

研究者リゾルバー α のコンセプト

蔵川圭¹, 武田英明¹, 高久雅生², 相澤彰子¹

¹国立情報学研究所 学術コンテンツサービス研究開発センター

〒101-0012 東京都千代田区一ツ橋 2-1-2

Tel: 03-4212-2372, Fax: 03-4212-2374, E-Mail: {kurakawa, takeda, aizawa}@nii.ac.jp

²物質材料研究機構

〒305-0047 茨城県つくば市千現 1-2-1

Tel: 029-859-2813, Fax: 029-589-2400, E-Mail: TAKAKU.Masao@nims.go.jp

概要

本論文は日本の研究者の Web 上の情報を集約して提供する情報基盤となる研究者リゾルバーについて紹介する。これを実現するために、まず、研究者同定フレームワークとして研究者名に注目し、同姓同名や旧姓、漢字異体字の問題に対処する方法を提案する。ここでは科学研究費補助金の研究者番号を主とする研究者 ID 体系を採用する。科学研究費補助金の研究者番号を利用することで、すでに存在する研究課題のデータベースを活用することができ、研究者の属性として氏名（漢字、カタカナ、ローマ字）や所属履歴、業績リストなどを関連付けることができるからである。さらに、Web 上の研究者に関するリソースへリンクを張るために研究者リソースを同定する方法を提案する。研究者の名前に加えて、所属を考慮して同定する方法である。研究者リゾルバーの α 版を実装しており、現在 47 大学の研究者総覧を対象として 22,322 リンクを実現しており、全体の 15 パーセントの研究者を網羅している。また、このほかに Web 上の研究者リソースのリンク先を提示するために、研究者名と所属を用いて他のサービスへクエリをリンクとして用意している。日本語論文検索サービスとしての CiNii や、図書所蔵情報検索サービスの WebCatPlus, GoogleScholar である。科学研究費補助金のデータベースサービス KAKEN についてはダイレクトリンクを提供している。本システムは実験的な提供であるが、いくつかの日本の機関リポジトリからすでに著者リンクとしてリンク先の対象となっている。

キーワード

研究者同定, 研究者 Web リソース, リンキング, KAKEN 研究者番号

The concept of Researcher Name Resolver (alpha)

Kei KURAKAWA¹, Hideaki TAKEDA¹, Masao TAKAKU², Akiko AIZAWA¹

¹Research and Development Center for Scientific Information Resources,

National Institute of Informatics

2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-0012, JAPAN

Phone: +81-3-4212-2372, Fax: +81-3-4212-2374, E-Mail: {kurakawa, takeda, aizawa}@nii.ac.jp

²National Institute for Material Sciences

1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, JAPAN

Tel: +81-29-859-2813, Fax: +81-29-589-2400, E-Mail: TAKAKU.Masao@nims.go.jp

Abstract

This report introduces Researcher Name Resolver which aims at providing users researcher directory in Japan, and giving linking points for researcher's web resources. To implement this function, we made the researcher identification framework to solve issues of name, such as the same family name and personal name, maiden name, and kanji character variants. The framework adopts the researcher numbers of KAKEN (the most famous national grant of Japan) as built-in IDs. The numbers are related personal name (Kanji name, Katakana name, and Romanized name) and the last affiliation among the related KAKEN projects. With these attributes, we tried to identify researchers in university researcher directories (currently 47 universities). As a result, 22,311 links are identified, which is about 15 % of all the researchers. Each personal page has links to query with the name for other services, such as CiNii: a Japanese article search, WebCat Plus: a book holdings search of Japanese universities, and Google Scholar. It has direct links to the KAKEN grant search service as well. Our system shows experimental web resource linking mechanism from the researcher point of view.

