Progress in Mathematical Information and Knowledge Bases

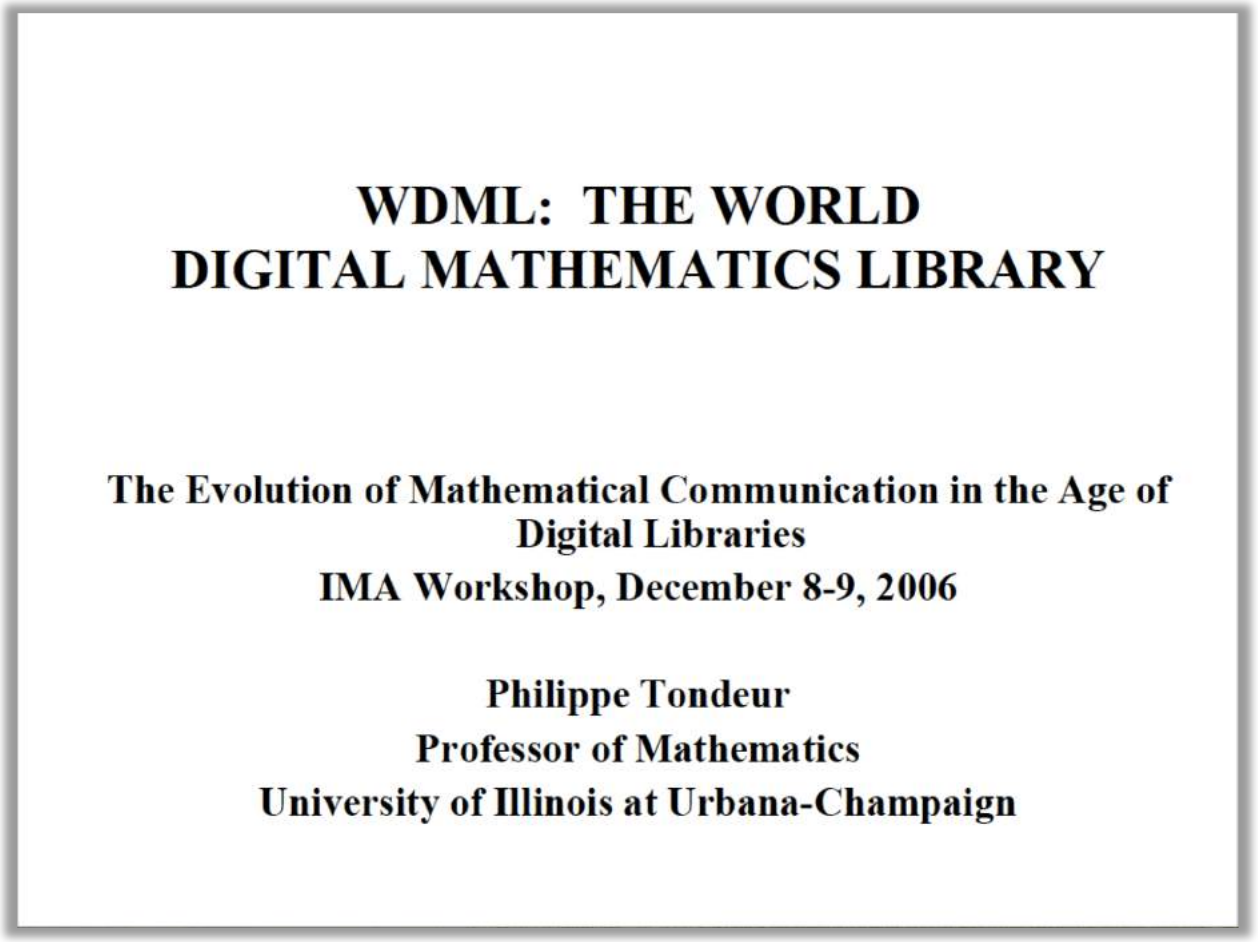
Stephen M. Watt
David R. Cheriton School of Computer Science

UNIVERSITY OF
WATERLOO

DML, MKM, UVP

- What we have heard in the session.
- Significant progress in MKM, so more ambitious goals.
- Must still satisfy basic needs.
- Current practical challenges + most significant opportunities.

