server として用いた Django は、Python で実装されたオープンソースの Web フレームワークであり、以下の機能を内蔵する。

- ・柔軟な O/R マップ
- ・強力なテンプレートエンジンで MVC 分離を実現
- ・セッション管理
- ・ユーザ認証
- ・国際化

実行環境の OS として Fedora Core 5 を使用したが、Django が動作する環境であれば本システムの動作に支障はないと思われる。

表 3-1 ソフトウェア構成

区分	ソフトウェア名称/バージョン
OS	Fedora Core 5
Web Application Server	Django 0.95
Web Server	Apache 2.2.2
Data Base	SQLite 3
開発言語	Python 2.4.4

3.2 機能構成

本システムの機能構成を表 3-2 に示す。

表3-2 機能構成

No	機能名称	概要
1	検索・連携機能	著者名ナレッジベースより、検索条件に合致する著者の情報を検索する。また、著者名ナレッジベースに登録されているリンク先情報にしたがって、他のサイトへ遷移する。 検索・連携を実行するインターフェースとして、GUIとAPIの2種類が用意されている。
2	GUI (中間窓)	以下の Web 画面を提供する。 ・検索画面 ・検索結果画面
3	検索 API	URL パラメータで検索条件を指定することにより、検索を実行する。リダイレクトの指定が有効かつ検索結果が一件に定まる場合は、他システムのサイトへ直接リダイレクトする。 それ以外の場合は、検索結果画面を表示する。

3.3 検索・連携機能

著者名ナレッジベースより、検索条件に合致する著者の情報を検索する。検索がヒットした場合は、検索結果を画面に表示するかリンク先の URL へ直接リダイレクトする。

(1) 著者名検索条件

検索条件として、表3-3の条件を指定することができる。

表3-3 検索条件

No	名称	説明
1	著者名	著者名の一部または全体を指定する。氏名の「姓」と「名」の間に、一つ以上の区切り文字を挿入する。「姓」と「名」の入力順序は考慮しない。
2	ターゲット	ターゲット（リンク先）を一つ以上指定する。
3	エンコード	著者名の文字コードをエンコーディング名で指定する（デフォルト：'utf-8'）。検索 API を介して検索する場合のみ指定可能である。
4	リダイレクト	検索後のリダイレクトの実行の可否を指定する（デフォルト：'false'）。検索 API を介して検索する場合のみ指定可能である。

著者名に指定された文字列のうち、区切り文字として識別する文字の一覧を表3-4に示す。

表3-4 区切り文字

文字種	区切り文字
空白文字	半角：スペース, 改行文字, タブ文字 全角：スペース
記号	半角： ;:./ 全角： ; : . , / .

(2) 著者名検索手順

著者名の文字列を指定されたエンコーディング名を使用し文字コード変換し、一つ以上の区切り文字で分割する。分割された部分文字列を検索条件とし、まず著者名を検索する。著者名でヒットしなかった場合、引き続いて著者別名で検索を行う。

検索条件と著者名（著者別名）がマッチする条件は以下の通りである。

- ・分割された各部分文字列と著者名（著者別名）が部分一致すること。
- ・すべての部分文字列と著者名（著者別名）がマッチすること（AND条件）。

(3) 他サイト連携

他システムへの連携には、以下の2つのケースがある。

- ・検索結果画面の検索結果一覧に列挙されたリンクをクリックした際に遷移する。
- ・検索結果画面を介さずに直接リダイレクトする。

3.4 GUI (中間窓)

(1) 検索画面

検索条件を入力し検索対象を選択して、検索を実行する画面である（図3-1）。新規に著者名ナレッジベースを検索する場合に使用する。検索結果は、次項に示す検索結果画面に表示される。