Keywords

Researcher Identification, Researcher's web resource, linking, KAKEN researcher number

1. はじめに

オープンアクセスを前提とした機関リポジトリが広まるに従って、Web上に存在する論文や著書などの研究成果は増大している。研究成果の効率的な獲得を実現するために、研究ワークフローにおけるWeb上の研究リソースへのアクセスの頻度もあわせて増大している。ここでは、研究者という観点から、研究者の業績としての研究論文や関連する情報について、特にWeb上に存在する学術コミュニケーションにかかわる資源を対象として、日本の研究者の情報を集約して提供する情報サービス基盤を提供することを目指している。

研究者や著者という観点から、研究論文や著書を整理して提供するサービスはすでいくつか存在する。たとえば、Thomson ReutersのResearcherID[1]、Elsevier ScopusのAuthor Identifierは商業出版物を対象を限定して実証されている例である。MicrosoftのLibra[2]はインターネット上に存在する会議録やジャーナルを対象として、著者にIDを付与して著者単位の検索を可能とした論文検索サービスである。

このほかに、本の著者を対象とした典拠がある。1980年代のOPACの開発初期から行われ、現在にいたって付随するサービスが提供されている。たとえば、LCのAuthorities[3]、NDL OPACやNACSIS-CATの著者名典拠、OCLCのWorldCat Identities[4]である。

日本では2005年から運用が開始された、大学などの研究機関が研究成果を発信する機関リポジトリ(以下、IRという)においても、研究者を中心にとらえたシステム連携がなされている。たとえば、島根大学のSWANと研究者総覧との連携[5]、金沢大学・早稲田大学・九州大学の業績データベース連携

プロジェクト[6]がある。いずれも、研究者が中心エンティティとなっている業績DBからIRへ書誌メタデータの登録、IR側のコンテンツの充実と利用の促進を狙ったものである。

以下では、Web上に存在する研究者の情報を集約して提供するサービス基盤について、その研究者同定のためのフレームワーク、Web上の研究者資源との同定およびリンク、さらに、それらに基づくプロトタイプシステムである研究者リゾルバー α を紹介する。その上で、アプリケーションサービスの方向性について議論し、最後に結論と展望を述べる。

2. 研究者同定フレームワーク

研究者や著者を区別するとき、その名前に基づくことが一般的である。しかしながら、以下のような共通の解決すべき対象が存在する。

- A) 同姓同名
- B) 旧姓
- C) (漢字)異体字

そのため、研究者や著者を区別して管理するとき、一般的に記号からなるIDを付与する。研究者リゾルバーでは、科学研究費補助金の研究者番号を基礎とし、次のようなID番号体系を採用している。

表 1 ID 番号体系

ID (13 桁の番号)	説明
10000xxxxxxxxxx	KAKEN ID (8 桁で識別) をベースとした自動登録番号
200xxxxxxxxxxxx	ReaD ID (10 桁で識別) をベースとした自動登録番号
3xxxxxxxxxxxxxx	オリジナル登録番号

科学研究費補助金の研究者番号は、文部科学省が1939年から開始した研究助成制度において、申請する研究者に原則として生涯一つだけ付与される番号である。研究助成はすべての研究分野・多様な研究組織・研究段階を対象とした、日本で最大・唯一の国の研究費配分制度である。多くの日本の研究者はこの研究助成に申請しているため、この研究者番号を持っている。登録されている研究者は約15万人である。

科学研究費補助金に申請し採択された研究課題について記述されたデータベースがある。このデータベースは、国立情報学研究所によってWeb上に公開され[8][9]、文部科学省の採択研究課題データベースと研究者自らの報告書によって構成されている。研究課題ごとに、研究題目、研究課題番号、研究分野、研究種目、研究機関、研究実施期間、助成金額、キーワード、研究代表者や研究分担者、研究成果概要、研究業績、特許情報が列挙されている。特に研究者について、研究者番号と合わせてその氏名が、漢字、カタカナで表記され、いくつかはローマ字でも表記される。研究者には、課題実施時の所属研究機関が併記される。

日本の研究者の情報をWeb上で提供するサービスとして、Japan Science and Technology Agency (JST)が運営するReaD (Directory Database of Research and Development Activities)がある。ReaDは日本の研究機関に所属する研究者を対象としており、研究者の氏名や所属機関、学歴、学位、職歴、研究課題、研究業績などが列挙されている。基本的に研究者自身の自由登録制であり大学の教員や研究スタッフだけでなく、大学院生も含まれる。登録されている研究者は約22万人である。

3. 研究者Webリソースリンク

Web上に存在する研究者に関する情報を特定するとき、氏名表記の一致を探してみることが最初の方法として有効である。GoogleなどのWeb検索エンジンを用いてWeb上のリソースに対して全文検索してみることは広く一般的に行われている。しかしながら、先に指摘したように氏名表記だけでは研究者を同定することはできない。そのため、完全に統制されたIDを研究者に付与するのである。

