OpenMathMap: Interaction

Carmela Acevedo, Michael Kohlhase

Jacobs University Bremen

Abstract. Mathematics is a fast growing discipline where an average of 120 thousand papers are published per year. We seek to provide better interfaces to interact with this huge collection of documents. To study new techniques that allow users to intuitively assess and interact with mathematical publication collections we use OpenMathMap (OMM) – a computer-generated map which represents Mathematical Subject Classifications (MSC) class as “countries”, “states”, and “cities”. We present a platform that allows to build interactions with semantical meta-data based on OMM and show a variety of information services taking advantage of the semantic and hierarchical properties of the dataset at hand. We evaluate the platform to determine how successful the map is as an intuitive complement for search. The map proved to be a useful interface for interacting with and understanding the results that were being displayed.

1 Introduction

We currently live in the times of big data. It has become a great challenge for humans to discern information that interests them. Currently the most common tool for going through a large amount of data are search engines. These systems provide results organized by how relevant they are without considering the relationship between them. Grouping the results by a common topic could guide the users to interpret them and provide an overview of the underlying structure of the information.

Organizing search results brings two challenges to the surface. One of them is how to organize objects hierarchically in a way that provides a better overview of the entire set of search than by listing them individually. The second challenge is to visualize and display this information so that it is intuitive enough to guide the user to discover new connections between results by diving into a topic of interest. Maps could be a potential answer to these problems. People are familiar with maps from a young age. We learn how to read and navigate them. They display the geographical world in a global, hierarchical way and at the same time help us comprehend our surroundings. Having global and local information simultaneously provides the information context. We will investigate whether a map can provide a hierarchical overview of non-geographical information that would promote discovery and investigation.

In the last fifteen years there have been an average of more than 120 thousand publications in mathematics per year [Zen13a]. The last version of the Mathematical Subject Classification (MSC [MSC12]) has been decided upon jointly between Mathematical Reviews (MR) and Zentralblatt für Mathematik (Zbl). All publications have been assigned a main MSC class they belong to and a number of secondary classes related to it. With this information a map of mathematics was generated. This is a good opportunity to research the possibility of using this map to retrieve information related to these publications. The OpenMathMap contains all first and second level MSC classes while showing how they relate to each other[DKL13; DK13; Dör13]. We seek to take advantage of the fact that two classes that are close to each other contain publications that are related. This interactive platform could serve as an overview for users to navigate and discover information.

We will study the use of the interactive map as a tool for a semantic approach to search. To illustrate this we consider the scenario of a student that is undecided as to what to focus on later on in her career in mathematics. She wants to be able to narrow down her choices to a handful. For this she needs information about which areas are thriving at the moment, what they are about and which authors are active in them. We will evaluate how integrating interactive features to the map facilitate information retrieval in order to answer these specific questions. Finally, we prove that by using maps for search the user gains more information about the global topic and at the same time finds answer more easily. The current state of the map and further information can be found at http://map.mathweb.org.

2 State of the art

The Paperscape project uses maps to visualize the arXiv dataset. Each circle in their map represents one paper [GK14]. The distance between them is related to how similar they are, the size of each circle is directly related to how many citations each specific publication has.

Dave Rusin created a Math Atlas in the year 2000 [Rus02]. This atlas shows a circle for each of Zentralblatt's MSC classes [MAT14]. These have a size depending on the amount of papers published in them. They are positioned according to their relation to the rest of the other categories (Fig. 1). The distance between classes is directly related to how many papers have been published under both categories. This is an almost purely visual representation of mathematics. The only possible interaction with the map is clicking each circle which redirects the user to another page describing the corresponding MSC class.

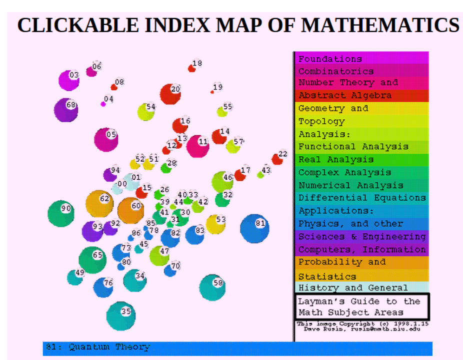


Fig. 1: Rusin's Math Atlas [Rus01]

A similar attempt to Rusin's is done by the OpenMathMap project. It uses the Zentralblatt dataset to display all of the MSC classes including their similarity and amount of publications.