(2) 検索結果画面

検索結果に合致した結果の一覧を表示する画面である（図3-2）。

検索結果として、著者名と著者の所属部署などの付加情報が一覧形式で表示される。リンク先ターゲットが登録されている場合は、それも列挙する。

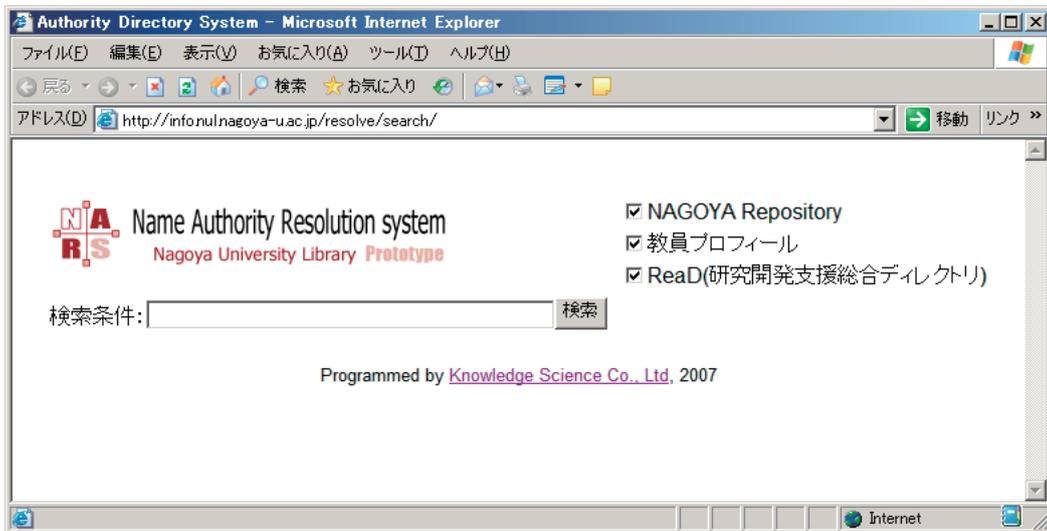


図 3-1 検索画面



図 3-2 検索結果画面

3.5 API

URL パラメータにより検索条件を指定することで、検索を実行できる API である。

URL の書式は、以下の通りである。

http://<ドメイン名>/resolve/?<URL パラメータ>

URL パラメータとして指定可能な検索条件の一覧を表 3-5 に示す。

表 3-5 検索条件

No	名称	書式	説明
1	著者名	ar=<文字列>	著者名の検索条件を指定する。 値：任意の文字列
2	ターゲット	cl=<文字列>	検索対象のターゲットを指定する。 複数指定する場合は、 "cl=<文字列>" の指定を 複数繰り返す。 値：ナレッジベースに登録したターゲットの ID
3	エンコード	enc=<文字列>	著者名の文字コードを指定する。 省略した場合のデフォルト値は、 'utf-8' である。 値：'utf-8', 'sjis', 'enc-jp'
4	リダイレクト	rdct=<真値値>	リダイレクトの実行可否を指定する。 'true' を指定しかつ検索結果が 1 件の場合に限り、 リダイレクトを実行する。

4. システムの適用

4.1 学術機関リポジトリ：NAGOYA Repository への実装

本システムは、18 年度に著者名典拠ディレクトリシステム (Name Authority Resolution System) として、プロトタイプ版を公開した (<http://info.nul.nagoya-u.ac.jp/resolve/search/>)。これは 3.4 に説明した本システムの検索画面であり、具体的には複数のデータベース (NAGOYA Repository と名古屋大学教員プロフィール) をターゲットとして、検索結果に表示させるところまで実現した。これは構想した機能の一部を実装したもので、全く別のデータベースをあたかも統合検索するようにして情報提供することができる。ただし、このサービスだけでは、著者名の問題を解決しながら異なるシステム間を相互リンクするという、本システム

の特徴を十分生かしたことにはならない。

そこで 19 年度には、NAGOYA Repository に本システムへのリンクボタンを付ける機能を実装した。メタデータの著者名フィールドの横に  というボタンが (リンクリゾルバのアイコンのように) 現れ、これをクリックすると本システムの検索結果画面が (リンクリゾルバの中間窓のように) 現れるというものである (図 4-1)。この実装にあたっては、DSpace の改造を行ない、著者名が表示されるタイミングで 3.5 に示した書式の URL を生成し、リンクがボタンで表示されるようにしている。

これにより、ユーザは図 4-2 のような 2 つのリンクを、容易にたどることができるようになる。すなわち、著者名の表記の揺らぎを気にすることなく、その著者の登録された論文を漏れなくリストできるようになり、また著者名の表記が異なる別システムであることを意識することなく、著者に関する追加情報を得ることができる。

このようにシステムの改造が最小限でよく、DSpace の側で教員プロフィール用のデータを維持する必要もないことが本システムの特徴である。また、リンクを拡張する (ターゲットを追加する) ことがあっても、その際には DSpace の改造は必要ない。

