Continuing toward a Global Digital Mathematics Library {The International Mathematical Knowledge Trust}

Patrick D. F. Ion [1] Olaf Teschke [2]

[1] MR AMS ret'd & University of Michigan, MI USA pion@umich.edu

[2] zbMath, Berlin, Germany olaf.teschke@zbmath.org

10 January 2018 / JMM 2018 — Special Session 83A

GDML — Global Digital Mathematics Library

What is it?

- Global — for all the World, from all the World
- Digital — using current technology
- Mathematics — for a specific subject, especially research
- Library — a knowledge base

GDML

Perhaps better

Worldwide Information System
for
Digitally Organized Mathematics

History

Ancient

- Great Library of Alexandria in the Mouseion founded ca. 323 BCE by Ptolemy.
  - Archimedes (287-212 BCE)
  - Eratosthenes (276-195 BCE)
  - Apollonius (262-190 BCE)
  - Aristarchus of Samos (310- 230 BCE)
  - Hero (ca. 10 CE-70 CE)
  - Hypatia, daughter of Theon, the last director of the Mouseion lynched by a rabble in 415 CE
  - New Bibliotheca Alexandrina
- ca. 700 years

Bibliotheca Alexandrina

Alexandria

History

Recent

- Pasigraphy: E. Schröder, G. Peano at ICM 1897
- Georg Valentin's mathematical bibliography to 1928
- Paul Otlet and Henri La Fontaine:
  "Repertoire Bibliographique Universel" (RBU) from 1895
  Mundaneum 1924 to ca. 1941 in Mons
- Vannevar Bush imagined Memex in 1945 (Shannon)

History

Otlet

- (1895 +) A highly advanced index card machine:
  “a moving desk shaped like a wheel, powered by a network of hinged spokes beneath a series of moving surfaces. The **machine would let users search, read and write** their way through **a vast** mechanical **database** stored on millions of 3×5 index cards.This **new research environment would** do more than just **let users** retrieve documents; it would also let them **annotate** the **relationships** between one another, the **connections** each [document] has with all other [documents], **forming** from them what might be called **the Universal Book**.”