3 Preliminaries

In the scope of this project we are dealing with mathematical publications. All the data was provided by Zentralblatt and it was later put together into one cohesive system with the aid of mapping tools.

Zentralblatt is a service that provides reviews and publication content from pure and applied mathematics [Wik13; Zen13b]. It contains a comprehensive database of all mathematical publications dating from 1755 which contains more than 3 million bibliographical entries. In order to keep an organized structure of these publications they have decided upon a Mathematical Subject Classification jointly with the Americans Mathematical Society's Mathematical Review. Each published paper is given a primary MSC class and several accompanying secondary classes.

Zentralblatt's portal provides the opportunity to search their database. However, this search does not take advantage of the relationship between its results. We are seeking to provide an interface that will give a visualization of such results.

4 Scenario and Use Cases

We will focus our research on one specific scenario from which we develop three use cases. The goal is to evaluate whether a map is a viable search tool.

4.1 Scenario

We consider a mathematics student, Stephanie. She is looking to find more information about the master's programs she is admitted to. There is a set of subjects she would like to study but she wants to know which is the most current. Since she is very interested in research she wants to know how these places rank in publications. It is important for her that the respective departments and their authors are active in the topics she wants to focus on.

From this scenario we extract three use cases that we will be focusing on:

4.2 Current Areas of Mathematics

She wants to know which areas of mathematics are publishing the most in the past years. The OpenMathMap provides the backbone to display all of this information at once. We will solve this problem by providing an overlay to the map that shows how active areas of mathematics currently are.

4.3 Relevant author's information

The map can be used to provide author's information in a similar way as before. By providing an overlay to the map that displays this information she could look at it all at once and analyze it better.

4.4 Result filtering using semantic information

To investigate about the topics she is interested in she searches for publications. The results of a search can be displayed on the map to show a general distribution of them. Knowing the geographical location of the classes can further help filter results that are of special interest.

5 Use cases implementation

In order to reach a platform that provides the mentioned use cases we have to start by developing some basic services.

We started with a tile map being served by a Javascript framework [Lea14]. We obtain the OpenStreetMap (OSM) [OSM14] XML file that was used to generate the map. This file contained all the countries' boundaries and is uploaded to a geometrical database (PostGIS) [Pos14]. This database provides tools for finding the intersection between two points and the centroid of a polygon which are extensively used later. At the current state of the system it is enough to generate a new OSM file and upload it to the database in order to reuse it for a newly generated map. This database also contains information about MSC classes and authors obtained from the Zentralblatt dataset.

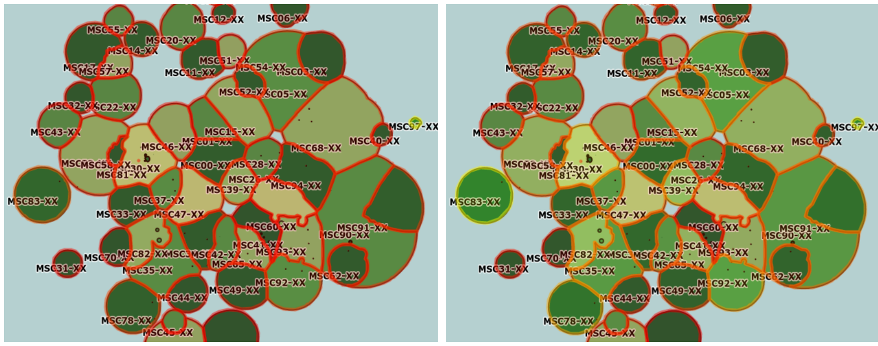
After having this available and being served by a framework (Django in this case) we create navigation opportunities for the user. The user can click a point to know which country it belongs to. Searching can also be done under the following categories: author, class and formula or text in a paper's abstract.

Search results from all these modes result in markers appearing on the map. Each marker when clicked launches a popup with the title of the marker and a link to the Zentralblatt page displaying the result. These markers are positioned in the centroid of the class they belong to.

The information provided by the search is retrieved from the database in all cases except of the formula search. In this case we use a service provided by Zentralblatt which allows for the search of formula's in LaTeXformat and text in abstracts or titles of publications.

To show the current areas of mathematics (4.2) we use an overlay on the map. The coloring provides an extra degree to the map with more information. Values are interpolated between red (`'#ff000'`) and yellow (`'#ffff00'`) mapped between 0 and 100 respectively. The number that is provided for each class is something that we had to investigate. It was necessary to narrow down the information about the year of publication in each class to a single number.

