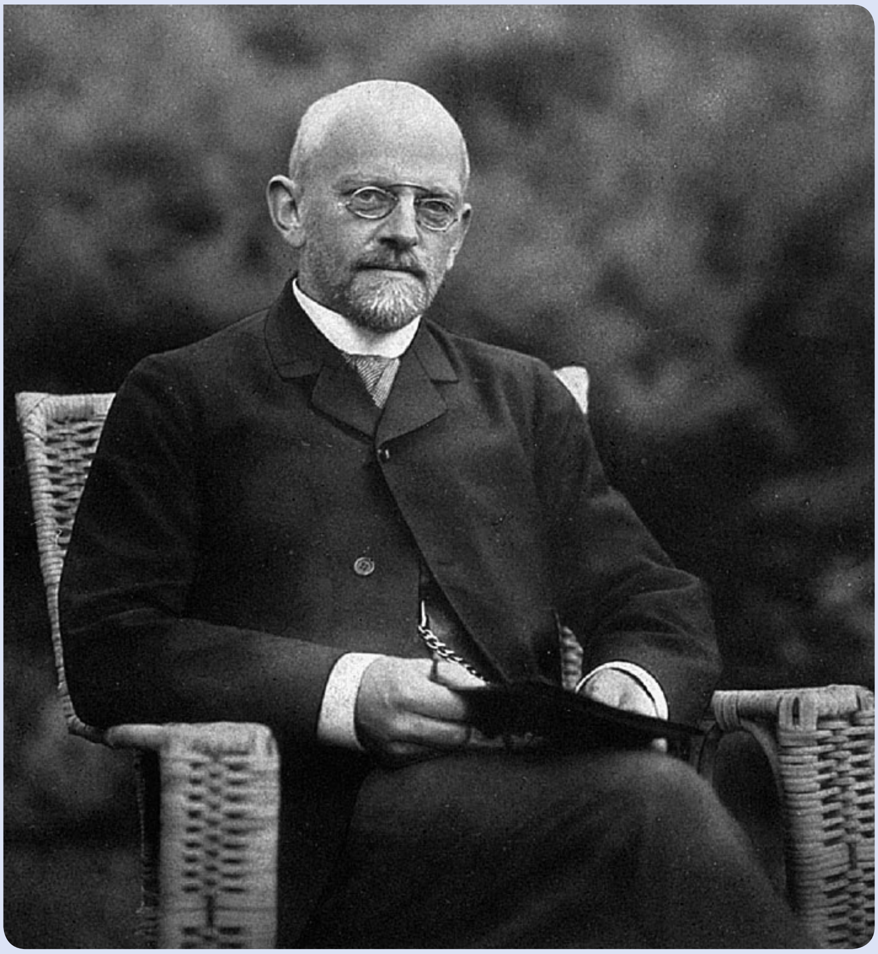
A Focus on Mathematics

New Directions of Integrated Knowledge Management in Mathematics –
Continuing Almost 150 Years of Reviewing Services

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure



"The tool implementing the mediation between theory and practice, between thought and observation is mathematics. Mathematics builds the connecting bridges and is constantly enhancing their capabilities. Therefore it happens that our entire contemporary culture, in so far as it rests on intellectual penetration and utilization of nature, finds its foundation in mathematics."

David Hilbert (German mathematician, 1862-1943)

Title Photo
Möbius band in Mihama (Japan). Sculpture by Keizo Ushio.
Published with the kind permission of the artist.

Table of Contents

Preface

Klaus Hulek
Editor-in-Chief of ZBMATH

While anthropologists are still divided on how much mathematics is encoded in the carvings of the 20,000-year-old Ishango bone, there is ample evidence that the early Sumerian cities already used advanced techniques of arithmetic, geometry and algebra around 4,000 B.C. Since then, such diverse civilizations as Ancient Babylon, Egypt, Greece, Rome or China and India, and indeed our modern societies shaped by global technology, have stored the essence of their scientific and technological knowledge in the language of mathematics. Today, essentially all developments affecting our modern life have their roots in mathematical insights: the structure of integrated circuits, efficient search engines, the cryptographic procedures that enable data transfer in cell phones and electronic banking, the statistical tools that allow for the analysis of complex data, the optimization routines which empower logistics and production, or the methods which underlie tomography or weather forecasting. Preceding most of these applications, Hilbert's quote above applies perfectly to them and appears to be truer than ever.

Accordingly, mathematical knowledge has grown tremendously, far beyond the limitations of an individual brain. From early times, its longevity and consistency required a comprehensive collection using appropriate storage media, such as clay tablets, papyrus, parchment, paper, or electronic chips. It was probably no coincidence that, in 1931, Otto Neugebauer, who first explored the mathematics contained in Babylonian cuneiform scripts and the Rhind papyrus, also founded the journal "Zentralblatt für Mathematik und ihre Grenzgebiete", with the aim of covering mathematical research literature by means of expert reviews. Due to the merger with "Jahrbuch für die Fortschritte der Mathematik" founded in 1868, we are now able to celebrate 150 years of mathematical reviewing. This service has evolved into the zbMATH database, which is today the largest and longest-running reviewing service in mathematics. The first article in this volume gives a short overview of this development (cf. pp. 5-9).

Research publications still form the core of mathematical research, but the landscape has changed significantly during the last decade. The major recent trends are traced in the second article (cf. pp. 10-14). We must not, however, forget that it is the people behind the papers that do the mathematics. Hence, an important part of zbMATH is its author database. In order to create this database, the challenges of author disambiguation outlined in the third article have to be met (cf. pp. 15-18).

Additionally, an increasing amount of information is available in media outside of traditional publications. Prominent examples are mathematical software and formulae. Their indexing in zbMATH opens up new search dimensions, as addressed in two dedicated articles (cf. pp. 23-27). But these are only the first instances of the future Global Digital Mathematics Library as pursued by the IMU. The very practice of mathematics will be significantly influenced by these pivotal changes in the field of mathematical information and communication. We discuss several of these aspects in general, while considering possible social implications in more detail and taking several clichés about mathematicians into account (cf. pp. 19-22).

Today, perhaps even more than in the past, upcoming developments will influence not only mathematical research but also impact on our everyday lives. Such developments are indeed relevant to a broader audience beyond the mathematical community.

We wish you an interesting and exciting read.

Klaus Hulek

From a Monthly Compilation of Printed Reviews to a Multifaceted Database: a Short History

Olaf Teschke

Around the middle of the 19th century, a new culture of scientific publishing took shape: The old habit of writing monographs or letters to colleagues was subsequently replaced by communicating ongoing research through journals such as Gergonne's *Annales de mathématiques pures et appliquées* (founded in 1810) or Crelle's *Journal für die reine und angewandte Mathematik* (founded in 1826). This came along with a hitherto unknown growth of published results, and, in consequence, the need to maintain an overview of the state of the field. Though the numbers were still small compared to the situation today (from about 1,000 articles in 1870 to about 3,000 in 1900), the shear size of new results scattered in different journals and languages called for a new approach to gather the relevant research information.

The *Jahrbuch über die Fortschritte der Mathematik* attempted to fulfill this task from 1868 by establishing a peculiar approach: In the form of annual volumes, the *Jahrbuch* catalogued the bibliographic information of all literature published in that year, grouped by subjects and endowed with independent reviews by experts. For many decades, this provided a successful and vivid form of scientific communication. By browsing through the pages of the old volumes one is often reminded of modern discussions in social networks: the culture of reviewing led to an intense exchange which served to validate or, sometimes, to disprove results. Indeed, the efforts of the *Jahrbuch* editors still bear fruits today: Due to the longevity of classical approaches[1], the information carefully collected back then helps to discover relevant results which are often older than a century.

However, at the end of the 1920s it became obvious that the format of annual volumes led to a growing delay (which finally resulted in the discontinuation of the *Jahrbuch* after WW II) and the mathematical community required a service with more promptness. The initiative for the foundation of a new mathematical reviewing journal came from Otto Neugebauer (the first editor-in-chief), Richard Courant, and Harald Bohr, together with the publisher Ferdinand Springer. In 1931, they founded the *Zentralblatt für Mathematik und ihre Grenzgebiete* as an international journal providing early reviews of the entire world literature in mathematics and related areas, publishing 18 volumes until 1938.

In 1933, shortly after the Nazi party rose to power, law restrictions and propaganda against Jews forced many Jewish mathematicians and political opponents to emigrate. Neugebauer – who had been member of the Social Democratic Party of Germany and spent also some time in Leningrad studying Babylonian scriptures – experienced the animus against his Jewish colleagues and him and asked for a leave, which was permitted. With the support of Bohr, he took up a professorship in Copenhagen in 1934, from where he continued his work for *Zentralblatt*. The Nazis tried to gain influence on the editorial policy by attempts of the Prussian Acade-


Otto Neugebauer

[1] Some examples are given in the contribution "Changes in the Publication Landscape", p. 10.

my to join *Zentralblatt* with the *Jahrbuch*. After a series of incidents Neugebauer eventually gave up his position as editor-in-chief in 1938; many other members of the editorial board followed his resignation. Neugebauer emigrated to the USA in 1939 where he accepted a position at Brown University. Together with the secretary of the American Mathematical Society, who was the dean of the Brown University at that time, he founded a new reviewing journal, the *Mathematical Reviews*, while the *Zentralblatt* editorial office was forced to join the *Jahrbuch* office, the management of which had been transferred in the meantime to Harald Geppert and Ludwig Bieberbach, both devoted Nazis and active members of the NSDAP.

Erika Pannwitz

By his double founding role for the two still running reviewing services in mathematics, Neugebauer succeeded to establish a culture which prevails in many aspects today: Rooted in the *Jahrbuch's* tradition to enable communication among scientists, reviews of publications primarily aim at providing the essential information to decide whether or not a publication might be relevant for the user's research. Devoted to the unity of mathematics, and open to its applications, the corpus should give an impartial view of the current state of approved research. Though the progress of research specialization brings different communication habits about (e.g., due to the significantly shorter half-life of publications in mathematical physics, articles from this area would find less often a reviewer, since scientists are satisfied with its description by the abstract), still more than 50% of the articles in core mathematics indexed in our database are additionally reviewed by an independent expert mathematician. Another indication for the viability of this approach is that many editors and reviewers who left in the 30s quickly renewed their efforts after World War II. Hence, the collapse of Nazi-Germany led only to a temporary suspension of the editorial work. *Zentralblatt* came to life again in 1947, with Hermann Ludwig Schmid as the new editor-in-chief, who played an important role in the reconstruction of *Zentralblatt*, reviving contacts with former colleagues and finally succeeding in restarting publication. Following Schmid's appointment as professor in Würzburg in 1953, Erika Pannwitz, who had worked as an editor for both the *Jahrbuch* in the 1930s and *Zentralblatt* since 1947, took over the editorship.

Editorial work at that time comprised almost all steps in creating the final volume: To scan the journals for relevant articles, assign them to reviewers, edit the incoming reviews, insert references to other papers via register volumes, group everything into content sections in alphabetic order, and read galley proofs and correct errors in several cycles. Additionally, the diversity of languages (much larger at that time) had to be handled, including a translation of the titles. The increasing number of publications caused a serious problem in the pre-computer age since manual workflows were not sufficiently scalable. The only chance to handle the publication masses was to enlarge the editorial board but budget limitations led this approach quickly to its limits. It is hence no exaggeration to say that the prevention of an exploding backlog, adding to the gap created by the war, was only possible due to heroic individual efforts.

Politics in divided Germany did not help either. The construction of the Berlin Wall in 1961 led to a division of *Zentralblatt* into basically two separate editorial offices, causing many complications for the editorial board and the publisher Springer. Walter Romberg was in charge of the Eastern editorial board, while Pannwitz continued as editor-in-chief of the Western office. A sophisticated apportionment of fields and journals, somehow reflecting the divisions in science created by the Cold War, enabled a collaboration which allowed to reduce friction, while both sides benefited from the others' strengths. This remarkable German-German cooperation lasted until 1977 and resulted in *Zentralblatt* regaining a leading position in mathematical reviewing, despite the complicated political constellation.

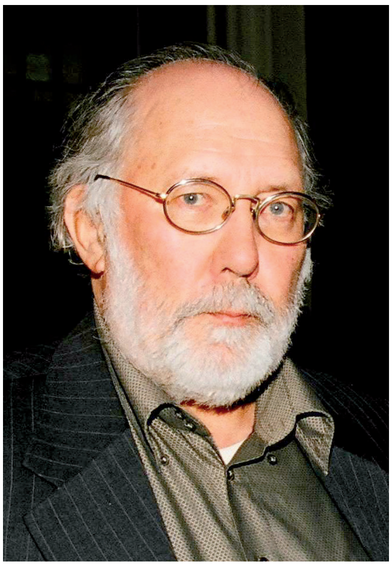
Walter Romberg

During the 1970s the annual production of mathematical publications doubled, reaching a size that could no longer be handled in a traditional way. One landmark in organizing the vast amount was the introduction of the Mathematical Subject Classification (MSC)[2] which greatly improved on the coarse scheme defined by field chapters. It did not only come along with the creation of in-depth grouping and content-based identifiers, but allowed also for a much finer definition of reviewer interests and more precise assignments. Though still guided by the printed volumes, content organization took a shape suitable for modern databases. Simultaneously, advances in computer science allowed for a modernization of the editorial work when *Zentralblatt* entered several partnerships with various scientific institutions to benefit from their computing facilities. Bernd Wegner, who took over the position of editor-in-chief for the West Berlin office in 1974, further developed the ideas of technically advancing the service. For instance, recording of the text was now performed on magnetic tapes and, from these, the printed version of *Zentralblatt* was produced. Due to the reorganization of all information and documentation centers in West Germany, *Zentralblatt* was eventually integrated into one of these, namely the Fachinformationszentrum Energie, Physik, Mathematik (the current FIZ Karlsruhe – Leibniz Institute for Information Infrastructure) in Karlsruhe. Around 1980, a bibliographic database was available which was able to manage the inflow of now around 50,000 new documents per year, as well as keeping track of the reviewers and the status of reviews.

