OpenMathMap: Accessing Math via Interactive Maps

Jan Wilken Dörrie, Michael Kohlhase

Computer Science, Jacobs University Bremen
`<first initial>.<last name>@jacobs-university.de`

Abstract. World Math literature is growing at an alarming rate (3.3M journal articles today increasing by 120k a year). While much of that can be retrieved online, we lack technologies to navigate and understand the space of math literature. The OpenMathMap project wants to develop and deploy novel interfaces that empower interested parties to find their way. We conjecture that such maps can act as cognitively adequate access mechanisms to many large-coverage MKM systems.
The first concrete interface is an interactive map generated from publication data. We have developed a prototype map generation service based on MSC classifications and deployed the maps resulting from ZBMath data in OpenStreetMap. It is accessible at `http://map.mathweb.org/`.

1 Introduction

In the information age fueled by the Internet, the problem of information and knowledge foraging changed from retrieving documents to finding out about them. In particular, navigating the space of available documents efficiently becomes an important subtask.

Even in science, the times where single individuals could have an overview over all of science are long past. Even in the Renaissance polymaths like Leonardo da Vinci were considered a rare exception. The scientific community has developed various tools to work around this problem: encyclopedias, survey articles, classification systems, and review services. But with the proliferation of scientific publication – 50 million articles in 2010 [Jin10] with a doubling time of 8-15 years these tools start collapsing under the sheer mass of information. Internet-age tools like search engines, bibsonomies, and citation databases solve (part of) the information retrieval and navigation problems by providing word-based search and

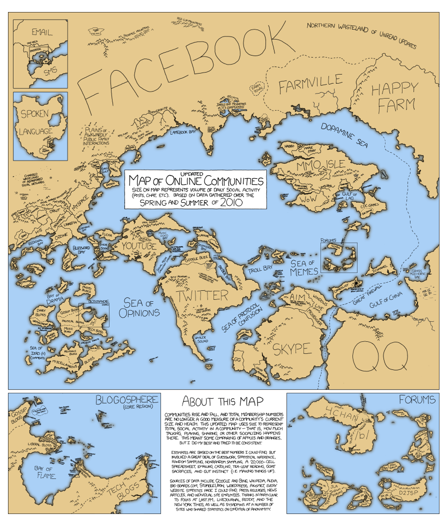


Fig. 1: Map of Online Communities, XKCD 2010 `http://xkcd.com/802/`

browsing along citations. Note that these tools are “myopic” in the sense that they only give very local view of the immediate surroundings of a word or document.

Classification systems like the Math Subject Classification (MSC, see [Msc]), take a more global stance, but they lack user interfaces that give information foragers an intuitive sense of direction and locality that is so helpful to humans in navigation tasks. In the MathSearch project we are currently rethinking access to mathematical knowledge and resources. As a first experiment, we are building a global, map-based navigation service for mathematics.

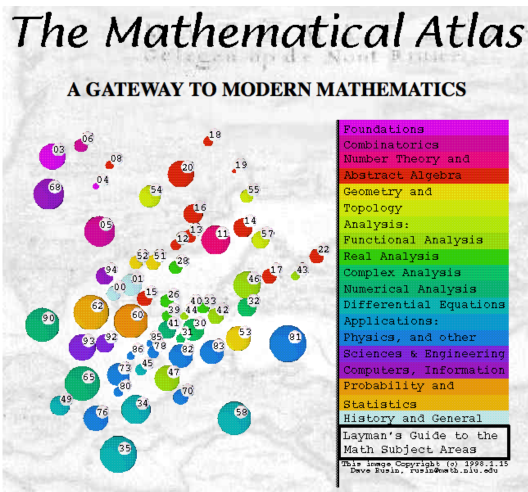


Fig. 2: Dave Rusin’s Math Atlas

The main idea is that humans are very skilled in spatial navigation and in particular have learned to use map representation to navigate spaces and locate targets. Concretely, we want to create a map of mathematics like the one in Figure 1 used to visualize usage patterns of online communities. We want to base the map on ideas from Dave Rusin’s Math Atlas [MathAtlas] (created 1998, last updated 2001, see also Figure 2), which uses topics from the Math Subject Classification for map regions and calculates the positioning and relative sizes from topic interconnections and the numbers of publications.

Acknowledgements Work on the concepts presented here has been partially supported by the Leibniz association under grant SAW-2012-FIZ_KA-2. The authors are indebted to Wolfram Sperber for the publication data for Zentralblatt Math and Patrick Ion for initial discussions and to Lars Linsen for supervision on data visualization matters.

2 Creating a Map from MSC Data

In the creation of the map we made use of the 2010 Mathematics Subject Classification [Msc] jointly developed by the American Mathematical Society and Zentralblatt Math. The result are 63 top level classes, 528 second level classes and 5607 third level classes summing up to 6198 classes in total. Zentralblatt MATH provided us with the metadata for 3.3 million articles in mathematics.

Map Geometry The first step in map creation is to compute the geometry from the publication data. In the current incarnation, the geometry should adequately represent the relative sizes and proximities of the MSC classes, where we define the similarity of two classes as $s(i,j) = |\mathrm{MSC}_i \cap \mathrm{MSC}_j|/|\mathrm{MSC}_i \cup \mathrm{MSC}_j|$.

For the initial version of the map geometry (see Figure 3), we calculate the similarity between every pair of top-level MSCs and obtain a similarity matrix of size 63×63. We applied multidimensional scaling (MDS) to obtain two-dimensional coordinates for each MSC. Computations were executed via Matlab's `mdscale` method, which takes a n by n (dis)similarity matrix D and the number of dimensions p as argument and returns a $n \cdot p$ - sized configuration vector Y.