We have to pick a value that is related to the change in publication behavior over time. At first we picked the average year of publication per class. This,



(a) Normalizing with maximum (b) Normalizing without outliers

Fig. 2: Normalizing values considering outliers.

however, could have the problem that classes with very different behavior had the same average. Then we tried to use the maximum which did not provide enough different values. Finally we decided on comparing the amount of publications that have occurred in a class during the last 5 years with the rest of the publications. We calculate the percentage that this amount represents which offers an insight on how active this class is compared to the past.

Since we scale all values between the minimum and the maximum outliers are a problem, concentrating values in the mid range. Because of this we disregard them and color them as either the minimum or the maximum (Fig. 2)

A single marker symbolizing an author is not enough to encode their trajectory throughout time (4.3). We are already positioning the marker on the class for which the author has contributed the most. Each author can display two different overlays.

The “Classes” overlay is meant to provide an understanding of which classes this author has published in. We create an overlay over the classes this author published under. Classes with a high value are encoded yellow and red those with a low value interpolating everything in between.

The “Timeline” overlay seeks to offer a timeline for the author’s work. This is similar to the first use case where we try to display the progress of the classes over time. This case is simpler since most authors do not have many publications. In this case the average does offer a meaningful intuition of the authors publication behavior. Commonly there also are no outliers because authors publish during their active careers which is a short window of time.

When searching results appear all over the map. Our user is only interested in her top MSC classes and hence would like to filter the results (4.4). Since there are potentially a lot of results a solution that lets the user pick markers individually is not viable. We also want to take advantage of the fact that similar classes are geographically near in the map. After obtaining the results a tool can be offered. This tool will let the user draw on top of the map. This will generate a layer with geometrical objects that encapsulate the desired results. After a

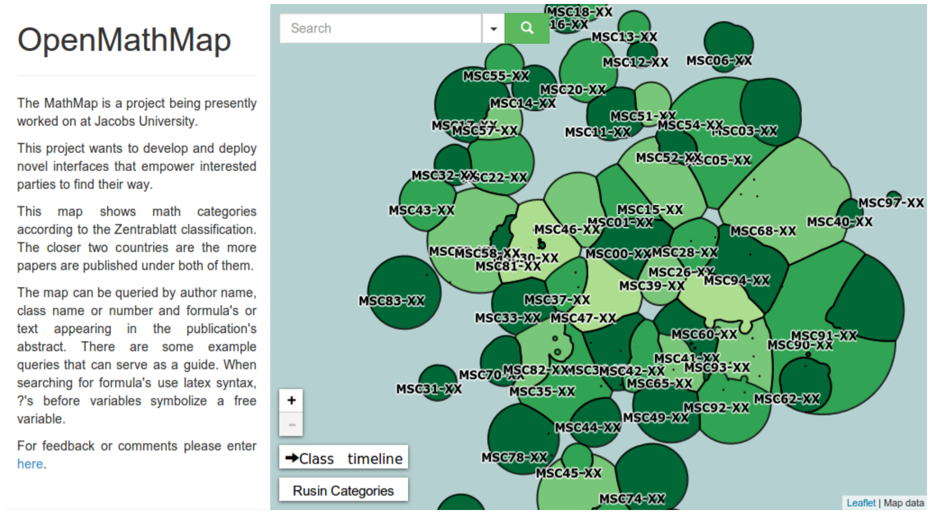
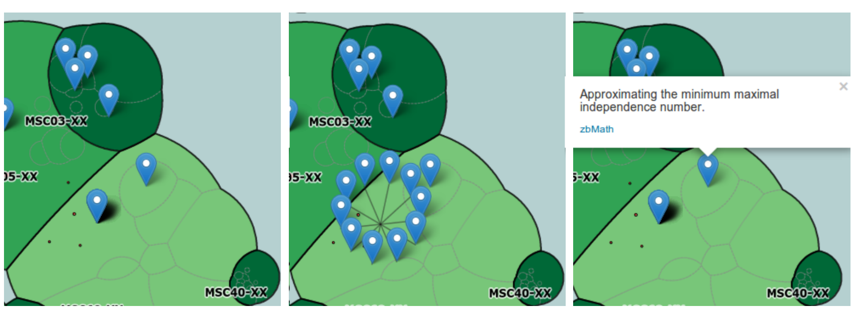


Fig. 3: General layout.

subset of the results has been selected they can confirm their selection and only the selected ones would remain on the map. This is done by going through all the markers and checking if they are inside one of the drawn polygons. Since the map can only display a limited amount of information we offer a link to the Zentralblatt site with information about all the individual results. This link is updated throughout the filtering process to refer only to the relevant markers.



