Classification and Dependency Visualization of the Articles of the Mizar Mathematical Library

Shotaro **Suzuki**[1], Masahiro **Nagasaki**[1] and Kazuhisa **Nakasho**[1,*]

[1]*Yamaguchi University, 2-16-1 Tokiwa-dai, Ube, Japan*

Abstract

The Mizar Mathematical Library (MML) is a collection of mathematical documents formalized by the Mizar system. Visualizing the interrelationships among the MML articles can illuminate their structure and connections, but the scale and intricacy pose significant challenges. In our research, we introduce a method to illustrate these MML dependencies: we sort the MML articles according to the classifications in the Encyclopedic Dictionary of Mathematics. Moreover, we are exploring the feasibility of utilizing generative AI to automate this sorting process, aiming to lessen the need for manual labor. Finally, we also discuss a new algorithm for rendering categorized dependency graphs.

Keywords

Mizar, Mizar Mathematical Library (MML), Mathmatics Subject Classification (MSC), Encyclopedic Dictionary of Mathematics, automated classification, compound graph layout, ChatGPT

1. Introduction

The Mizar language is a formalized language designed for writing mathematical theorems and proofs, while the Mizar System refers to the software suite utilized to validate and enhance proofs created in this language [1]. The Mizar community's primary endeavor is the ongoing development of the Mizar Mathematical Library (MML), comprising over 1,400 documents known as articles [2]. With over 3.1 million lines and growing at a rate of 100,000 lines per year, the MML presents a vast and intricate dataset. The interrelationships among MML articles increase in complexity each year, rendering the MML increasingly challenging to comprehend, employ, and maintain. Visualizing MML dependencies in an easily digestible format would significantly aid users in understanding the MML. Consider a use case where a user wants to identify an article in order to cite a theorem. If the user knows the articles that contain related theorems, the user can identify the article that contains the theorem by relying on its dependencies. If the fields to which the articles belong are visually displayed, it will be easier to narrow down the search to articles that contain theorems. Therefore, the need for research and development of such visualization tools is becoming increasingly urgent.

At present, the dependencies among MML articles are represented using both hierarchical and clustering layouts [3]. While the hierarchical layout excels at illustrating graphs with a

hierarchical structure, the clustering layout is designed to depict clusters of nodes. Nevertheless, the force-directed algorithm applied in the study did not yield an optimal configuration using the clustering layout. Consequently, there is a need for a fresh approach.

In this study, we initially categorized MML articles based on the table of contents from the Encyclopedic Dictionary of Mathematics [4]. To diminish the need for time-consuming manual categorization, we also explored the potential of automating this process using ChatGPT. Moreover, we developed a novel algorithm for arranging these categorized articles into a dependency graph.

2. Related Works

2.1. Automated Classification of Mathematical Papers

M. Schubotz et al. (2020) [5] developed the AutoMSC system to categorize mathematical articles according to the primary sections of the Mathematics Subject Classification (MSC). AutoMSC employs machine learning algorithms to glean classification models from training datasets and subsequently assigns classification labels to new papers. This streamlines the peer-review process for digital libraries and services, enhancing both speed and accuracy, and thereby affords researchers more time for their core investigations. ChatGPT, a state-of-the-art Large Language Model (LLM), has demonstrated impressive performance across a broad range of tasks within natural language processing [6]. Its potential for high performance in text classification has also been noted [7, 8, 9]. However, to date, no published research results exist regarding the use of LLMs for classifying mathematical papers.

2.2. Visualization of Library Dependencies

Visualization of dependencies in software libraries is used in a wide range of fields such as software maintenance, reverse engineering, and reengineering [10]. Visualization of dependencies by graphs is one of the best means to support program comprehension by facilitating an intuitive understanding of the relationships between components and the overall structure.

