Author profile pages in zbMATH – improving accuracy through user interaction

Helena Mihaljević-Brandt,* Fabian Müller,† Nicolas Roy‡

June 13, 2014

A solid and distinctive online record of a scientist's research achievements is nowadays a must, in particular for a researcher at an early career stage looking for a better university position, applying for a grant or a promotion, or trying to establish new contacts and build partnerships and projects. In the global world of science there are already numerous services offering research profile sites, such as Google Scholar, Scopus or ORCID.

In the field of mathematics, however, the community still relies a lot on the traditional services MathSciNet and zbMATH, which stand for a high level of quality control regarding data inclusion and assignment. However, the enormously growing number of mathematical publications and new mathematicians begs the question of how to design author profile pages for mathematicians which fit the new trends of communication in the era of social networks, while keeping (or better increasing) the existing level of reliability.

In our contribution we plan to present some ideas from the development team of zbMATH addressing this topic.

1 Introduction

A highly visible online presence of a scientist's research activities has become almost inevitable to advance one's own academic career, even for a field like mathematics with communities of manageable size and a rather traditional communication structure. For researchers at an early career stage, a solid record of scholarly contributions is undoubtedly essential for getting a new job, a promotion or the funding for a new project. A

---

*FIZ Karlsruhe, hmb@zentralblatt-math.org
†FIZ Karlsruhe, fabian@zentralblatt-math.org
‡FIZ Karlsruhe, roy@zentralblatt-math.org

distinguishable profile increases the chance of correct attribution, recognition and citation of a scientist's impact and provides the possibility to find others working on a similar topic, establish new contacts, and build partnerships and projects.

It is therefore not surprising that various online providers offer platforms for researchers with the goal to increase the visibility of their scientific activities. Many search engines have recognized the potential of such services, Google Scholar being probably the most successful example. Also the social media hype of making a profile of oneself publicly available has been brought into the scientific world with enormous success as shown, e.g., by ResearchGate. Last but not least, seeing the need for a global and sustainable authorship administration, the initiative ORCID[1] has the promising potential of becoming the standard author identifier.

In the field of mathematics, there are two large traditional indexing and reviewing services, namely MathSciNet and zbMATH, as well as some regional bibliographic services such as Math-Net.ru for Russian literature, which offer profiles for the authors indexed in their databases.

Author profiles in zbMATH

zbMATH, known to many under its former name Zentralblatt MATH, provides access to bibliographic data and content information of publications from all areas of mathematics and its applications. Currently, about 3.4 millions publications are indexed, going back to 1755 and with an annual growth of $\sim 120,000$ new records.

zbMATH indexes only peer-reviewed literature. All items undergo an internal editorial process, and many of them, in particular in the core fields of mathematics, are reviewed additionally by an independent expert in the corresponding field. This editorial procedure ensures a high integrity level of the mathematical content, helping, e.g., to detect maleficent behaviour such as "plagiarism" or "non-sense" papers[2].

The authors of a new publication record in zbMATH are either identified with existing profiles, or new profiles are being created. The first step of this procedure is solved algorithmically by analysing the name string, but also additional information such as the co-authors. The post-processing involves also manual correction which is often initialized through user requests. The zbMATH author identification team receives such requests on a daily basis, showing an increasing interest of the mathematical community in the availability of reliable author profiles.

The quality control of the indexed documents and of the authorship identification explains why zbMATH (and similar services) enjoy high confidence of the mathematical community, despite the numerous alternative services mentioned before. Furthermore, since zbMATH aims at a complete coverage of all literature, and hence all authors, in mathematics, this is the right place to obtain an overview of the entire publishing activity in this field, where all mathematicians since 1868 are visible and not only those able or

---

[1]The Open Researcher and Contributor ID (ORCID) is a recent non-profit initiative to assign unique identifiers to authors of research publications (`http://orcid.org`).

