

Tips from the Experts

Author Identification Systems

A. Ben Wagner

Sciences Librarian

Science & Engineering Library

University at Buffalo

Buffalo, New York

abwagner@buffalo.edu

Copyright 2009, A. Ben Wagner. Used with permission.

Abstract

Many efforts are currently underway to disambiguate author names and assign unique identification numbers so that publications by a given scholar can be reliably grouped together. This paper reviews a number of operational and in-development services. Some systems like ResearcherId.Com depend on self-registration and self-identification of a researcher's articles. Some database producers are using a combination of computer algorithms and manual intervention to assign author identification numbers and thereby cluster publications as records are entered into their systems. Searchers doing author name searches can then use guided search features to review and select these clusters. Web of Knowledge offers Author Finder while Scopus has a similar Author Identifier feature. Various government agencies and non-profit organizations are also implementing or planning to implement additional solutions such as CrossRef's ContributorID.

The Challenge at Hand

Author name disambiguation and the association of scholarly works with the correct author have long been a problem for those wishing to develop a comprehensive list of publications for individuals. Given all the other advances in information retrieval and web technologies, some scholars are wondering if we finally have the tools needed to tackle this problem. Facets of this problem include:

- Inconsistent name formats caused by the authors themselves or editors
- Various transliteration systems, especially where different non-Roman alphabet names result in the same transliterated Roman alphabet name.
- Legal name changes
- Cultural variants in the position of surnames
- Compound or hyphenated names

The sheer volume of scholarly materials

- Highly similar names sometimes even doing similar work at the same institution.
- The large number of common names, especially certain surnames in many cultures.

Any solution likely revolves around a universal author identification number. This article discusses efforts underway to establish what would be described by catalogers as an authority list of authors which can be linked to a definitive set of papers published by each author. We will focus on systems that allow authors to self-identify their own publication list. In addition, we will briefly discuss some attempts by database vendors to uniquely identify authors as records are processed for inclusion in the database or during the formulation of a search query.

Despite the power of computers and the ingenuity of these efforts, the standard caveat of searching applies: Make no assumptions; try every variant you can think of; and carefully screen the results.

ResearcherID.Com

In January 2008, Thomson Scientific, successor to the Institute for Scientific Information, announced a completely free author identification system called ResearcherID <<http://www.researcherid.com/>> (Thomson Reuters 2008). It is advertised as a global, multi-disciplinary scholarly research community.

During the free registration process, each scholar is assigned a unique identification number and a permanent URL to one's personal profile. You can then record the institutions you have worked for, research areas of interest, descriptive text, keywords, role (e.g., academic researcher, student, or librarian), and contact information. You have full control over what information appears in your public profile. See, for example, this author's profile at <<http://www.researcherid.com/rid/B-3784-2009>>.

Most importantly, one can import one's publications list from Web of Knowledge, EndNote/EndNote Web, or the generic RIS citation format produced by many other personal citation managers. Though ResearcherID can readily be used by researchers without Web of Knowledge (WOK) access, there are certain advantages to being a WOK subscriber. Citations imported from WOK have a link back to the database record and you can view your citation metrics, including the h-index. This metrics display is very similar to the Web of Science Citation Report feature. Finally, a new tool will show authors which articles have already been added to their publication list when they're searching WOK. One can maintain up to three different publication lists per profile.

Anyone can search the registry and view public profiles to find collaborators, review publication lists and explore how research is used around the world. For example, searching for 'Wagner, A' leads to four records, one of them being this author's profile.

At this point in time, there is no independent verification of authorship for articles in ResearcherID. Though there have been no reports of false claims, this may be a concern for some. On the other hand, any vetted system would require large amounts of time and resources which would mean it could not be free and potentially widely adopted. This system is no different than so many other tools and information on the Internet, requiring a mixture of individual responsibility, an honor system, community policing, and for critical work, verification by the searcher. Perhaps of greater concern would be researchers that intentionally or unintentionally register multiple times.

Resources at their web site include:

- A colorful fact sheet: http://isiwebofknowledge.com/media/pdf/ResearcherIDFS_web.pdf
- Web tools like an embedded research badge:
<http://isiwebofknowledge.com/researcherid/ridlabs/>
- Online tutorials: http://scientific.thomsonreuters.com/training/rid/#recorded_training

AuthorClaim

AuthorClaim <<http://authorclaim.org/>> is an open source solution with the same objectives as ResearcherID, a free author registration system that allows scholars to associate their publications with a profile. The system is basically a clone of the older RePEc Author Service <<http://authors.repec.org/>> that has registered a majority of active economics researchers, over 20,000 at last report. Both services were created by the same person, Thomas Krichel, who teaches at the Palmer School of Library and Information Science at Long Island University.

During registration, one provides an e-mail address, full name, and institutional affiliations. Name variations are generated by the system which can be edited by the registrant.

The system then searches a set group of databases for candidate articles. As of May 30, 2009, six databases are harvested: arXiv.org, CiteSeerX@PSU, Current Index to Statistics, DBLP (computer science), PubMed, and E-LIS (library & information science). In the case of this author, three of my 10 refereed articles were found, not surprising given the nature of the databases searched. Unfortunately, there currently is no mechanism to manually add missing publications. This likely would discourage many scholars from participating. On the plus side, by using open access databases excepting one, Current Index to Statistics, all publications are linked back to source database records.