研究者のIDの統制はデータベースごとに固有である。このデータベースごとに統制されたIDをID集合と呼ぼう。我々の提供する研究者リゾルバーのID集合も統制されたID集合の一つである。

リソース同士の関係を定義するとき、研究者リゾルバーのID集合と他のID集合とのマッチングを行う。ここで想定するID集合は、以下の2通りである。

1. 大学研究者総覧
2. ReaD

大学の研究者総覧とは、各大学における教員や研究者に関するWeb上に公開された情報のリストである。日本の多くの大学で研究者総覧は整備されており、研究者の氏名や所属部局、学歴、職歴、担当授業科目、研究業績、学会活動実績、連絡先などが列挙されている。ほとんどの場合システム内部に固有の研究者ID集合を持っている。

研究者リゾルバーのID集合と他のID集合とのマッチングのために、氏名表記だけでは同定のための必要十分条件とはなりえない。ここでは、以下のように2つのルールに基づいて同定を試みた。

1. 漢字氏名の一致 \wedge 所属の一致 \wedge 所属内でユニーク名
2. KAKEN 研究者番号の一致

科学研究費補助金のデータベースから、KAKEN 研究者番号に紐づけられた最新の所属機関名が取得できる。アクティブな研究者の多くがこの助成金を獲得しているため、所属機関名と実際に所属する大学の研究者総覧とが一致することが多い。そのため、所属機関の中でユニークな氏名であれば同一人物として判定する。

研究者総覧には、KAKEN 研究者番号をデータとして持っているものがある。番号が一致した場合、氏名表記は所属に関する情報とは関係なく、完全に同一人物であると判定する。

4. 研究者リゾルバー α の実装

研究者リゾルバーの α リリースを実装して、<http://rms.nii.ac.jp>にて公開している。公開先URLにアクセスすると、研究者を検索するフィールドをもつ画面が表示される。画面の例を図1に示す。

トップページにはフィールドが一つ配置されており、検索キーワードを入力後、検索対象として研究者名、科研費研究者番号、研究者リゾルバーID、研究キーワード、研究分野名のいずれかを指定して検索を実行する。すると、該当する研究者の一覧が表示されるので、そのうちの研究者を一人選択することで、該当する研究者のWeb資源へのリンクが列挙されたページに遷移する。

研究者のページには、先に述べた大学研究者総覧やReaDに対し研究者IDの同定ができた場合に該当ページへのリンクが列挙される。現在は47大学の研究者総覧を対象として同定を試み、22,311人、全体の約15パーセントの研究者ページにリンクが張られている。ReaDへは、80,008人、全体の約55パーセントのダイレクトリンクを実現している。また、研究者氏名（和・英）および最終所属履歴を組み合わせて、日本語論文検索サイトCiNii、日本の大学図書館所蔵図書検索サイトWebcat Plus、ReaD、Google、学術情報検索サイトGoogle Scholarへの検索問い合わせを行うリンクを用意している。科学