[2]For an enlightening review in zbMATH see, e.g., `https://zbmath.org/?q=an:1202.51019`

willing to use "social-like networks". The alternative services are not able to offer such a broad coverage for the field of mathematics, and presumably this won't change in the near future. Even ORCID, the service with the most promising potential to become the standard author identifier, has yet only 146,064 non-empty profiles across all scientific fields [3], among which only a few thousand are actually mathematicians [4].
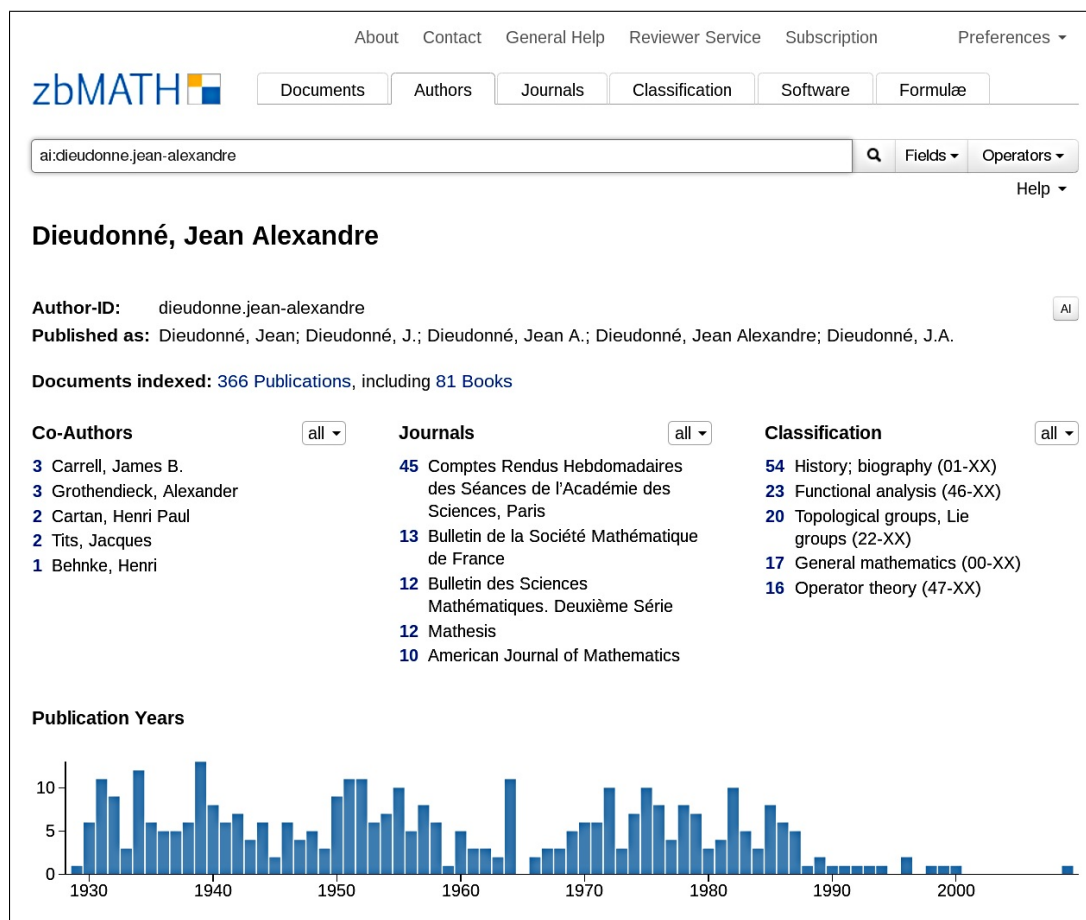

Figure 1: The author profile of Jean Alexandre Dieudonné showing his numerous publications filtered according to co-authors, source, mathematical subject and publication year, respectively.

Why is author disambiguation difficult?

The analysis of authorship has to be achieved with an enormously growing number of publications and people: currently, there are more than 3.4 millions publications corresponding to ~ 5.5 millions authorships in zbMATH, for an estimated number of 860,000

---

[3]This corresponds to the official ORCID statistics from June 6 2014 (`https://orcid.org/statistics`).

[4]We found only $\sim 3,000$ mathematicians in ORCID by matching DOIs from the publications available in the ORCID Public Data File from October 2013 with zbMATH.

author identities. The trend of maintaining profile pages and registering at various scientific networking platforms (ResearchGate, ORCID, Google Scholar, ...) is quite recent, and still not intensively picked up by mathematicians. Although the identifiers or links to these services would obviously be a great help for the identification of authorship, they are at the moment almost never available in the publications. Information like affiliation, email, etc. is a rather standard part of a publication's metadata, but these tend to change, leaving practically no other information but the name for the author identification[5]

The task of attributing an exact set of publications to a certain author based only on person name is far from trivial, due to several reasons:

- *Incompleteness:* The available data may be incomplete; some parts of the name may be missing (e.g., middle name, second familiy name) or abreviated (use of initials). In earlier years, it was also not uncommon to publish under the family name only.