Around 1900 - Otlet's Vision

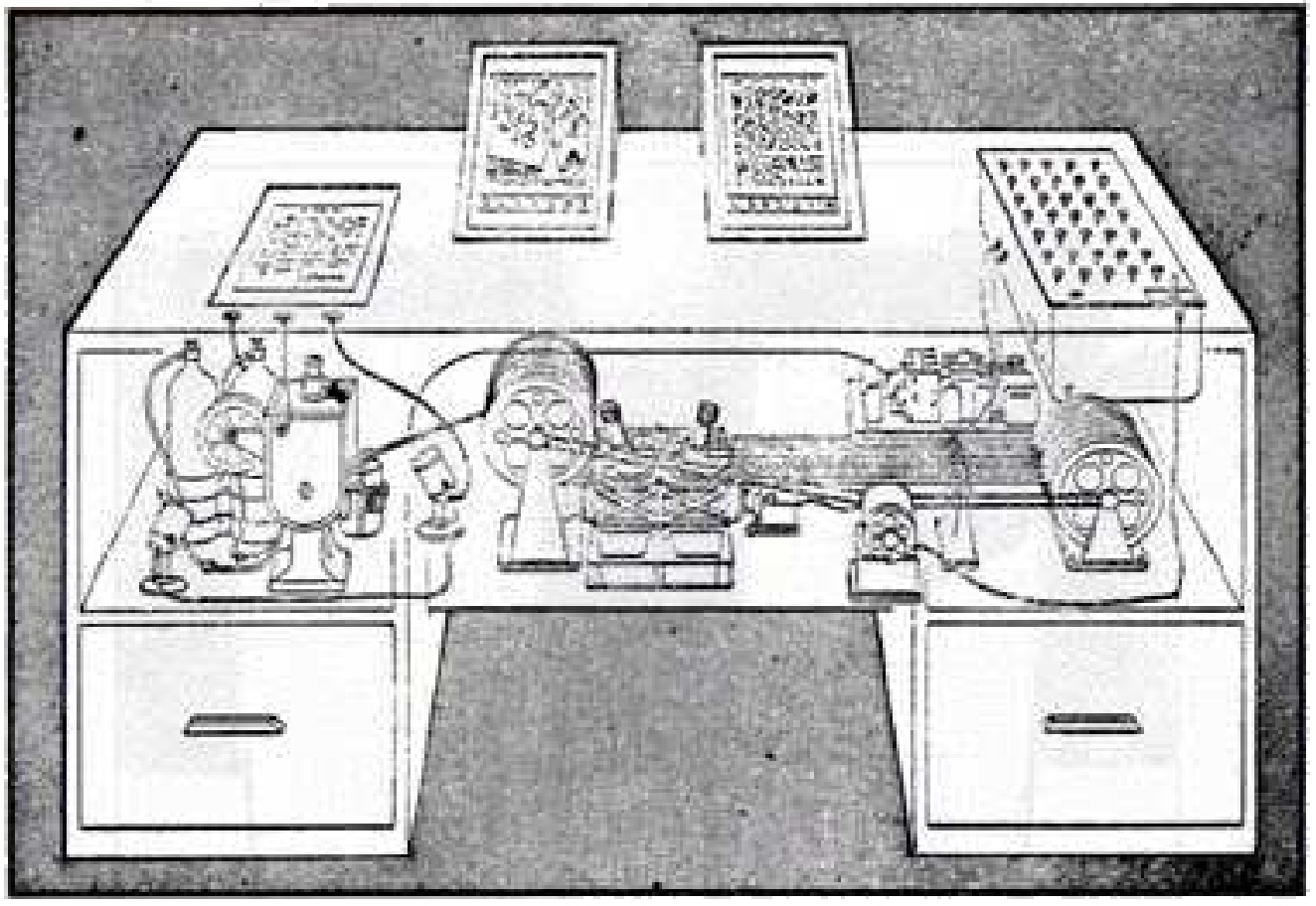
History

Otlet

- (1934) Otlet suggests plans for a global network of electric telescopes that would allow people to search and browse through millions of interlinked documents, images, audio and video files. He described how people would use the devices to send messages to one another, share files and even congregate in online social networks. He called the whole thing a “réseau”.
- Otlet described a networked world where “anyone in his armchair would be able to contemplate the whole of creation”.