4.2 今後の活用構想

本システムのターゲットは、現在、独立行政法人科学技術振興機構が提供する研究開発支援総合ディレクトリ (ReaD) を加え 3 種類になった。今後も各種システムに拡張することを検討している (表 4-1)。このように多様なシステムを自在にリンク参照できれば、学内の研究成果を有効にアピールでき、著者にとってもユーザにとっても有意義である。また NAGOYA Repository の登録作業においても有効であると考えている。

表 4-1 ターゲットと識別子

	システム名称 (識別子)
現在のターゲット	NAGOYA Repository (日本語表記氏名) 教員プロフィール (氏名) ReaD 研究開発支援総合ディレクトリ (研究者コード)
検討中のターゲット	名大の授業 (担当教員名) OPAC (著者名典拠レコード ID)

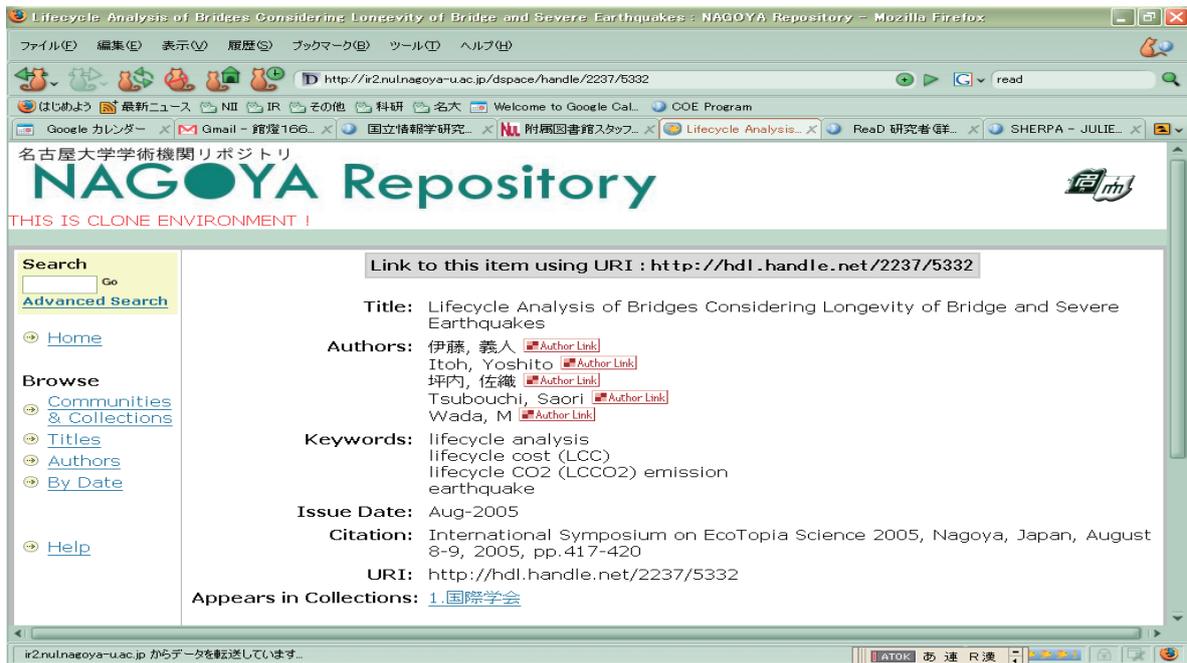


図 4-1 NAGOYA Repository のメタデータ

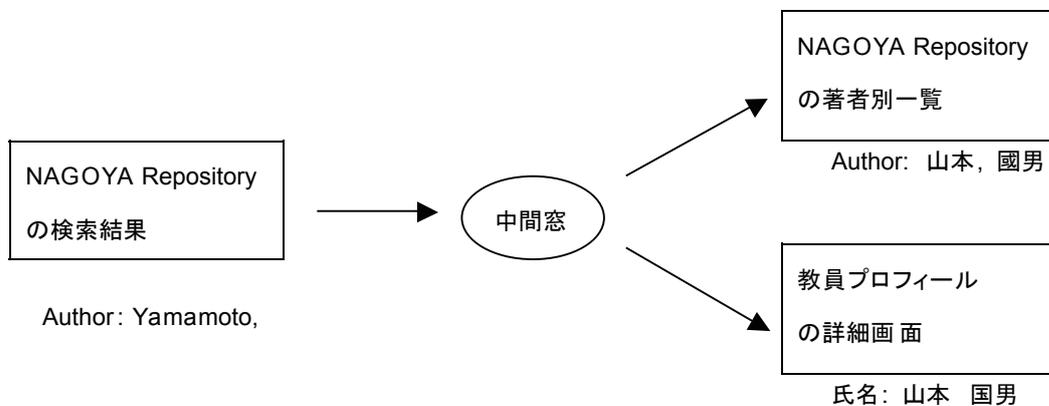


図 4-2 リンク概念図

また教員プロフィールシステムからも本システムへのリンクを準備しており、相互参照が可能となる予定である。

本システムは、特定のシステムに依存しないため、他機関でのシステム連携に移植することもきわめて容易である。その意味で CSI 事業（領域 2）のプロジェクトである「多様な情報資源の統合・

提示」においても、本システムの役割は明確である。例えば九州大学が目指している学内の各種データベースをゆるやかに結合していくという構想において、著者名で結合する際の道具として活用できるかもしれない²⁾。

5. 今後の展開

5.1 検索・連携の精度

本システムは著者名をキーとした、システム間の柔軟なリンクをめざしており、固定的なシステム間連携を開発するものではない。したがって、次に示すような、やや不便と思えるケースも生じることになり、課題として認識している。

例えば NAGOYA Repository に実装した著者リンクのボタンは、学内者であれば何らかのリンク先を表示することになるが、共著の学外者で、リンク先の情報がない場合にも、その著者リンクのボタンが現れる。これは両システムが緊密に連結されているわけではないことの必然である。

また同姓同名の二人がいた場合、中間窓の結果一覧には両者が並列される。そのとき本システムは、検索者が専門などの情報から自分で判断して正しい一人を選ぶことを想定しているが、例えばここで NAGOYA Repository へのリンクを選んだ場合、せっかく正しい一人を選んでいても、NAGOYA Repository にリンク遷移したあと、再び同姓同名の両者のアイテムが区別されずに並んでしまう。ReaD のように著者がコード管理されていれば起こらない問題だが、どの程度リンク先のシステムで著者名が管理されているかによって検索結果の精度が左右される。頻発するケースではなく、避けがたい問題ではあるが、検索者にとって不親切な遷移になる。

