

科研費データベースにもとづく 研究者情報ブラウジングツール

高久雅生[†] 相澤彰子[‡] 大山敬三[‡]

[†] 情報・システム研究機構 新領域融合研究センター

[‡] 国立情報学研究所 コンテンツ科学研究系

{masao, aizawa, oyama}@nii.ac.jp

概要

本稿では国立情報学研究所が提供している科学研究費補助金データベースによる成果報告書、実績報告書を元にした、研究者情報のブラウジングツールの開発について述べ、学術情報リンケージとの関わり、Webを始めとするその他の情報源とのリンケージ・統合、および、今後の課題について議論する。

1 はじめに

近年の学術研究の興隆とともに、研究活動の活発なやりとりが行われている。研究活動の主な成果は研究発表、論文、著書、特許等の文献として公表され、その成果をさらに発展させた継続研究や発展的研究がなされることで、その学術研究分野自体の発展が加速される。また、これらの発展に欠かせない、人材の育成、研究資金助成等の活用も盛んになされている。

科学研究費補助金は日本における最大の学術研究助成金であり、2006年現在、文部科学省と日本学術振興会がその審査、交付事業を行っている。

国立情報学研究所では、科学研究費補助金の採択課題の基本データとともに、研究実績報告、研究成果報告書の電子化とデータベース化を進め、2004年10月からWeb上において成果物データベースとして公開し、無償で利用できるようになっている[1]。以下、このデータベースを科研費DBと呼ぶ。

この科研費DBを使えば、ある研究者が交付を受けた課題とその概要、時期、交付金額、その成果となった研究報告などの情報を閲覧できる。この科研費DBにおいて研究者個人の情報を見る場合には、研究者の氏名で検索し、その研究者が関わった研究課題の情報を閲覧する形を取る。しかし、このアプ

ローチでは同姓同名の研究者が存在した場合に、複数人の研究課題が混ざってしまう。そのため、その把握に困難を生じたり、他の情報源との統合・比較といったさらなるリンケージ・活用に障害となることがある。

そこで、筆者らは研究者の情報に着目し、その際、科研費申請時に研究者に付与される研究者番号をキーとすることにより、研究者個人の情報をまとめて提供できるようになり、また、他の情報源とのリンケージも合わせて行うことにより、より効果的な情報提供が行えるものと考え、これら手法によるブラウジングツールを試作した。まず、科研費DBに登録されている研究課題報告書のデータを、研究者個人の情報をまとめる形で再編成し、ブラウジングツールとして一覧できるように改良したプロトタイプシステムを開発した。プロトタイプシステムにおいては、基本的情報を提供するだけでなく、Web等に存在する研究者情報とリンケージすることにより、これまでまとめた情報を入手できなかった研究者情報について、様々な情報源からの内容を統合し、その全体像が分かりやすいよう提示したり、一部の情報については可視化を行うといった機能も盛りこんだ。本稿ではこのプロトタイプシステムの詳細および設計方針について報告するとともに、リンケージ手法、Web上の情報の統合手法、研究者情報

およびその学術コミュニケーションにおける意義について考察する。

1.1 関連研究

研究者情報の提供は、これまでもいくつかのシステム、サービス、研究などの形でなされてきている。これらは、研究者情報サービスという形を取ったもの、Web における多様な情報を活用するものの2つに大別される。本節ではそれらについて概説する。

学術コミュニケーションにおける人材育成や研究者同士による学術交流の観点から、その基本的な情報として研究者情報を公開し、産学連携、研究交流に活かしていくことが重要となっており、研究者情報サービスはこの流れから提供されているものがある。例えば、この趣旨で公開されているデータベースとして、研究開発支援総合ディレクトリ ReaD [2] があり、調査票などによる登録に基づき、研究者の氏名、所属組織、研究分野、研究歴、研究業績、メールアドレス、ホームページなどの情報を登録、公開している。研究者人材データベース JREC-IN [3] では研究者の求人、求職といった人材活用を主な対象として、求職研究者による情報登録と提供が行われている。

海外における例としては、DBLP Computer Science Bibliography と呼ばれる計算機科学分野の論文書誌情報のデータベースがよく知られている [4][5]。DBLP では、計算機科学分野の主要な論文誌、国際会議プロシーディングスに掲載された論文の情報を元にした書誌データベースを構築し、それらの著者氏名を索引付けし、共著関係をキーとしてリンクすることにより、研究業績と共著者ネットワークのブラウジング機能を提供している。