An example of a system for visualizing dependencies in interactive theorem provers (ITPs) is the work of Y Bertot et al. (2000) [11]. This is a visualization of the dependencies of Coq [12] by Graphviz [1]. In this graph, definitions and theorems are represented by nodes, and dependencies are represented by edges. Currently, this tool is not available due to lack of maintenance, but a similar tool, coq-dpdgraph[2], is available. coq-dpdgraph visualizes the dependencies of Coq's HoTT library[3]. J. Alama et al. (2012) [13] extracts dependencies between Coq and Mizar library and performs a quantitative comparative analysis of library functionality. R. Marcus et al. (2020) [14] proposed the TGView3D system, which renders the dependencies of a formalized mathematical library as a 3D graph. By using 3D graphs, this tool succeeded in effectively visualizing the dependencies of formalized mathematical libraries, which tend to be more complex than ordinary software libraries.

---

[1]https://graphviz.org/
[2]https://github.com/coq-community/coq-dpdgraph
[3]https://github.com/HoTT/Coq-HoTT/wiki

H. Furushima et al. (2022) [3] proposed an approach to visualize the dependencies among MML articles. Their study employed a hierarchical layout within the Sugiyama framework [15] and a clustering layout using the Graphviz sfdp layout[4]. However, due to the strong dependencies within the MML, numerous edge crossings are evident in the hierarchical layout. In a clustering layout, nodes with closer dependency graph distances are expected to be placed closer together. In Figure 1, the nodes with closer graph distance are highlighted around the yellow-green node, and the grayed-out nodes are farther away. However, the grayed-out nodes are also located close to the yellow-green nodes. Consequently, it fails to deliver an expected clustered layout.
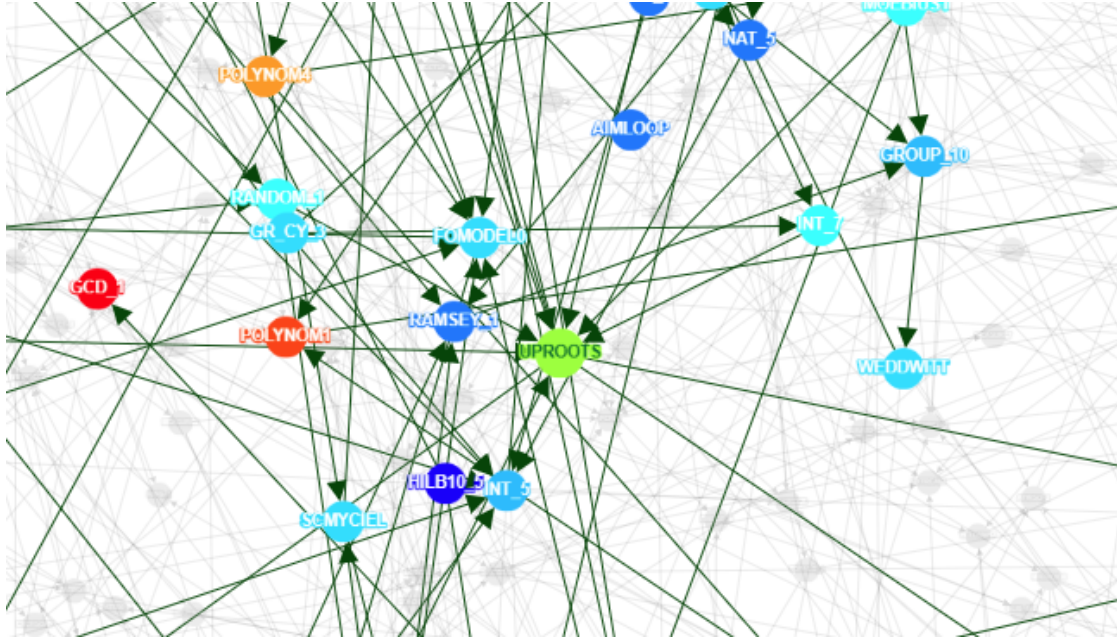

Figure 1: Clustering layout of the dependency graph of the MML (previous approach)

3. Methods

3.1. Classification Using the Encyclopedic Dictionary of Mathematics

Approximately 450 MML articles are categorized according to the Mathematics Subject Classification (MSC), a scheme adopted by mathematical databases such as MathSciNet and zbMATH. MSC provides a valuable means for classifying cutting-edge findings in contemporary mathematics. However, given the elementary and foundational nature of many definitions and theorems within the MML, there are numerous instances where MSC classification may be less applicable. The Encyclopedic Dictionary of Mathematics [4] provides a comprehensive repository of terms, theorems, formulas, and concepts across diverse mathematical fields, offering valuable insights for a broad readership, from novices to experts. For example, the construction of real numbers