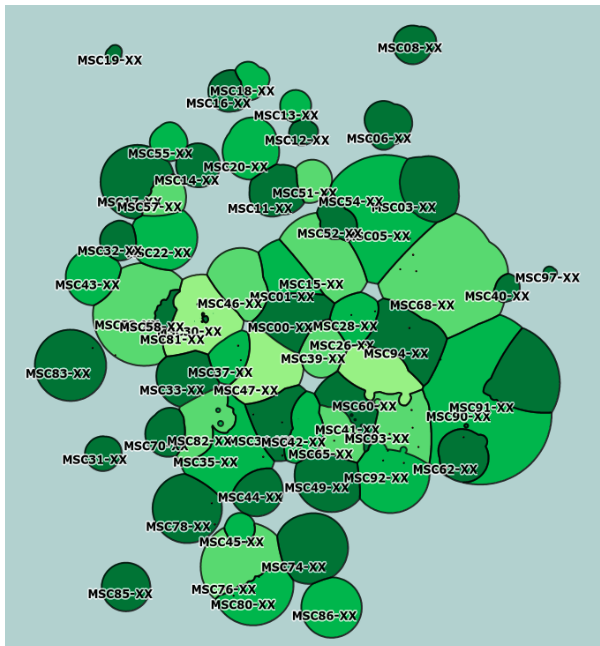


Fig. 3: Geometry of the Math Subject Classifications

To visualize the size of a given MSC class in terms of "map area", we have to assign any given point in 2D space to a MSC class. We use a radial basis function whose origin is given by MDS and obtain the map geometry in figure 3.

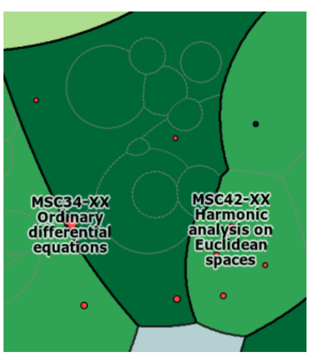


Fig. 4: Adding Settlements

As the MDS computation becomes intractable for larger similarity matrices we opt for a hierarchical approach to determining finer-grained map geometries (taking second-level and leaf MSC classes into account). Here we apply the same procedure as above, but add "boundary classes" from the neighboring MSCs.

Next we populate map geometry with "cities", "towns", and "villages": we simply view every classified paper as an "inhabitant", and compute the "center of gravity" of (the MDS coordinates of) its MSC codes. As the number MSC combinations is finite, this will yields a finite number of settlements, which can be visualized by size; see the red dots in Figure 4 on the left.

Mapmaking & Deployment The next step is to convert the geometry data from the last section into a map that has the features we are used to. Note that the color coding in Figure 3 only shows the "elevations" of the radial basis functions we used for computing areas/borders and should not be conserved in the computed map. This frees one dimension, the "terrain height", for visualizing additional information. We are currently experimenting with encoding the "activity level" making research hotspots peaks that can serve as landmarks in the map.

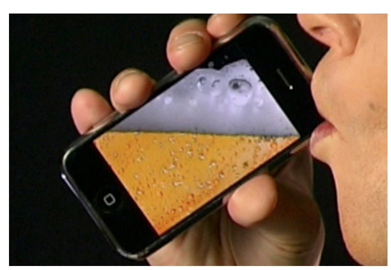
Interactive Services & Mashups Having our map deployed on OSM already gives us some base-level interactivity: zooming, shifting, and name-based search. Additional location-based interactions can be adding custom JavaScript to the pages served by OSM subject to availability of date. One immediate example is the generation of custom queries for publication databases like Zentralblatt Math [ZBMath]. Another service might be to localize mathematicians by their publication record and give them "home address" according to their primary research topic (based on the center of gravity of their publications. Similarly, research trajectories of mathematicians could be plotted on the map by computing yearly centers of gravity. Finally, we could use the math maps as a target for mashups of external services. For instance, the search results of a mathematical search engine could be shown by localizing them on the OpenMathMap service.

3 Conclusion & Future Work

We have presented a novel access method to mathematical knowledge and resources that makes use of the highly evolved cognitive skills of spatial representations in humans. We have implemented a first prototype (`http://map.mathweb.org/`) that deploys maps computed from mathematical publication data in a standard map server and instruments it with information services. This prototype is just a first step we want to use in experimentation in human-oriented access methods to mathematics. We could imagine that connections between mathematical areas could be implemented as roads, highways or air/sea connections (possibly depending on their salience), important theorems could be entered/visualized as landmarks, and finally, we could imagine to go from interactive map servers to much more immersive environments (from Minecraft to second life).

Finally, we acknowledge that the motivation for the OpenMathMap project was a cognitive question, which we have answered with a technical system.

Even though first feedback from mathematicians ranged from puzzled to enthusiastic (with an emphasis on the latter), we will have to systematically evaluate whether OpenMathMap-like systems and services can help with mathematician's day-to-day navigation problems and access tasks, or if OpenMathMap is essentially the equivalent to the iPhone beer app, a useless, but fun gadget.



References

[Jin10] Arif Jinha. "Article 50 million: an estimate of the number of scholarly articles in existence". In: *Learned Publishing* 23.3 (2010), pp. 258–263. DOI: 10.1087/20100308.

[MathAtlas] Dave Rusin. *The Mathematical Atlas, a Gateway to Modern Mathematics*. URL: http://www.math-atlas.org/ (visited on 11/18/2009).

[Msc] *Mathematics Subject Classification MSC2010*. 2010. URL: http://msc2010.org (visited on 11/16/2011).

[ZBMath] *Zentralblatt MATH*. URL: http://www.zentralblatt-math.org/zbmath/ (visited on 06/12/2012).