他方では、現在の Web 環境の成熟とともに、研究者自らが情報発信を行おうという動きも大きな流れとして存在する。岡本 [6] は「研究の過程や成果を公開する」ことの重要性について述べ、研究者個人ホームページの活用事例とその情報発信ツールとしての可能性を示している。また、研究者本人がホームページなどの形で提供する以外にも、所属組織や学協会が作成する研究者総覧、研究者紹介といった

形で、研究者情報はその提供主体、提供の形態、場所を問わず様々な形でオンラインで提供、発信されている。

このような Web 上の情報源は雑多であり、そのままでは信頼ある情報源として活用したり、他の情報源とリンケージするためには様々な単位での情報の同定を必要とする。これらに対応する技術として、Web 上の情報の自動抽出や自動分類といった研究が盛んになされている [7][8]。これらの内、研究者情報に焦点をあてているものとして、以下がある。松尾ら [9] は、Web 上における氏名の共起情報を元にした研究者コミュニティ内の人間関係の自動抽出手法を提案している。Wang ら [10] は Web 上の研究者情報を網羅的に収集するための手法を提案している。

2 研究者情報

学術情報の流通において、研究者の情報が必要される場面が多々生じる。必要とされる研究者情報とは、典型的には以下のようなものが挙げられる。

- 氏名（例: 高久雅生; Masao Takaku）
- 年齢（例: 30）
- 所属、職位（例: 情報・システム研究機構; 非常勤研究員）
- 研究分野（例: 情報科学）
- 発表文献の一覧（例: 『科研費 DB にもとづく研究者情報ブラウジングツール』(2006), ...）
- ホームページ、メールアドレス（例: [http://research.nii.ac.jp/~masao/;](http://research.nii.ac.jp/~masao/)
masao@nii.ac.jp）
- 研究歴、学歴、学位など

これらの情報を必要とするのは、一つには研究者自身であり、自身の研究の広報・普及活動などに活用したり、研究助成や所属組織内外の予算獲得における申請書・報告書といった様々な形で記載を要する場面が多い。もう一方では、関連分野の研究者や、当該研究者の実績を見てその評価を行いたい人事担当者や評価担当者がこれらの情報を探し、これらの情報を元にその研究内容の理解や、実績の把握を行

うことができる。

科研費 DB においては、研究者一人一人につき固有の研究者番号が申請時に付与される枠組みがあり、報告書の基本データとの対応を取ることで、上記の基本情報のうち、氏名、所属、職位、研究分野、発表文献と、各課題との対応を得ることができる。さらに、研究者個人に留まらず、研究課題に対する研究分担者の情報を合わせて得ることにより、それらの研究者間の関係を得ることができる。一方、科研費報告書に記載されない情報については、科研費 DB データからだけでは得られないため、例えば研究者のホームページ等についての情報は、別途 Web 上の情報源などを使って取得する必要が出てくる。

3 研究者情報ブラウジングツール

本節では試作した研究者情報ブラウジングツールについて述べる。

研究者情報ブラウジングツールは、科研費 DB における研究者情報を、研究者番号をキーとして抽出し、さらにそれを、研究者個人という観点から並べかえ、課題報告書に記載された情報とその属性情報とを相互にリンクした形で閲覧できるようにしたツールである。また、Web 上に存在する他の情報源とも関連付けて表示できる機能もある。

本ツールの対象として扱うデータは科研費 DB のデータから取り出した情報の一部であり、102,506 人分の研究者情報が含まれている。これは、採択課題の研究代表者だけでなく、その研究分担者として研究者番号が付与されている者を含む人数である。なお、これは主に 1985 年から 2004 年までの総数 290,456 件の科研費採択課題の情報に基づくものであり、科研費助成金の採択を受けていない研究者は含まれていない。また、報告書記載の発表文献総数は 2,320,832 件あり、このうち、後述の正規化処理で重複を除くと 1,893,724 件となる。

ブラウジングツールは、Perl 言語による Web ベースのアプリケーションとして構築し、Sun Solaris サーバの Apache HTTP サーバ上で CGI 環境で動作する。