研究費補助金の研究者番号が割り当てられた研究者は該当する KAKEN の研究者ページへのリンクがある。

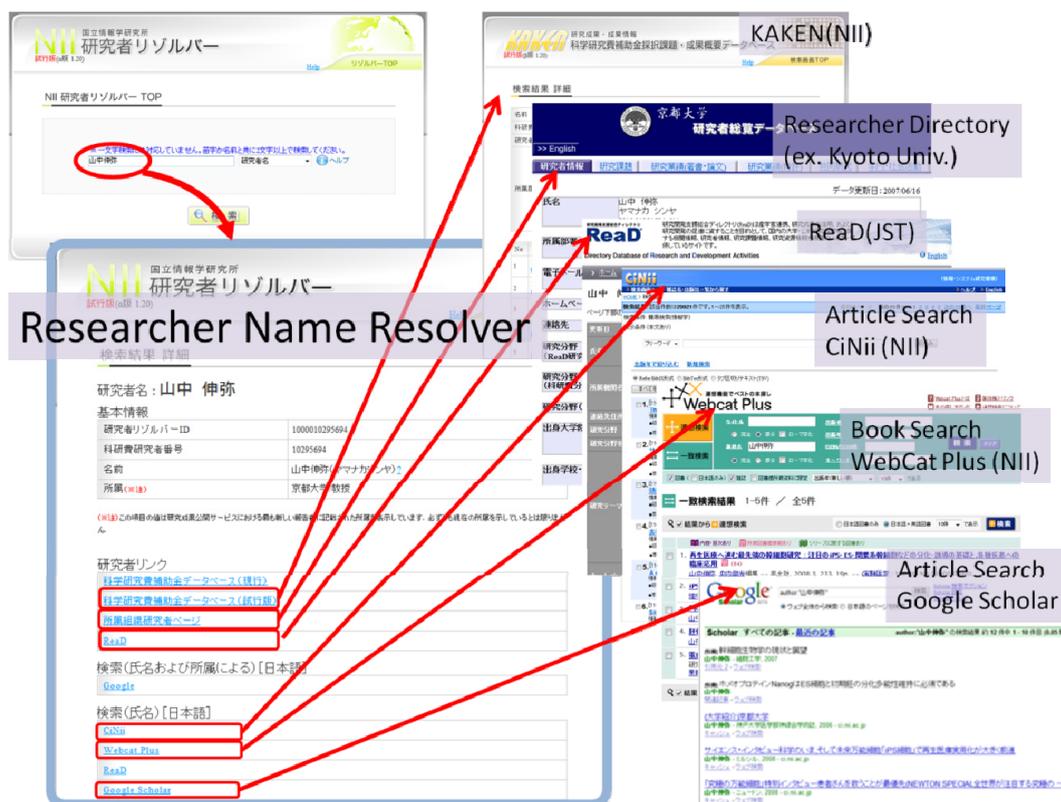


図 1 研究者リゾルバーαと Web リソースリンク

研究者リゾルバーαと機関リポジトリとのシステム間連携が、現在のところDSpaceベースの京都大学学術情報リポジトリ(KURENAI: Kyoto University Research Information Repository)[10]と静岡大学リポジトリ[11]とで行われていることを確認している。京都大学のリポジトリでは、書誌ページに記述されたdc.contributor.authorフィールドの著者名(漢字, ローマ字)を用いて、研究者リゾルバーαに著者名での検索問い合わせのためのリンクを配置している。静岡大学のリポジトリでは、dc.contributor.author, dc.contributor.alternativeの著者名(漢字, ローマ字), dc.contributor.transcriptionフィールドの著者名カタカナを用いて、著者名での問い合わせのためのリンクを配置している。

5. アプリケーションサービスの方向性

研究者エンティティに注目した Web リソースリンクの技術がアプリケーションとしてどのような方向性を示すことができるのであろうか。アプリケーションが前提とするサービスシナリオは以下の研究者ワークフローにおけるニーズを満たそうとする。

研究者の基本的な仕事の成果は研究論文として現れる。研究論文を生産するために関連した研究論文を網羅的に調査整理し、これから書こうとする研究論文を知識体系の中に位置づける。中には研究論文に関係した付属的な資料やデータについて収集したり参照したりする。すでに述べたようにこれら研究論文は Web 上で提供されるようになって、その数は増加の一途をたどっている。

たとえば Web を介して研究論文を取得するとき、キーワードを指定して Web 検索エンジンを利用したりデータベースを指定した検索サービスを利用したりする。ある分野の研究論文を網羅的に調べたいとき、キーワードによる検索は有効である。また、同一の学会誌に掲載された論文を網羅的に閲覧し、知識体系を理解しようとする。このとき、知識体系そのものの理解だけでなくある研究者の研究プロセスや研究の発展性を知ろうと試みる場合がある。より詳しく書かれた研究論文を探したり、展開された研究論文を追ったりするのである。研究論文を読む時には、研究論文を知識体系の中で探索する一方で、研究者に着目して展開を追うことの両方を行き来する。

現在の Web 検索エンジンでは、研究論文そのものの探索と研究者に注目した探索の双方について、いまだチャレンジングな課題を残している。研究論文に限ってみても Web 上には様々なデータベースが存在し、それらは分散データベースとして相互に無関係な状態で公開されている。Web という分散データベースの性質上人手によって全体の統制を取ることは困難であり、事実上それらの関係をつなぐための技術が必要とされている。基礎となるのは同じものと異なるものを区別する同定技術である。さしあたり書誌同定と研究者同定の Web を前提とした技術であろう。これによって Web 上の研究論文を論文そのものと著者とに分けて整理して、精度の高い検索サービスを提供することが可能となる。