Another major breakthrough was the introduction of TeX. Today, it is hard to imagine the enormous efforts required for setting mathematical texts in lead, with handwritten addition of complicated diagrams and exotic letters. Hence, *Zentralblatt* benefitted extremely from the introduction of Donald Knuth's ingenious typesetting system. Up to this day, precise TeX-coded data are the heart of the service, allowing for very recent applications like MathML display or formula search. Overall, the enhanced infrastructure formed the basis to coordinate special technical and editorial workflows, and created the capacity to include additional information, and keep pace with the growth of literature.

The natural next step was to make the enormous amount of information available to the users beyond the traditional printed volumes. The first release of the database as an offline version on CD-ROM called CompactMATH was published in 1990. The transition of *Zentralblatt* to a service accessible through the Internet was accomplished in 1996; the database was named MATH and subsequently renamed zbMATH. In 2004 all records from the *Jahrbuch* were digitized and incorporated as an extension to the database. Moreover, the complete bibliographic data of Crelle's Journal (*Journal für die reine und angewandte Mathematik*) was added from its first issue of 1826. This makes zbMATH unique as the most comprehensive source of mathematical information from 1826 to the present.

Further evolutions were again influenced by politics: After the end of the Cold War, and with the subsequent beginning of European integration, many co-operations took shape which allowed for an even better coverage of the regional literature. By now, there exist ten external editorial units, ranging from mostly Eastern Europe to China. This was also reflected by another major change in the editorial structure: the



Bernd Wegner

[2] See the article "Some Facets of Mathematical Knowledge Management and Communication", p. 28.

European Mathematical Society (EMS) joined as a zbMATH stakeholder in 1999. Since then, the EMS has had an enormous impact on the further development of the service: many ideas were generated from the joint feedback and communication coming from both the reviewer community and EMS members. Among the many fruitful initiatives, the successful free access program and the book donation program for developing countries, both steered by the EMS-DC committee and relying on zbMATH infrastructure, should at least be mentioned.

With the database becoming the main service, the developments during the last decade were centred around enhancing and extending the depth, precision and completeness of the data. With the core service defined by the bibliographic data and reviews, many more facets were added successively, thereby greatly empowering the access opportunities of the gathered information. Linking to full texts of the publications was a natural follow-up to the immense digitization efforts: By now, there are not just almost 2 million DOI's[3] integrated, but also additional links to Open Access digital libraries like EuDML and Project Euclid, or repositories like the arXiv.

Today, searching goes far beyond documents. What started as a regularly printed author index with only limited use, has, since 2008, taken the shape of a full database of more than 800,000 mathematics people. In addition, author profiles were introduced in 2010, which accumulate the comprehensive author-related information derived from the data, including publication structure, research areas and coauthor networks. The key requirement here is, as for most of the other database features, the precision of the data: author disambiguation is a difficult task. Common heuristics, as employed by general search engines, will usually succeed in about 70-80% of the cases; but for a profile which combines several features this is far from ideal: since the errors cumulate, the profile accuracy would plunge quickly far below 50%. Hence, only massive investments in data quality, as well as a smooth integration of community feedback, can ensure a useful service at this stage.[4]

This is all the more true when it comes to the level of a citation index. In 2011, zbMATH started to integrate citation data, not just for a large part of the newly indexed documents (now around 120,000 per year) but also retrospectively. Today, there are about 8,500,000 citations available, going back as far as 1885. However, there is much more to be done: as in every reference database, even this large figure – though defining a representative part of the journals – is only a fraction (about 15-20%) of all potentially available citations. Hence, there is still much ongoing work to do to extend this basis, in particular collaborating with more publishers. But this is just the first step: citations need to be matched against the existing identifiers to gather reliable information. This is currently done on the document level with sufficient precision, but is steadily improved with a view to citation profiles (which, potentially influenced by cumulated errors, require utmost accuracy).

Recent research projects extend zbMATH information beyond classical bibliographic data, citations and reviews. Semantic tools enable the search for mathematical subjects and expressions which greatly refine the Classification by MSC[5] and even allow for combined formula search[6]. A completely new aspect comes from the growing importance of mathematical software both as a tool for and a result of mathematical research. As an outcome, the swMATH database for mathematical software has been built up since 2011[7].

Such a complex network of information requires a sophisticated interface, empowering users to navigate along the twisted paths through the data labyrinth after all, mathematicians employ specific search strategies[8]. This has been achieved by the introduction of the new zbMATH interface, a result of several initiatives after the appointment of Gert-Martin Greuel as new editor-in-chief in 2012. With the focus on enhancing the online service, the traditional printed service was discontinued. The interface, with new logo and look, allows now for multifaceted search approaches from many different viewpoints. By taking advantage of many opportunities offered by modern browsers,

[3] Digital Objects Identifier, ensuring a permanent link to full texts
[4] See the article "Author Profile Pages in zbMATH", p. 15.
[5] See the article "Some Facets of Mathematical Knowledge Management and Communication", p. 28.
[6] See the article "Mathematical Formula Search", p. 26.
[7] See the article "The Mathematical Software Portal swMATH", p. 23.
[8] See the article "Catering to Clichés: Mathematical Practices and Interfaces", p. 19.

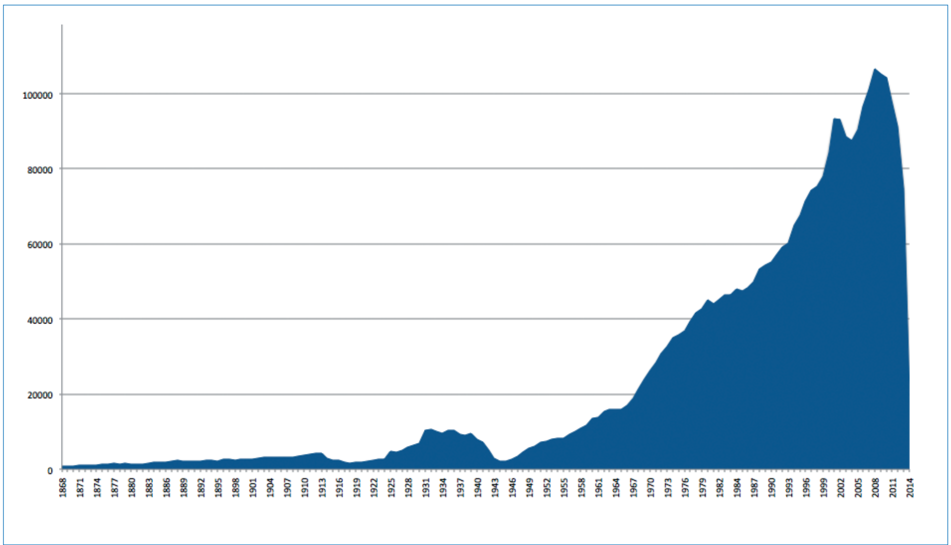
users are now quickly guided to the required information. The search is organized in different tabs documents, authors, journals, classification, software allowing to focus on a specific type of information. This structures the information in an easily and intuitively comprehensible manner: if you are interested in author profiles, then you use the author tab; whenever you click on an author's name, her or his author profile page is presented; a journal title links to the corresponding journal profile; and clicking on the number of documents a certain person has published in a certain mathematical area takes you to exactly those documents in our database where you can read reviews or abstracts of the articles and can access, in most cases, their full texts.

The user's options are greatly increased by the introduction of the filtering function which can not only be used to refine the initial search query according to authors, journals, mathematical fields and the publication year but also immediately provides a search profile which can be directly used for, e.g., evaluation purposes. In this way, it is not only possible, for instance, to start off with a rather unspecified query and to subsequently narrow it down, giving the interface also a browsing character, but also to phrase more complicated search queries such as "Who are the most prolific authors in the top 5 mathematics journals?".

The basis for doing so is a search engine which allows for extremely fast multiple parallel searches, and a MathML conversion which quickly displays formulae in modern browsers. Alternatively, the slower MathJax option can also be chosen as a preference, which is a general feature of the interface: users can choose several options (sorting, output format, number of results, etc.) according to their needs.

The integration of the swMATH database as a new facet connects mathematical software information to publications. This can be used from many viewpoints: e.g, one may find suitable software packages to solve specific problems by starting with a thematic search, or evaluate software as a non-classical research result by its citation profile in publications.

The qualitative demands of the academic community regarding coverage of mathematical literature have not changed over the decades since the publication of the first *Jahrbuch* and the foundation of *Zentralblatt*. The criteria of completeness, timeliness and objectivity remain a fundamental goal for mathematics reviewing organs. The incorporation of modern technologies to the core of the service, including linkage to complementary material, semantic enhancement, author disambiguation and inclusion of mathematical software, make zbMATH an indispensable tool for researchers in their search for accurate and high-quality information on mathematics publications.



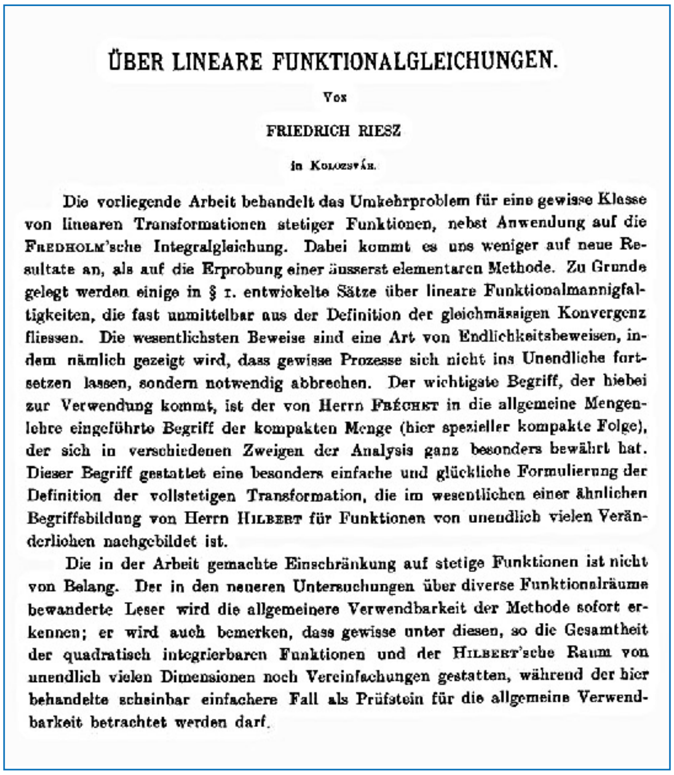
Growth of mathematical publications in zbMATH since 1868 (recent years not yet complete).

Changes in the Publication Landscape

Knowledge from almost 200 years of mathematics is gathered in zbMATH. How did the publication culture evolve, what are the recent developments, and how are they reflected in the database?

Gert-Martin Greuel, Dirk Werner

Mathematical publishing has a long history, but in the last two decades the process of producing and indexing mathematical literature has undergone significant changes. Ever since August Crelle founded his *Journal für die reine und angewandte Mathematik* in 1826, still nicknamed Crelle‘s Journal, more and more journals entirely devoted to mathematics have come into existence; the *Zentralblatt* database zbMATH covers more than one thousand of these plus another thousand periodicals of interdisciplinary character where mathematics is applied at a nontrivial level to problems in physics, economics, engineering etc., problems that can only be described using mathematics and that are
hence intrinsically mathematical.



ÜBER LINEARE FUNKTIONALGLEICHUNGEN.

Von

FRIEDRICH RIESZ

in Kolozsvár.

Die vorliegende Arbeit behandelt das Umkehrproblem für eine gewisse Klasse von linearen Transformationen stetiger Funktionen, nebst Anwendung auf die Fredholm'sche Integralgleichung. Dabei kommt es uns weniger auf neue Resultate an, als auf die Erprobung einer äusserst elementaren Methode. Zu Grunde gelegt werden einige in § 1. entwickelte Sätze über lineare Funktionalmannigfaltigkeiten, die fast unmittelbar aus der Definition der gleichmässigen Konvergenz fliessen. Die wesentlichsten Beweise sind eine Art von Endlichkeitsbeweisen, indem nämlich gezeigt wird, dass gewisse Prozesse sich nicht ins Unendliche fortsetzen lassen, sondern notwendig abbrechen. Der wichtigste Begriff, der hiebei zur Verwendung kommt, ist der von Herrn Fréchet in die allgemeine Mengenlehre eingeführte Begriff der kompakten Menge (hier spezieller kompakte Folge), der sich in verschiedenen Zweigen der Analysis ganz besonders bewährt hat. Dieser Begriff gestattet eine besonders einfache und glückliche Formulierung der Definition der vollstetigen Transformation, die im wesentlichen einer ähnlichen Begriffsbildung von Herrn Hilbert für Funktionen von unendlich vielen Veränderlichen nachgebildet ist.