13 Years Ago



WDML: THE WORLD DIGITAL MATHEMATICS LIBRARY

The Evolution of Mathematical Communication in the Age of Digital Libraries

IMA Workshop, December 8-9, 2006

Philippe Tondeur

Professor of Mathematics

University of Illinois at Urbana-Champaign

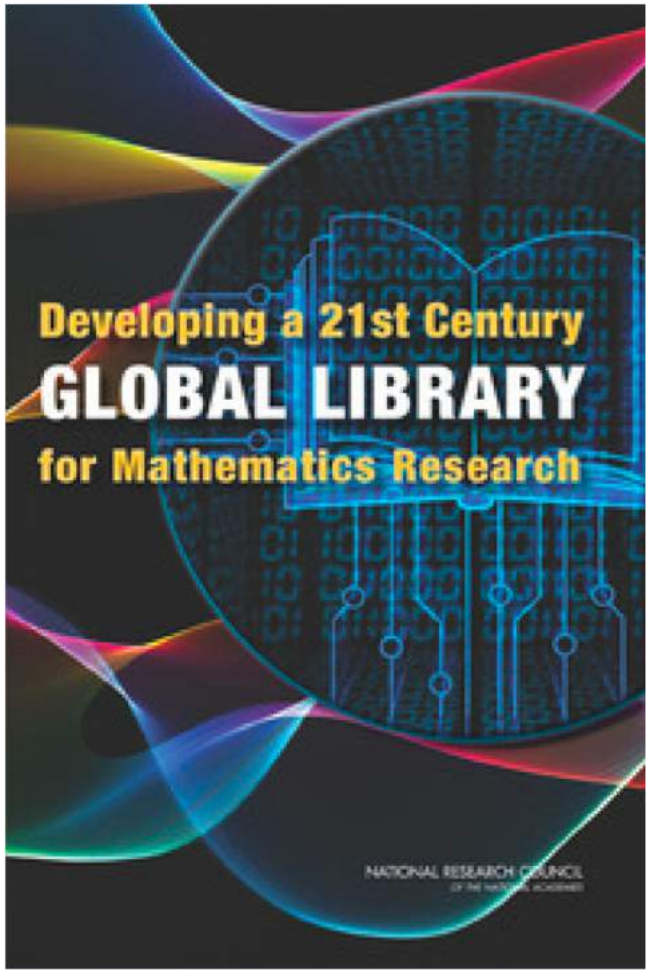
UNIVERSITY OF WATERLOO

Concept

- Digitize all past mathematical literature and link it to the present literature
- Initiators included:
  Bernd Wegner (ZentralblattMath),
  Keith Dennis (MathSciNet),
  Paul Ginsparg (arXiv), and others

UNIVERSITY OF WATERLOO

6 Years Ago



UNIVERSITY OF WATERLOO

SUMMARY 1

1 INTRODUCTION 8
Overview, 8
Study Definition and Scope and the Committee's Approach, 8
Structure of the Report, 11
Previous Digital Mathematics Library Efforts, 11
The Universe of Published Mathematical Information, 14
Conceptual Tools, 19
Current Mathematical Resources, 21
References, 26

2 POTENTIAL VALUE OF A DIGITAL MATHEMATICS LIBRARY 28
What Is Missing from the Mathematical Information Landscape?, 28
What Gaps Would the Digital Mathematics Library Fill?, 29
References, 53

3 ISSUES TO BE ADDRESSED 55
Developing Partnerships, 55
Engaging the Mathematics Community, 58
Managing Large Data Sets, 59
Open Access, 65
Maintenance, 67
References, 70

UNIVERSITY OF WATERLOO

Next steps towards a Global Mathematics Library

2014 x

Ingrid Daubechies ingrid@math.duke.edu via uwo.ca 7/23/14

to watt, m.kohlhase, sojka, Vladimir, pitman, bruno.buchberg., Peter.Paule, Jeb

Dear Colleague,

This in an invitation to attend a 2-hour meeting to discuss development of a Global Digital Mathematics Library (GDML). This meeting will be held during and at the same venue as the upcoming International Congress of Mathematicians (ICM) of the International Mathematical Union (IMU); more precisely, it will take place in the afternoon of August 17, 2014, in or near the COEX Conference Center in Seoul, S. Korea. Exact details will be communicated later, and will depend, in part, on the number of people who can attend in person or remotely, and the times convenient to them.

This proposed meeting is organized by the Committee on Electronic Information and Communication (CEIC) of the IMU. It is in response to the report by the National Academy of Sciences of the USA (available from http://arxiv.org/abs/1404.1905 or http://www.nap.edu/catalog.php?record_id=18619) -- and, in particular, the section "Constitution of the Digital Mathematics Library Organization" starting on page 80.

UNIVERSITY OF WATERLOO

The meeting's goal is to assess interest of organizations and individuals in contributing to this effort and to discuss setting up a small organization that would

* coordinate efforts of several existing DML efforts, some started after the IMU's call for a World Digital Mathematics Library (WDML) several years ago;
* encourage the development of tools and a platform that would add functionality to these DMLs beyond the simple availability of documents, and that would be useful to mathematical researchers worldwide;
* help build a consortium of institutions to support the GDML.

UNIVERSITY OF WATERLOO

- IMU Working Group established to
  - Develop a concrete roadmap
  - Incremental budget
  - Organize proposals to funders

- Members:
  Thierry Bouche (EuDML)
  Bruno Buchberger (Risc LINZ)
  Patrick Ion (Math Reviews (ret.), MathML)
  Michel Kohlhase (Jacobs, OpenMath)
  Jim Pitman (Berkeley)
  Olaf Teschke (ZbMath)
  Stephen Watt (ORCCA, MathML, doc analysis)
  Eric Weisstein (MathWorld, Wolfram Research)

UNIVERSITY OF WATERLOO

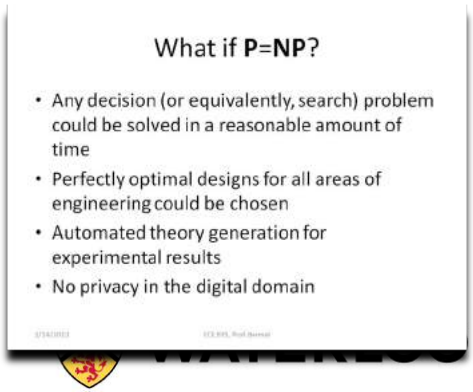
What is a Library?