---

[4]https://graphviz.org/docs/layouts/sfdp/

is one of the important topics in formalization, but there is no corresponding entry in MSC. However, the Encyclopedic Dictionary of Mathematics has a detailed entry on the construction of real numbers. The advantage of the Encyclopedic Dictionary of Mathematics is also that the detailed contents always correspond to the table of contents, so there is less confusion in classification. Hence, we chose to classify MML articles based on the table of contents from the Encyclopedic Dictionary of Mathematics. This classification involves a two-tiered process, first categorizing into *divisions*, then further into *items* within those *divisions*. For instance, an article featuring theorems on matrix calculus would be classified under the *item* "2. Matrix" within the *division* " III. Algebra". We obtained the above classification by asking ChatGPT questions in corresponding two steps. The ChatGPT prompts were entered manually. The prompts and sample responses are included in Appendix A.

3.2. Visualization of Dependency Graph of the MML

To efficiently visualize the dependencies among Mizar articles categorized by our method, we endeavor to position articles within the same *division* or *item* as close to each other as possible, while preserving the hierarchical relationships of the dependencies to the greatest extent. Graphs with these dual relationships - dependency and nested group - are termed compound graphs, for which various layout methods have been suggested [16, 17]. However, current methods fall short when it comes to visualizing compound graphs with many connections, such as those seen in the MML. Therefore, we crafted a new algorithm to enhance the visibility of these relationships. In the graphs created by our algorithm, each *item* from the Encyclopedic Dictionary of Mathematics is presented as a rectangle, with the articles affiliated with each *item* displayed within the corresponding rectangle. Both the rectangles and the articles within a rectangle are vertically aligned according to dependencies as much as possible. The procedure for constructing the layout with our algorithm is outlined below:

1. Apply the Sugiyama framework [15] to each *item* to generate a small graph of the articles, and calculate the size of a rectangle sufficient to encapsulate this small graph.
2. Apply the Sugiyama framework to all articles in the MML to determine the centroid of the articles associated with each *item*. These coordinates are then used as the initial positions for the rectangles.
3. Adjust the rectangles to eliminate overlaps.
   a) Shift rectangles along the y-axis if they overlap and their top and bottom y-coordinates are in the same order, respectively. That is, if we have a pair of rectangles A and B with top and bottom y-coordinates given as (y_1, y_2) and (y_3, y_4) respectively, the condition for moving can be expressed as $(y_1 - y_3) \cdot (y_2 - y_4) > 0$. In this step, rectangles are shifted in a direction that reduces overlap while preserving positioning of y-coordinate of rectangles. However, the y-axis distances between the rectangles after the adjustment are set to a constant.
   b) Shift rectangles along the x-axis if they overlap. In this step, all overlaps are resolved while preserving the order of the x-coordinates of the rectangles' centroids.

The Sugiyama framework delivers a layout that minimizes edge intersections while preserving vertical alignment. Subsequently, our algorithm generates a layout that keeps the vertical