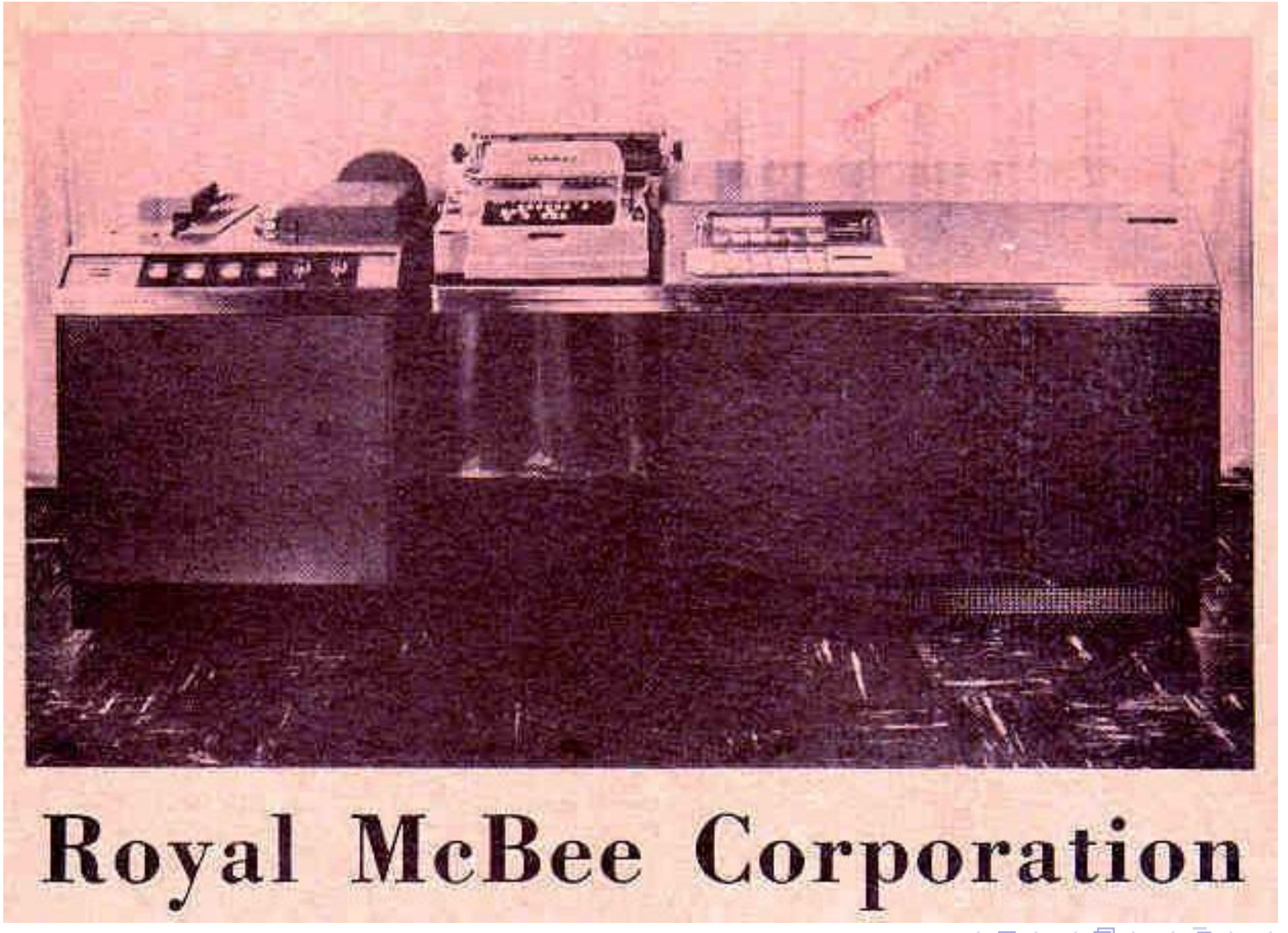
Around 1940 - Otlet's End

Vannevar Bush - Memex - 1945



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

Royal McBee LGP30 - 1958



History

WDML

- Late 1990's: initial vision
- 1998: WDML endorsed by the International Mathematical Union (IMU)
- 2001: IMU issues "Call to All Mathematicians to Make Publications Electronically Available"
- 2000's: large digitization projects
- 2006: IMU Report Digital Mathematics Library: "A Vision for the Future"

History

WDML

- 2010: European Digital Mathematics Library (EuDML)
- Digital Public Library of America launches with support of Sloan Foundation
- 2011: Alfred P. Sloan Foundation funds WDML workshop at NAS November, 2012
- 2013: NAS/NRC Report “The Mathematical Sciences in 2025”, January
- 2014: NAS/NRC Report “Developing a 21st Century Global Library for Mathematics Research” , March [Daubechies, Lynch]

GDML

- 2014: Seoul ICM Meeting, August
- Creation of GDML WG
- 2015: Recognized as WG of IMU CEIC:
  Committee on Electronic Communication and Information