Die in der Arbeit gemachte Einschränkung auf stetige Funktionen ist nicht von Belang. Der in den neueren Untersuchungen über diverse Funktionalräume bewanderte Leser wird die allgemeinere Verwendbarkeit der Methode sofort erkennen; er wird auch bemerken, dass gewisse unter diesen, so die Gesamtheit der quadratisch integrierbaren Funktionen und der Hilbert'sche Raum von unendlich vielen Dimensionen noch Vereinfachungen gestatten, während der hier behandelte scheinbar einfachere Fall als Prüfstein für die allgemeine Verwendbarkeit betrachtet werden darf.

Figure 1: F. Riesz‘s 1916 *Acta Mathematica* paper

There are at least two features that are particular to the mathematical literature. One is, obviously, the extended use of formulas, which traditionally were difficult and costly to render in print, the other is its longevity. As for the latter we would like to mention Frigyes Riesz‘s paper *Über lineare Funktionalgleichungen* published in *Acta Mathematica* in 1916 (see Figure 1); this is a fundamental paper in functional analysis in which Riesz developed the spectral theory of compact operators on Banach spaces. More interestingly, he did this in a hitherto unsurpassed way: his arguments have not been simplified in the last 100 years and are reproduced in all contemporary textbooks.

This is just one example of an article of timeless value, and the mathematical literature is full of them. It is important for a mathematician to have these treasures at one’s disposal, ideally accompanied by an expert‘s opinion. The zbMATH database provides exactly this. It has incorporated the data from the *Jahrbuch über die Fortschritte der Mathematik*, its predecessor and for some time its competitor founded in 1868, and gives access to such information at a mouse click (Figure 2).

In fact, a contemporary view on such classical works would clearly be welcome, and zbMATH offers a few of them, for example the one in Figure 3. One of our most experienced and dedicated reviewers has set out to review all the material in Crelle‘s Journal from its inception until 1868, when the *Jahrbuch* started, and we are looking forward to the results of this truly herculean endeavour.

Returning to the wealth of formulas in mathematical works it is probably safe to say that Donald Knuth‘s

Riesz, F.

Über lineare Funktionalgleichungen. (German) JFM 46.0635.01

Acta Math. 41, 71-98 (1916).

Die Arbeit behandelt eine umfassende Klasse der in ihren Eigenschaften den linearen Integralgleichungen 2. Art mit stetigem Kern analogen linearen Funktionalgleichungen

$$B[\varphi] = \varphi - A[\varphi] = f$$

nach einer Methode, die lediglich die allgemeinen charakteristischen Eigenschaften der linearen Funktionaltransformationen benutzt und von der Art der Objekte dieser Transformationen (der Elemente des zugrunde gelegten unendlichdimensionalen Raumes) im wesentlichen unabhängig ist; die Entwicklungen werden durchgeführt für den Raum der stetigen Funktionen $f(x)$ von x im Intervalle $a \leqq x \leqq b$. Dabei wird als Norm $\|f(x)\|$ verwendet das Maximum von $|f(x)|$, als Distanz zweier Funktionen die Norm ihrer Differenz und demgemäß als Konvergenzbegriff für eine Funktionenfolge der der gleichmäßigen Konvergenz; als *kompakte* Funktionenfolge im Sinne von *Fréchet* hat daher eine solche zu gelten, bei der jede Teilfolge eine gleichmäßig konvergente Teilfolge enthält. Das Haupthilfsmittel bilden einige Untersuchungen über lineare Mannigfaltigkeiten von Funktionen (die im Sinne gleichmäßiger Konvergenz abgeschlossen sind), die sich auf Distanzabschätzungen von verschiedenen linearen Mannigfaltigkeiten angehörenden Funktionen beziehen, und die die bei dem vorliegenden Normbegriff nicht anwendbaren Verfahren der orthogonalen Projektion (*Bessel*sche Ungleichung oder dgl.) ersetzen; wesentlich ist ferner die Bemerkung, daß nur in einer Mannigfaltigkeit endlicher Dimensionszahl jede beschränkte Funktionenfolge (mit beschränkter Norm) kompakt ist.

Eine Funktionaltransformation $f_1 = A[f]$ heißt *linear*, wenn sie distributiv und beschränkt (d. h. $\|A[f]\| \leqq M\|f\|$ mit festem M für alle f) ist; sie heißt *vollstetig*, wenn sie jede beschränkte Funktionenfolge in eine kompakte überführt. Für solche vollstetigen Transformationen A wird die Funktionalgleichung (1) untersucht. Die letzte Bemerkung des vorigen Absatzes führt zu der Erkenntnis, daß die Lösungen der homogenen Gleichung $B[\varphi] = 0$ eine lineare Mannigfaltigkeit endlicher Dimensionszahl bilden; das gleiche gilt für die iterierten Gleichungen $B^n[\varphi] = 0$, und da sich zeigt, daß von einem hinreichend großen $n = \nu$ an hier keine neuen Funktionen mehr hinzutreten können, ergibt sich so die endlichdimensionale Mannigfaltigkeit der Hauptfunktionen – wie sie in der Theorie der Integralgleichungen heißen -, die Verf. als "Nullelemente" bezeichnet. Analog läuft die Untersuchung der durch die Transformationen $g = B^n[\varphi]$ aus dem Raum aller φ entstehenden linearen Mannigfaltigkeiten, die für $n \geqq \nu$ miteinander identisch sind ("Kernelemente") und für die Lösung der inhomogenen Gleichung charakteristisch werden. Durch die Unterscheidung $\nu \geqq 0$ ergibt sich dann die *Fredholm*sche Alternative; allgemein läßt sich jede Funktion eindeutig als Summe eines Kern- und eines Nullelementes und entsprechend A eindeutig als die Summe zweier Transformationen darstellen, von denen die eine A_1 alle Nullelemente, die andere A_2 alle Kernelemente in Null überführt. Dann ist $E - A_1$ eindeutig umkehrbar, während die zu $E - A_2$ gehörigen Gleichungen dieselben Lösungsverhältnisse aufweisen, wie (1). Da A_2 eine Transformation der endlichdimensionalen Mannigfaltigkeit der Nullelemente in sich darstellt, so ist damit die Untersuchung von (1) auf ein algebraisches Problem in endlichvielen Dimensionen zurückgeführt.

Eine Transformation "vom Integraltypus" $K[f] = \int_a^b k(xy)f(y)dy$ mit stetigem $k(x, y)$ erweist sich leicht als vollstetige Transformation; die entwickelten Sätze ergeben dann ohne Schwierigkeit die gesamte Auflösungstheorie der Integralgleichung 2. Art sowie die Abspaltung des Hauptbestandteiles des Kernes und seinen Ausdruck durch die Hauptfunktionen.

Reviewer: Hellinger, Prof. (Frankfurt a. M.)

References:

[1] C. Arzelà, "Sulle funzioni di linee", Memorie d. R. Accad. d. Scienze di Bologna, serie 5, t. V (1895), S. 225–244.

Figure 2: Review of Riesz's paper 1 in *Jahrbuch*

Liouville, Joseph

Memoir on the integration of a class of transcendental functions. (Mémoire sur l'intégration d'une classe des fonctions transcendantes.) (French) ERAM 013.0476cj

J. Reine Angew. Math. 13, 93-118 (1835).

Joseph Liouville was in his early twenties when he considered the integration of some algebraic functions as a first step towards integration in finite terms. Laplace had discussed this problem en passant, and Abel and Poisson also gave contributions to this topic. The main result is given as Theorem on p. 94 of the paper under review:

"Si l'intégrale $\int P\,dx$ est une fonction finie de x, y, z, on pourra toujours poser:

$$\int P dx = t + A \log u + B \log v + \cdots + C \log w,$$

$A, B, \ldots, C$ étant des constantes et $t, u, v, \ldots, w$ des fonctions algébriques de x, y, z."

In other words: This theorem predicts directly from the integrand what the antiderivative looks like. The crucial fact is that this gives us the prior knowledge that only logarithms of terms that appear already in the input and no other new functions can appear in the output of such an integral.

The theorem given on p. 108 of the paper under review generalizes this to general elementary functions containing algebraic, exponential and logarithmic functions and arbitrary compositions of them, and is nowadays called *Liouville's Theorem* (from differential algebra):

"Soit P une fonction rationnelle de la variable indépendante x et des deux quantités y, z déterminées par deux équations différentielles de la forme $dy = p\,dx, dz = q\,dx$, p et q étant aussi des fonctions rationnelles de x, y, z: si la valeur de $\int P\,dx$ est une fonction finie de x, y, z, on pourra toujours poser

$$\int P dx = t + A \log u + B \log v + \cdots + C \log w,$$

les fonctions $t, u, v, \ldots, w$ étant non seulement algébriques, mais encore rationnelles par rapport à x, y, z, et $A, B, \ldots, C$ désignant des constantes."

Liouville was well aware of the fact that these theorems can be used to prove that some functions, for example elliptic integrals, *cannot* be written in finite terms.

In some first papers preceding the paper under review he had worked out his famous theorem for algebraic extensions. At the beginning of his studies an algebraic function was something explicit like $y = \sqrt[5]{x}$ for him, but then he introduced the notion of a general algebraic function given by a polynomial equation with integer coefficients like $y^5 - x = 0$.

Starting with the rational functions such algebraic extensions can be made iteratively, and Liouville then speaks of functions of the first, the second etc. type. In modern language he had discovered (and introduced) extension towers of differential fields.

Note that the proofs given by Liouville were purely analytic, whereas in modern investigations these questions are treated completely algebraically. *M. Bronstein* [Symbolic integration I: Transcendental functions, Springer (1997; Zbl 1059.12002)] states Liouville's theorem in the modern algebraic way, D denoting the derivative operator:

Figure 3: Review of a 1835 paper by J. Liouville from zbMATH, 2009

public domain software TeX and its extension LaTeX by Leslie Lamport have revolutionised the typesetting of mathematical texts. When Knuth checked the galley proofs of the second edition of Volume 2 of his *Art of Computer Programming*, he was more than disappointed by the mediocre quality of the typography, and he entered into his diary, "I have to solve the problem myself." The rest is history: In order to get a formula like

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$$

using LaTeX, one simply types

```
$$\frac{1}{\sqrt{2\pi}}
\int_{-\infty}^\infty e^{-x^2/2} dx = 1$$
```

Incidentally, Knuth has chosen the $-symbol to indicate math mode for good reason: Typesetting mathematical formulas was extremely difficult and expensive in the old days.

All this has had repercussions on zbMATH which is today (of course) produced on the basis of TeX, since it possesses, apart from the aesthetical appeal of its output, the advantage of being a highly structured language. The vast majority of our reviewers use TeX or LaTeX to submit their contributions (but ironically this brochure is not produced in TeX). However, the way we display reviews in a web browser is not by means of TeX itself, but through MathML (Mathematical Markup Language), an extension of HTML ideally suited for mathematical texts, which is capable of capturing both their contents and structure. The MathML code is of course generated from the original TeX source code. An ongoing research project uses MathML for formula search that allows users of zbMATH for instance to search the database for formulas like $x^2 + y^2 = z^2$, regardless of whether the appearance of the formula is as above or $a^2 + b^2 = c^2$ or $m^2 + n^2 = p^2$.

Having said this, we have already embarked upon the other major revolution: the advent of the Internet. It has never been easier to disseminate results of mathematical research than today; many authors post their preprints on their homepages or on the arXiv. Incidentally, one of the most important works in the recent mathematical literature, 2006 Fields medallist Grigori Perelman's proof of Thurston's geometrisation conjecture including the Poincaré conjecture has appeared only on the arXiv (Figure 4). For these truly exceptional papers, whose details were meticulously elaborated by the mathematical community, we have suspended the requirement that articles have to be published in peer reviewed journals in order to get indexed and reviewed in zbMATH.

Another new feature are mathematical blogs. This is typically a forum for informal discussions, but in at least one case this has led to deep results. We are thinking of the polymath project initiated by 1998 Fields medallist Tim Gowers (Figure 5), where a massive online collaboration including, among many others, 2006 Fields medallist Terence Tao has produced a new proof of the density Hales-Jewett theorem.

Returning to the more traditional way of disseminating mathematical research by means of books and journals we can observe a profound effect of the use of TeX on the pricing of monographs. Books of the same size, published in the same series and covering similar topics cost the same today as 30 years ago whereas prices for fiction paperbacks have increased by a factor of 3 in this period. The same cannot be said about journals, which have profited in the same way from the new opportunities, if not more. Much has been written about the cost of knowledge, and we won't enter this territory here.

However, there have been efforts to break the pricing spiral that have a great impact on the publication landscape in general. Ever since electronic journals were introduced, the idea of open access (OA) publishing has come up. Whereas traditional publishers charge libraries (i.e., readers) by subscription fees, OA journals can freely be read on the Internet. On the other end they might or might not charge the authors. There are (some) OA journals run by enthusiasts with basic support from universities and learned societies that are free on both ends, and there are publishing houses whose business model is to generate revenue from what is called *article processing* charges. In this category there are lousy journals, good ones and more recently even top-notch high quality journals launched by prestigious publishers.

