Connecting Islands: Bridging zbMATH and DLMF with Scholix, a blueprint for connecting expert knowledge systems

Howard S. Cohl, Moritz Schubotz and Olaf Teschke

This paper reports on the recently launched zbMATH Links API. We discuss its potential based on the initial link partner, the National Institute of Standards and Technology Digital Library of Mathematical Functions. As the API provides machine readable data in the links, we show how one can use data from both sources for further analysis. To exemplify the simplicity, we also show how one can use zbMATH's link data in Jupyter notebooks.

1 Introduction

As reported in the last EMS Magazine (formerly Newsletter of the European Mathematical Society) [7], zbMATH Open provides application programming interfaces (API) to make zbMATH data machine accessible. We described the OAI-PMH API which enables the harvesting of zbMATH Open metadata. In contrast, this issue focuses on links between zbMATH Open and third parties. Our zbMATH Links API, available from http://purl.org/zb/14, provides a machine-readable interface for links between academic literature and other resources. To make this API interoperable with various information systems, we rely on the Scholix API standard [2]. Scholix, which is short for *A Framework for Scholarly Link eXchange*, is a long-running initiative supported by partners such as the Research Data Alliance, Crossref, and DataCite amongst many others, which aims to exchange information on research data and related scholarly articles. By exporting our data in a Scholix-compliant manner, we ensure that our data gets integrated into the worldwide ecosystem of open data. In this regard, it is not only important to export individual data sets, but also to explicitly annotate the links between different data sets in a standardized, machine readable format.

The zbMATH Open team is currently in the process of linking zbMATH Open reviews and abstracts with various partners such as

1. NIST Digital Library of Mathematical Functions (DLMF) https://dlmf.nist.gov [3, 5],
2. The On-Line Encyclopedia of Integer Sequences https://oeis.org [3],
3. The arXiv https://arxiv.org[1] [8],
4. MathOverflow https://mathoverflow.net [4],
5. and many others.

The first step, establishing links between zbMATH Open and DLMF has now been completed.

In [6] we described the details of the zbMATH Links API interface and analyzed the current links in the DLMF; statistical analysis of metadata was obtained by combining both data sources. For instance, we can analyze the distribution of Mathematics Subject Classification (MSC) classes in DLMF chapters, or the average age of the referenced publications. Let us now explain how one can proceed to generate any similar analysis in a very short time, using simple tools.

2 Jupyter notebook demonstration

One way to use data from the zbMATH Links API is via Jupyter notebooks jupyter.org. Jupyter notebooks are interactive notebooks that can be run in the browser and thus do not require any setup or configuration. In contrast to other interactive notebooks by commercial publishers, Jupyter notebooks are based on free and open source software, which implies that one is not bound to a specific vendor. Recently, Jupyter notebooks have become increasingly popular and are being used to create easily reproducible scientific workflows [1]. For this demonstration, we use Jupyter with Python and employ the library pandas pandas.pydata.org for data aggregations as well as plotly to create plotly.com interactive visualizations. In Figure 1, we create an interactive version of the MSC distribution of the articles linked in the DLMF as described in [6]. As shown, the visualization can be created in eight lines of code and fetches

[1] The mention of specific products, trademarks, or brand names is for purposes of identification only. Such mention is not to be interpreted in any way as an endorsement or certification of such products or brands by the National Institute of Standards and Technology, nor does it imply that the products so identified are necessarily the best available for the purpose. All trademarks mentioned herein belong to their respective owners.

EMS MAGAZINE 120 (2021) — DOI 10.4171/MAG-35

```
import pandas as pd
pd.options.plotting.backend = "plotly"
msc = pd.read_json(
    'https://zblink.formulasearchengine.com/links_api/statistics/msc/')
msc = msc.rename(columns={msc.columns[0]:'msc',msc.columns[1]:'count'})
msc = msc.set_index('msc')
msc.head(2)
```