研究論文の Web 上のロケーションを提供する検索サービスに関して言えば、すでに紹介した Microsoft Libra のほかに、CiteSeerX[12] や DBLP[13]、Google Scholar[14] がある。これらは、基本的に研究論文の Web 上のロケーションを提供し、研究者をエンティティとして捉えて検索結果を提供している。Web 上のリソースへの網羅性と同定精度に差があり、対象とする言語は英語を基本としながらも多言語への対応をしているサービスもある。

6. 結論と展望

本論文は日本の研究者の Web 上の情報を集約して提供する情報基盤となる研究者リゾルバーについて紹介した。これを実現するために、まず、研究者同定フレームワークとして研究者名に注目し、同姓同名や旧姓、漢字異体字の問題に対処する方法を提案した。ここでは科学研究費補助金の研究者番号を主とする研究者 ID 体系を採用した。科学研究費補助金の研究者番号を利用することで、すでに存在する研究課題のデータベースを活用することができ、研究者の属性として氏名（漢字、カタカナ、ローマ字）や所属履歴、業績リストなどを関連付けることができる。さらに、Web 上の研究者に関するリソースへリンクを張るために研究者リソースを同定する方法を提案した。研究者の名前に加えて、所属を考慮して同定する方法である。研究者リゾルバーの α 版を実装しており、現在 47 大学の研究者総覧を対象として 22,322 リンクを実現しており、全体の 15 パーセントの研究者を網羅している。また、このほかに Web 上の研究者リソースのリンク先を提示するために、研究者名と所属を用いて他のサービスへクエリをリンクとして用意した。日本語論文検索サービスとしての CiNii や、図書館蔵情報検索サービスの WebCatPlus、Google Scholar である。科学研究費補助金のデータベースサービス KAKEN についてはダイレクトリンクを提供した。本システムは実験的な提供であるが、いくつかの日本の機関リポジトリからすでに著者リンクとしてリンク先の対象となっている。

今後は、Web 上の研究者リソースのリンク先を増やすために研究者総覧の対象範囲を広げることを予定している。また、Web サービス API を用意し、RDF/XML によってシステム間連携を容易にすることを予定している。

参考文献

- [1] Thomson Reuters, ResearcherID, <http://www.researcherid.com/>, (accessed 2009-02-16).
- [2] Microsoft Libra Academic Search, <http://libra.msra.cn/Default.aspx>, (accessed 2009-02-16).
- [3] Library of Congress Authorities, <http://authorities.loc.gov/>, (accessed 2009-02-16).
- [4] OCLC WorldCat Identities, <http://orlabs.oclc.org/Identities/>, (accessed 2009-02-16).
- [5] 福山栄作, 高清水直美, 中井陽子, 昌子喜信. 島根大学学術情報リポジトリ SWAN におけるセルフアーカイビングの実際:大学評価情報データベース連携と登録支援機能を中心として. 情報管理. 2008, vol.51, no.4, p.260-272.
- [6] 業績データベース・機関リポジトリ連携プロジェクト (金沢大学, 早稲田大学, 九州大学), <http://www.lib.kanazawa-u.ac.jp/kura/achievement/index.html>, (accessed 2009-02-16).
- [7] 小野真由美, 井上創造, 星子奈美. 研究者業績情報システムと学術情報リポジトリの連携. デジタル図書館. No.33, 2007
- [8] KAKEN: the database of grants-in-aid for scientific research under the ministry of education, culture, sports, science and technology of Japanese government, <http://seika.nii.ac.jp/>, (accessed 2009-02-16)
- [9] KAKEN (beta): beta release implementation of KAKEN, <http://seikaplus.csc.nii.ac.jp/>, (accessed 2009-02-16)
- [10] KURENAI: Kyoto University repository, <http://repository.kulib.kyoto-u.ac.jp/dspace/>, (accessed 2009-02-16)
- [11] SURE: Shizuoka University repository, <http://ir.lib.shizuoka.ac.jp/>, (accessed 2009-02-16)
- [12] CiteseerX(beta), <http://citeseerx.ist.psu.edu/>, (accessed 2009-02-16)
- [13] DBLP, <http://www.informatik.uni-trier.de/~ley/db/>, (accessed 2009-02-16)
- [14] Google Scholar, <http://scholar.google.com>, (accessed 2009-02-16)