Of course, there have always been journals of high and of poor quality. But with the subscription model a certain quality control was exerted by library committees who decided which subscriptions to pay for; this is different with OA journals. On the other hand,

Perelman, Grisha
Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. (English) Zbl 1130.53003
arXiv e-print service, Cornell University Library, Paper No. 0307245, 7 p., electronic only (2003)
MSC: 53-02 53C44 53C21 57M40 57R60 Reviewer: Gérard Besson (Grenoble)
BibTeX arXiv

Perelman, Grisha
Ricci flow with surgery on three-manifolds. (English) Zbl 1130.53002
arXiv e-print service, Cornell University Library, Paper No. 0303109, 22 p., electronic only (2003)
MSC: 53-02 53C44 53C21 57M40 57R60 Reviewer: Gérard Besson (Grenoble)
BibTeX arXiv

Perelman, Grisha
The entropy formula for the Ricci flow and its geometric applications. (English) Zbl 1130.53001
arXiv e-print service, Cornell University Library, Paper No. 0211159, 39 p., electronic only (2002)
MSC: 53-02 53C44 53C21 57M40 57R60 Reviewer: Gérard Besson (Grenoble)
BibTeX arXiv

Figure 4: Search for G. Perelman's arXiv papers

Polymath, D.H.J.
A new proof of the density Hales-Jewett theorem. (English) Zbl 1267.11010
Ann. Math. (2) 175, No. 3, 1283-1327 (2012).
MSC: 11B25 05D10
BibTeX Full Text DOI

Polymath, D.H.J.
Density Hales-Jewett and Moser numbers. (English) Zbl 1239.05189
Bárány, Imre (ed.) et al., An irregular mind. Szemerédi is 70. Dedicated to Endre Szemerédi on the occasion of his seventieth birthday. Berlin: Springer (ISBN 978-3-642-14443-1/pbk; 978-963-9453-14-2; 978-3-642-14444-8/ebook). Bolyai Society Mathematical Studies 21, 689-753 (2010).
MSC: 05D10 Reviewer: Igor Vladimirov Protasov (Kyiv)
BibTeX Full Text DOI

Filter results by ...

Authors
Polymath, D.H.J. (2)

Journals
Annals of Mathematics. Second Series (1)

Classification
05-XX (2)
11-XX (1)

Publication Year
2012 (1)
2010 (1)

Figure 5: Search for papers by D.J.H. Polymath

the bundling policy of big publishers makes it difficult for library committees to keep track of the variety of journals of different quality that are in the offered package.

The number of OA journals covered in zbMATH has soared from 180 in 2005 to more than 500 today. Unfortunately, not all of them fall into the "good" or "excellent" category. More than 7,000 reviewers have signed up to contribute to zbMATH, yet we do not have enough reviewers on our roster to have every mathematical article reviewed, especially those that do not appear to have much substance. But occasionally we are in a position to solicit an expert's opinion on a paper that is deemed insignificant by the zbMATH editors. Often, our first impression is corroborated by the reviews; for example, one reviewer has written, "The presentation of the paper is very poor. The statement of Theorem 3.1 is wrong." Another one has said, "The poor reference list and the partly less than stringent mathematical formulations (cf., e.g., the text of Theorem 1) indicate that the author is not very familiar with the recent literature on (...)."
One more example: "The authors (...) conclude the article with a fixed point theorem that is, essentially, Banach's contraction mapping principle. Unfortunately, this article contains a plethora of typographical errors, which makes it somewhat difficult to read." Actually, some reviewers refuse to write at all on papers without any substance and return the manuscript to us right away. All the above examples refer to articles published by the same publishing house. This publisher also released a paper entitled "A complete simple proof of the Fermat's last conjecture", which needs no further comment.

These examples indicate that sometimes a publisher's claim that all their articles are peer reviewed has to be taken with a pinch of salt (maybe they just have a different notion of peer). When we have gathered sufficient evidence about a journal that disproves this claim we have to discontinue indexing it. These publishers sacrifice quality for turnover; every accepted paper, irrespective of its quality, means revenue (typically, article processing charges are in the range of $200 to $500). Thus, one is reminded of Frank Zappa‘s album title “We‘re only in it for the money“. . .

A most obvious example of non-existing quality, indeed non-existing content, in a scholarly publication is provided by a (hoax) paper accepted for publication in the OA journal “Advances in Pure Mathematics“ that just consists of a random collection of mathematical phrases generated by the software *mathgen*; not a single sentence in this paper makes any sense. This was a trial balloon to probe the greatest lower bound of the quality of an accepted contribution. (It is zero, as proved by this incident.) It is a little disconcerting to learn from an article in *Nature* magazine that in Computer Science more than 100 papers compiled by mathgen‘s older sister *SCIgen* got published by well-known publishers.

One more problem in indexing scholarly writing is plagiarism. Recently, several politicians in Germany were involved in scandals concerning their doctoral dissertations when it was found out that large portions of these were copied from other sources. (As an aside, mathematics or indeed science was never a topic of such a dissertation.) Still, there are a number of cases of plagiarism in the mathematical literature as well, and although our readers and reviewers together with some internal routines help us to detect them, we certainly do not have a complete picture. However, when we come across identical papers, we point this out. Figure 6 contains this editorial remark: “Apart from the title, the paper is identical to [A. Conflitti and M.J. Schlosser, J. Nonlinear Math. Phys. 17, No. 4, 429-443 (2010; Zbl 1223.33006)], also including the email address and url of M. J. Schlosser.” Note that the only difference in the title is the typo fuction for function.

Let us end on a positive note. Another feature of present-day mathematical research is the use and development of mathematical software. The new database swMATH on mathematical software (www.swmath.org) systematically analyses all items in zbMATH and provides links between a software package and all available publications that describe, use, or cite the software; for more on this see the article on page 23.

Finally, we try to look to the future. Whether mathematicians can benefit from the future development of electronic publishing depends to some extent on the agreement on standards; we have already mentioned MathML. On the basis of such standards we can use semantic tools for analysing mathematical texts, in particular for the development of metadata schemes for mathematical publications (e.g., finding additional references, similar papers, ...). They can also provide (semi-) automatic methods for creating a controlled mathematical vocabulary, keywords and key phrases. The use of MathML as a presentation and content format allows for the development of new methods of content analysis, in particular for formula search; see the article on page 26 for more details.

Lal, Shobha; Verma, Kiran

A basic hypergeometric equation in context to noncommutative hypergeometric function. (A basic hypergeometric equation in context to noncommutative hypergeometric fuction.) (English)

Zbl 1266.33002

Int. J. Math. Sci. Appl. 2, No. 2, 493-507 (2012).

Editorial remark: Apart from the title, the paper is identical to [*A. Conflitti* and *M. J. Schlosser*, J. Nonlinear Math. Phys. 17, No. 4, 429–443 (2010; Zbl 1223.33006)], also including the email address and url of M. J. Schlosser.

MSC:

33C05 Classical hypergeometric functions, ${}_2F_1$
33D15 Basic hypergeometric functions of one variable, ${}_r\phi_s$
34G10 Linear ODE in abstract spaces
34K30 Functional-differential equations in abstract spaces
46H99 Topological algebras, normed rings and algebras, Banach algebras
47A56 Functions whose values are linear operators

Figure 6: zbMATH comment on a duplicated paper

Author Profile Pages in zbMATH

What does the zbMATH author database show, and how does it work?

Helena Mihaljević-Brandt, Nicolas Roy

A solid and distinctive online record of a scientist's research achievements is nowadays almost indispensable to advance one's own academic career, even for a field like mathematics with communities of manageable size and a still rather traditional communication structure. For researchers at an early career stage, a solid record of scholarly contributions is undoubtedly essential for getting a new job, a promotion or the funding for a new project. A distinguishable profile increases the chance of correct attribution, recognition and citation of a scientist's impact and provides the possibility to find others working on a similar topic, establish new contacts, and build partnerships and projects.

Nowadays, most researchers use a personal homepage within the website of their university to present their research and teaching activities. However, the maintenance of such a site can be rather tedious, in particular when moving from one university position to another. Furthermore, not every scientist is sufficiently skilled or interested in designing an attractive website. Some university homepages even suffer from being badly ranked by major search engines, leading to a lower online visibility.

It is therefore not surprising that various online providers offer tools for researchers with the goal to increase the visibility of their scientific activities. Many search engines have recognized the potential of such services, Google Scholar being probably the most prominent example. Also the social media hype of making a profile of oneself publicly available has been brought into the scientific world with enormous success as shown, e.g., by ResearchGate, a kind of "Facebook for researchers". Besides a presentation of the scientific records of a researcher, this service helps to find and establish new contacts and potential collaborations with other researchers, especially based on subject similarities.

One could become intimidated by this growing jungle of services and associated author identifiers. The need for a global and sustainable authorship administration has been recognized by the initiative ORCID[1], which has the promising potential of becoming the future standard author identifier.

In the field of mathematics, the two large traditional indexing and reviewing services, MathSciNet and zbMATH (as well as some regional bibliographic services such as Math-Net.ru for Russian literature) also offer profiles for the authors indexed in their databases.

Author profiles at zbMATH

A user browsing the new interface of zbMATH will immediately find a dedicated tab for author search queries, where one can look for detailed information on a particular mathematician. Entering, for example, the name "jean dieudonné" (or "dieudonne", for instance) into the search field will lead the user to this famous mathematician's author profile, as shown in Figure 1.

Every author profile in zbMATH displays the author's identifier, together with name variations occurring in her or his publications. The assigned publications are broken down according to co-authors, journals and mathematical subjects. Furthermore, the author's contributions are also displayed in chronological order as a clickable diagram for easy visualization of his scientific output. The author profile provides direct links to additional information on related objects in the zbMATH database such as co-authors, joint papers, or documents published in a certain journal. For

[1] The Open Researcher and Contributor ID (ORCID) is a recent non-profit initiative to assign unique identiers to authors of research publications (http://orcid.org).

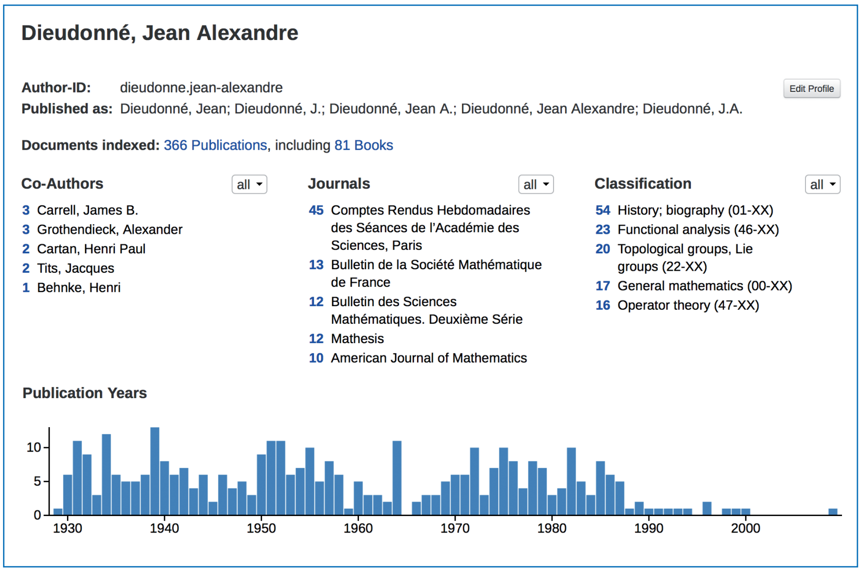


Figure 1: The author profile of Jean Alexandre Dieudonné showing his publications filtered according to co-authors, source, mathematical subject and publication year, respectively.

instance, in the author profile of Dieudonné displayed in Figure 1, the number 3 in front of his co-author "Grothendieck, Alexander" is linked to the list of their joint publications. We recently started to connect our profiles to those of other services, among others to ORCID, the Mathematics Genealogy Project, and to the Biographies of Women Mathematicians of the Agnes Scott College. Also, the profile of an author who has written reviews to publications in our database offers direct links to his or her valuable contributions.

Seeing the increasing prolificness of online services offering researcher profiles, one might wonder about the legitimacy and benefit of having an (additional) author profile at zbMATH. A striking argument for a mathematician is certainly the fact that zbMATH aims at a complete coverage of all mathematical publications at a research level. Currently, about 3.4 million publications are indexed, from all areas of mathematics and its applications, going back to 1755 and complete since 1868. Moreover, our database contains records that are not even available online or rather difficult to retrieve using standard web-search engines, such as journal compilations from universities with a rather low budget or publishing in languages other than English.

From the perspective of data reliability, it is important to mention that zbMATH indexes only peer-reviewed literature. All items undergo an internal editorial process, and many of them, in particular in the core fields of mathematics, are reviewed additionally by an independent expert in one of the corresponding fields. This editorial procedure ensures a high integrity level of the mathematical content, helping, e.g., to detect malicious behaviour such as plagiarism or non-sense papers[2]. The quality control of the indexed documents and of the authorship identification explains why zbMATH (and similar services) enjoy a high level of confidence from the mathematical community, despite the numerous alternative services mentioned before.

Last but not least, our service indexes the works of *all* mathematicians, including those who were active long before the internet era, and hence have neither a homepage nor a ResearchGate or Google Scholar profile. Our database currently counts ~ 860,000 author identities, which makes zbMATH a central place for retrieving information about publication authorships in the entire mathematics community.