alignment of articles within an *item*, and additionally, maintains the vertical alignment of all articles to a certain extent. Phase 3a is necessary to form a graph with a reasonably uniform aspect ratio; without phase 3a, the overlap is always enlarged along the x-direction, resulting in a graph that is flat horizontally. Unclassified articles are placed outside of rectangle because they belong to abstract layers. A part of the MML dependency graph, as depicted using our algorithm, is illustrated in Figure 2. Figure 3 shows an enlarged view of the item rectangle for trigonometric functions. The graph was rendered using the JavaScript library, Cytoscape [18].
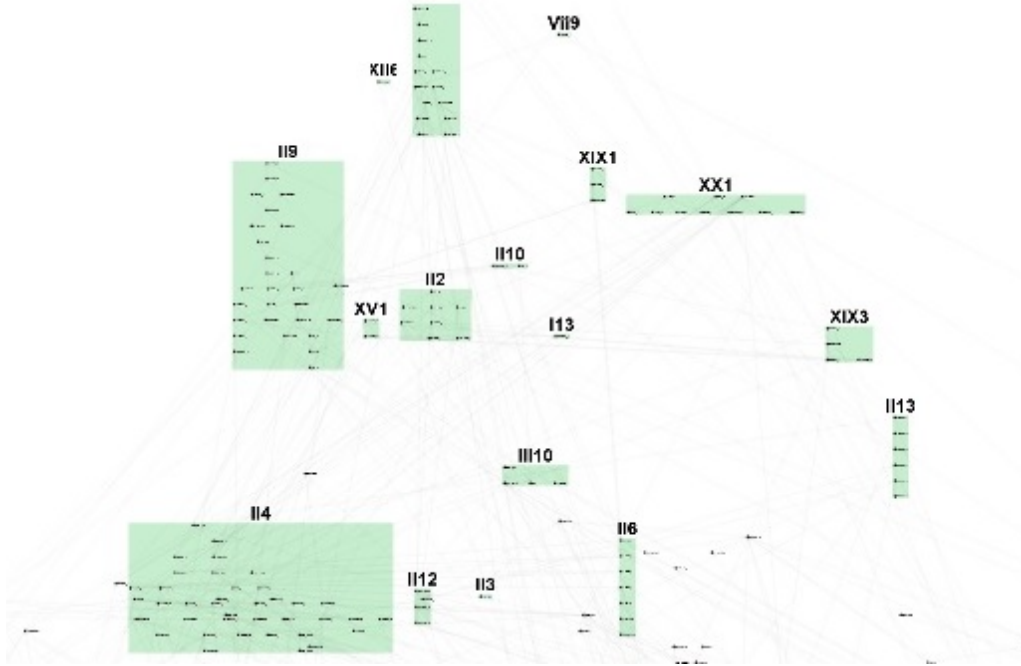

Figure 2: Visualization of the MML dependency graph with our algorithm

4. Evaluation

4.1. Classification accuracy of ChatGPT

We evaluated the classification accuracy of ChatGPT for MML articles by comparing its results with 493 manually classified articles. The findings showed that ChatGPT achieved an accuracy rate of 46.7 percent (230/493) for the *divisions* and 27.2 percent (135/493) for both the *divisions* and *items*. ChatGPT's misclassifications can be grouped into three main patterns:

The first pattern involves cases where the term "formalization" is frequently used in an MML article's abstract, which results in a misclassification into the *item* "17. Proof Theory" within the division " I. Mathematical Logic and Foundations of Mathematics".

The second pattern represents situations where the abstract is overly simplistic, thus not providing enough information for effective classification.
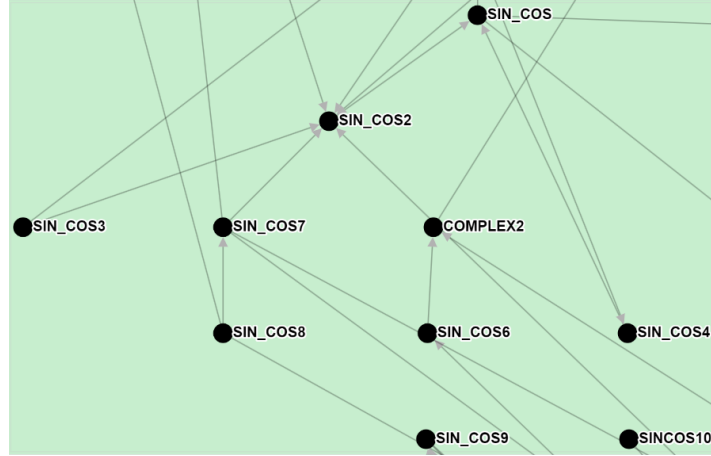
Figure 3: The dependency graph in the item for trigonometric functions

The third pattern occurs when the *division* is correctly identified, but there are discrepancies between the ChatGPT and manual classification results because the articles address multiple *items*. In these instances, ChatGPT proposed multiple classification candidates, and in most situations, the correct answer was among them.

Conversely, the articles that were correctly classified often had a robust abstract or focused on theorems that epitomize each *division* or *item*.