Registration is completed and verified by an e-mail containing an activation URL. Instead of an identification number, one receives a permanent, public URL to one's profile. See, for example, the author's profile <<http://authorclaim.org/profile/pwa1/>>.

The biggest drawback is that there is no way to browse or search profiles, even by name. Unless researchers have publicized their AuthorClaim URLs, you have no idea if they have registered. Since this service is based on open source software <<http://acis.openlib.org>>, Dr. Krichel hopes others will provide this capability. One could use the ftp protocol to download the public profiles for all registrants, but most researchers probably won't want to bother with that.

Vendor-specific Author Disambiguation Systems

MathSciNet, Web of Science, and Scopus all attempt to disambiguate author names, relying to a great degree on computer algorithms. Such programs are time-consuming to develop and maintain, and the algorithms are far from perfect. These efforts will be briefly discussed. They will be mostly of interest to current subscribers of these services who can directly test these features and come to their own conclusions about their efficacy.

From its beginning in 1940, MathSciNet had an editorial policy to uniquely identify and track authors. This started as a manual 3x5 card operation. They now run a sophisticated program that automatically identifies 80% of current authors with the other 20% requiring some level of staff intervention ([TePaske-King and Richert 2001](#)). Any searcher undoubtedly would desire that this level of authority control was universally practiced by all database producers. However, even within the narrow confines of mathematics, the process is very expensive and reportedly requires a large number of staff.

Web of Science uses a combination of guided selections during the search process and direct notifications from users regarding errors and omissions. The challenge for any such approach is to minimize both Type 1 errors (including publications that are not by a given author) and Type 2 errors (missing publications that are in fact by a given author). Like recall and precision, it is difficult to do both well at the same time.

Web of Science's guided assistance is called "Author Finder" and appears on the main search query screen. One enters a surname, first and middle initials. One gets a list with a record count for various levels of specificity such as Weinstein, B, Weinstein B*, Weinstein BA*, etc. After choosing the desired set, the results are successively limited by a pick list of broad subject areas, e.g., physical sciences, and then a ranked list of institutions. Obviously, this feature is most effective when one has researched the work history of the targeted author.

Scopus calls its feature Author Identifier <<http://www.info.sciverse.com/scopus/scopus-in-detail/tools/authoridentifier>>. This web page links to a tutorial that demonstrates the capabilities. As records are processed, computer algorithms identify authors and assign them a unique number. This number is used to group together all publications by a given author.

To search authors, one simply enters the last name and an initial or first name. A list of the potential author names will be presented along with name variants and number of publications. An Author Details page gives additional information such as citing references and citation metrics like the h-index. Search results can be further narrowed by affiliation, city, country, and subject area. Enserink (2009) reports that Scopus errs on the side of caution, providing multiple clusters for what, upon further investigation, turn out to be the same person. However, a searcher can temporarily group a number of separate authors into a single entry.

In Development

Some discussions have begun about using [OpenID](#) as a researcher\publication identification system ([Neylon 2009](#)). OpenID is an open standard that allows users to create a single digital identity and thereby log in to many different services automatically. It has been suggested that OpenID information become a standard part of article submissions or that the registration be expanded to include publication lists in addition to the institutional affiliations already captured. Though there is a degree of third party verification by virtue of the requirement to register with an OpenID service provider, there is no central party verification. As with ResearcherID, there still is nothing to prevent a given person from having multiple OpenID's, provided they have multiple e-mail addresses.

It will also be interesting to see how the [UK Names Project](#) develops. They are currently running a pilot project designed to provide UK institutional and subject repositories with a service that will reliably and uniquely identify individuals and institutions. In April 2009, they have received funding to expand the current prototype into a more comprehensive pilot system.

CrossRef is developing a system called ContributorID. Since CrossRef provides the infrastructure for persistent Digital Object Identifiers (DOIs), ideally every article as it is accepted for publication would not only be assigned a DOI, but each author would be assigned/report their unique ContributorID. Hence, a permanent link would be established between every paper and its authors. Very little information has been released about this effort, though a brief news item appeared in [CrossRef's quarterly newsletter](#).

It would be difficult to even list the universities, publishers, non-profit organizations, and national agencies and governments developing or implementing their own systems. Since 2007, Netherlands has assigned each researcher a Digital Author Identifier. The National Institutes of Health, specifically the National Center for Biotechnology Information (NCBI), is working on

a system so that it can link grants and resulting publications.

Clearly over time, certain systems will prevail or merge. For now the hope of a single worldwide researcher i.d. number is still unfulfilled, but many projects are underway and are rapidly evolving.

References

Enserink, M. 2009. Scientific publishing: are you ready to become a number? *Science* 323 (5922):1662-1664.

Neylon, C. (2009, January 20). A specialist OpenID service to provide unique researcher IDs? Message posted to: *Science in the Open*.

<http://blog.openwetware.org/scienceintheopen/2009/01/20/a-specialist-openid-service-to-provide-unique-researcher-ids/>

TePaske-King, B. and N. Richert. 2001. The identification of authors in the Mathematical Reviews Database. *Issues in Science and Technology Librarianship* (31). [Online]. Available: <http://www.istl.org/01-summer/databases.html> [Accessed: June 25, 2009].

Thomson Reuters. 2008. Thomson Scientific launches Researcher ID.com to associate a researcher with their published works. [Online Press Release]. Available: <http://scientific.thomson.com/press/2008/8429910/> [Accessed: June 25, 2009].

[Previous](#)

[Contents](#)

[Next](#)