Author disambiguation – a multifaceted challenge

The trend of maintaining profile pages and registering on various scientic networking platforms (ResearchGate, ORCID, Google Scholar, ...) is quite recent, and still not intensively picked up by mathematicians. Although the identifiers or links to these services would obviously be a great help for the identification of authorship, they are almost never available in the publications at the moment. Information like affiliation, email, etc. is a rather standard part of a publication's metadata, but these tend to change, leaving practically no other information but the name for the author identification[3].

[2] For an enlightening review in zbMATH see, e.g., https://zbmath.org/?q=an:1202.51019
[3] zbMATH is also involved in a project on text mining using techniques from NLP aiming at the extraction of mathematical vocabulary, which we hope to apply as a tool also for author disambiguation.

The task of attributing an exact set of publications to a certain author based only on a person's name is far from trivial, due to several reasons:

Incompleteness: The available data may be incomplete; some parts of the name may be missing (e.g., middle name, second family name) or abbreviated (use of initials). In earlier years, it was also not uncommon to publish under the family name only. For instance, all information we have on the authors of the article "Erwiderungen auf die Antwort des Dr. Münter" are the surnames Franke and Schmidt[4].

Synonyms: For a single author, one can face a great variability in names, in particular due to different spellings and transliterations. For instance the family name of the famous Russian mathematician *Пафнутий Львович Чебышёв* has been transliterated in at least 10 different ways (Čebišev, Tschebyschew, Chebyshev,Tchebichef, etc). On the other hand, name variations resulting from a name change such as, e.g., after marriage, pose an additional obstacle for the identification task. For example, the mathematician "Yvonne Choquet-Bruhat" has also published under the family names "Fourès-Bruhat", "Bruhat and Fourès".

Homonyms: The same name may refer to multiple individuals. This is particularly acute in the case of certain Eastern Asian names; for example, in China the surnames "Wang", "Li" and "Zhang" account for more than 20% of the population. A search in zbmath.org of all authors with name matching "Y. Wang" would give more than 800 results, suggesting the high level of ambiguity that a paper published under "Y. Wang" could have.

This explains why author name disambiguation is a longstanding research topic with high relevance for bibliometric studies and publication retrieval, and why the field of entity recognition has gained a lot of interest recently.

Author disambiguation at zbMATH

The authors of a new publication record in zbMATH are either identified by means of existing profiles, or new profiles are being created. The first step of this procedure is solved algorithmically by analysing the name string. For example, for an article published as "Alan Weinstein", the algorithm will find two candidates in our author database, namely "Alan D. Weinstein" and "Alan M. Weinstein", with identifiers weinstein.alan-d and weinstein.alan-m, respectively. If, as in this case, the information about the name is not sufficient to identify the author, the algorithm will try to use additional information such as the co-authors. For example, suppose that the article under consideration was co-written by "Victor Guillemin". Then, the algorithm will consider all publication records of both author entities weinstein.alan-d and weinstein.alan-m, and detect that weinstein.alan-d already has other reliably assigned records co-written by "Victor Guillemin", allowing in this case for a disambiguation of the authorship of the article.

After such an algorithmic step, a post-processing procedure involves also manual correction which is often initialized through user requests. The zbMATH author identification team receives such requests on a daily basis, showing an increasing interest of the mathematical community in the availability of reliable author profiles.

In recent years, we have invested a lot of energy into both the algorithmic and manual way of handling the author identification task. The regular improvements of our algorithm, together with an intensive manual curation for targeted groups of authors, have allowed for a signicant enhancement of a part of our data. The 3.4 million publications currently indexed at zbMATH correspond to ~ 5.6 million authorships, which are distributed between our ~ 860,000 author identities. More than 7% of these attributions have been checked manually, treating some particularly difficult cases and providing a gold-standard test suite for further algorithmic development. Within the remaining 93% of automatically attributed authorships, only 6% correspond to ambiguous attributions (i.e., the algorithm cannot decide between several candidates). Compared to the statistics from two years ago, we have managed to decrease the amount of ambiguous assignments by more than 2%, despite an absolute data growth of ~ 7%.

As outlined above, the problem of author disambiguation is too difficult to be tackled completely algorithmically. On the other hand, the amount of publications incorporated every year into zbMATH (~ 120,000 new items) and other information services is too big to solve this task manually. We thus decided to open up this process at least partially to community input. At present, requests to zbMATH to clean up author profiles or correct mistakes in attribution of publications are usually presented by researchers via email.

[4] https://zbmath.org/?q=an:02637903

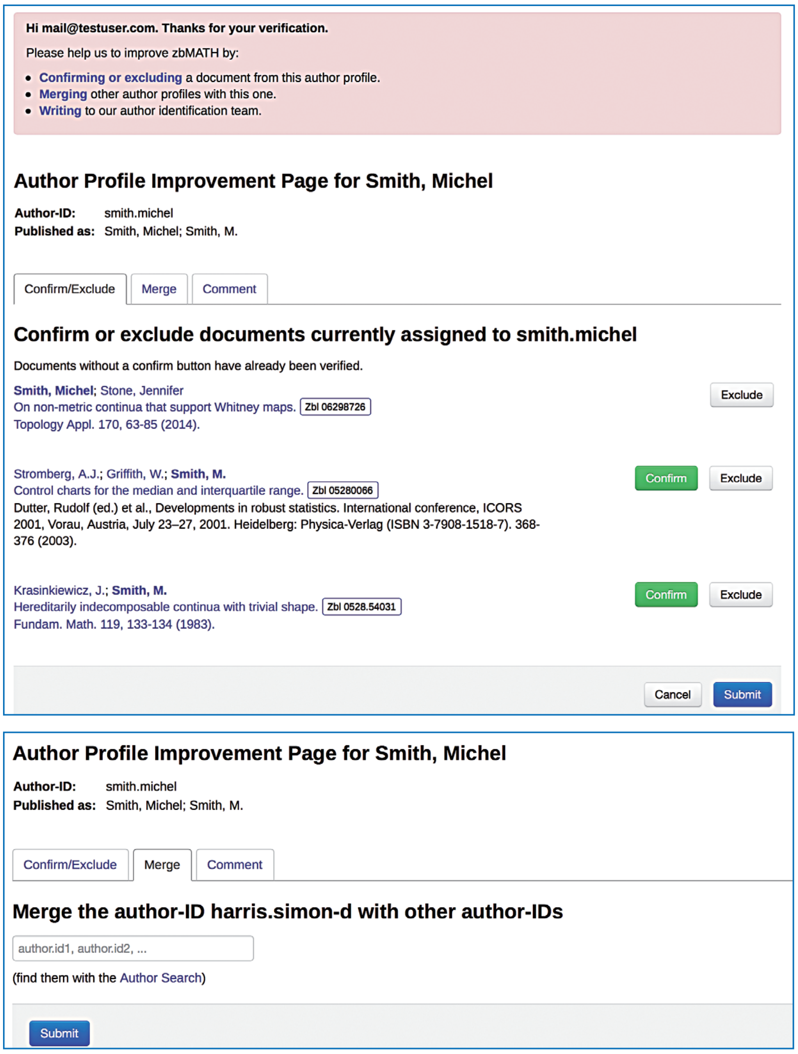


Figure 2: User interface for editing of author profiles

This approach is evidently error-prone since, e.g., it requires hand typing of author or document identifiers which are highly sensitive to typos. Also, it is not very efficient or scalable because of the amount of time required for both writing and processing such an email request. Therefore, we decided to design a publicly accessible interface[5] that allows users to improve the quality of publication data in author profiles (see Figure 2 for a prototype). Through this interface, the user has the opportunity to confirm correctly assigned publications, or exclude incorrectly assigned publications, and resolve in this way ambiguities in algorithmic assignment. She can also merge different profiles that use different names (or name variants) for the same person. The target users of this interface are of course the authors themselves, but also their associates (colleagues, students, co-workers, etc.), with the obvious aim of maintaining their own author profiles up-to-date. But the zbMATH author interface is a service for science itself and the community as a whole. People generally interested in quality of information content (e.g., librarians or historians) also have naturally the possibility to participate in this enhancement of information.

Future developments

For a couple of months we have been testing a prototype of the author disambiguation interface among different user groups (zbMATH editors, mathematics students, librarians). Based on very positive feedback we are confident that the public version of the interface will enjoy high usage in the mathematical community.

Given the expected success of our interface, we envisage to enhance it by additional features, which would require secure authentification. Among the planned extensions would be the possibility to add comments or to tag the content of publications, or to upload and match additional metadata and sources such as references or photos.

In addition to the presentation of the achieved publications of an author, it is often highly relevant and interesting to visualize the scientific impact of these publications. One of the standard sources for such bibliometric information are the references of books and research articles, which we will soon include also on the level of author profiles.

[5] https://zbmath.org/author-profile/edit/

Catering to Clichés: Mathematical Practices and Interfaces

What is special about a mathematician's approach when searching for information, and how could it be ideally supported?

Andrea Kohlhase, Fabian Müller

Professional mathematicians do have a certain reputation of being very smart, unsocial, and ultraconservative. Like most clichés, this one probably contains a grain of truth. But since most people are judging from the outside world without knowing their commonly used mathematical practices and modes of operation, we ask ourselves just how much truth lies in this cliché.

By *mathematical practices* we will understand in this article the actions taken by a professional mathematician (hereafter denoted simply *mathematician*) while "doing mathematics". More particularly, we use the term *practice* in the sense of Lave and Wenger, who introduced the concept of *Communities of Practice* (CoP)[1] in 1991 as the context in which learning takes place and knowledge is produced. By now it is a well-established analytical tool in various fields.
In the same spirit, Brown suggests that

> *"People don't learn to become [mathematicians] by memorizing formulas; rather it's the implicit practices that matter most. Indeed, knowing only the explicit, mouthing the formulas, is exactly what gives an outsider away. Insiders know more. By coming to inhabit the relevant community, they get to know not just the 'standard' answers, but the real questions, sensibilities, and aesthetics, and why they matter."* [2]

Note that "mathematical practice" is thus not meant in opposition to "mathematical theory" here, but rather refers to community-inherent schemes of action and communication.

In this article we want to take a closer look at several clichés about mathematicians as a result of their specific mathematical practices and how to deal with these when designing interfaces for mathematicians.

Are Mathematicians Smarter than other People?

We assume that mathematicians are generally as smart as other academic people (it would be very disturbing if not). But they are analysts at heart. In particular, they tend to be rather reflective in their mathematical practices and – note – also quite intuitive.

For example, let us look at one of the essential mathematical means of expression: formulae. What is a mathematical formula? According to Wikipedia, in mathematics it is "*an entity constructed using the symbols and formation rules of a given logical language*". Even though there are multiple mathematical Communities of Practice, which use a partially different set of symbols and slightly varying formation rules, there is a common understanding of how to encode several levels of information into formulae by extending the linear form of text. Whenever we read a formula, we have to decode the information contained therein. A mathematician can do this easily (at least within her area of specialty), which indicates that the cognitive processes required are embedded into mathematical intuition.

From the outside, this might lead to the impression of mathematicians being smarter than other people. Unfortunately, most such intuitions are not (and can never be) formalized in writing such as a text book

[1] *Jean Lave* and *Etienne Wenger*. "Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives S.)". Cambridge University Press, 1991.
[2] *John Seely Brown*. "Learning in the digital age". Available at http://net.educause.edu/ir/ library/pdf/FFPIU015.pdf, 2005. Accessed on 2011-05-11.

for instance. Therefore, it is quite hard to elicit them and draw conclusions for mathematical software and interfaces. It is made even more difficult by mathematicians' tendencies to think in terms of "truths" and "falsities", due to the fact that at the core of mathematical identity is the concept of *proof*, understood as a process which ascertains reason[3]. As a consequence (and in contrast with other disciplines), "truth" or "reason" is not a question of passion but of logic. In particular, there is no obvious method for uncovering mathematical practices. As part of the Leibniz-funded "MathSearch" project[4], zbMATH sparked a new approach towards formalizing such hidden practices.[5] In what follows, we will look at some of the findings and concrete interface suggestions.

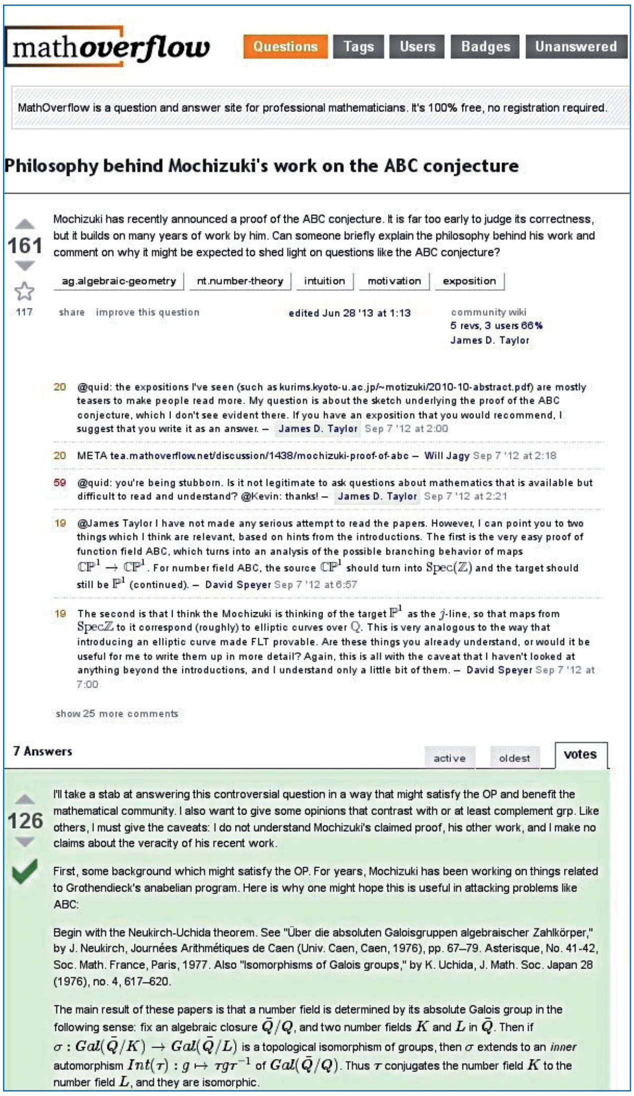


Figure 1: Current developments in mathematics being discussed on mathoverflow

Are Mathematicians Unsocial?