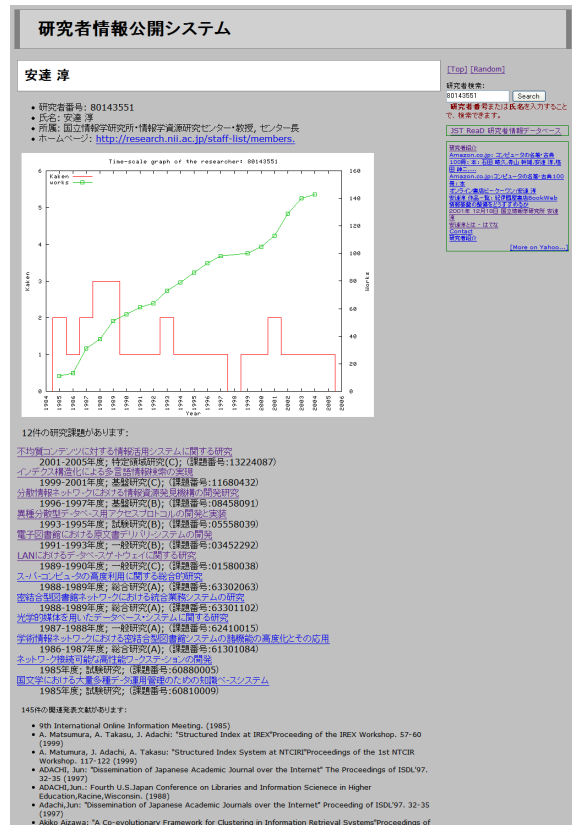


図 1 研究者情報表示画面

各研究者の情報ページとして、「安達淳(研究者番号:80143551)」を例に取って説明する(図1)。研究者情報ページにはまず、研究者の基本的な属性である氏名・所属・ホームページが表示される。その下に、研究者の科研費補助金の取得状況と報告書記載データに基づく研究発表の推移を示すグラフがあり、さらにその下に、各研究課題の情報とその詳細へのリンク、各研究報告書に記載されている研究発表の情報が載せられている。また、右側のサイドメニューには、外部情報源からの情報として、科学技術振興財団が運営する研究者情報データベース ReaD の当該研究者のページへのリンクがある。さらにその下に、Web 検索エンジン Yahoo!において当該研究者氏名で検索した結果を Yahoo! Search Web Services [11] から取得しリンク表示している。

研究者の活動を時系列に沿って可視化する一手法として、その年度毎の課題採用件数、報告書記載の

研究発表成果の累計をグラフにプロットし、表示している。このプロットは X 軸に各年度を示し、Y 軸として、課題採用件数を度数分布（赤色）で示すと同時に、各年の研究発表件数の累計を折れ線グラフ（緑色）として表示している。このプロットを見ることにより、研究者の論文成果発表と科研費採択課題それぞれの伸びや相関が一目で眺めることができ、また、経年的な推移を確認できる。

一方、こういった個々の研究者情報に辿りつくためのアクセス手段について、以下、簡単に説明する。図 2 にツールの初期画面を示す。

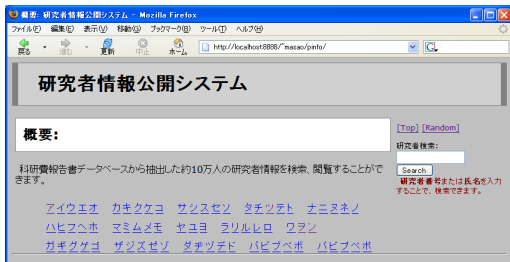


図 2 研究者情報ブラウジングツール初期画面

右側に常に表示されるサイドメニューとして、以下の機能を用意している。

- Top: 初期画面へ戻るリンク
- Random: 登録されている研究者をランダムに選択し、表示する機能
- 検索欄: 研究者番号もしくは氏名を入力して、該当する研究者を検索・表示する機能

また、左側の欄には、氏名の読みの始めの音を 50 音順に並べ、その読みから始まる研究者の検索が行えるようにしている。

氏名検索を行うと、その入力キーワードが氏名の一部にマッチする研究者が表示される。読み、アルファベット表記が登録されている研究者については、これらを入力することでも検索できる。図 3 に検索キーワード「安達」で氏名検索した例を示す。検索件数と同時に、個別研究者の氏名、研究者番号が示され、各研究者情報のページへのリンクが張られている。