4.2. Layout Quality of Compound Graph

Our algorithm managed to maintain vertical alignment for 76 percent of the dependencies between articles. Conversely, the KLay layout [17], a standard feature of Cytoscape.js, managed to achieve this alignment in only 43 percent of cases. This result signifies that our approach facilitated superior quality visualization compared to one of widely recognized algorithms.

Moreover, our new layout demonstrates that the rectangles representing *items* within the same *division* are situated reasonably close to each other, indicating that articles with similar semantic relationships are positioned near one another.

The graph of 493 articles that were manually classified and drawn with our algorithm is available on GitHub Pages[5].

5. Future Works

Through the manual categorization of the MML, we have found that classification, in accordance with the table of contents of the Encyclopedic Dictionary of Mathematics, is suitably effective for formalized mathematical libraries. However, ChatGPT's accuracy in MML classification was only 27.2 percent, indicating ample opportunity for improvement. This lack of precision was frequently tied to issues with abstract content. We anticipate that accuracy might be

---

[5]https://qiuyueluzi.github.io/compound_graph/compound_graph/hierarchical_graph.html

enhanced by feeding MML articles to ChatGPT to generate abstracts in natural language. It is also worth investigating whether the methods used in MSC classification could be applied to the Encyclopedic Dictionary of Mathematics to improve its accuracy.

For graph drawing, it is necessary to modify the appearance of the graph by directly displaying the names of *divisions* and *items* instead of indices. We aim to integrate it into a graph application within the emwiki system[6], following the manual classification of all MML articles.

Acknowledgments

References

[1] A. Grabowski, A. Kornilowicz, A. Naumowicz, Mizar in a Nutshell, Journal of Formalized Reasoning 3 (2010) 153–245.

[2] G. Bancerek, C. Byliński, A. Grabowski, A. Korniłowicz, R. Matuszewski, A. Naumowicz, K. Pąk, The Role of the Mizar Mathematical Library for Interactive Proof Development in Mizar, Journal of Automated Reasoning 61 (2018) 9–32.

[3] H. Furushima, D. Yamamichi, S. Shigenaka, K. Nakasho, K. Wasaki, An Integrated Web Platform for the Mizar Mathematical Library, in: International Conference on Intelligent Computer Mathematics, Springer, 2022, pp. 141–146.

[4] K. Itō, Encyclopedic Dictionary of Mathematics, volume 1, MIT press, 1993.

[5] M. Schubotz, P. Scharpf, O. Teschke, A. Kühnemund, C. Breitinger, B. Gipp, AutoMSC: Automatic Assignment of Mathematics Subject Classification Labels, in: International Conference on Intelligent Computer Mathematics, Springer, 2020, pp. 237–250.

[6] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, et al., Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models, arXiv preprint arXiv:2304.01852 (2023).

[7] T. Kuzman, I. Mozetic, N. Ljubešic, ChatGPT: Beginning of an End of Manual Linguistic Data Annotation? Use Case of Automatic Genre Identification, ArXiv, abs/2303.03953 (2023).

[8] M. M. Amin, E. Cambria, B. W. Schuller, Will Affective Computing Emerge From Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT, IEEE Intelligent Systems 38 (2023) 15–23.

[9] B. Zhang, D. Ding, L. Jing, How Would Stance Detection Techniques Evolve After the Launch of ChatGPT?, arXiv preprint arXiv:2212.14548 (2022).

[10] R. Koschke, Software Visualization in Software Maintenance, Reverse Engineering, and Re-engineering: A Research Survey, Journal of Software Maintenance and Evolution: Research and Practice 15 (2003) 87–109.

[11] Y. Bertot, O. Pons, L. Pottier, Dependency Graphs for Interactive Theorem Provers, Technical Report, Technical Report RR-4052, INRIA, 2000.

---

[6]https://em1.cs.shinshu-u.ac.jp/emwiki/release/graph/

[12] Y. Bertot, A Short Presentation of Coq, in: Theorem Proving in Higher Order Logics: 21st International Conference, TPHOLs 2008, Montreal, Canada, August 18-21, 2008. Proceedings 21, Springer, 2008, pp. 12–16.

[13] J. Alama, L. Mamane, J. Urban, Dependencies in formal mathematics: Applications and extraction for Coq and Mizar, in: International Conference on Intelligent Computer Mathematics, Springer, 2012, pp. 1–16.