- Different views:
  - A building in Alexandria
  - Collection of books and periodicals
  - Subscriptions to access digital books, journal back files and databases
  - Collection of pdf page images
  - A set of indices into recognized standard collections
  - Mechanisms to use collections to answer questions
  - A combination of the above

UNIVERSITY OF WATERLOO

Where is Mathematical Knowledge Today?

- People
- Research Journals
- Textbooks and Monographs
- Informal literature



Where is Math Knowledge Online?

Digital Library of Mathematical Functions

MathWorld™

WIKIPEDIA
The Free Encyclopedia

MathSciNet

LMFDB

zbMATH
the first resource for mathematics

StatisticFinder

Google Scholar

Follow

Paul Erdős
Mathematics
number theory, combinatorics, probability, set theory, mathematical analysis
No verified email - Homepage

Citation indices	All	Since 2011
Citations	64368	21713
h-index	109	59
i10-index	762	329

Co-authors View all...

Ralph Faudree, András Sárközy, Janos Pach, Laszlo Lovasz

Different Meanings for “Math Literature”

- Formal literature
- Informal literature
- Research monographs
- Expository work (surveys, texts, “user guides”)

- Specialized collections addressing specific topics
  - OEIS, DLMF, HOL Light libraries, eCF, functions.wolfram.com

UNIVERSITY OF WATERLOO

Different Meanings for Knowledge

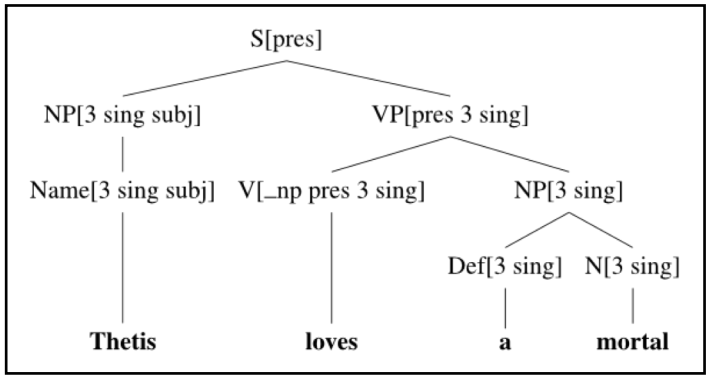
Not all useful mathematical knowledge is given by true statements.

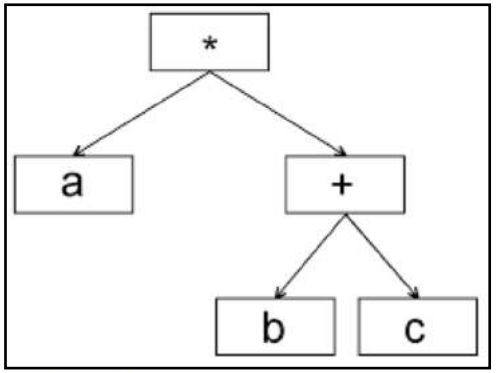
- Conjectures that turn out to be false.
- Proofs with flaws.
- Application of results where conditions are not verified.
- Approximations and probabilistic statements.

UNIVERSITY OF WATERLOO

Capturing Math

- Page images *or* born-digital pages
- Structure analysis (DAS)
- Natural language and math recognition
  - Identification of glyphs.
  - Parsing
- Semantic capture





UNIVERSITY OF WATERLOO

Different Meanings for "Semantic Capture"

- Library Science:
  Bibliographic reference graph, keywords, subject class'n
- Document Analysis:
  OCR / Doc structure recovery from born digital docs
  Word spotting, equation parsing
- Shallow capture:
  Natural language processing
- Deep capture (requires fill-in and correction):
  Formalized mathematics
  Capture of narrative
- Synthesis:
  Linking series of papers/volumes/areas of mathematics

UNIVERSITY OF WATERLOO

Different Meanings for "Semantic Capture"

- Capture exactly what was written, as written
  (level of formality, level of correctness)
  Useful parts of wrong, incomplete or superseded work
- Augment what was written
  correction, formalization
- Capture definitions only
- Capture defns + theorem statements only
- Capture defns + statements + proofs
- Capture defns + statements + proofs + meta statements

UNIVERSITY OF WATERLOO

Born Digital Documents

- Will be the majority

- Good TeX and Bad TeX (everywhere)

…$\{ n | (n+1)^2 $ isn't perfect $\}$ …

…$\{n | (n+2)^2 \text{isn't perfect} \}$…

- Use of semantic macros (e.g. DLMF)
  \BesselJ{\nu}@{z} vs J_\nu(z)

UNIVERSITY OF WATERLOO

Different Meanings for Basic Concepts

- 1391 Chaucer: "*equation*"
- 1557 Recorde: "="
- 1637 Descartes: "x"

1. $x^4 - a = 0$
2. $xa = ax$
3. $8 + 2 = 10$

A. 1,2,3 "equation"
B. 1 "équation", 2 "identité", 3"rélation"

- English, Latin A
- Chinese, French, German, Russian,… B

- 1 vs 2 undecidable

What is an Equation?

Solomon Marcus
Romanian Academy of Sciences
Bucharest, Romania

Stephen M. Watt
University of Western Ontario
London, Canada

Abstract—**The equation concept is one of the most fundamental in mathematics, yet the word "equation" means different things to different people. It is used not only with various precise technical meanings, but also as a metaphor for complex situations. We review some of the history of the equation and its present meaning and use in a variety of settings. Although some languages make distinctions among ideas such equations that are always satisfied versus those that may not be, or between equations with variables and relations without, we observe that it may not be possible to decide into which of these cases a particular equality falls.**

Keywords—equation; history of mathematics; mathematical terminology, mathematical translation

I. Equations Everywhere

There are few concepts in modern mathematics that are more fundamental or more used than that of the equation. The term is so basic that most users of mathematics, and indeed most mathematicians, seldom think about what it means. We often discuss equations without explicitly establishing the domain of discourse and we come to conclusions without ever agreeing what we are talking about.

This mathematical word now transcends its technical meaning and has taken hold in everyday language as a metaphor for complex relationships or as an appeal for an act of solution. Two years ago, U.S. President Barack Obama launched the slogan *Change the Equation*, in order to stimulate innovative programs in mathematics and in science. In mass-media, in advertising, the word *equation* is used as a symbol for some problematic situations, asking for the right answer.

Nobody could exclude Chancellor Merkel from the political equation of the next years; the astrological equation of personality; the equation of life and death; Federer's equation; the equation of a quiet life; the equation of success: work + risk + tenacity; money is included in the equation of happiness; Annan and Syria's insoluble equation; human equation; the project Equation has been launched at 29 November 2011; Equation – an integrated, independent Belgian communication agency; title of a film: *The love in equation; Afghanistan's role in the equation of big powers.*

These are only a few examples turned out by a cursory web search. They show something about the public perception of mathematics. Just because it is difficult for most people to understand mathematics, these people infer that it deserves to be admired and considered as a universal pattern. But at the same time and for the same reason, many people feel intimidated by mathematics and they are ready to expect from it much more than it can really done. On the other hand, the lack of understanding generates sometimes the opposite effect, the suspicion that mathematics manipulates people; a doubt emerges about the real utility of mathematics. This doubt is visible in the attitude of many governments, which claim from mathematics immediate practical applications, in contrast with its very nature: mathematics is a long distance enterprise; in most cases, its impact is not visible next day. But we can see today's impact of mathematics done some decades, centuries and millennia ago. The word *equation*, by its multiple aspects and by its huge metaphorical capacity, concentrates and articulates to a large extent several controversial situations, symptomatic for the ambiguous status of mathematics.

II. Equations as a Trap

Until two years ago, both of us were convinced that *equation* in English had the same meaning as *equation* in French, *Gleichung* in German and *ecuație* in Romanian. This belief stopped in September 2010, when Marcus was shocked by the statement at page xiv of the book *Where Mathematics Come From* [1], claiming that Euler's famous relation between e, π, and i is an equation, despite the fact that it includes no variable. How is it possible to have such a mistake in a famous book, opening new ways in connecting cognitive sciences and mathematics? This was Marcus' state of mind at that moment.

As it happened, this observation occurred in Timişoara during the 2010 International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. When Marcus expressed this wonder to Watt, his reply came as another shock: "I am sorry, in English any equality is an equation." Accross from us was a professor from Japan, who confirmed what Watt said.

Back in our home cities, we both began to pursue the question further. For *Larousse* and *Encyclopaedia Universalis* only equalities including one or several variables can be equations, while in English dictionaries any equality is an equation, as Marcus confirmed from Professor Sergiu Rudeanu. Watt took this relay further, interacting with people of different countries, bringing a more detailed picture of the situation. But the great surprise came when we approached the Latin