- *Synonyms:* For a single author, one can face a great variability in names, in particular due to different spellings and transliterations, but also through name changes such as after marriage.

- *Homonyms:* The same name may refer to multiple individuals. This is particularly acute in the case of certain Eastern Asian names; for example, in China the top three surnames (Wang, Li and Zhang) account for more than 20% of the population.

This explains why author name disambiguation is a longstanding research topic with high relevance for bibliometric studies and publication retrieval, and why the general field named entity recognition has gained a lot of interest recently (see, e.g., [1, 2, 3]).

2 Community input

As outlined above, the problem of author disambiguation is too difficult to be tackled completely algorithmically. On the other hand, the amount of publications incorporated every year into zbMATH and other information services is too big to solve this task manually. We thus decided to open up this process at least partially to community input.

Experience shows that mathematicians take an active interest in grooming and optimizing their research profiles, both the ones they create for themselves (ORCID, ResearchGate, Google Scholar, ResearcherID) and those that were initiated automatically (Scopus Author Identifier)[6]. This correlates with results from the recent study [4], indicating that mathematicians like to be empowered when using mathematical search

---

[5]zbMATH is also involved in a project on text mining using techniques from NLP aiming, among others, at the extraction of mathematical vocabulary, which could also be used as a tool for author disambiguation.

[6]Some institutions even publish guidelines on how to increase visibility via research profiles, e.g. Utrecht University Library at `http://libguides.library.uu.nl/profiles`.

interfaces, i.e., they like having full control over what is included in a search result and the way it is presented.

At present, requests to zbMATH to clean up author profiles or correct mistakes in attribution of publications are usually presented by researchers via email, an approach that is evidently error-prone and not very scalable. Providing an interface to perform these tasks in a defined way is an obvious solution.

On the other hand, in the interest of keeping reliability as high as possible such a solution cannot be fully automated. Hence it would be desirable to include user input, but find a good combination between algorithmic attribution, user correction and manual curation by trustworthy personnel. The latter is already facilitated by an internal author disambiguation interface used by the zbMATH staff.

A public interface for author disambiguation

In order to include community input, we are faced with the task of designing a publicly accessible interface that allows users to improve the quality of publication data in author profiles. In particular, when dealing with a single author's list of publications they should be able to

- confirm correctly assigned publications,

- exclude incorrectly assigned publications,

- resolve ambiguities in algorithmic assignment,

- merge profiles that use different names (or name variants) for the same person, and

- supply links to profiles in other services and personal homepages.

Possible users of this service would be

- zbMATH authors themselves, i.e. researchers in mathematics at every stage,

- their associates (colleagues, students, coworkers etc.), as well as

- people generally interested in quality of information content (e.g. librarians).

As this list implies, we do not want nor expect users to solely care for their own author profiles, but rather see this enhancement of information as a service for science itself and the community as a whole.

Moreover, there are different usage scenarios that come to mind and that should be supported by the interface: A casual user might notice a mistake in some author's profile and just want to quickly correct it. On the other hand, an author or one of her assistants might want to completely clean up a profile, i.e., ensure that all of the author's works (and hopefully only those) are attributed to her and only to her, all name variants are correctly recognized, and as much metadata as possible is associated with the profile. As

this can be a somewhat daunting task in the case of established researchers, the interface should allow for doing some partial work and picking up later where one left off.

To make the interface accessible and easy to use, our goal was to provide a clear overview of the data and the possible ways of action that the user could take. In particular, some delicate questions concerned the outward presentation of the internal states the data could have. For example, if algorithmic assignment of a paper turned up multiple possible authorships, should we present that fact to the user when browsing a single author's publication list? If some item has been manually assigned via another user's input before, should we allow reassignment (or reconfirmation) of the same?

Furthermore, while the user should get some immediate feedback after taking an action, it must be clearly communicated that the changes will not be incorporated automatically, but rather are subject to an internal review process.

Implications for our design

As shown in Figure 2 (which shows a preliminary version of the interface), we decided to place an unobtrusive box at the top of the page with an overview of the possible actions and links to the relevant sections of the page. There is no real authentication as of now[7], but we do ask for an email address. This in principle allows for further enquiries by our author identification team, and may be turned into a real (but still lightweight) system of authentication later on.

Articles that have been manually assigned before are distinguished from others by the lack of a green `Confirm` button. They still have a red `Exclude` button as we are indeed interested in cases where users think that someone got it wrong before. Every click sends off an immediate AJAX request queuing the action for internal review, and the respective button afterwards appears pressed. We still provide an explicit submit button, which does nothing but show a reassuring message that the input has been received and will be reviewed internally. At a later stage we plan to introduce filter and sorting capabilities, e.g., to show only non-manually assigned publications or only those with ambiguities.

Finally, we provide the option to write a free text message to our author identification team, as there are bound to be some requests that cannot be reasonably handled using an all-purpose interface. In this way we at least receive this information in a defined way and together with some identifying metadata.

The user input subsequently passes through an internal checking procedure involving algorithmical and manual analyses before the corresponding data modifications are executed in the database. This step ensures that the quality standard of the data and hence zbMATH's authoritative status is preserved.

---

[7]This is primarily to lower the hurdle for people to actually use the interface. On the other hand, the question of finding a reliable authentication system allowing a mathematician to identify herself as the author of a given set of publications is far from trivial.

Hi mail@testuser.com. Thanks for your verification.

Please help us to improve zbMATH by:

- **Confirming or excluding** a document from this author profile.
- **Merging** other author profiles with this one.
- **Writing** to our author identification team.

Author Profile Improvement Page for Smith, Michel

Author-ID: smith.michel
Published as: Smith, Michel; Smith, M.

| Confirm/Exclude | Merge | Comment |

Confirm or exclude documents currently assigned to smith.michel

Documents without a confirm button have already been verified.

Smith, Michel; Stone, Jennifer
On non-metric continua that support Whitney maps. [Zbl 06298726]
Topology Appl. 170, 63-85 (2014).

    [Exclude]

Stromberg, A.J.; Griffith, W.; **Smith, M.**
Control charts for the median and interquartile range. [Zbl 05280066]
Dutter, Rudolf (ed.) et al., Developments in robust statistics. International conference, ICORS 2001, Vorau, Austria, July 23–27, 2001. Heidelberg: Physica-Verlag (ISBN 3-7908-1518-7). 368-376 (2003).

    [Confirm] [Exclude]

Krasinkiewicz, J.; **Smith, M.**
Hereditarily indecomposable continua with trivial shape. [Zbl 0528.54031]
Fundam. Math. 119, 133-134 (1983).

    [Confirm] [Exclude]

    [Cancel] [Submit]

Author Profile Improvement Page for Smith, Michel

Author-ID: smith.michel
Published as: Smith, Michel; Smith, M.

| Confirm/Exclude | Merge | Comment |

Merge the author-ID harris.simon-d with other author-IDs

[author.id1, author.id2, ...]

(find them with the Author Search)

[Submit]

Figure 2: Preliminary version of the author disambiguation interface.

3 Further ideas

For a couple of months we have been testing a prototype of the author disambiguation interface among different user groups (zbMATH editors, mathematics students, librarians). Based on the very positive feedback we are confident that the public version of the interface will enjoy high usage in the mathematical community.

Provided the expected success of our interface, we envisage to enhance it by additional features, which would require reliable authentication. Among the planned extensions would be the possibility to add comments or to tag the content of publications, or to upload and match additional metadata and sources such as references or photos.

References

[1] A. A. Ferreira, M. A. Gonçalves and A. H. Laender, *A brief survey of automatic methods for author name disambiguation.* SIGMOD Record, 41, No. 2, 15–26 (2012)

[2] A. Fahrni and M. Strube, *Jointly disambiguating and clustering concepts and entities with Markov logic.* in: Proceedings of COLING-2012. 815–831 (2012)

[3] Y. Qian, Y. Hu, J. Cui, Q. Zheng and Z. Nie, *Combining machine learning and human judgment in author disambiguation.* in: Proceedings of CIKM-2011 (2011)

[4] A. Kohlhase. *Search interfaces for mathematicians.* in: Proceedings of CICM-2014. Lecture Notes in Computer Science (2014)