また、個別の研究課題情報の閲覧もできる。図 4

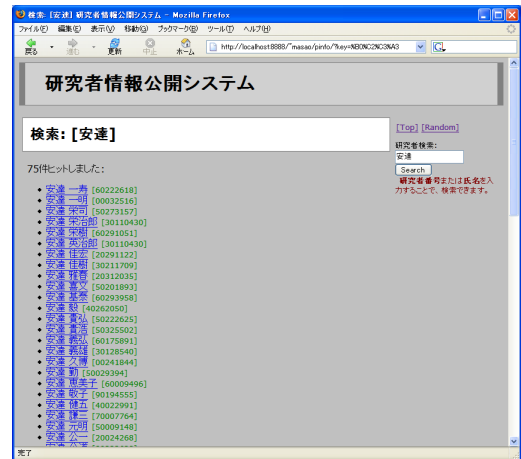


図 3 研究者氏名検索結果画面

に研究課題の表示例を示す。研究課題に表示される情報は、元の科研費 DB とほぼ同じで、研究課題名、課題番号、研究代表者、キーワード、共同研究者、研究の概要、発表報告文献などが表示されている。このうち、研究代表者、研究分担者の個別研究者は、各研究者用ページへのリンクとなっており、それら研究者の情報へと辿れる。

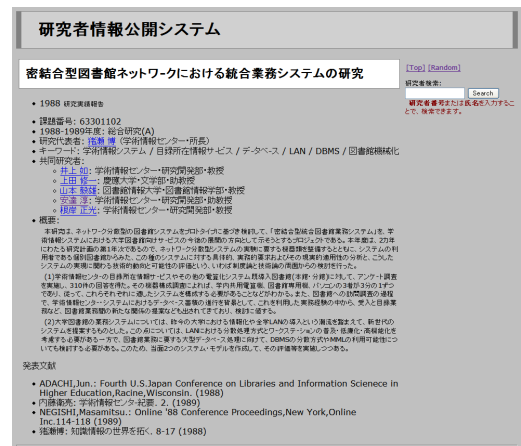


図 4 研究課題情報表示画面

4 考察

4.1 リンケージ

3 節で述べたように、本ブラウジングツールでは、研究者情報ディレクトリ ReaD を外部情報源として採用し、研究者情報を表示した際に、ReaD データ

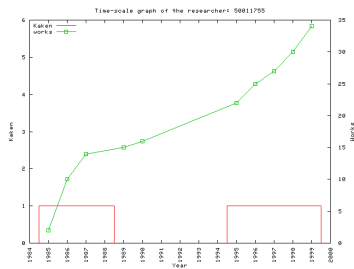


図5 プロット例1

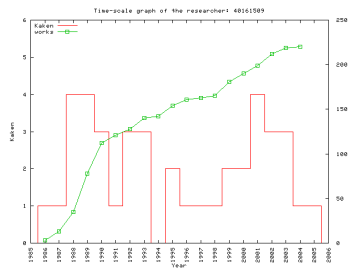


図6 プロット例2

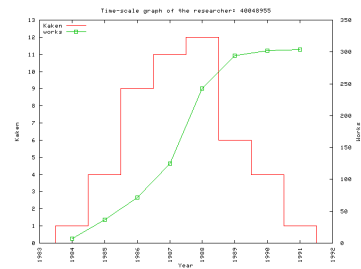


図7 プロット例3

ベース上の当該研究者の情報ページへのリンクを張ったり、ホームページの情報を抽出して、ページ内にリンク表示させたりする機能を設けた。また、Web 上一般のサーチエンジンである Yahoo! に対して Web Services を経由して研究者氏名での検索を行い、Web 上に存在する当該研究者の情報を容易に辿れるようにした。

これらのように、研究者情報を Web 上の情報源と結び付けて表示することにより、その研究者についての様々なコンテキストを提供することができる。例えば、既存の Web 上にはその研究者が提供している自身の研究室における研究紹介のページや、書店サイトが提供している著書や著者紹介のページ、またニュース記事等でとりあげられたことがあればその記事やコメント、さらには、レビューや掲示板といった形での議論など、様々なコンテンツがある。科研費 DB のようなある程度まとまった研究者情報に、これら Web 上の様々な情報源と結び付けるリンケージを加えることにより、その研究者が行う活動がより広範囲に理解でき、さらには、これまでは見えなかった活動の全体像が見えてくる。

