Researcher Name Resolver: A framework for researcher identification in Japan

Kei Kurakawa[1], Hideaki Takeda[1], Masao Takaku[2], and Akiko Aizawa[1]

(1) Research and Development Center for Scientific Information Resources, National Institute of Informatics, Japan
(2) National Institute for Materials Science, Japan

Introduction

Institutional repositories with the aim of open access are gradually spreading in academia, and more and more research articles and academic books are being archived on the web. In particular, researchers are accessing more and more electronic articles, papers, and books on the web. This paper describes an information service that firstly provides researcher name authority on the web, and secondly gathers the web locations of academic information resources and organizes them for individual researchers (especially researchers working in Japan).

There already exist several services that arrange articles, papers, and books by author. For example, for registered researchers, Thomson Reuters provides a bibliographic list with an h-index research performance indicator and journal impact factor, called ResearcherID [1]. Elsevier features a function, called Author Identifier, that arranges articles by author on Scopus, a proprietary search service for articles on the web [2]. Microsoft has experimentally released a search service, called Academic Search [3], for all resources, especially openly accessible articles of conference proceedings and journals on the web. It automatically arranges articles by author, conference name, and journal title.

Authority files in bibliographic records for books are also important information linked to researchers. National and university libraries have provided OPAC (Online Public Access Catalogue) and related services since the 1980s. For instance, the Library of Congress (LC) released Authorities [4], and OCLC (Online Computer Library Center) has released a beta service, WorldCat Identities [5], which shows book author profiles. In Japan, National Diet Library (NDL) and the National Institute of Informatics (NII) provide the authority files of OPAC.

Japanese university libraries and research institutes started to archive academic electronic resources in Institutional Repositories (IRs) in 2005. Some of these IRs enhanced service integration functions for authors. Shimane University developed a system integration mechanism linking its IR, called SWAN, and its faculty and staff directory [6]. The joint library group of Kanazawa University, Waseda University, and Kyushu University have a system integration project linking databases cataloguing the research achievements of their faculty and their IRs [7]. These systems are intended to promote efficient data entry and usage of IRs.

The above author related services apparently need name authority to identify author or creator of works. Martin Enserink surveyed the whole researcher identification framework in the current systems[8]. ResearcherID of Thomson Reuters is the front runner of researcher identification system. In the context of IRs, E-Prints has its own name authority file. arXiv is implemented to provide author publication listings, that is expected to work with a social network system, i.e., Facebook[9]. The JISC funded names project implements name authority service on the web which provides UK researcher information records in several machine readable formats to external systems[10]. In the late 2009, Thomson Rueters and Nature publishing group lead to propose ORCID (Open Researcher and Contributor ID), non-proprietary and individual researcher identification system that ought to be a catalyst for all over existing identification systems[11].

In the following sections, we will explain our service platform which firstly provides researcher name authority on the web, secondly aggregates the locations of researcher's resources and arranges them by individual researcher. The Researcher Name Resolver (alpha) is based on the researcher identification framework and web resource linking mechanism.

Researcher identification framework

Identifying researchers or authors generally requires names. Nevertheless, we come across the following name issues especially in regard to Japanese.

A) The same family name and personal name
B) Maiden name usage
C) Kanji character variants

To organize researchers or authors, we usually assign symbolic IDs for each person. The Researcher Name Resolver adopts a numbering ID system (see Table 1) partly based on the researcher number belonging to Grants-in-aid (KAKENHI) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and on the ReaD ID system of the Japan Science and Technology Agency (JST).

Table 1 Numbering System for Researcher Name Resolver

ID (13 digit number) for Researcher Name Resolver	Description
10000xxxxxxxx	ID assignment for KAKENHI registrants. xxxxxxxx is fulfilled with KAKENHI researcher ID (8 digits).
200xxxxxxxxxx	ID assignment for ReaD registrants, excluding KAKENHI registrants. xxxxxxxxxx is fulfilled with ReaD ID (10 digits).
3xxxxxxxxxxxx	ID assignment for none KAKENHI and ReaD registrants.

The KAKENHI dates from 1939 when it was established by MEXT's predecessor the Ministry of Education, Science and Culture. It uses eight digits to register a researcher applying for a grant. The number is unique and given to a researcher for his or her whole academic life. KAKENHI covers all research fields and all stages of research activities. It is the only nationwide grant in Japan and the biggest among Japanese research budget items. Most researchers in Japan have applied for this grant and have been registered (currently 175,677 are registered in KAKEN as mentioned later).

The public can access awards information through a database with a search service on the web, which we call KAKEN [12]. The database was constructed and is managed by NII, and its information consists of award profiles compiled by MEXT and documents reported by researchers. The description for each award includes project title, project number, project year, research field, research category, institute, research period budget, keywords, project members including principal investigators and co-investigators, abstracts of annual reports, research achievements, and industrial property information. The project members are listed by their Kanji name (Chinese characters), Katakana equivalent (Japanese phonetic script), alphabetical equivalent, and registered ID. Kanji name, Katakana name, and ID are mandatory. The affiliations of the members (for the project year) are listed as well.