UNIVERSITY OF WATERLOO

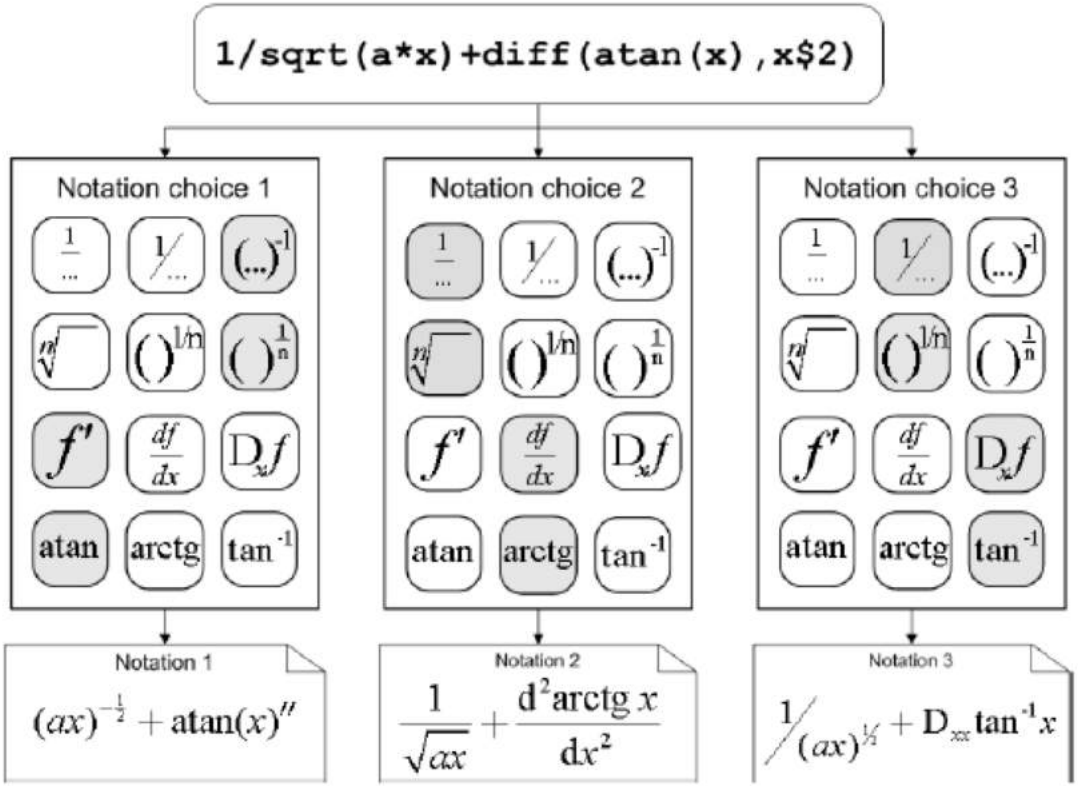
Different Things for the Same Notation

- J_ν
  - Bessel function
  - Angular momentum
  - 4-currrent

UNIVERSITY OF WATERLOO

Different Notations for the Same Thing

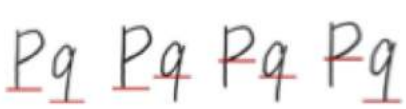
- TeX ⇔ MathML, Presentation ⇔ Content



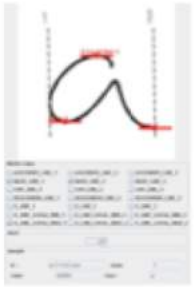
UNIVERSITY OF WATERLOO

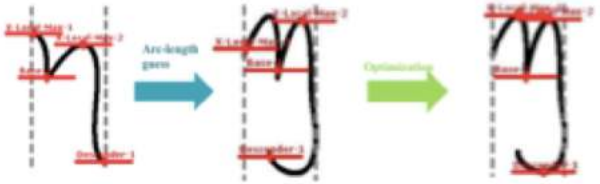
Disambiguation

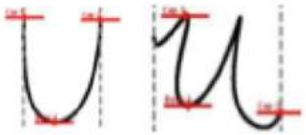
$$e^x = \int e^x dx = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$



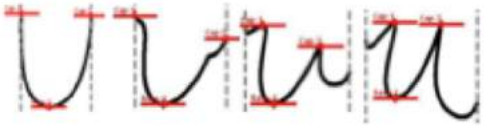
$$f(s) \approx \sum_{i=0}^{d} c_i P_i(s)$$

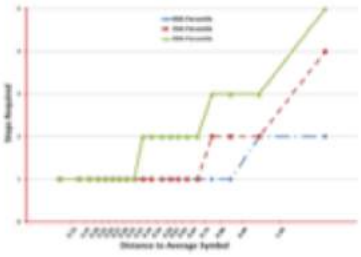






$$C(t) = (1-t)\bar{C} + tC_{targ}$$





ITY OF
RLOO

Different Levels of Formalism

- Exist in the literature
- “Flexi-formalism” (Kohlhase)
- The de Bruijn factor (Freek): ~4
- “A little inaccuracy saves a world of explanation.”

Clovis on the Alleged Romance of Business

1924

"IT is the fashion nowadays," said Clovis, "to talk about the romance of Business. There isn't such a thing.

. . .

It was the unledgered wanderer, the careless-hearted seafarer, the aimless outcast, who opened up new trade routes, tapped new markets, brought home samples or cargoes of new edibles and unknown condiments. It was they who brought the glamour and romance to the threshold of business life, where it was promptly reduced to pounds, shillings, and pence; invoiced, double-entried, quoted, written-off, and so forth; most of those terms are probably wrong, but a little inaccuracy sometimes saves tons of explanation.

H. H. Munro ("Saki")
1870-1916

UNIVERSITY OF WATERLOO

IMKT: Technical and Organizational Questions

- Technical:
  - How can the existing literature repositories be united?
  - What forms of semantic representation are most achievable and useful for mathematical knowledge?
  - How can mathematical OCR and natural language processing be used in a semi-supervised machine learning bootstrap process?

Initial Example Problems

- Concordance of the vocabularies for special functions
- FAbstracts
- Formal Harmony
- Document Analysis

UNIVERSITY OF WATERLOO

Initial Example Problems

- Collection of PDFs
- Concordance of the vocabularies for special functions
- FAbstracts
- Formal Harmony
- Document Analysis
- Machine learning for semantic capture

Initial Example Problems

- Collection of PDFs: EuDML, Euclid, Jstore + zbMath, Hathi
- Concordance of the vocabularies for special functions: TBD
- FAbstracts: Hales
- Formal Harmony: Dowek
- Document Analysis: DAS
- Machine learning for semantic capture: Blei and Lafferty

UNIVERSITY OF WATERLOO

MIDAS 2020

- Ingrid Daubechies: *Genesis of the current IMKT*
- Katja Berčič: *Math data*
- Mila Runnwerth: *Research infrastructure*
- Bruce Miller: *Writing mathematics*
- Steve Damelin: *Reading mathematics*
- Mitchel Keller: *Genealogy of mathematics*

UNIVERSITY OF WATERLOO

MIDAS 2020

- John Harrison: *Automated reasoning*
- Tom Hales: *Controlled natural language*
- Gilles Dowek: *Configurable logic for proof collections*
- Richard Zanibbi: *Math-aware search*
- Stephen Watt: *Where we are*

UNIVERSITY OF WATERLOO

MIDAS 2020

- Panel:
  - *Who are the clients for a GDML?*
  - *What is the data science of proofs?*
  - *What are the barriers to universal access to mathematics?*

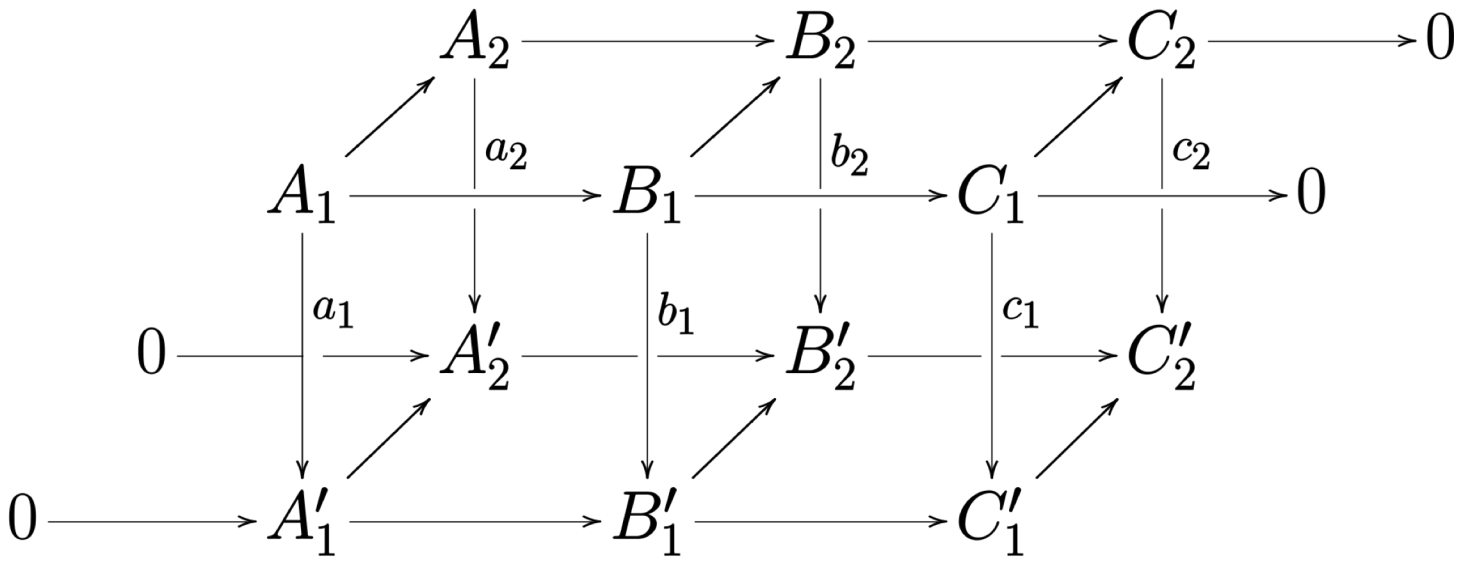
UNIVERSITY OF WATERLOO

DML, MKM, UVP

- Technical questions are “solved enough.”
- Coordination, ecosystem, sustainability.
- Users’ needs *vs/et* developers’ vision.

- Maple *vs* Axiom

“Math Is Not About Numbers”

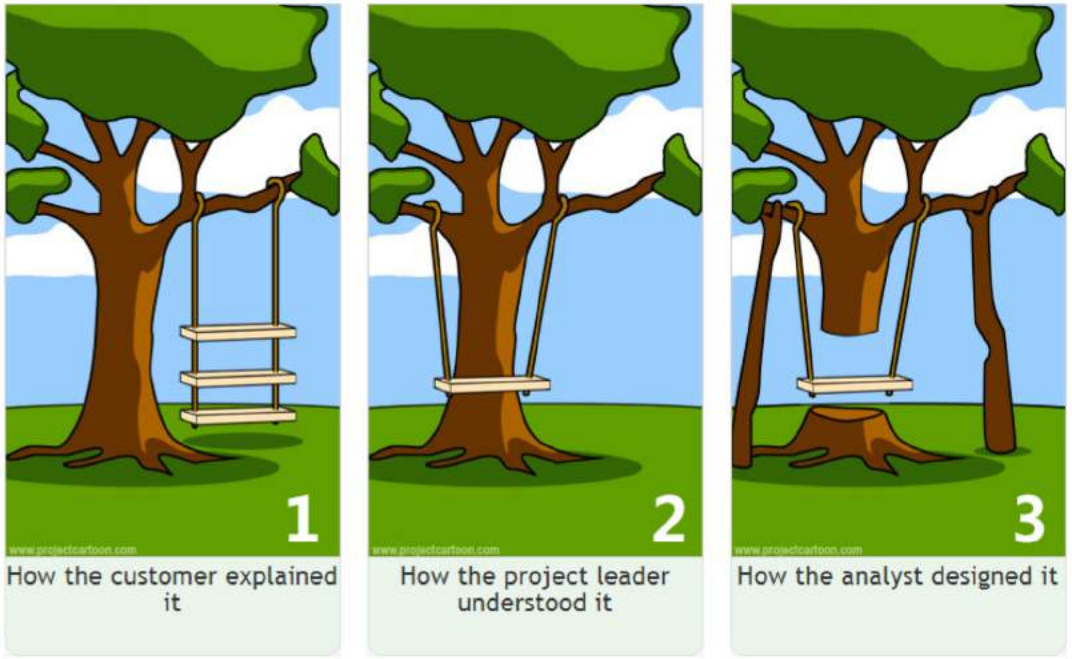


UNIVERSITY OF WATERLOO

But Numbers Matter

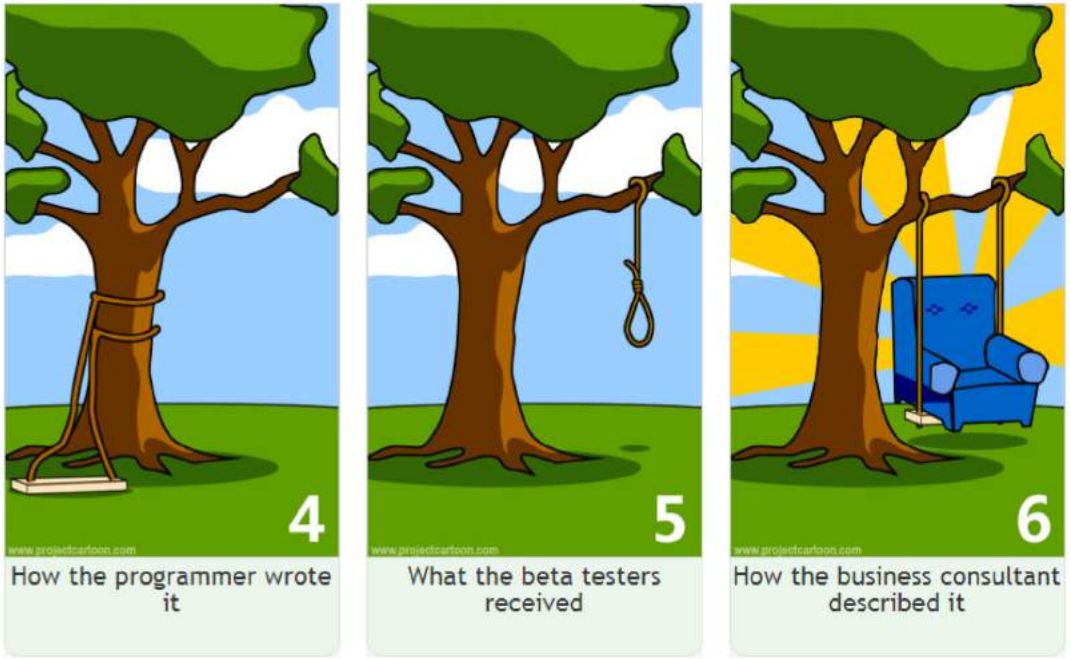
UNIVERSITY OF WATERLOO

A Cautionary Tale



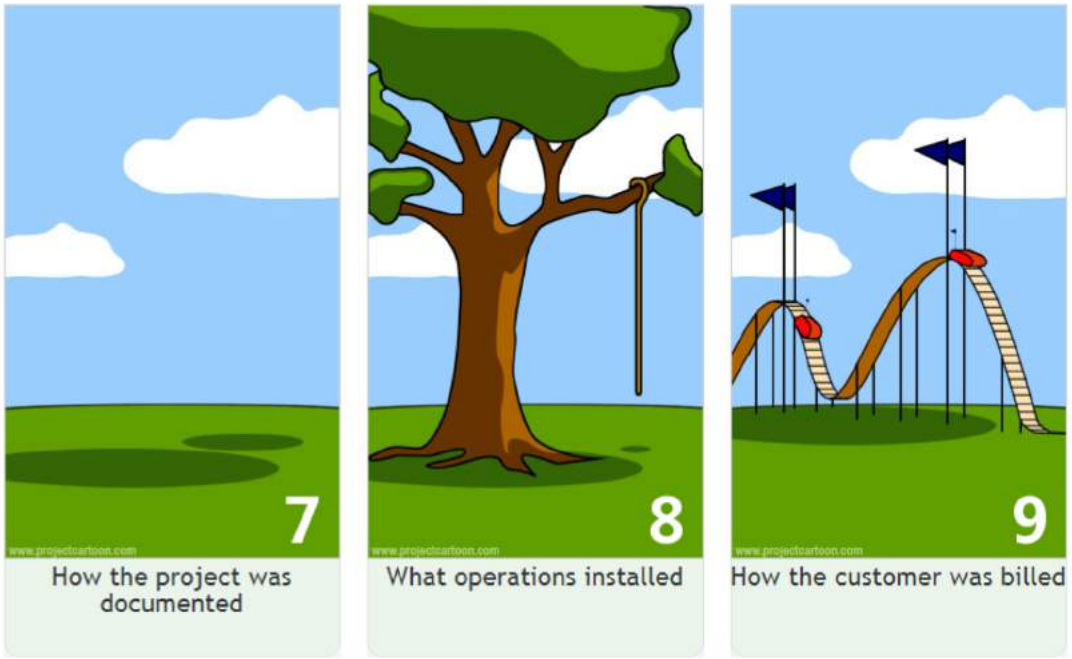
UNIVERSITY OF WATERLOO

A Cautionary Tale



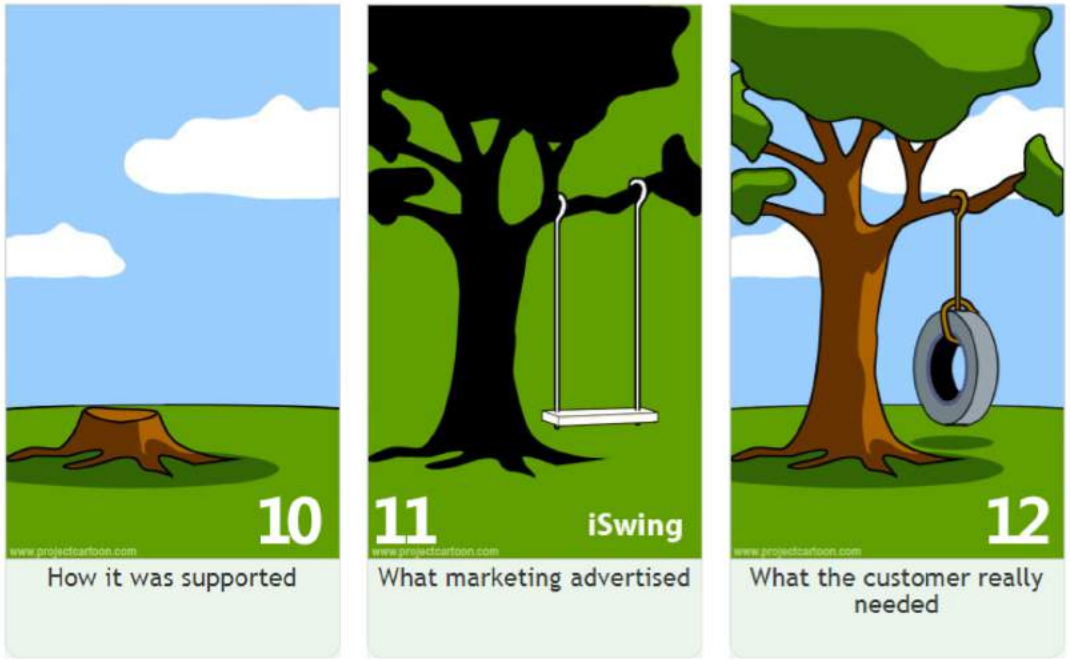
UNIVERSITY OF WATERLOO

A Cautionary Tale



UNIVERSITY OF WATERLOO

A Cautionary Tale



UNIVERSITY OF WATERLOO

IMKT: Technical and Organizational Questions

- Organizational:
  - How can we build upon existing research projects around the globe?
  - How can we most effectively engage relevant commercial enterprises, including publishers and software companies?
  - How can these efforts be brought to the public in a coherent and sustainable fashion?

DML, MKM, UVP

- The components we have seen in this session are a valuable tool set for mathematicians and users of mathematics.
- Open interfaces will allow these to work together and for new participants in an ecosystem.
- Providing these as a public service cannot rely solely on volunteer efforts.
- Business model is required.

Economic considerations

- If you build it, they will come.

  Field of Dreams (1989)

Economic considerations

- ~~If you build it, they will come.~~

  ~~*Field of Dreams (1989)*~~

- If you fund it, they will come.

Newton's Laws

1. An object at rest will remain at rest.

2. The velocity of an object changes when exposed to an external force. $F = m\,a$

UNIVERSITY OF WATERLOO

Economic considerations

- National funds for disposition of publicly funded research
  - Germany, Canada,…
- Philanthropic support.
- This can get the ball rolling.

Newton's Laws

1. An object at rest will remain at rest. An object in uniform motion will remain in uniform motion.

2. The velocity of an object changes when exposed to an external force. $F = m\,a$

UNIVERSITY OF WATERLOO

Business Models

- Continued operation requires sustainable funding model.
  - $2-$3 million USD/yr.
- Fragile to rely on philanthropy.
- “Public Good” model.
  - Long-term support contracts with national agencies.
- “Femium” model.
  - $3/month requires 100,000 subscriptions.

Newton's Laws

1. An object at rest will remain at rest. An object in uniform motion will remain in uniform motion.
2. The velocity of an object changes when exposed to an external force. $F = m\,a$
3. For every action, there is an equal and opposite reaction.

UNIVERSITY OF WATERLOO

Economic considerations

- Continued operation requires sustainable funding model.
  - $2-$3 million USD/yr.
- Fragile to rely on philanthropy.
- “Public Good” model.
  - Long-term support contracts with national agencies.
- “Femium” model.
  - $3/month requires 100,000 subscriptions.
- “Spotify” model.
  - Publishers as partners.

UNIVERSITY OF WATERLOO

Conclusions

- “Good enough” is the seed.
- Allow flexibility in precision.
- Machine interfaces nourish the ecosystem.
- Require acceptable financial model for bootstrapping and sustainability.