GDML Working Group

Austria 1, Canada 1, France 1, Germany 2, USA 3

- Thierry Bouche (Université Joseph Fourier, Grenoble, France)
- Bruno Buchberger (Johannes Kepler Universität, Linz, Austria)
- Patrick Ion (AMS & UM, Ann Arbor MI, USA)
- Michael Kohlhase (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)
- Jim Pitman (University of California, Berkeley CA, USA)
- Olaf Teschke (zbMATH, Berlin, Germany)
- Stephen Watt (University of Waterloo, Waterloo ON, Canada)
- Eric Weisstein (Wolfram Research Inc, Champaign IL, USA)

GDML

Mission

To construct, as a global public good, an open knowledge base encompassing the results of the world's mathematics through collaborations deploying both present and new technology, and to foster a supporting community.

GDML

Goals

- To enhance openness and accessibility of all mathematical knowledge world-wide, present, past and future.
- To serve research mathematics, education and the scientific and technological use of mathematics.
- To be a resource for developing tools to promote use and development of mathematics.
- To facilitate creation, dissemination and archiving of semantically annotated mathematical material.
- To encourage the collaborative development of services based on semantic annotation.

GDML

Role

The GDML tries to achieve its goals by building collaborations. The effort involves the creation of standards and indications of best practices, encouraging the instantiation of such standards with content, and making such content openly available.

Issues

Categories

- Organization, Governance & Community
- Corpus & Collection
- Tools & Services
- Knowledge Management

Issues

Organization, Governance & Community

- *Chicken and egg*: International Mathematical Union, WG
- International — legal, communication: examples
  - HathiTrust, DPLA, JSTOR, COS, ...
- Mathematics as a Universal Language: math community

Issues

Corpus & Collection

- What must a GDML include?
- How big is our literature?
- What progress is there in digitization, both in quality and quantity?
- Where are digital math collections today, whether records of print, newer document types or software and data?
- What are copyright and other restrictions on our mathematical legacy?

Issues

Corpus & Collection

- Boundaries
  - Advanced Research Mathematics (mostly)
  - Applied Mathematics (the theoretical)
  - Any natural language (mostly English presently)
  - MR $\cup$ zb ?? [What is MR $\cap$ zb ?]
  - Legacy material versus broad present
- Ownership
  - Publishing is a business
  - Mathematics is a branch of knowledge
  - Mathematical facts are not patentable
  - Much publication metadata is public
  - Collections of such are not intrinsically held to be public
- Competition to be replaced by collaboration

Math Literature

- Formal literature
- Informal literature
- Research monographs
- Expository works (surveys, tutorials, user guides)
- Specialized collections for specific topics:

Where is Mathematical Knowledge?

- People
- Research Journals
- Textbooks and Monographs
- Informal literature
- Datasets
- Software
- Web sites

Math Knowledge

Variety

- Conjectures that turn out to be false.
- Proofs with flaws.
- Proofs sketches and analogies.
- Application of results where conditions are not verified.
- Approximations and probabilistic statements.
- “most of these terms are probably wrong, but a little inaccuracy sometimes saves tons of explanation” [H. H. Munro (“Saki”) 1870-1916]

Web

Article Repositories

- Publisher archives [Science Direct, SpringerLink . . .]
- JSTOR
- HathiTrust
- arXiv
- EuDML
- NUMDAM
- Gallica
- Göttinger Digitalisierungszentrum
- RusDML
- Private Lists

Web

Indexes, Reviews, Author Databases

- Inspec
- MathSciNet [Mathematical Reviews]
- zbMATH [Zentralblatt; Jahrbuch FM]
- Math Genealogy Project
- Google Scholar
- ResearchGate
- Wikipedia
- Mathworld

Web

Specialized Tools and Databases

- OEIS: Online Encyclopedia of Integer Sequences
- DLMF: : Digital Library of Mathematical Functions
- DRMF: Digital Repository of Mathematical Formulas
- DDMF: Dynamic Dictionary of Mathematical Functions
- Online Integral Calculator
- Inverse Symbolic Calculator
- Atlas of Finite Simple Groups
- LFMFDB: L-functions and modular forms database
- Combinatorial Statistic Finder
- A Catalogue of Lattices
- Encyclopedia of Triangle Centers

Web

Proof Libraries

- Mizar Mathematical Library
- Archive of Formal Proofs
- MetaMath Proof Explorer
- ...