(a) Markers. (b) Marker expansion. (c) Marker popup.

Fig. 4: Marker overview.

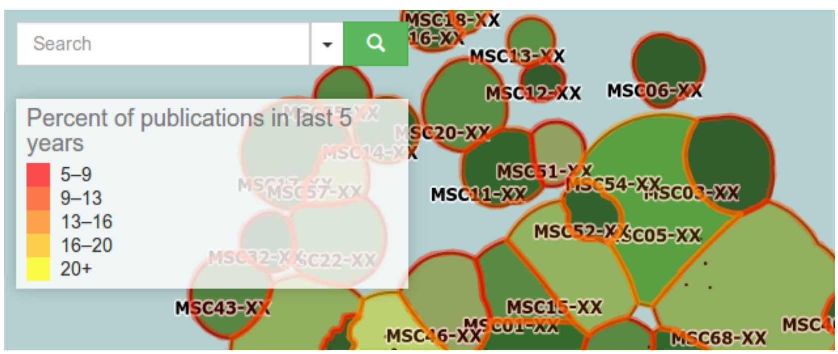
6 User Interface

In order to have all of these features we need an interface for the user to interact with the map. There are several components that need to be created to support all of the use cases. They are afterwards integrated to provide all of the desired interaction.

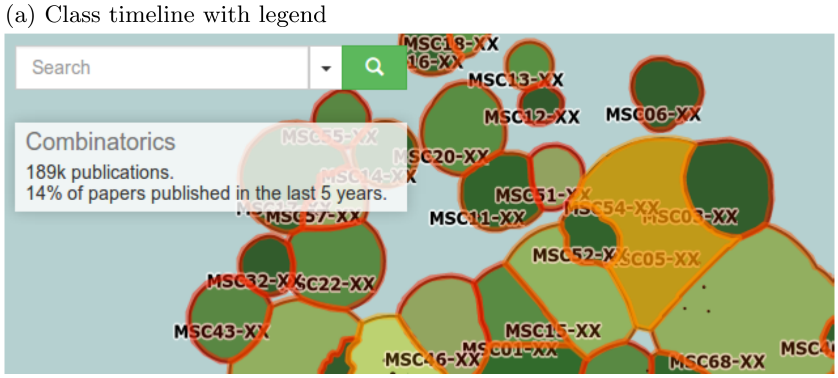
The map is displayed by the Leaflet service which provides zooming from previously generated tile images of the map.

A search bar has been placed on the top left corner of the actual map area (Fig. 3). This integrates it and informs the user that the search results will be in the scope of the map. A mode is selected by placing a shortcut followed by a ":" in front of the query. For example "a: Carmela Acevedo" will search for an author named "Carmela Acevedo".

The results are displayed as markers on the map (Fig. 4a). Clicking on overlapping markers displays all of them in a circle. In this way markers can maintain their position while the user has the ability to navigate them all (Fig. 4b). All markers have a popup containing the title of the respective object and a link to Zentralblatt with the specific results data (Fig. 4c).