5.2 外部データの自動収集とコンバート

システムを超えてリンクを実現するためには、著者名ナレッジベースを作成することが必要である。学内の研究者を対象としているので、国レベルの著者名典拠のような膨大な人数を扱うわけではないが、それでもシステムごとに識別子を格納しなければならず、更新作業も必要である。

そのためターゲットごとにデータを一括投入する仕組みが必要である。また更新に際して、差分データが送信されてくるような機能が外部システムとの間で実現すれば最善である。

5.3 オープンソース化による展開

本システムはオープンソースとしての公開を予定している。システムに依存せず独立して導入ができ、最小限のカスタマイズで既存システムからのリンクも可能になる。今後、複数の機関で導入が進めば、開発コミュニティを形成し、開発情報を共有することができるようになるであろう。将来的には著者名ナレッジベースを共有することも考えられる。

また「著者名解決」として開発したシステムであるが、著者名に限らず何らかのリゾルバとして応用することも可能である。

6. まとめ

業績データベースとの緊密な連携を開発するケースが増えている。登録の手間を省くなど注目すべきものだが、開発面の負担は決して軽いではない。一方、利用面にしぼって各種のシステムを相互リンクしていくことを考えれば、本システムは十分効果的だろう。システムに依存せず、開発が最低限ですむため、既存システムの大幅な改造が困難である場合、連携すべきシステムがたくさん存在する状況などで有効である。

連携が困難と思われた多様なシステムが、著者名を介して自由に相互リンクされ、新しい発見を促していくことを期待したい。

謝辞

本システムは国立情報学研究所の次世代学術コンテンツ基盤共同構築事業（平成 18～19 年度）の委託事業として開発された。

参考文献

- 1) オープンアクセスジャパン. “日本における機関リポジトリ”. <http://www.openaccessjapan.com/2006/01/post-38.html>, (参照 2009-01-15).
- 2) 荒木寛幸. “学内各種データベースとのゆるやかな結合による統合検索”. 国立情報学研究所, 平成 18 年度 CSI 委託事業報告交流会, 2007-07-03. <http://www.nii.ac.jp/irp/info/2006/debrief/4-4kyushu.pdf>, (参照 2009-01-15).