Web

Software Systems

- Geogebra
- Pari/GP
- ProofWeb
- Wolfram|Alpha ...
- Maple
- Mathematica
- Flyspec
- Coq,
- ...
- Guide to Available Mathematical Software
- GitHub

Digitization

Knowledge from Documents

1. Assemble document collections
2. Capture page sets (born-digital vs. scanned)
3. OCR of text and formulas
4. Capture metadata / index and link
5. Semantic capture
6. Apply knowledge tools

Issues

Corpus & Collection

- Materials: just discussed
- Cataloging: metadata standards; EuDML
  - zbMATH, MathSciNet
  - EuDML, Beebe, other public aggregations
- Authority, Trust, Provenance: current standards?
- Reproducible Research Standard - V. Stodden
- Crowd sourcing
- Annotation and personal collecting: Mendeley, Bibsonomy

Issues

Tools & Services

- Multilingual: Unicode
- Formulas: MathML, OpenMath, TEX / LATEX
- Multiform: XML for description, or whatever's needed
- Listings; Annotation: lack of full support; W3C Annotation
- Data-mining: LDA; NLP+: MathWordNet
  Blei & Lafferty; Zanibbi & Giles
- Corpus structure: graph analysis & visualization;
  — simplicial complex homology; persistent homology

Issues

Knowledge Management

- Classification: MSC 2010 in SKOS (Linked Open Data)
  MSC 2020 Revision
- Ontology
- Semantic Intermediate Abstraction Language
  - between basic markup and formalization
  - previous attempts — flexiformality
  - Part of Math tagging
  - semantic search, . . .
- Previous attempts
  - Automath (1960), . . .
  - Maple, Mathematica
- Issues of proof
  - Computer Assisted
  - Four-Color, Kepler-Hales, Odd-Order, . . . Theorems
  - JVM and chip verification

Status

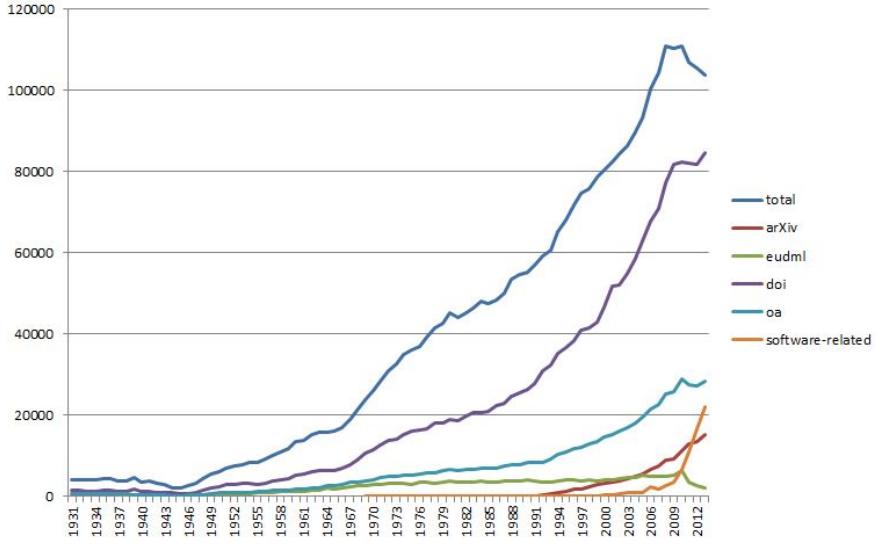
Content aggregation (I)