Mathematicians are people. Just like everyone else, they crave for the love of their family, the amity of their friends, and the respect of their peers. However, more so than any other profession they have a reputation for being solitary, often peculiar and sometimes outright antisocial.

For example, in the petition for divorce that Richard Feynman's wife filed against him, it was stated that he was

"[...] constantly working calculus problems in his head as soon as awake, while driving car, sitting in living room, and so forth, and that his one hobby was playing his African drums. His ex-wife reportedly testified that on several occasions when she unwittingly disturbed either his calculus or his drums he flew into a violent rage, during which time he attacked her, threw pieces of bric-a-brac about and smashed the furniture."[6]

Be that in reality as it may, we see things from the perspective of modern tools for the working research mathematician from which we will now look at this question.

Social media and the Web 2.0 are on the rise everywhere, and mathematics is no exception. Junior (and sometimes senior[7]) mathematicians put their questions on mathoverflow and jointly develop and discuss answers, researchers groom their professional profiles on ResearchGate and Google Scholar, and the AMS publishes its announcements on Facebook, Twitter and Google Plus. But how social are these activities? Do they express mathematicians' need to make friends, or are they just tools to get the job done?

[3] *Bettina Heintz*. "Die Innenwelt der Mathematik. Zur Kultur und Praxis einer beweisenden Disziplin". Springer Verlag, Wien, 2000., 210.
[4] See article " Mathematical Formula Search" on page 26.
[5] This resulted in *Andrea Kohlhase*. "Design of search interfaces for mathematicians". Submitted to MathUI 2014 (Mathematical User Interfaces), 2014.
Andrea Kohlhase. "Math web search interfaces and the generation gap of mathematicians". 2014.
Andrea Kohlhase. "Search interfaces for mathematicians". In *Stephan Watt, James Davenport, Alan Sexton, Petr Sojka*, and *Josef Urban*, editors, "Intelligent Computer Mathematics 2014, Lecture Notes in Computer Science". Springer, 2014. accepted.
[6] http://gizmodo.com/5916502/richard-feynmans-fbi-files-make-fascinating-reading
[7] http://mathoverflow.net/users/766/terry-tao?tab=xquestions

Our findings indicate the latter. According to our studies mentioned in footnote 1, mathematicians appreciate community interaction but view it primarily as a tool, that is, something which enables them to attain their goal of solving a problem or proving a theorem for instance. It seems that while the social factor is acknowledged and for most even a pleasant side effect, few use such platforms for the express motive of making friends or building up professional networks. In short, although establishing social interactions for pleasure or career is certainly something that occurs to mathematicians, it is not something they would strive for when using a mathematical service in their day-to-day work.

So what exactly do we mean by social interaction being perceived as a tool? As mentioned in the introduction, the term *practice*[8] is an umbrella term for common customs that are shared within a community and at least partially even serve to define that community. Many of these practices – from doing a proof by induction to structuring a research paper – cannot be communicated in writing. Rather they must be learned by mimicking the actions of seniors, analyzing and exercising them until they become internalized. This acquiring of skills that mark one as a member of an "in group" is precisely where social media can be of help to mathematicians. Posing questions on stackoverflow is a great preparation for mastering the delicate skill of asking pertinent questions at research-level seminars. Solving problems with colleagues on a blackboard forces one to phrase one's thought in a manner that is understandable to others, thus laying the groundwork for writing readable papers. In this way, social media can be considered essential to mathematical practice.

Figure 2: Chalkboard in a Math Classroom (published by Derek Bruff)

Thus one should not so much consider these media as *social networks* with the connotation of making acquaintances that can turn into friends or boost one's career, but rather as research nets which exhibit a network infrastructure but are valued mainly for their efficiency in aiding research. Of course, this does not mean that mathematicians are coldly calculating and consciously seeking out other people only as tools to be exploited. It does however suggest that including a facebook link for example to a mathematical web service is by itself unlikely to provide much of an added value.

Of course, with the generation of digital natives coming of age, things are bound to change in the future. The first author analyzed the data collected in her studies, which were mentioned in footnote 5, partitioned according to seniority of the participants. One of the findings was that among younger mathematicians, socially oriented services like ResearchGate are connotated in the same way as e.g. personal relations to colleagues. Thus tools that facilitate networking behaviour may soon start to play a major part in mathematical practice, and mathematical services should take care to accommodate that need. An interface like zbMATH, which already heavily features user-generated content in the form of articles and book reviews, could "socialize" itself by allowing comments on articles and reviews, maybe even up- and downvotes in keeping with the tide of time.

Even the existing author profiles could be enriched by personal data and discussion. MathSciNet already allows for uploading a photograph – maybe a friends list will be next...

Are Mathematicians Ultraconservative?

A common misapprehension with respect to mathematicians concerns their seemingly ultraconservative attitude. For example, who else still uses blackboards with chalk and sponge as their favorite teaching and communication tool? In Figure 2 we can see a modern math classroom at Vanderbilt published by Derek Bruff[9]. He even comments: *"The math building (...) was renovated a couple of years ago. The chalkboards weren't replaced by white boards, as is popular these days. The Vanderbilt math faculty (including me) are very fond of our chalkboards!"*

[8] as defined in Jean Lave and Etienne Wenger. "Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives S.)". Cambridge University Press, 1991.
[9] Derek Bruff. Stevenson center 1210, 2013. found on flickr.com.

Indeed, even if the whiteboard has frequently replaced the blackboard, the smartboard (even if conveniently available) has not. But what are the underlying reasons? Obviously, using the blackboard means editing text, formulae and drawing sketches, which could be done just as easily with a smartboard. Therefore, can the real cause be that mathematicians like to keep their old and beloved tools and do not want to switch tools easily, that is, they are just ultraconservative people?

It turns out that assessing such a practice as ultraconservative might be too hasty. In the mentioned studies it was shown that familiarity with a math search interface is not an evaluation schema of mathematicians. In particular, for them, using a tool like a blackboard does not rest on them being familiar with it. In the study it also became apparent that they appreciate function over form in interfaces.

Therefore let us look at what functionality the blackboard interface really has to offer. Rather surprisingly, it serves to accomplish quite a handful of tasks: these include wiping clean, getting chalky hands and clothes, using color-coding for expressing relationships or enhancement, using a scratch pad for a side note, and collaborating with a shared visual focus. Each of these properties is woven into mathematical practices and does have its very own value e.g. in the workflow of collaborative work. For example, wiping clean not only deletes but in fact irreversibly destroys previous work, and thereby stresses the relative trust in this action. As a cognitive tool it is invaluable, as it can help to clear the mind, getting rid of a false thought both mentally and physically. In contrast, merely striking it out connotates reevaluation of a fact without losing it. With a smartboard, a mathematician doesn't have such fine-granular options.

Future design of mathematical interfaces should thus scrutinize the hidden functionalities of mathematical work practices in order to attract mathematicians to use them. Reviewing services like MathSciNet or zbMATH may be better off if they recognize e.g. that mathematicians do not care so much for *authors* as for *people*. For instance, they want to get to know other mathematicians who can assist them in achieving their very own, individual mathematical challenges. "Using" them in their daily work until recently required face-to-face meetings, in order to assess individuals with respect to their specific value and idiosyncracies. With the Internet, such real-life meetings do not seem as necessary any longer. But there is still a long way to go from "author profiles" towards "people finder".

Indeed, mathematicians are not ultraconservative in their work practices: They use the tools at hand much more deliberately and precisely than expected. More math-specific search services like mathematical software search and formula search, both of which zbMATH features, empower mathematicians and will be very helpful once they are integrated in everyday mathematical practices. Interestingly, in the interviews conducted by the first author[10], mathematicians indicated that they simply do not believe in the finding capability of a general formula search. Again, the underlying reason is not that they are ultraconservative, but that they do not know yet what to expect. A solution for zbMATH's formula search might result in an utmost transparency: Where does data come from, in what ways can it be manipulated, how does the computer support work and what is consequently to be expected in the result?

Conclusion

We have looked at some of the most common clichés one encounters when dealing with society's perception of mathematicians. Misunderstood mathematical practices are at the core of these clichés. Based on the results of the zbMATH studies we can now start to design better mathematical user interfaces and services. More studies like the ones presented here need to be conducted to more deeply understand the underlying reasons for these practices.

With the facts we have now, we can already deduct that blindly following mainstream trends to include social media everywhere is not likely to provide much added value to mathematicians, and that clear exposition of the inner workings is imperative when introducing a new interface to the mathematical world. Moreover, it becomes quite clear that mathematicians' needs and preferences are quite removed from that of the "average" user. It remains a challenge as well as a necessity to bear this fact in mind and align design decisions with their requirements – even if that means occasionally reinforcing clichés.

[10] See also *Jin Zhao, Min-Yen Kan*, and *Yin Leng Theng*. "Math information retrieval: User requirements and prototype implementation". In Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '08, pages 187–196, New York, NY, USA, 2008. ACM

The Mathematical Software Portal swMATH

The outcome of today's mathematical research goes far beyond publications. A considerable amount of knowledge condenses in the form of program code. The portal swMATH connects publications with a database of mathematical software, allowing for completely new ways to approach problems and document research results.

Gert-Martin Greuel, Wolfram Sperber

If a company faces the problem of designing a new class of circuits, say, for a sensor, or if engineers are working on integration of new components in a complex system on a chip, the first thing they need is a mathematical model and then the software for experimentation or, possibly, for solving the arising mathematical problems. In most cases one would use standard tools like simulators but it can happen that one encounters new phenomena whose analysis requires the help of specialized mathematical software.

But how can we find suitable software? This problem has several perfidies, such as: Where can we search for software? How should we formulate the search (in mathematical terms)? How should we decide which software out of several similar ones is suitable for our problem? The first and standard attempt is to use a search engine or to look for appropriate functions in one of the well-known and established general purpose software systems such as Matlab, Mathematica or Maple. However, quite often the results of the search are not satisfactory, particularly if the problem is rather specific. The large search engines, such as Google, Bing or Yahoo, are not at all suited for a search for specialized questions related to mathematical software. This is mainly due to the fact that links between a mathematical question and the software packages which are able to solve it are usually very poorly, if at all, represented on the Web.

In order to develop an all-inclusive information and search service for mathematical software, the Mathematisches Forschungsinstitut Oberwolfach and FIZ Karlsruhe, together with a group of institutions with expertise in mathematical software, joined forces in a project – SMATH – funded by the German Leibniz Association. The result of the project, the portal swMATH, is freely available on www.swmath.org and integrated into the zbMATH database as the search facet www.zbmath.org/software/.

The publication-based approach to mathematical software

Mathematical software and publications are closely interconnected, although, as mediums to present mathematical knowledge, they are clearly different. Ideas and algorithms often initially appear in publications and are then subsequently implemented into software packages. The implemented software is then used, in applied as well as in pure mathematics, to solve problems either numerically or symbolically, and the solution is subsequently published in a journal. On the other hand, developing mathematical software is itself a rapidly growing branch of mathematics that decisively relies on mathematical theories. As a result, mathematical software allows realization and computational verification of mathematical theories and it also helps to formulate new mathematical conjectures. Thus, the use of software can inspire new research and lead to novel mathematical results.

However, the medium "software" differs in many aspects from mathematical publications, which is the classical form of presenting mathematical knowledge: software has a temporary character and is developed dynamically, leading, for instance, to releases of different versions. It also depends on technical factors, different performance parameters are possible and its life-span is often limited. Hence mathematical software may be rather considered as "experimental" or "interactive" mathematics. On the other hand, there is an increasing linking between mathematical publications and mathematical software, since more and more publications describe a software package or report on using software to solve a certain problem. This connection between mathematical publications and mathematical software was one of the main ideas behind the information service swMATH.

Within the project our approach for acquiring information on software was to systematically use information from publications indexed in zbMATH. For this we combined heuristic methods and manual evaluation, which has led to the discovery of thousands of links to software in the zbMATH data.

During the development of our heuristics we learned that there are roughly two types of mathematical publications referring to software: publications whose main content is the software itself ("standard publications") and those where software is applied as a tool ("user publications"). As a consequence, while a standard publication can be utilized to obtain important characterizations of the software in question, a user publication contributes, in particular, information on research fields and problem types to which the software package can be applied. It is exactly this content information which we utilised to establish a similarity metric on the set of software packages in swMATH.