一方、その他の情報源として書誌データベースがあり、これとのリンケージにより、当該研究者の執筆論文の追跡、調査がさらに容易に行えるようになることが期待できる。この機能は现阶段のブラウジングツールでは未実装であるが、国立情報学研究所がサービスしている NII 論文情報ナビゲータ CiNii データベース [12] のデータとのリンケージにより、より詳細で高品質な情報へとリンクすることができ、科研費 DB におけるデータと相補的に利用することにより、より精細な情報を提供できる可能性が

広がる。しかしながら、もう一方では、これらの新たなデータベースとリンケージを行う際には、科研費 DB にあったような一意の研究者番号が付与されていないデータとなるため、同姓同名などの同定の問題を解決する必要があり、高精度のリンケージエンジンが必要とされる [13]。

4.2 可視化手法

3 節で述べたように、本ブラウジングツールには各研究者の情報を時系列に沿ってプロットして提示する機能がある。これは、テキストだけでは把握しにくい時系列的な推移を分かりやすく提示することにより、その把握を支援することを目的としたものである。また、単に把握、理解を助けるだけでなく、これらの情報を使った更なる分析も可能となるものと考えている。

図 5-7 に個別研究者の時系列課題・発表プロットの一例を挙げる。典型的には、課題採択数が膨大になれば、当然それに伴う発表の数も増える。また、課題採択の傾向も比較的切れ目なく科研費を取得している研究者もいれば、科研費取得の頻度の高くない研究者ではプロットがまばらに分散するなどの違いが一目で分かる。これらを眺めていくだけでも、科研費を取り巻く様々な典型的なパターンが見えてくる。例えば、開始当初より切れ目なく研究課題を採択され、その中で発表された文献も多数にのぼる研究者や、採択件数の伸びが年数に従って増加傾向を示すなど、その典型的パターンも見てとることができる。

一方で、これらは単に科研費報告書に記載された論文数をプロットしたものであり、個々の研究者による研究発表を網羅するものではない点に注意する

必要がある。例えば、科研費採択の無かった年度については当然発表件数が0件となり、発表実績数の累計は伸びを示さない。また、年度報告に間に合わない年度途中などの最新情報は追加されず、発表実績数の累計は伸びを示さない。さらに、この論文数は課題報告書記載のもの全てをプロットしたものであり、例えば、研究分担者による発表文献を始め、当該研究者がそもそも著者として名を連ねていない発表文献も含まれている。このように、個々の研究者の情報という点からはプロットそのものが正確性に欠ける面があり、前節で述べた論文書誌データベースなどの情報を用いて補正を行う必要がある。

論文発表年の同定は、科研費 DB における報告書記載データの書式に沿って、末尾の括弧内のデータを取り出すことにより行っている。例えば、「相澤 彰子: "情報空間における双対的クラスタリングの試み" 人工知能学会第 48 回人工知能基礎論研究会資料, SIG-FAI-A104. 85-90 (2002)」のような報告書記載の発表文献の参照があった場合、末尾の年表記「(2002)」の部分を抽出している。また、例えば印刷中の段階の論文などを参照することにより、各年度間にまたがって複数回参照されている論文もあるが、これらについては、簡単なルールベースの簡易同定処理を行った。このルールは、「et al,etc,in press,to appear,印刷中,ほか,他」といった文献参照中で異表記の多い文字列を予め削除した上で、空白類や大文字・小文字を正規化し、同一文字列となった書誌事項があった場合、それらは重複書誌であるとみなすような処理を行った。このため、これらの単純なルールで除ける重複書誌以外では、重複書誌でありながら表記揺れの関係などにより多くの文献が多重にカウントされてしまっている。このため可視化・ブラウジングツールにおける正確性は確保されていないので、文献数等の解釈に際しては注意が必要である。これら同定処理については、表記揺れがあった場合でも書誌同定できるような、より頑健なリンケージエンジンを組み込み、さらに正規化処理を行うことにより、正確性を上げることが可能であると思われる。この点については、今後の課題として残っている。

4.3 個人情報

科研費は公的研究助成金であり、その成果が広く還元されることは研究者にとっても、一般社会においても重要なこととなると思われる。その一方で、科研費 DB に登録された情報を他の情報源と統合し、リンクをつなげることで、必ずしもこれまでは見えてこなかった情報とをリンクしてしまうことによる弊害も考えられる。例えば、研究者の個人情報保護もその一つである。一般に研究者は個人の名とその連絡先などを公表し、言論に対して責任を負う体制で仕事を進めているが、これらの研究者情報といえども、個人情報保護の観点から、研究者個人の同意無しに、それらの情報を統合したデータを自由に公開することは難しい面もあると思われる。