- Corpus estimation: Based on a large sample, MR $\cap$ zb $\sim$ 60% of zb / 66% of MR ($\pm$5% matching error)
- Total since 1868 about 190,000 books with average $\sim$350 p., about 3.9 million articles with average $\sim$14.4 pages
- Makes up about 120 million pages of mathematics, almost evenly distributed between books and articles (note: this ration has changed significantly through the years!)
- Note: Consistent with older estimates (70-100 million pages some years ago by Keith Dennis); older items are relatively small in numbers much require high digitization efforts (e. g., Göttingen bequests collection)

Status

Digitization

Various levels of digititation have been achieved:



- Scan/pdf ($\sim$ 80% of documents, $\sim$ 60% of pages)
- Open available pdf ($\sim$ 20% of documents, $\sim$ 10% of pages)
- Open available LaTeX, XML, MathML ready for content analysis, formula processing ... ($\sim$ 5% of documents, $\sim$ 2.5% of pages)

Status

Content aggregation (II)

Beyond literature, mathematical information is aggregated in increasingly diverse form

- Mathematical software: GAMS, swMATH, repositories . . .
- Research data collections: OEIS, DMLF, LMFDB, Manifold Atlas Project, Electronic Geometry Models, ATLAS of Finite Group Representations . . .
- Oral and visual mathematics (conference videos, collections of slides, visualizations . . .)
- Discussion/Collaboration platforms (MathOverflow, Polymath, Encyclopedia of Mathematics, Wikidata . . .)

Examples

techniques follow those in the previous paper. One uses a Buchstab decomposition of the sieve function, and attempts to estimate as much as possible asymptotically, using mean value estimates for products of Dirichlet polynomials. Sieve tricks such as `rôle reversals' produce some minor improvements, but the main advantage over the earlier work comes from the deployment of {\it N. Watt}'s mean value result [J. Number Theory 53, 179-210 (1995; Zbl 0837.11050)]. This is used in the form

$$\int_{-T}^{T}\Big|\sum_{m\le M}a_m m^{-1/2-it}\Big|^2\Big|\sum_{n\le N}n^{-1/2-it}\Big|^4\,dt \ll T^{1+\varepsilon}(1+M^2T^{-1/2})\max_m|a_m|^2,$$

valid for $N \ll T$ and any fixed $\varepsilon > 0$. Classically the factor $1 + M^2T^{-1/2}$ would have to be replaced by M. A great deal of notation is taken from the previous paper, and the exposition is extremely compact. Thus the reader should not attempt the present paper without a thorough understanding of the principles described in the original article. Reviewer: Roger Heath-Brown (Oxford)

MSC:

Cited in **6** Reviews
Cited in **86** Documents

11N05 Distribution of primes
11N36 Applications of sieve methods

MathOverflow questions:
Upper bound for $p_{n^2} - p_{(n-1)^2}$?

Keywords:
primes; differences; upper bound; mean value; Dirichlet polynomials

MATRIX | AI | CI | PDF | BibTeX | XML | Cite | Full Text: DOI | 87 6 MPG SUB

© 2017 FIZ Karlsruhe GmbH Privacy Policy Legal Notices Terms & Conditions W3C

By the results of

Baker, R.C.; Harman, G.; Pintz, J., **The difference between consecutive primes. II**, Proc. Lond. Math. Soc., III. Ser. 83, No.3, 532-562 (2001). ZBL1016.11037.

the number of primes in the interval $[x, x + x^{0.525}]$ is $\gg x^{0.525}/\log x$ for x large enough (see the final inequality of the paper); in particular, $p_{(n-1)^2} + p_{(n-1)^2}^{0.525} \geq p_{n^2}$ for n large enough. This gives a bound of the form $O(n^{1.05}\log^{0.525} n)$. Any improvement on this bound would likely improve the Baker-Harman-Pintz bound on large prime gaps, so I doubt one can do much better using what is in the literature. Of course, the situation is much better on RH or even on LH (the Lindelöf Hypothesis). (Also, one can show the expected bound of $O(n \log n)$ for almost all n, with a fairly small exceptional set, using the known results on primes in short intervals on average.)