JST provides a Japanese researcher directory service on the web, called ReaD (Directory Database of Research and Development Activities). ReaD is known to all researchers in Japan, and its researcher profiles list name, affiliation, educational background, professional history, projects, achievements, etc. Registration is free and the directory includes about 200,000 researchers, staff, and graduate students. The profiles are updated by the registrants.

Web resource linking

To associate information resources on the web with a specific researcher, one could attempt to match resources with the characters of the name by means of a web search engine such as Google. However, as mentioned above, the characters of a name are not sufficient for identifying resources associated with a specific researcher. Hence, we assign an ID to each researcher.

ID control has to be effective only on each database. Here, we call the IDs of an individual database a set of IDs. The IDs of the Researcher Name Resolver belong to such a set.

For defining relationships among resources, we match the set of IDs in Researcher Name Resolver

and other sets of IDs in the following researcher databases:

1. Directories of faculty and research staff in a university
2. ReaD

Most universities in Japan provide faculty and research staff directories. These directories list name, affiliations, educational background, professional experience, classes, research field, keywords, research achievements, academic society activities, social activities, contact information, and so on. They commonly have a set of IDs to identify researchers.

To find the relationships between a set of IDs in the Researcher Name Resolver and a set of IDs in a university directory, we define the following rules:

If each ID in the Researcher Name Resolver and a university researcher directory

1. has the same Kanji name characters and belongs to the same affiliation and is a unique name in the affiliation
2. or, has the same KAKENHI researcher ID,

these two IDs indicate the same researcher.

The latest affiliations of researchers are extracted from the KAKEN database. Most active researchers tend to have received this grant, so the latest affiliation extracted from KAKEN database can often be matched with the current affiliation of the researcher. Therefore, it is possible to identify two IDs for the same researcher.

Some of the university directories have a KAKENHI researcher ID data field. In such case, the ID of the directory can be completely matched with the ID of the Researcher Name Resolver.

Alpha release of Researcher Name Resolver

We have developed the alpha version of the Researcher Name Resolver and have released it at http://rns.nii.ac.jp. A user accessing the URL will see the top page with a keyword search field, as shown in Fig. 1.

The field can be filled with keywords such as researcher name, KAKENHI researcher ID, Researcher Name Resolver ID, research keyword, and research field. The objective search field can be chosen by a pull-down menu. Researchers fitting the search conditions are listed by clicking the search button after inputting a keyword in the field. Each researcher's page contains a name, latest affiliation, and links to resources on the web. The links include ones to university research directories and ReaD. Currently, 47 university directories are covered. The system has 22,311 links, about 15% of all registered researchers. There are 80,008 links to ReaD, or about 55% of all registered researchers. The page also has links for sending a query with a researcher name (English or Japanese) and the latest affiliation for the CiNii (a Japanese journal article search service provided by NII), Webcat Plus (another NII search service for book holdings of Japanese universities), ReaD, Google, and Google Scholar. It has a link to the KAKEN database as well.

The Researcher Name Resolver also give OpenSearch API for the search service and provide researcher related attributes and links in RDF/XML for the semantic web. This will give it the ability to integrate systems such as institutional repositories and researcher directories.

Currently, the Kyoto University repository (KURENAI) [14] and Shizuoka University repository (SURE) [15] link to our system. Both repositories are D-Space based. KURENAI has links on each repository item page for a query with an author name (Kanji or alphabetical name) extracted from the dc.contributor.author field. SURE has a link on each author page for a query with Kanji or alphabetical name extracted from dc.contributor.author or dc.contributor.alternative field or a Katakana name extracted from dc.contributor.transcription field on an item page.

Conclusions and future work

This report introduced the Researcher Name Resolver that provides users with name authority service of researchers in Japan and giving links to researcher web resources. To implement its functions, we made the researcher identification framework to resolve name issues, such as researchers with the same family name and personal name, maiden name, and kanji character variants. The framework uses the researcher numbers of KAKENHI and ReaD to construct some IDs. The numbers of KAKENHI are related to personal names (kanji, katakana, and alphabetical) and

