Chapter 5

The digital shadow of mathematics and its ramifications

Howard S. Cohl, Moritz Schubotz

1 The natural habitat of mathematics

Mathematics is ubiquitous. It is the incredibly creative and widely spoken language of logic. Mathematical knowledge resides in the minds of mathematicians and those who want or need to use mathematics. They have learned mathematics through the process of thought; by listening to people speak and through conversations; through the reading of books; and more recently by browsing on the internet; by listening and watching videos; and through written and computational practice. Through practice, understanding is accomplished and has led to concept extension and generalization. Ultimately, through writing, publication and presentation, dissemination is obtained and one hopes to achieve the global blossoming, conceptualization and description of mathematical notions. This continuing process generates a wealth of data which should be shared globally, but is in practice subject to restrictions to access. Through refinement and use, important results precipitate and become more widely available. The use of computers has greatly facilitated this process.

Mathematical functions and the operations they satisfy are widespread. The so-called special functions are mathematical functions which are so useful and have appeared so often (in applications) that they have been given special names. As well as special functions, there are also special constants, numbers and special sequences of numbers (see the On-Line Encyclopedia of Integer Sequences[1]). There is also a large collection of mathematical objects or operations which have commonly appeared and these have been given names as well. The names of these special objects summarise and provide an organisational structure to mathematics.

What are special function names? These names are often ascribed to the discoverer or to a person who greatly exploited their use, or simply to a description of their action, or sometimes, out of the void. Special functions arise in a variety of contexts. Historically, some of the most common special functions arose in areas of classical analysis and natural mathematics such as in the study of the figure of the

[1] https://oeis.org (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.)

earth by Pierre-Simon Laplace, through the separation of linear partial differential equations. Special functions include classical orthogonal polynomials, Bessel functions, associated Legendre functions, the gamma function, and elliptic integrals. Even more esoteric functions arise such as parabolic cylinder functions, Mathieu functions, Lamé functions, ellipsoidal harmonics, elliptic functions and so on. This is the tip of the iceberg. There are many more and today one can find an excellent summary of the most important ones in the NIST Digital Library of Mathematical Functions [1] in which there are 36 chapters, each focusing on their most important properties. More generally, Scharpf et al. [3] define the term of *Formula Concept* "as a collection of equivalent formulae with different representations." The question and description of equivalence must be pinpointed so that a wider audience may understand the conversation. Special function and number data is a subset of the much larger collection of mathematical knowledge.

Nowadays one can peruse the constantly evolving collection of mathematical knowledge by visiting and examining the online arXiv preprint server. This dataset gives a good sample of the breath and depth of mathematical knowledge which is constantly evolving. In an even more refined collection, the journals of mathematics and mathematical physics provide an even more carefully curated collection of information. There are as well collections of monographs published by the mathematical science publishers. Together, we have focused on the mathematical knowledge associated with the real and complex analysis of special functions and numbers. However, there exist alternative and more abstract mathematical knowledge such as that which is connected with group theory, abstract algebra, number theory, differential geometry, topology, graph theory, category theory, set theory, type theory, logic, and so on. There is often a deep underlying connection between these fields which all have footprints in the entirety of mathematical knowledge.

Of supreme wealth has been those mathematicians who explore the mathematical terrain through their research – those who have discovered and revisited areas, and have provided extensions, generalizations – new results. Usually these individuals provide the benefit of sharing their discoveries through publication of journal articles and perhaps in monographs. In the future, AI may more significantly play the role of these mathematicians, but there exist significant obstacles to this transition [2, 5].

2 The ongoing and future invasion of mathematics into the digital space

In order for there to be full computational access to the data associated with mathematical knowledge, one must transform this data into a form in which it is understandable by a machine. In order to accomplish this, one should enhance the machine so that it is clever enough to understand and use the data. This is the problem of semanticisation or semantic augmentation [4] and it lies at the heart of the problem. We must develop a confidence that correct mathematical meanings may be inferred by the machine. With the mathematics of special functions, this journey is well underway, and our special route is through the preparation of mathematical documents, the most common way to spread and communicate mathematical description.

The most common method for mathematical data to be entered into the literature is connected to the problem of typesetting mathematical information. In today's literature, the most common method for typesetting mathematical information is with the use of TeX or LaTeX. These are programming languages which center around typesetting mathematical expressions. Even though LaTeX produces readable presentation, the content may be shrouded. In order to remedy this, more precise methods for describing mathematical syntax is necessary. For now, we enjoy communication to and through Computer Algebra Systems (e.g., Mathematica, Maple, MATLAB, Reduce, Magma, SageMath, SymPy, etc.). In the case of online mathematical content, the use of XML or MathML is powerfully opportune. Our team captures this semantic data initially though the use of LaTeX and important metadata connected with the content is provided to the user.

Even more thorough prescription for describing mathematical content on a machine is provided through languages used to develop formalized mathematics – such as those used in Automated or Interactive Theorem Proving (e.g., Lean (proof assistant), Isabelle (automated theorem prover), Coq (interactive theorem prover)). As one moves further in this direction, the ability for humans to read the mathematics starts to fade away, but the ability for computers to process such information is greatly enhanced. This is the question of human comprehension vs. the question of machine comprehension and the ability to rid oneself of ambiguity while enhancing precision – a principal goal for the field of mathematical knowledge management.

3 Our conclusion and the eventual payoff

Many features of mathematics are clear to its readers. However, in reality there are many assumptions which the reader understands without explanation. This becomes apparent, when looking at theorem proving systems and digital mathematical compendia which describe the semantics of mathematics. Semanticisation is the horizon where humans and description meet and will play a fundamental role for the forthcoming evolution of mathematical knowledge. Once there is a critically large machine-readable collection of mathematical content, through the bootstrapping process, artificial intelligence should be able to ascertain missing semantic through the same process that humans use. When mathematical knowledge is fully accessible to machines, only our imagination will provide a boundary to possible routes of mathematical exploration in the digital realm. We have only just begun.

References

[1] NIST Digital Library of Mathematical Functions, http://dlmf.nist.gov, Release 1.2.0 of 2024-03-27 visited on 14 March 2024

[2] A. Greiner-Petter, T. Ruas, M. Schubotz, A. Aizawa, W. I. Grosky, and B. Gipp, Why machines cannot learn mathematics, yet. In *Proceedings of the 4th joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL 2019) co-located with the 42nd international ACM SIGIR conference on research and development in information retrieval (SIGIR 2019), Paris, France, July 25, 2019*, edited by M. K. Chandrasekaran and P. Mayr, pp. 130–137, CEUR Workshop Proceedings 2414, CEUR-WS.org, 2019

[3] P. Scharpf, M. Schubotz, H. S. Cohl, and B. Gipp, Towards formula concept discovery and recognition. In *Proceedings of the 4th joint workshop on bibliometric-enhanced information retrieval and natural language processing for digital libraries (BIRNDL 2019) co-located with the 42nd international ACM SIGIR conference on research and development in information retrieval (SIGIR 2019), Paris, France, July 25, 2019*, edited by M. K. Chandrasekaran and P. Mayr, pp. 108–115, CEUR Workshop Proceedings 2414, CEUR-WS.org, 2019

[4] M. Schubotz, *Augmenting mathematical formulae for more effective querying & efficient presentation*. Ph.D. thesis, Technical University of Berlin, Germany, 2017, https://d-nb.info/1135201722 visited on 14 March 2024

[5] M. Schubotz, N. Meuschke, M. Leich, and B. Gipp, Exploring the one-brain barrier: A manual contribution to the NTCIR-12 MathIR task. In *Proceedings of the 12th NTCIR conference on evaluation of information access technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016*, edited by N. Kando, T. Sakai, and M. Sanderson, pp. 309–317, National Institute of Informatics (NII), 2016