share cite improve this answer

edited Aug 25 at 13:07 Todd Trimble ♦ 40.7k ● 5 ● 145 ● 238

answered Aug 11 at 23:56 Terry Tao 52.6k ● 16 ● 227 ● 318

add a comment

Using unconditional results cited by Dusart in this paper in their simple versions obtained by Robin and Salvy which are of the form

$$p_k \leq k\left(\log k + \log\log k - 1 + \frac{\log\log k - 2}{\log k}\right),$$

and

$$p_k \geq k\left(\log k + \log\log k - 1 + \frac{\log\log k - 2.1}{\log k}\right),$$

for n large enough (the first holds after $k \geq 688383$, the second after $k \geq 3$) if my arithmetic is correct we have the inequality

$$p_{n^2} - p_{(n-1)^2} \leq (1+\varepsilon)\,4n^2\log n$$

see an example newsletter

Related

11 Is there a known bound in prime gaps?
2 Effective upper bound on large prime gaps; or, what is the first prime after a googolplex?
2 Upper bound on Chen primes in an interval?
7 On a sum involving prime numbers
3 Error analysis needed for more refined estimates (than Šalát-Znám) of the sum of prime powers
14 Is there a known primitive recursive upper bound on the nth "Zhang prime"
5 Do small prime (bounded) gaps imply larger (but still bounded) gaps?
1 An estimation of p_n
1 Bound for $|p_n - \mathrm{li}^{-1}(n)|$
3 Best provable and unconditional lower and upper bounds for Brun's constant

Encyclopedia of Mathematics

Navigation: Main page | Pages A-Z | StatProb Collection | Recent changes | Current events | Random page | Help | Project talk | Tools | What links here | Related changes | Special pages | Printable version | Permanent link | Page information

Page | Discussion

Median algebra

2010 Mathematics Subject Classification: *Primary:* 08-XX [MSN] [ZBL]

A set with a ternary operation $\langle x, y, z\rangle$ satisfying a set of axioms which generalise the notion of median, or majority vote, as a Boolean function.

The axioms are

1. $\langle x, y, y\rangle = y$
2. $\langle x, y, z\rangle = \langle z, x, y\rangle$
3. $\langle x, y, z\rangle = \langle x, z, y\rangle$
4. $\langle\langle x, w, y\rangle, w, z\rangle = \langle x, w, \langle y, w, z\rangle\rangle$

The second and third axioms imply commutativity; it is possible (but not easy) to show that in the presence of the other three, axiom (3) is redundant. The fourth axiom implies associati the two

- $\langle x, y, y\rangle = y$
- $\langle u, v, \langle u, w, x\rangle\rangle = \langle u, x, \langle w, u, v\rangle\rangle$

also suffice.

In a Boolean algebra the median function $\langle x, y, z\rangle = (x \vee y) \wedge (y \vee z) \wedge (z \vee x)$ satisfies these axioms, so that every Boolean algebra is a median algebra.

Birkhoff and Kiss showed that a median algebra with elements 0 and 1 satisfying $\langle 0, x, 1\rangle = x$ is a distributive lattice.

References

- Garrett Birkhoff, S.A. Kiss, "A ternary operation in distributive lattices", *Bull. Amer. Math. Soc.* **53** (1947) pp. 749-752 Zbl 0031.25002
- John R. Isbell, "Median algebra", *Trans. Amer. Math. Soc.* **260** (1980) pp. 319-362 Zbl 0446.06007
- Donald E. Knuth, "Introduction to combinatorial algorithms and Boolean functions", ser. The Art of Computer Programming 4.0 (2008) pp. 64-74 ISBN 0-321-53496-4 Zbl 1178.68372