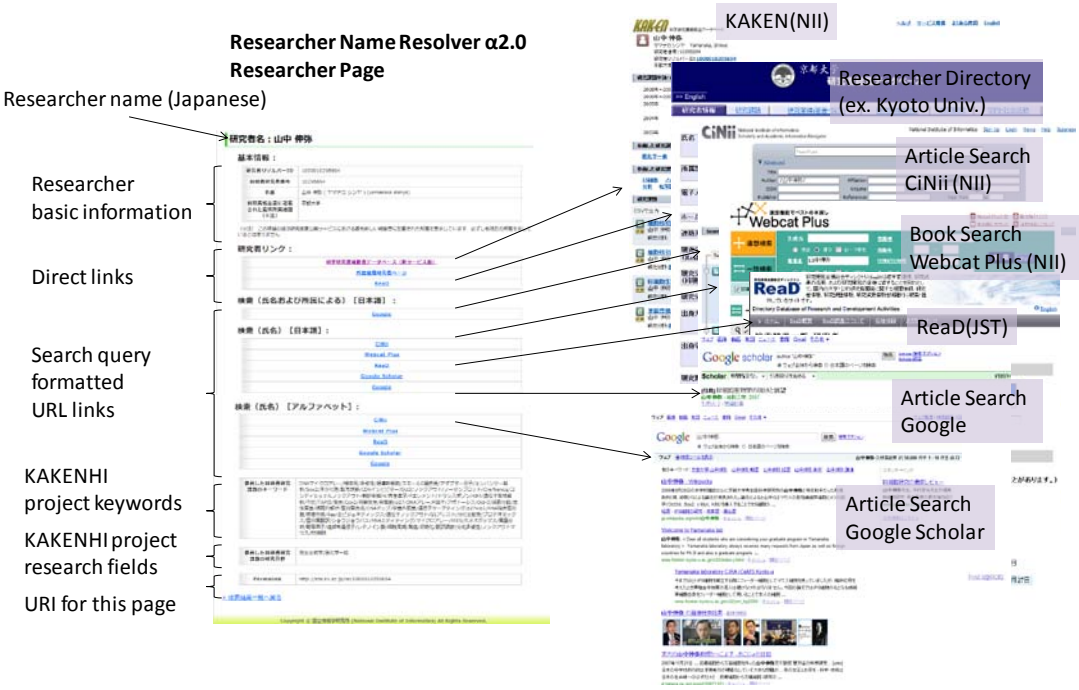

Fig. 1 Resarcher Name Resolver (α 2.0)

the last affiliation among the related KAKENHI projects. With these attributes, we tried to identify researchers in university researcher directories (currently 47 universities). As a result, 22,311 links have been identified, which is about 15% of all researchers. It also has 80,008 links to ReaD, which is about 55% of all registered researchers. Each personal page has links to queries with the name for other services, such as CiNii: a Japanese article search, Webcat Plus: a book holdings search of Japanese universities, ReaD, Google and Google Scholar. It has direct links to the KAKEN grant search service as well. Our experimental web resource linking mechanism is oriented towards the researcher.

In the future, to expand its coverage and enrich its attributes, other researcher authorities will be inter-connected. We will expand the target university researcher directories to cover all researchers in Japan.

References

[1] Thomson Reuters ResearcherID, http://www.researcherid.com/, (accessed 2010-03-01).

[2] Scopus Author Identifier, http://info.scopus.com/etc/authoridentifier/, (accessed 2010-03-01).

[3] Microsoft Academic Search, http://academic.research.microsoft.com/, (accessed 2010-03-01).

[4] Library of Congress Authorities, http://authorities.loc.gov/, (accessed 2010-03-01).

[5] OCLC WorldCat Identities, http://orlabs.oclc.org/Identities/, (accessed 2010-03-01).

[6] SWAN: Shimane University Web Archives of kNowledge, http://www.lib.shimane-u.ac.jp/0/collection/repo/, (accessed 2010-03-01).

[7] A Project on Data Sharing for Achievement Database and Institutional Repository, (Kanazawa University, Waseda University, Kyushu University), http://www.lib.kanazawa-u.ac.jp/kura/achievement/index.html, (accessed 2010-03-01).

[8] Martin Enserink, "Are You Ready to Become a Number?," *Science*, Vol. 323, no. 5922, pp. 1662 - 1664, (2009), DOI:10.1126/science.323.5922.1662

[9] Simeon Warner, Author Identifiers in Scholarly Repositories. *Open Repositories 2009*, extended abstract, 4 pages, (2009)

[10] Amanda Hill, What's in a Name? Prototyping a Name Authority Service for UK Repositories. In *Proceedings of the 10th biennial International Conference of the International Society for Knowledge Organization (ISKO2008)*

[11] Editorial, Credit where credit is due. *Nature* 462, 825 (17 December 2009) | doi:10.1038/462825a

[12] KAKEN: the database of grants-in-aid for scientific research under the ministry of education, culture, sports, science and technology of Japanese government, http://kaken.nii.ac.jp/, (accessed 2010-03-01)

[13] KURENAI: Kyoto University repository, http://repository.kulib.kyoto-u.ac.jp/dspace/, (accessed 2010-03-01)

[14] SURE: Shizuoka University repository, http://ir.lib.shizuoka.ac.jp/, (accessed 2010-03-01)