5 まとめ

本稿では、科研費 DB に基づくブラウジングツールについて紹介し、研究者情報の可視化・ブラウジング手法について考察した。

今後、研究活動の学際的な展開、人材育成、人材交流といった面からも、研究者情報の効率的な活用はますます重要な課題となってくるとと思われる。筆者は、本稿で述べた提供およびリンケージの手法を発展させることが、これらの課題に対し有効な手立てとなると考えている。

一方、異種 DB との統合や、さらなる Web 上の情報源との統合に向けては、多くの課題が残されている。一つには、異種情報源の統合に有効なリンケージエンジンの開発が今後の課題としてあり、人名の同定、書誌の同定といった既存の研究との連携を視野に入れた上で、合わせて取り組む必要がある。また、Web 上の情報源との統合に向けては、一般のサーチエンジンを利用するだけでなく、研究者というドメインを特定した上での「エンティティ検索」技術の開発が必要となる。これらの点について、今後の課題として開発、検討を続けていきたい。

参考文献

- [1] 科学研究費補助金データベース, オンライン, <http://seika.nii.ac.jp/>.

- [2] ReaD 研究開発支援総合ディレクトリ. オンライン. <http://read.jst.go.jp/>.
- [3] 研究者人材データベース JREC-IN. オンライン. <http://jrecin.jst.go.jp/>.
- [4] Michael Ley. DBLP computer science bibliography. Online. <http://dblp.uni-trier.de/>.
- [5] Michael Ley. The DBLP computer science bibliography: Evolution, research issues, perspectives. In *Proceedings of 9th International Symposium on String Processing and Information Retrieval (SPIRE 2002)*, LNCS 2476, pp. 1–10. Springer, 2002.
- [6] 岡本真. これからホームページをつくる研究者のために: ウェブから学術情報を発信する実践ガイド. 築地書館, 東京, 2006. 274p. (ISBN 4-8067-1355-X).
- [7] Chia-Hui Chang, Mohammed Kayed, Mohab Ramzy Girgis, and Khaled F. Shaalan. A survey of web information extraction systems. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 10, pp. 1411–1428, 2006.
- [8] 芳鐘冬樹, 井田正明, 野澤孝之, 宮崎和光, 喜多一. ウェブ文書からの情報抽出に関する研究の概観: シラバスデータへの適用に向けて. *大学評価・学位研究*, No. 1, pp. 133–143, 2005.
- [9] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満. Web 上の情報から人間関係ネットワークの抽出. *人工知能学会誌*, Vol. 20, No. 1, pp. 46–56, 2005.
- [10] Yuxin Wang and Keizo Oyama. Combining page group structure and content for roughly filtering researchers' homepages with high recall. *情報処理学会論文誌 (TOD)*, Vol. 47, pp. 11–23, 2006.
- [11] Yahoo! Search Web Services. オンライン. <http://developer.yahoo.com/search/>.
- [12] NII 論文情報ナビゲータ CiNii. オンライン. <http://ci.nii.ac.jp/>.
- [13] 相澤彰子, 大山敬三, 高須淳宏, 安達淳. レコード同定問題に関する研究の課題と現状. *電子情報通信学会論文誌*, Vol. J88-D-I, No. 3, pp. 576–589, 2005.

A Browsing Tool for Researcher Information Based on *Kakenhi* Database

Masao Takaku[†] Akiko Aizawa[‡] Keizo Oyama[†]

[†]Transdisciplinary Research Integration Center,
Research Organization of Information and Systems

[‡]Digital Content and Media Sciences Research Division,
National Institute of Informatics
{masao, aizawa, oyama}@nii.ac.jp

Abstract

Today, researcher information is essential for academic communication. The Grants-in-aid for Scientific Research (*Kakenhi*) is one of the largest research fund in Japan. National Institute of Informatics (NII) has built and served Kakenhi database service based on the outcome reports on Kakenhi fund. In this paper, we report a browsing tool for researcher information based on this Kakenhi Database. The tool provides and visualizes some of useful data, so it allows users to browse information for individual researcher connected with Kakenhi reports. We describe the tool's development and its features, and discuss linkages with other academic resources and/or Web-based resources, and future directions.