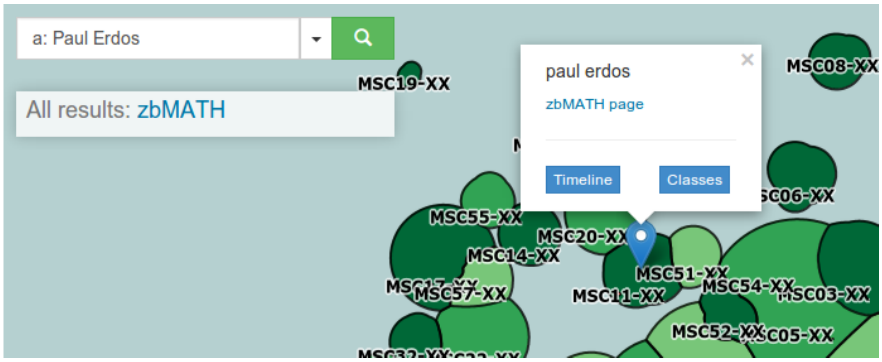


(a) Class timeline with legend

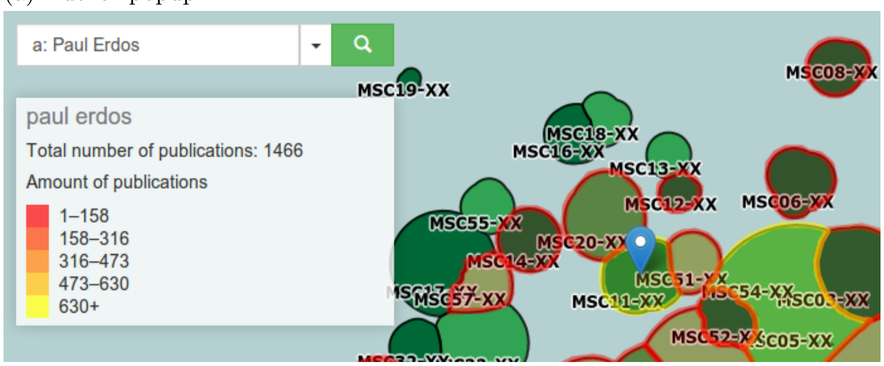


(b) Legend for specific class on hover.

Fig. 5: Timeline.



(a) Author popup.



(b) Author classes.

Fig. 6: Author information.

For the current areas of mathematics use case (4.2) we provide a button “Class timeline” which generates the country overlay (Fig. 3). Below the search there is a HTML div containing a legend for the overlay (Fig. 5a). Whenever a user hovers over a country it is updated to contain specific information about the respective country. This helps to better understand the colors and provides exact information to the user (Fig. 5b).

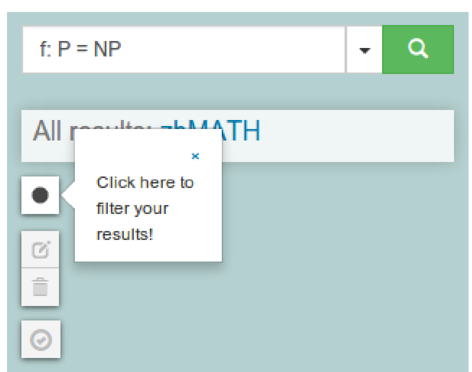
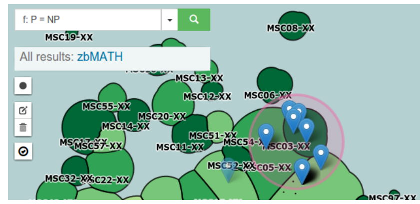


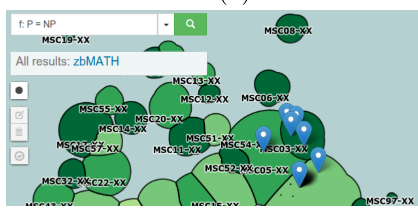
Fig. 7: Filtering toolbar.

To provide more information about the author (4.3) we use overlays that can be visualized by clicking one of the two buttons in their marker popup. Each button prompts the appearance of

(a) Map with results.



(b) Drawn circle.



(c) Filtered markers.

Fig. 8: Filter workflow.

a respective overlay. This is done in the exact same way as in the general case also with a legend for extra information (Fig. 6).

When a search that requires filtering (4.4) is performed we add an extra toolbar to the map below the information area. It contains four buttons: draw, edit, remove and confirm (Fig. 7). These buttons contain hover states to guide the user. After obtaining results users can draw circles around markers they desire to keep. By confirming they obtain only the selected results (Fig. 8).

7 Evaluation

In order to determine if this system is a successful aid for searching an evaluation was performed. This process sought to compare the present solution to previously existing systems.

7.1 Methods

The system was evaluated five mathematically inclined university students in the process of obtaining either an undergraduate or graduate degree. After a given time to get familiar with the system they were asked to perform three tasks both in the map and the Zentralblatt search portal. The tasks were:

1. Find the most currently active class.
2. Find a specific author and describe their publishing behavior.
3. Search for a formula and obtain results in one specific class.

They finally filled out a survey comparing the ease for obtaining information in both systems.

7.2 Results

All the participants grasped the concept of the map during the beginning of the evaluation and worked diligently through the tasks [Ace14]. They found the map to be a useful metaphor to display the information with. After having the results and data available on the map they were able to successfully analyze it and understand how results were related to each other.