Article | Discussion

08-XX

08-XX General algebraic systems {Info & Stats}

- **08-00** General reference works (handbooks, dictionaries, bibliographies, etc.)
- **08-01** Instructional exposition (textbooks, tutorial papers, etc.)
- **08-02** Research exposition (monographs, survey articles)
- **08-03** Historical (must also be assigned at least one classification number from Section 01)
- **08-04** Explicit machine computation and programs (not the theory of computation or programming)
- **08-06** Proceedings, conferences, collections, etc.
- **08Axx** Algebraic structures [See also 03C05]
- **08Bxx** Varieties [See also 03C05]
- **08Cxx** Other classes of algebras

Main Page
MSC2010 Top
Recent changes
Toolbox
What links here
Related changes
Special pages
Printable version
Permanent link
Page information

GDML

2016

- JMM Special Session on Mathematical Information in the Digital Age of Science, Seattle Jan 9-11 2016
- Semantic Representation of Mathematical Knowledge Workshop, Fields Institute February 3–5 2016, with Wolfram Research as Sloan grant recipient
- Applied for and received Sloan grant to found an International Mathematical Knowledge Trust (IMKT)

GDML

2017

- Foundation of IMKT based in Waterloo ON, Canada [July]
  - Boards: Governance and Scientific Advisory
  - Work groups
- Short term: Outreach, seed projects, coordination
- Long term: Make available the “totality” of mathematical knowledge in digital form employing human- and machine-usable knowledge tools
- Initiatives
  - FAbstracts
  - Special Function Concordance
  - FHarmony
  - Document analysis: *n*-gram studies
- FAbstracts and FHarmony at Big Proof, Cambridge 10–14 July 2017

Initiative

FAbstracts

- FAbstracts means formal abstracts.
- Extract the main results from published mathematical papers into language both human and machine readable.
- Each mathematical term used should be defined in language both human and machine readable.
- Ultimately the statements and definitions should be so precise that they can be translated in a fully automated way into statements and definitions in a proof assistant.
- The language that is used for the FAbstracts should be so expressive that ordinary mathematicians should be able to understand entries.
- A start could be made on such a project by choosing a suitable area for which there is already a good basis of formalized results available.

Initiative

FHarmony

- FHarmony is a harmonization project for formal systems. This is related naturally to the previous necessity for communication between parallel efforts in FAbstracts.
- FHarmony is concerned with the technical aspects of constructing bridgework and crossovers between formal frameworks. For instance, how similar or different are different formalizations of the Jordan Curve Theorem, say. Perhaps this is a particularly good example because of the controversy over the history of its proof.
- HoTT libraries for Coq, HOL Light, Agda, Lean, ...

GDML

2018

- JMM Special Session on Mathematical Information in the Digital Age of Science, San Diego, Jan 9–11 2018
- ICM 2018, 1–9 August 2018, Panel on Digital Libraries

Future

- Organization, Governance & Community
  - Community building, Asian, US and European Trust entities,
  - Web presence and Wiki on the initiatives
- Collection Development
  - Collaboration with EuDML, arXiv and Euclid
  - Collaboration with Wikipedia, WikiData
  - Contact with potential Asian partners
- Tools & Services
  - Mathematical Object Identifiers (MOI)
  - Proposal toward open access book identification
  - Digitization Catalog & DML documents and wiki
  - IMU Proceedings
- Knowledge
  - Initiatives
  - Stacks Project?
  - Machine Learning results: Lafferty & Blei; Zanibbi & Giles
  - Collaboration with WRI

Our Mission

To construct, as a global public good, an open knowledge base encompassing the results of the world's mathematics through collaborations deploying both present and new technology, and to foster a supporting community.

Grand Challenge

To extract and mechanize the world's mathematical knowledge.

- Mathematical knowledge appears in the literature, data and code.
- Mathematical knowledge management tools become useful and necessary when dealing with large corpora.
- Collecting the entirety of published mathematics and applying MKM is within reach.