msc	count
33	491
65	351

```
fig=msc.plot(y='count',kind='bar')
fig.show()
```

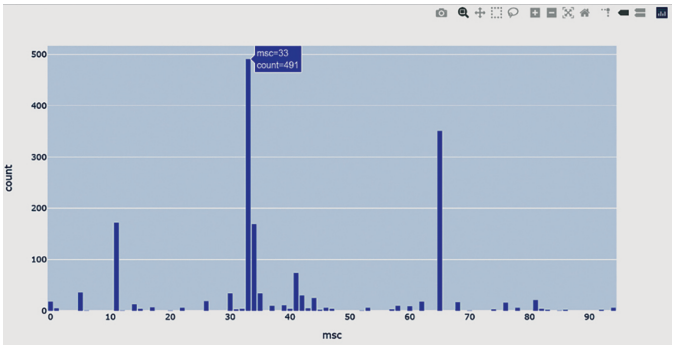

Figure 1. Jupyter notebook running on https://mybinder.org fetching and visualizing data from the zbMATH Links API

data online from our API. In pandas there is built-in support to read from our API endpoint. Thus one can use real-time results from our API without any more effort than loading any other resource from the local file system. The source code with additional examples and further links to the interactive visualization is available from https://github.com/zbMATHOpen/LinksApiJupyterDemo.

3 Conclusion and outlook

We have shown how easy it is to use the data obtained from our zbMATH Links API. While currently, DLMF links are only accessible via this API, additional links are currently in the process of being generated. Moreover, trusted third parties will be able to add new links to their respective services. Additionally, conformity with the Scholix scheme ensures that content aggregators such as OpenAIRE, DataCite, and others can integrate our data into their systems and workflows.

As of the publication of this article, the zbMATH Links data is not yet displayed on the zbMATH Open user-interface. The integration of the data and the API within our user-interface is scheduled for the second half of 2021.

References

[1] M. Beg, J. Taka, T. Kluyver, A. Konovalov, M. Ragan-Kelley, N. M. Thiéry and H. Fangohr, Using jupyter for reproducible scientific workflows. *Computing in Science Engineering* **23**, 36–46 (2021)

[2] A. Burton, M. Fenner, W. Haak and P. Manghi, Scholix metadata schema for exchange of scholarly communication links. DOI 10.5281/zenodo.1120265 (2017)

[3] K. Hulek, F. Müller, M. Schubotz and O. Teschke, Mathematical research data – an analysis through zbMATH references. *Eur. Math. Soc. Newsl.* **113**, 54–57 (2019)

[4] F. Müller, M. Schubotz and O. Teschke, References to research literature in QA forums – a case study of zbMATH links from MathOverflow. *Eur. Math. Soc. Newsl.* **114**, 50–52 (2019)

[5] *NIST Digital Library of Mathematical Functions*. dlmf.nist.gov, Release 1.1.1 of 2021-03-15, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. (2021)

[6] M. Petrera, D. Trautwein, I. Beckenbach, D. Ehsani, F. Mueller, O. Teschke, B. Gipp and M. Schubotz, zbMATH Open: API solutions and research challenges, arXiv:2106.04664 (2021)

[7] M. Schubotz and O. Teschke, zbMATH Open: Towards standardized machine interfaces to expose bibliographic metadata. *EMS Magazine* **119**, 50–53 (2021)

[8] O. Teschke, Green, gold, platinum, nickel: On the status of open access in mathematics. *Eur. Math. Soc. Newsl.* **110**, 60–63 (2018)

Howard S. Cohl is a mathematician in the Applied and Computational Mathematics Division at the National Institute of Standards and Technology in Gaithersburg, Maryland, USA. He is the Technical Editor for the DLMF project and the project leader for the NIST Digital Repository of Mathematical Formulae seeding and development projects. Cohl is a research mathematician whose main interests lie in the areas of special functions and orthogonal polynomials. He also has been exploring mathematical knowledge management and the digital expression of mostly unambiguous context-free full semantic information for mathematical formulae.

howard.cohl@nist.gov

Moritz Schubotz is a senior researcher for mathematical information retrieval and decentralized open science at zbMATH. He maintains the support for mathematical formulæ in Wikipedia and is an off-site collaborator at NIST.

moritz.schubotz@fiz-karlsruhe.de

Olaf Teschke is Managing editor of zbMATH and Vice-chair of the EMS Committee on publications and electronic dissemination.

olaf.teschke@fiz-karlsruhe.de