Participants considered that the overlays providing extra information were very useful. Filtering results was rated to be easier on the map than on the Zentralblatt search. Users did have technical problems with specific tools and buttons that hindered at moments their possibility of obtaining results.

7.3 Discussion

There were some technical problems related to the implementation of the system. Users could not easily find buttons which was related to the fact that they were outside of their field of vision (the map) and too small. The filtering tool seemed too complicated because it demanded too many steps.

We can consider that the use of the map to display the information was successful. Users were able to understand it and moreover they were able to draw conclusions from it. Objects that were displayed in the map were quickly interpreted by the participants. This shows us how useful the map can be in hierarchical search scenarios where results have a semantical relation.

8 Conclusions and Future Work

In this project we have investigated the potential of maps as search tools for large hierarchical information databases. For this we developed a framework that enables users to interactively perform three search tasks: obtain general information about the current developments in mathematical publications, find fields of interest of particular authors and filter results according to their semantical relations. We have evaluated this framework on five users and concluded that our tool is more helpful to answer these questions than typical search engines. Therefore we prove that maps are a useful search tool because they provide an overview of a topic while letting users investigate areas without losing track of the big picture. The best tools to facilitate hierarchical interactive search are yet to be found.

To improve the map and provide more interaction possibilities, queries should be able to be combined and compared in order to provide more specific results. We are currently not fully taking advantage of the hierarchical nature of the map. A new map could be generated that displays further levels and even individual objects. The map could also be integrated with previous search systems, such as the Zentralblatt search, to complement each other. Finally, new maps should be used with the existing framework in order to evaluate whether the results are related to the map or the interactive framework component.

Acknowledgements Work on the concepts presented here has been partially supported by the Leibniz association under grant SAW-2012-FIZ_KA-2. We are indebted to Wolfram Sperber from ZBMath for support with the data and to Jan Wilken Dörrie and Andrea Kohlhase for insightful discussions.

References

[Ace14] Carmela Acevedo. *OpenMathMap: Interaction*. 2014. URL: http://map.mathweb.org/media/openmathmap-interaction.pdf (visited on May 22, 2014).

[DK13] Jan Wilken Dörrie and Michael Kohlhase. *OpenMathMap: Accessing Math via Interactive Maps*. 2013.

[DKL13] Jan Wilken Dörrie, Michael Kohlhase, and Lars Linsen. “OpenMathMap: Accessing Math via Interactive Maps”. In: (2013).

[Dör13] Jan Wilken Dörrie. “OpenMathMap: Acessing Math via Interactive Maps”. In: (2013).

[GK14] Damien George and Rob Knegjens. *Paperscape*. 2014. URL: http://www.paperscape.org/ (visited on Jan. 25, 2014).

[Lea14] Leaflet. *An Open-Source JavaScript Library for Mobile-Friendly Interactive Maps by CloudMade*. 2014. URL: http://leaflet.cloudmade.com/ (visited on May 9, 2014).

[MAT14] Zentralblatt MATH. *Zentralblatt MATH*. 2014. URL: http://www.zentralblatt-math.org/msc/ (visited on Jan. 25, 2014).

[MSC12] American Mathematical Society. *MSC2010*. 2012. URL: http://www.ams.org/mathscinet/msc/pdfs/classifications2010.pdf (visited on Dec. 7, 2012).

[OSM14] OpenStreetMap. *OSM XML*. 2014. URL: http://wiki.openstreetmap.org/wiki/OSM_XML (visited on May 9, 2014).

[Pos14] PostGIS. *Spatial and Geographic objects for PostgreSQL*. 2014. URL: http://postgis.net/ (visited on May 9, 2014).

[Rus01] Dave Rusin. *About the MathMap image*. 2001. URL: http://www.math.niu.edu/~rusin/known-math/collection/mathmap.html (visited on Dec. 2, 2012).

[Rus02] Dave Rusin. *Information about the Mathematical Atlas collection*. 2002. URL: http://www.math.niu.edu/~rusin/known-math/collection/index.html (visited on Dec. 2, 2012).

[Wik13] Wikipedia. *Zentralblatt MATH*. 2013. URL: http://en.wikipedia.org/wiki/Zentralblatt_MATH (visited on Feb. 26, 2014).

[Zen13a] Zentralblatt. *zbMATH*. 2013. URL: http://zbmath.org/?q=py:201* (visited on May 9, 2014).

[Zen13b] Zentralblatt. *zbMATH About*. 2013. URL: http://zbmath.org/about/ (visited on Feb. 26, 2014).