[14] R. Marcus, M. Kohlhase, F. Rabe, TGView3D: A System for 3-Dimensional Visualization of Theory Graphs, in: International Conference on Intelligent Computer Mathematics, Springer, 2020, pp. 290–296.

[15] K. Sugiyama, S. Tagawa, M. Toda, Methods for Visual Understanding of Hierarchical System Structures, IEEE Transactions on Systems, Man, and Cybernetics 11 (1981) 109–125.

[16] K. Sugiyama, K. Misue, Visualization of Structural Information: Automatic Drawing of Compound Digraphs, IEEE Transactions on Systems, Man, and Cybernetics 21 (1991) 876–892.

[17] C. D. Schulze, M. Spönemann, R. von Hanxleden, Drawing Layered Graphs with Port Constraints, Journal of Visual Languages and Computing, Special Issue on Diagram Aesthetics and Layout 25 (2014) 89–106.

[18] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks, Genome research 13 (2003) 2498–2504.

A. Example of Classification by ChatGPT

Examples of ChatGPT prompts and responses are presented below.

> **User**: I would like to categorize a mathematical paper with the following abstract by department.
> "This is the first part of the axiomatics of the Mizar system. It includes the axioms of the Tarski Grothendieck set theory. They are: the axiom stating that everything is a set, the extensionality axiom, the definitional axiom of the singleton, the definitional axiom of the pair, the definitional axiom of the union of a family of sets, the definitional axiom of the boolean (the power set) of a set, the regularity axiom, the definitional axiom of the ordered pair, the Tarski's axiom A introduced in [?] (see also [?]), and the Frænkel scheme. Also, the definition of equinumerosity is introduced."
> Which of the following departments can this paper be classified into? I. mathematical logic and foundations of mathematics, II. set and topology, III. algebra, IV. theory of numbers, V. group and theory of representation, VI. algebraic geometry, VII. geometry, VIII. differential geometry, IX. topology, X. analysis, XI. complex analysis, XII. functional analysis, XIII. theory of differential equations, XIV. special functions, XV. numerical analysis, XVI. application analysis, XVII. theory of probability, XVIII. mathematical statistics, XIX. discrete mathematics and combinatorics, XX. mathematics in information sciences, XXI. optimization theory, XXII. mechanics and physics, XXIII. history of mathematics
> **ChatGPT**: This paper is classified into "I mathematical logic and foundations of mathematics". This paper is about the axiomatic foundation of the Mizar system and describes the axioms of Tarski-Grothendieck set theory. The concept of "equicardinality" is also introduced in the paper.
> **User**: I would like to classify this paper more specifically. Which of the following categories would be the most appropriate? 1. mathematical logic and foundations of mathematics, 2. semantics of formal systems, 3. formal systems and proofs, 4. computable function, 5. model theory, 6. stability theory, 7. non-standard analysis, 8. o-minimal theories, 9. axiomatic set theory, 10. forcing, 11. large cardinals, 12. descriptive set theory, 13. recursion theory, 14. decision problem, 15. degrees of unsolvability, 16. constructive ordinal, 17. proof theory, 18. G¨odel's incompleteness theorems, 19. non-standard models of arithmetic, 20. type theory and λ-calculus, 21. Herbrand's theorem and the resolution principle, 22. non-standard logics, 23. paradox
> **ChatGPT**: This paper is likely to be classified into "9. axiomatic set theory". The paper discusses the axiomatic foundations of set theory and describes the axioms of the Mizar system and Tarski-Grothendieck set theory.

As described above, we first classified the responses by the *divisions* of the Encyclopedic Dictionary of Mathematics, and from the responses, we further classified the responses into more detailed items. If ChatGPT responded to more than one *division* or *item* in the classification process, it was prompted to respond with the most appropriate *division* or *item*. An example is shown below.

CHATGPT: This paper can be classified more specifically into the following categories: 12. curvilinear integral and surface integral 13. measure theory 14. integration theory

USER: Thank you for your response. Which of the following categories would be the most appropriate?

CHATGPT: The most appropriate category for this paper is "13 measure theory".