The portal swMATH

The result of the SMATH project is the open access portal swMATH, which is now maintained and further developed by FIZ Karlsruhe. The underlying database contains information on more than 7,000 software packages connected to more than 70,000 software-relevant publications. Currently, it is the world's largest database on mathematical software.

The swMATH service enhances the zbMATH portal as a valuable additional search facet. Publication records in zbMATH containing relevant software information provide links to swMATH and vice versa. Each package indexed in swMATH is assigned a unique identifier which can be used for referencing software. The importance of unique identifiers has long been recognised for publications (DOI) and even for authors[1] but there is as of yet no such global solution for software, making it difficult to automatically analyse references to software in publications.

Every record supplies a short description of the software, relevant key phrases in form of a weighted cloud, Mathematical Subject Classification codes indicating the application areas of the software, a list of publications referencing the software, as well as information on similar software and hence a content-based recommendation system. Information about versions, licences, hard- and software dependencies, programming languages, etc. are provided as far as possible.

The user can access the portal through an easy one-line or a structured search form, one can browse the records by names, keywords and MSC classes, and search results can be ranked based on the occurrence of the software in zbMATH references.

Concluding remarks

The swMATH service is designed as a community-based initiative whose main purpose is to provide comprehensive and reliable information on mathematical software. It will be maintained and extended by FIZ Karlsruhe and for this we need input from the community for improving the service.[2] We believe that the service is not only useful for the user of mathematical software. The documentation of literature referring to a software is also a valuable source of information for the authors of the software since it shows them where and how their software is used. We hope that this will give the developers of software more recognition for their valuable work.

[1] See article "Author profile pages in zbMath", p. 15.
[2] Users can provide information by using the forms under "Contribute" or under "Add information to this software".

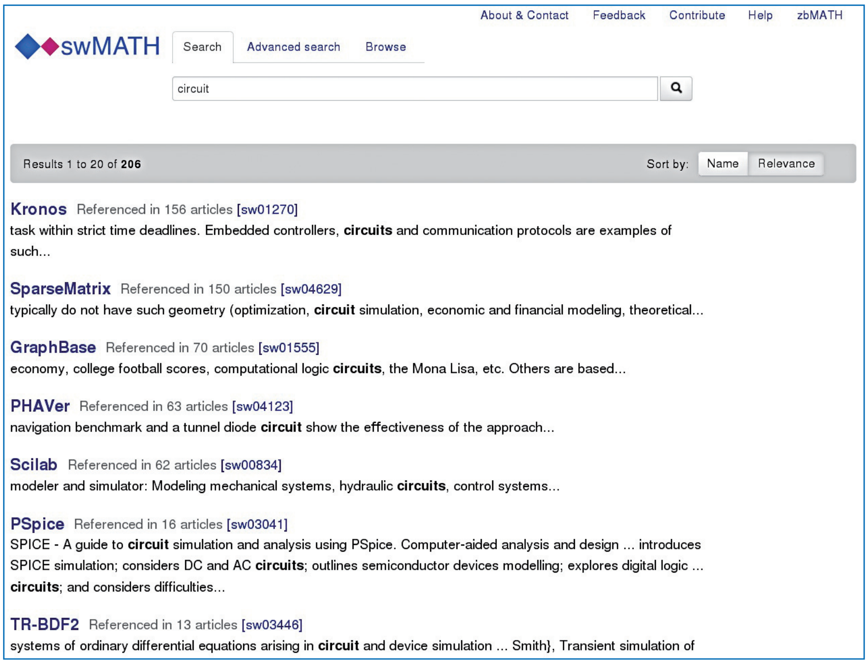


Figure 1:
Searching for software applicable to circuits yields more than 200 software packages, “Kronos” being the one with the most citations in zbMATH

Figure 2:
A software in swMATH typically appears with a short description and additional content-based information such as a keyword-cloud, citing articles and relevant MSC classes.

Mathematical Formula Search

How does one find a mathematical formula?

Michael Kohlhase

Have you ever encountered an unfamiliar symbol in a paper and immediately wanted to know more about the object it denotes? Or an expression in a calculation for which you would like to analyze relevant literature? Or have you gotten stuck in a proof and wanted to know which identities are applicable so that you can progress?

A traditional approach to such situations would be to consult an expert in the field, and this is in many cases certainly still a good idea. But you may not know the right person, and even with an expert available one can hardly be sure that she or he will cover the vast complexity of modern development in mathematics. In particular, the retrieval of non-English literature remains a difficulty here. You could also post your question to a forum like http://mathoverflow.net but, again, you have to rely on luck that the right person comes across it. Maybe you would even be desperate enough to attempt to employ a search engine like Google or Bing even though you know that they are specialized in finding word occurrences in documents. But formulae are not words and so results from traditional search engines are erratic. What we really need in the situations described above is a *search engine with formula search capabilities*.

The MathSearch project

To address this necessity and support mathematical research, the German Leibniz Association funds a collaborative research project by FIZ Karlsruhe and *Zentralblatt MATH* (zbMATH) and a group of computer scientists from the Jacobs University Bremen. The goal of the 3 year MathSearch project, which started in March 2012, is to develop tools for information retrieval and literature access for mathematics. A first prototype is already available on http://zbmath.org/formulae/ and ready to be explored by mathematicians (see Figure 1); subsequently, improved versions will be permanently integrated into the zbMATH interface as an additional search facet.

In situations where we partially remember a formula – e.g. the energy of a signal *s(t)* has something to do with squaring *s(t)* and integrating over it – we would like to search for formula schemata like

$$\int_{?a}^{?b} ?s(t)^2 \mathrm{d}t$$

where *?a*, *?b* and *?s* are query variables: wildcards that can be instantiated by the search engine (those are the parts we do not remember or do not care about). Similarly, if we are stuck in a proof, e.g. needing an approximation of the integral $\int_0^{2\pi} |\sin(x)\cos(x)| \mathrm{d}x$ from above, queries such as the one in Figure 1 could provide hints.

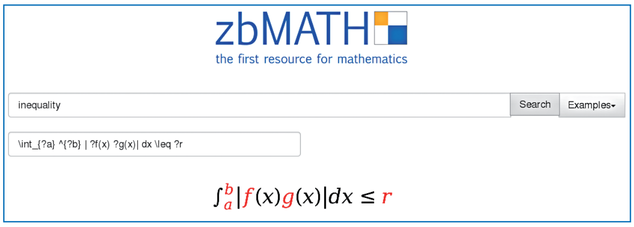


Figure 1: Searching for an applicable theorem

Since such search requests belong to the daily work of a mathematician, infrastructure services have already started to integrate formula search engines into their facilities. The NIST Digital Library of Mathematical Functions (http://dlmf.nist.gov/), for instance, offers a formula search for their content, which is highly standardized. The European Digital Mathematics Library EuDML (http://eudml.org/) also offers a formula search with matches based on similarity. However, none of the existing search engines employs a thoroughly *semantic* approach, trying to encode the complete mathematical *meaning* of the entered query; instead, the matchings are displayed according to their *structural similarity*. Within the MathSearch project we try to combine the expertise of the mathematical knowledge management group at the Jacobs University together with the broad knowledge of the zbMATH editorial board at FIZ Karlsruhe in order to build an *intelligent* search facility for mathematicians. The zbMATH database with its comprehensive and carefully edited content is certainly a very good source for such a service.

Mathematical knowledge retrieval

The problem of mathematical information retrieval and literature access has four facets: (i) *digitization* (only digital documents can be searched); (ii) *content extraction;* (iii) *search;* (iv) *result presentation*. The MathSearch project sidesteps the first by restricting itself to born-digital (LaTeX) documents: primarily the zbMATH database and the http://arXiv.org corpus, which leaves three problems for the MathSearch project.

Regarding (*ii*), note that mathematical documents are written in formats (usually LaTeX) optimized for formatting (visual layout) of formulae, not their functional structure, which interests us in a search. In the first example above, we want to find Parseval's theorem,

$$\frac{1}{T}\int_0^T s^2(x)\mathrm{d}x = \sum_{k=-\infty}^{\infty} |c_k|^2$$

even though the boundary variables have been renamed. Generally, we want to search for the functional structure of a formula, e.g. for a "binomial coefficient *n* choose *k*", modulo notation conventions like $\binom{n}{k}$ or $\mathcal{C}^n_k$. Of course, this content extraction problem is highly non-trivial, since it is riddled with ambiguities which can only be resolved from context. This problem rears its ugly head even for very simple formulae: $f(a + b)$ can be the product of a scalar f with a sum $a + b$ or the application of a function f to $a + b$ (invisible operator ambiguity), *sin x/y* can be $\frac{\sin(x)}{y}$ or $\sin\left(\frac{x}{y}\right)$ (scope ambiguity), and finally B_n can be the *n*-th Bernoulli or Bessel number (lexical ambiguity).

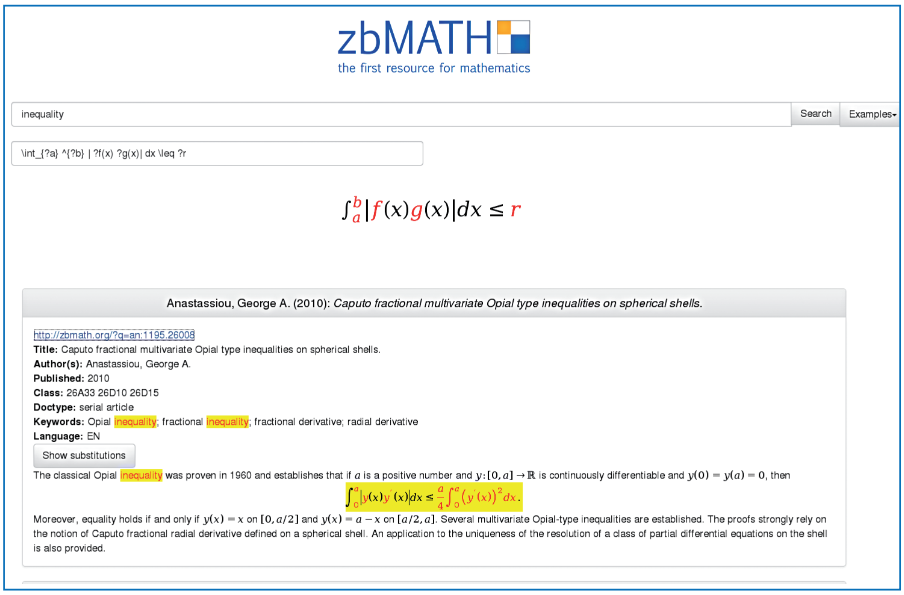


Figure 2: A search result for the query in Figure 1

For (iii) we note that we need sublinear processing algorithms. For an estimated 10-100 billion formulae occurring in the mathematical literature, anything less efficient would not lead to acceptable answer times. In the MathSearch project we currently employ substitution tree indexing, a technique borrowed from automated theorem proving, which has essentially constant answer times ranging from 3-70 ms (avg.=11ms). Although keyword search combined with unification queries for formulae seems to support most math information retrieval needs, combination with metadata search is still an open problem. For the input of formula queries we use LaTeX extended to deal with query variables (see the formula in the lower input box in Figure 1). We generate a visual preview of the formula query on the fly – here query variables are marked in red.

For result presentation (iv) we have implemented text snippet aggregation and formula highlighting. Figure 2 shows the results for the query in Figure 1 and highlights the formulae that match the query. As our search algorithm provides us with the substitution instance of the query variables, we can highlight them in colors. In the zbMATH application, results are ordered chronologically, but in general, ranking of mathematical search results is an open problem.

Ultimately, the development of math information retrieval systems will be less a problem of devising efficient search algorithms or disambiguation strategies, more a problem of cataloging notation conventions, understanding the use of context and identifier scoping in mathematical documents, and engineering query languages in which mathematicians feel comfortable expressing their information needs. Therefore we encourage the mathematical community to use our formula search engine prototype on http://zbmath.org/formulae/ and to give feedback on the search service in general and the search results in particular. This, and the analysis of the queries posed by the community will allow the MathSearch project to improve and calibrate the service.

Some Facets of Mathematical Knowledge Management and Communication

Wolfram Sperber, Ulf Schöneberg

Mathematical knowledge management is the attempt to systematically structure mathematical knowledge and make it accessible to the mathematical community. But what exactly is mathematics? Clearly, it is not a monolithic science and its facets underlie a dynamic development, which we do not intend to discuss here in detail.

Mathematics stems from different sources such as counting and trading, land measurement and geometry, calendar development and astronomy. Today, mathematicians develop new mathematical theories but also apply mathematical concepts to solve problems in other sciences, industry, services, etc. This is not a unidirectional process from theory to application; developments in other fields such as quantum mechanics or information and communication systems lead to new mathematical problems and questions.

The most important source of mathematical knowledge, both in the pre-digital age as well as today, is mathematical literature. For many decades, special institutions of the mathematical community, so-called reviewing services, have collected and indexed mathematical publications and hence provided comprehensive catalogues of mathematical developments and findings. The information provided by such services is not restricted to bibliographic data; the indexed publications are embedded into the canon of mathematics by being assigned to special subjects and reviewed by experts who provide an independent insight into the results and methods employed in a publication. It is therefore no surprise that the reviewing services have been named as the “memory of mathematics”. Furthermore, the overview provided by reviewing services allows to observe new trends in mathematical areas or to find colleagues who work in the same field, offering researchers concrete help prospering in their own research activities.

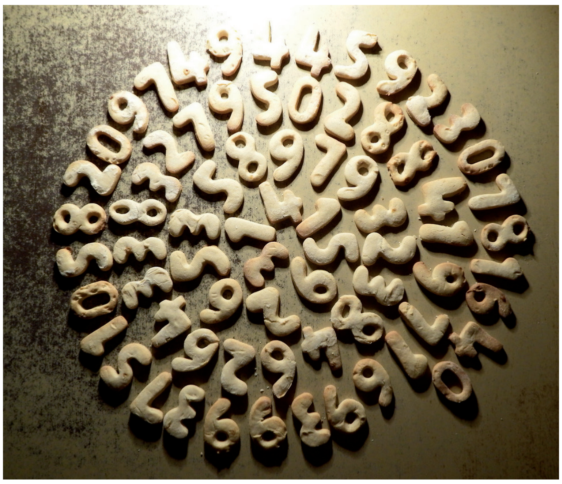


Figure 1: Irrational cookies forming a s∏ral

However, the Web has changed the rules of scientific communication and information completely, confronting the information services for mathematics with new challenges. In this article, we wish to discuss some of the trends and developments, starting with some remarks about the history and development of mathematical knowledge and mathematical communication.

Looking back

Language is the basis of all communication, and this is also valid for mathematics. The mathematical language combines natural language with a special mathematical terminology, which is inseparably connected with symbols for various mathematical objects such as operators or relations.

The current mathematical language is a product of the development of modern mathematics covering the period of the last 500 years. Before the 16th century, most mathematics was written out in words, the equality sign, for instance, was introduced in 1557 by the Welsh mathematician and physician Robert Recorde. The lack of symbols and syntactic rules had hampered the communication of mathematical topics. The invention of symbols allowed to define mathematical objects, operations, relations, and properties in a short and (more or less) unique way, and it also became possible to present and prove mathematical statements in a precise, short and elegant form. Syntactic rules allowed the concatenation of symbols to form complex terms and statements.

The culture of mathematical communication has also fundamentally changed. In the period from 1500 to 1800, it was usual practice to keep new mathematical discoveries and results secret and disputes between scientists were arranged as competitions.

The definition and standardisation of symbols and notations and the change of communication culture paved the way for a rapid development of mathematical research and communication. Moreover, the formalisation and standardisation of mathematical language has strengthened the international character of mathematics and has accelerated the communication between mathematicians. Consequently, mathematics boomed and became broader, foundations such as infinitesimal calculus and new branches such as abstract algebra were developed. This was reflected by a dramatic increase in the number of publications in mathematics. Mathematical literature was synonymous with mathematical knowledge and became a common medium for mathematical communication.

There was however a price to pay for this flood of information: it became more and more difficult to follow the progress of mathematics. So, in the mid-19th century, some initiatives were set up to develop and realise ideas for a better organisation of mathematical knowledge. One of the most important activities was the foundation of the *Jahrbuch über die Fortschritte der Mathematik* (JFM), the first reviewing and indexing service in mathematics worldwide. The JFM collected all mathematical publications, catalogued the bibliographic data such as author name, title and source, assigned each publication to a mathematical subject, and provided short descriptions of the mathematical content through expert reviews. Thus, JFM was the first tool for knowledge management in this field.

Arrival at the digital age

Computers have signicantly changed scholarly information and communication. All information, including mathematics, can be expressed in digital form, and it is this digitised information that can be processed by a computer. The fact that mathematics uses glyphs from both existing alphabets and artificial ones made the typesetting of mathematical publications in the pre-digital age difficult and expensive. However, the development of Donald E. Knuth's TeX, one of the world's most sophisticated digital typographical systems, changed this signicantly. Other markup technologies such as XML and HTML were subsequently combined with ideas from TeX, leading to MathML and similar technologies enabling a smooth presentation of mathematical information on webpages.

It goes without saying that the Web has revolutionised publishing, communication, access, handling and retrieval of information. While the digitisation allows for the typesetting of mathematical content in high quality, Internet and Web stand for easy and cheap publishing of information. Moreover, the Internet provides tools for a new level of scientific communication through formats such as preprint servers, blogs, wikis, and email forums.

The Web is the largest archive of data, information and knowledge in the history of mankind containing web pages, images, books, articles, software, simulations, games etc. To make this enormous amount of information accessible in an effective and easy way, new technologies had to be developed. Different concepts were devised for a machine-based encoding and processing of information which are combined under the label "Semantic Web", a collaborative activity of the World Wide Web Consortium (W3C). One of the main mechanisms for describing sources in the Semantic Web concept is the Resource Description Framework (RDF). This is a graph model for the presentation of information which extends linking by specifying the relation between the endpoints of a link. Another important structural framework for knowledge presentation related to the Semantic Web are ontologies, i.e., hierarchical representations of knowledge using a shared vocabulary, and the closely related Ontology Web Language (OWL), which contains vocabulary and grammar for presenting semantic relations and annotations. There are of course many more concepts and technologies used within the Web to make search and retrieval as easy and intuitive as possible, including heuristics, rule-based, or statistical methods and concepts such as Natural Language Processing (NLP) or Google's PageRank. Without always being aware of it, we benefit from the semantics in the Web on a daily basis. It is, for instance, not such a long time ago that search engines presented only a ranked list of links related to the query we typed into the search line. Nowadays, tagged XML structures are used in order to extract the right information from websites and arrange it in a logically correct way. A nice example for this is searching for the term "world cup" today during a world soccer tournament: instead of links to other websites, Google immediately tells us about all matches scheduled for today and the tentative results.

Specifics of mathematical knowledge and the Mathematical Knowledge Management initiative

Just like the question “What is mathematics?” is non-trivial, so, too, is the question “What is mathematical knowledge?”. Mathematical knowledge consists of theories, theorems, conjectures, examples, mathematical objects described by a definition, properties of mathematical objects, operators, relations, methods, proofs, mathematical models, mathematical software and simulations, visualisations, mathematics publications, etc. This list is by no means complete, and it covers elements of mathematical knowledge at different levels. For example, theorems and proofs belong to theories, and operators and relations can be part of formulae.

A group of mathematicians and information experts have formed a forum for Mathematical Knowledge Management (MKM) to discuss and develop new concepts and standards for presentation, semantic enrichment, and handling of mathematical information. Management of mathematical knowledge is not a *l’art pour l’art* activity, and the MKM initiative aims to develop useful and practice-oriented concepts and tools for presentation, content analysis, processing and retrieval of mathematical information. One of the main tools used in MKM is the Mathematics Markup Language MathML which allows an adequate presentation of mathematics, especially of mathematical formulae, but also a semantic markup of the mathematical knowledge contained therein. MathML, which is a recommendation of the W3C math group and has been integrated into HTML5, is also used within the zbMATH interface.

MKM and zbMATH

As mentioned, mathematicians and users of mathematics are, day in, day out, confronted with a flood of new mathematical results presented in publications, software, visualisations, etc. Typically, the information is provided in digital form on some Web server. In what follows, we present two methods of MKM which are currently used in zbMATH for content analysis and retrieval with the goal to provide to the mathematical community efficient tools to find the information which is relevant and useful for their work and research.

A top-down approach: Classification

One of the oldest scientific methods of knowledge organisation is classification, whose roots reach even to the ancient Greece and the work of Aristotle. Classification is part of our daily life, and we benefit from it mostly without noticing its existence, be it the search for a phone number in the yellow pages or the quick orientation in a supermarket, which works so well only because of the way the retail goods are organised. The classification of objects is the arrangement of these into classes or categories according to certain aspects or properties defined in a scheme. A class may be further devided into smaller classes, often called subclasses, which leads to a hierarchical structure that most classification schemes have in common.

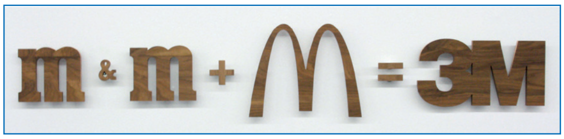


Figure 2: How syntax and semantics add up (by Nic Hess)

The most important classification scheme for mathematical publications is the Mathematics Subject Classification (MSC). It was developed in the 1970s by the American Mathematical Society and has since been maintained and further developed by the editorial boards of the two reviewing services *Mathematical Reviews* and *Zentralblatt MATH*. The scheme is revised every 10 years, the current version being the MSC2010[1]. The MSC is used by authors to classify their work, by librarians to categorize and present their holdings, and of course it plays a prominent role in indexing mathematical literature. Since the 70s, every publication indexed in zbMATH (that appeared since the foundation of the MSC) is classified by at least one MSC code, where the code listed as primary is meant to represent the main mathematical field of the publication’s content.

In its current form the MSC skeleton is a finely granulated three-level tree of mathematical subjects and application fields of mathematics covering more than 6,000 classes: 63 on the top level, 528 on the second, and 5,606 on the bottom level (leaf nodes). For example, the top level 35 stands for Partial differential equations, 35Q for Equations of mathematical physics and other areas of application, and 35Q30 for Navier-Stokes equations. As this example suggests, some of the MSC classes are partially overlapping,

[1] www.msc2010.org

e.g., 35-XX has a lot of objects in common with 76-XX (Fluid mechanics), but the focus is different (the major topics of 35-XX are mathematical aspects of partial differential equations, for 76-XX applications of the mathematical theories in fluid dynamics are more important). For characterising the similarities between MSC classes, the MSC was extended by adding various further relations.

The MSC2010 is presented in different formats including an SKOS version[2]. The acronym SKOS stands for Simple Knowledge Organization System[3] and is also a W3C standard. The essential advantage of SKOS over other formats is the availability of a widely standardized semantic markup of the elements which allows for machine processing of MSC codes (e.g., an automatic interpretation of MSC codes of Web objects). Further advantages of the SKOS version are the multidimensional extensibility, the improved presentation of the similarity relations, and an adequate presentation of mathematical glyphs.

It also provides in combination with other methods the chance to overcome some of the shortcomings of the MSC like the huge number of classes or the heterogeneous overall structure. Some classes describe only a mathematical object, other classes contain a special property, a quality aspect, such as a method, or a model, or an application area, or a combination of different features. Compare for example 35F20 (General theory), which is rather broad, to 35J92 (Quasilinear elliptic equations with *p*-Laplacian), a class dealing with a narrowly defined type of equation.

The MSC is an intellectually developed scheme relying on a broad discussion and consensus within the mathematical community, and for this reason it is very useful. It defines a top-down structure which is a valuable method for retrieval. But the MSC cannot be used for a more detailed search of information (below the level of the MSC). To accomplish this one has to evaluate the particular information of each publication.

A bottom-up approach: Key phrase extraction

The database zbMATH currently lists more than 3.4 million mathematical publications. The MSC classification alone is clearly not sufficient for searching for relevant articles within such a large corpus. Hence a "full text search" is integrated in the user interface of zbMATH covering the reviews resp. abstracts, titles, authors and sources of the indexed publications. Within a current research project, Natural Language Processing (NLP) methods are adapted and used for a machine-based extraction of key phrases. The key phrases extracted from the metadata of a record in zbMATH inform in more detail about the mathematical content of the publication than the MSC code.

The key phrases extracted from the zbMATH corpus can be used for the creation of a mathematical vocabulary, where the frequency of a key phrase can be utilised as a heuristic measure of its significance. Moreover, the publication records serve as a connector between MSC codes and mathematical key phrases, making it possible to enrich the MSC by assigning to each class the most frequent key phrases. A controlled vocabulary is also the first major step towards new retrieval functionalities such as a similarity search on the set of mathematical publications or an intelligent recommendation system. By combining the top-down and the bottom-up approaches we obtain new powerful methods and tools for content analysis, additional semantic annotation, and retrieval.

A foresight

The focus of the MKM initiative is the full spectrum of mathematical knowledge: presentation (e.g. MathML, format converters), enhanced content analysis (e.g. automatic classification and clustering, key phrase analysis, mathematical ontologies, formula analysis and search), and the use of mathematical software (e.g. computer algebra systems). Such tools support mathematicians in creating mathematical content in an adequate and standardised form as well as finding useful knowledge for research, development, and education. They immediately facilitate communication, e.g. by helping to find colleagues working in the same field to discuss problems regardless of geographic location, and cooperation, by, e.g., providing frameworks to commonly create mathematical documents and tools, within the mathematical community and beyond. Therefore, mathematical knowledge must be made available in a machine-understandable and machine-processable form. zbMATH is an active player in the MKM community as a developer as well as a user hoping to improve our search facilities for the mathematical community.

[2] www.msc2010.org/mscwork
[3] www.w3.org/2004/02/skos/

FIZ Karlsruhe – Leibniz Institute for Information Infrastructure
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Germany

Phone +49 7247 808-0
Fax +49 7247 808-259
contact@fiz-karlsruhe.de
www.fiz-karlsruhe.de

Berlin
FIZ Karlsruhe's information services in mathematics and informatics are produced at the Department «Mathematics and Computer Science» in Berlin.

FIZ Karlsruhe
Franklinstr. 11
10587 Berlin
Germany

Phone +49 30 399934 0
Fax +49 30 392700 9
editor@zblmath.fiz-karlsruhe.de
www.zentralblatt-math.org/zmath/en

A Focus on Mathematics
Editors: Klaus Hulek, Dirk Werner
Staff of Department Mathematics and Computer Science
Staff of Marketing Communications

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure