

MATH-AWARE SEARCH ENGINES:
PHYSICS APPLICATIONS AND OVERVIEW

Deanna Pineau

Aug 2016

University of Waterloo
Waterloo, ON

Mini-Thesis / Research Paper
In partial fulfillment of:
MSc Physics

Acknowledgments

This work is dedicated to my parents and to my partner in life.

I would also like to thank my friends and other family members for their support. For their effort (including reading this document), guidance and patience, I would like to thank my advisor, Brian McNamara (University of Waterloo), and my second reader, Erik Schnetter (Perimeter Institute). To all the students and professors whom I have met and learned with: thank you for knowledge, discussions, wisdom, and fun times. I would also like to thank the administrative staff for their efforts in making the clockwork tick.

Abstract

New search engines exist which take equations as their query, and return results which match either the query's mathematical meaning or its structural presentation. These results are drawn from scientific papers, online encyclopedias, and math discussion forums. Their scope includes some of physics, a lot of math, and some content from other natural sciences.

Development of math-aware search engines has progressed far in the decade or so that the field has been alive. The engines are able to parse some equivalencies of mathematics - such as commutativity - and to find similar (in addition to equivalent) formulae to the query. Their math interpretation abilities, excepting one encyclopedia-based engine (Wolfram Alpha), do not yet reach that of Computer Algebra Systems (CAS) like *Mathematica*. Developments are proceeding for journal-based engines to include the mathematical meaning-comparison features of CASs; hopefully, in time, the math rules encompassed by CASs will grow past their present limitations in this application as well.

Physics uses math to encode systems' evolutions. Sometimes, mathematical roadblocks are encountered in physics: a math problem is encountered for which no-one who knows about the problem also knows of a solution. Searching across disciplines, using equations, might help the researchers to locate a solution. The fact that jargon is avoided in the search phase allows this cross-disciplinary research to be efficient. Pre-made solutions are not the only thing which might be found: good approximations, worked-out work-arounds to using that equation, and analogous systems from which to improve understanding (or design experiments) might be found.

To bridge the gap between developers' publications on their search engines, and potential users' awareness of those same engines, this paper reviews their applicability to physics and presents the current math-aware search engines.

Contents

List of Tables	7
1 Introduction	8
1.1 What is math-aware search?	8
1.2 Ability to recognize and compare math	9
1.3 Database coverage	9
1.4 Applications to Physics	9
1.5 Paper layout	10
2 Physics Applications	11
2.1 Now Available (Although they improve with time)	12
2.1.1 Solve Math Problems	12
2.1.2 Design Models and Match them with Physical Systems	12
2.1.3 Find Analogous Systems for Experimentation	12
2.1.4 Pattern-find within physics	13
2.1.5 Find a specific format of an equation	13
2.1.6 Identify Special Functions	13
2.1.7 Look up partially-remembered theorems/laws	13
2.1.8 Look up the meaning of unfamiliar math notation	14
2.1.9 Preparing LaTeX Equations for Publication	14
2.2 Future Possibilities	15
2.2.1 Discover Alternate Interpretations via Different Equation Formats	15
2.2.2 Discover new solution methods via Different Equation Formats	15
2.2.3 Speed up computer programs' run-time	16
2.2.4 Custom "alerts" when new publications contain a specific equation	16
3 Extent of Engines' Abilities	17
3.1 Sequence of Events in an Engine's Operation	17
3.2 Database coverage	18
3.3 Extent of abilities to: extract math in-context	19
3.4 Extent of abilities to: match formulae	20
3.4.1 Mathematical meaning-based comparisons	20
3.4.2 Presentation-based (Structural) Comparisons	22
3.5 Future Developments	25
3.5.1 Field's identified directions	25

3.5.2	New tools proposed by developers	26
3.5.3	For Physics: Ideas for Useful Directions	26
4	Guide to Use	28
4.1	Choosing which Engine to Use	28
4.1.1	Presentation vs. Meaning-Based Search	28
4.1.2	Documents Searched: Field and Document Type	30
4.1.3	Query Features	30
4.2	Successful Queries	30
4.2.1	Making the query	31
4.2.2	Wildcards	32
4.2.3	Troubleshooting	32
4.3	Results	33
4.3.1	Easing the search, in results stage	33
4.3.2	Ranking Algorithms' Effects	34
4.3.3	Presentation Expectations	34
5	Available Math-Aware Search Engines	35
5.1	Engines Included: Working, Web-Available	35
5.2	Keeping Up-To-Date	35
5.3	Multi-Engine Interface	36
5.3.1	m_{in}	36
5.4	Web-Available Engines	36
5.4.1	LaTeXSearch	36
5.4.2	zbMATH	37
5.4.3	EuDML	38
5.4.4	NIST Digital Library of Mathematical Functions (DLMF)	38
5.4.5	Tangent	39
5.4.6	Uniquation	39
5.4.7	Wolfram Alpha	39
5.5	Ready for ArXiv deployment: Needs Server Space	40
5.6	In development for physics	40
5.6.1	$5e^{x+y}$	40
5.7	Build a Custom Equation-Searchable Database	40
5.7.1	Tangent	41
5.7.2	MathWebSearch (MWS)	41
5.7.3	Math Indexer and Searcher (MIaS)	42
6	Examples	45
6.1	m_{in}	45
6.2	LaTeXSearch	46
6.3	zbMATH	47
6.4	EuDML	48
6.5	DLMF	49
6.6	Tangent	50

6.7	Uniquation	51
6.8	Wolfram Alpha	52
7	Conclusion	53
A	Glossary	54
B	Nifty Math Tools	56
	Bibliography	59

List of Tables

5.1	Web-Available Search Engines for Math	43
5.2	Querying in Web-Available Engines	43
5.3	Downloadable Versions (& one Physics-Catered Upcoming Engine)	44
5.4	Querying in Downloadable Versions	44

Chapter 1

Introduction

What do the AdS/CFT Correspondence, dumb holes, and effective theories have in common?

All provide opportunities for better understanding and/or enable solutions to be found to troublesome math. The first two do that via linking together pairs of previously disparate phenomena, enabling calculation for quantum field theory from the former, experimentation upon black hole analogue systems from the latter, and insight into physics from both. [1–5] The latter - effective theories - builds applications and understanding through finding mathematical models which work for a certain setup, yet which do not have to connect with (and can, in fact, blatantly ignore certain parts of) all other known physics [6, 7].

There are new tools available which enable the finding of more discoveries such as these ones. They are math-aware search engines. Like keyword-based search engines, they let the user specify equations - and keywords - to be found in a database of documents. This gives researchers a new window into interdisciplinary search, by removing jargon and allowing direct searches in our highly-communicative language of math.

So far, their databases include journal articles, encyclopedias, and math forums. [8–17].

1.1 What is math-aware search?

The goal of these math-aware search engines is to retrieve mathematical formulae which are relevant to a queried formula. [18–21]

What is relevant is a subjective matter. [18] There are two camps at present: one which ranks equations’ relevance based on their structural similarities to the query equation(s) (called “presentation-based”), and another which retrieves equations which have the same mathematical meaning as does the query despite differing notation (called “meaning-based”). Each find not only equations which are exactly matching the query, but also those which share similarities with it. [19] For example, a retrieved result might contain only part of the query equation, or might append terms. [13, 18, 22]

The search engines are still developing, having been made primarily in the past decade or so. [18] They are well-enough developed to be useful, by providing a new window into our existing research. As an encyclopedia editor said: “we don’t even know how much we don’t know we know”. [23] These search engines help us find what we already know, by enabling us to search in one of our main languages: that of math.

1.2 Ability to recognize and compare math

Presentation-based search might match, for instance, $\log(\sqrt{z})$ and $\log(b)$. They allow variables' names to be changed, constants and variables to be exchanged, expressions to be substituted for variables, and algebraic terms to be added or removed. [13, 18, 22, 24, 25]

Meaning-based search, on the other hand, could ideally match $\sin(x)$ to $(1/2)i(e^{-iz} - e^{iz})$. [15, 22, 26] However, at present, search engines' knowledge of such math equivalencies vary [18] and all fall short of being able to match formulas based just on their meaning, ignoring their notation. [8, 22] A clear delimitation of what they can recognize would require a full understanding of their mathematics parsing algorithms, and is beyond the scope of this paper. However, a few limitations and examples are explicitly stated in the search engines' publications, and are included in this paper (Chapter 3). One simple example of what they can parse is commutativity: ab equals ba . Beyond that simple relation, the two web-available meaning-based search engines vary in their abilities. One can parse all mathematical equivalencies which *Mathematica* can (Wolfram Alpha) [14], which is still limited when trying to recognize special functions (such as that $T_n(z) = \frac{\delta_{n,0}}{2} + \frac{n}{2} \sum_{k=1}^{\lfloor (n/2) \rfloor} \frac{(-1)^k (n-k-1)! (2z)^{n-2k}}{k! (n-2k)!} + 2^{n-1} z^n$; $n \in \mathbb{N}$ [27] and $T_n(x) = \cos(n \cos^{-1}(x))$ [28] are the same function). Another (MathWebSearch) allows associativity, lets variables' names vary using wildcards, and generalizes some functions to $f(x)$ notation. [8, 29, 30] However, this engine does not parse all math to a standardized, mathematically-equivalent form: a search for $\lambda x(1-x+1-1)$ retrieves different results, for example, than does a search for $\lambda x(1-x)$. Math-aware search developers are investigating improving their mathematical comparison abilities with Computer Algebra Systems (such as *Mathematica*): the CAS could be used to simplify all database and queried expressions to a standard notation, using rules of math and definitions of functions, so that equations with different writings but the same meaning could be much better found. [22, 26]

1.3 Database coverage

The databases covered include physics and math papers [9–11], special function compendiums [12, 13], encyclopedias [14, 15], and math forums [17]. Not all physics publications are included, yet, but one math-aware search architecture is stated to be ready for full deployment on the arXiv.org content. [31] Some of Springer's articles are covered [11, 32], and many European math and mathematical physics journals are math-searchable [10].

1.4 Applications to Physics

It has been said, prior to the start of engines' development (roughly a decade ago), that it is easier to re-derive a medium-difficulty result in math than it is to find it in the literature. [23] The author commented that the same could probably be said for applied mathematics disciplines, of which physics is one. There is much lying around in the literature which each researcher knows little or nothing about. To find out what we already know, find connections between systems, discover solutions, and target searches and collaborations, the math-aware search engines fill a long-empty niche. Interdisciplinary and intra-field information finding

becomes simpler, because researchers don't have to know the field's jargon to find information: we can search using formulae, instead, to get a starting point.

1.5 Paper layout

Applications to the research of physics, from these search engines, are numerous - some proposed ones are included in Chapter 2. Readers who just want to use the engines can skip Chapter 3, which details the engines' math comparison abilities, limitations, and future developments. Successful searches and choice of engine are among the subjects overviewed in Chapter 4, which is a quick "Guide to Use". Following that is a usage-oriented summary of each math-aware search engine which is presently web-hosted, in Chapter 5. That chapter also summarizes the web-available engines' downloadable versions (when applicable), from which custom databases can be created. Furthermore, it presents a CERN project which is underway, to create math-aware search which is catered to physics' particular needs. In chapter 6, screenshots of searches with each search engine are included.

Good reference tables, for which math-aware search engines have each feature, how to query with them, and where to find them, are tables 5.1 and 5.2.

Finally, a glossary is located in Appendix A, followed by a partial list of other useful math tools which this author encountered during this research (Appendix B).

Chapter 2

Physics Applications

Equations encapsulate a crucial part of the knowledge in physics ¹. When applied to physical systems, they enable prediction and description of phenomena.

Finding information by searching for an equation based on its mathematical meaning is powerful. It can enable new solutions to physics problems by exposing math solution methods, alternative mathematical models, and work-arounds to using a certain equation. It can deepen our understanding, and provide new information for development of theories, by allowing researchers to more easily find connections or patterns, analogue systems, and alternative perspectives about a certain model or system.

Additionally, finding information based on equations' structural appearance enables users to look up equations of a specific form - which can carry special significance to a certain field - and to find new connections, solutions, and similar models or systems to those which the researcher is familiar with.

These math-aware search systems are an aid to collaboration, as well as a method of individual search across and within disciplines. They provide a way for each researcher to better locate information within our known mathematics. With their aid, we can enhance our interaction with the equations of our observations, models, and established theories.

This section is divided into those applications which are possible now, and those which require some significant developments. Most of the applications herein are the ideas of this author, except when noted otherwise, although it is questionable whether some of them are *implied* in other publications which are cited elsewhere in this work. Ideas included here are intended to give physicists a start of an idea on how these math-aware search engines can be used in everyday research.

¹Although some researchers think we could eventually do physics without using any math, as discussed in submissions (including winning entries) for the 2015 FQXi essay contest, "Trick or Truth: the Mysterious Connection Between Physics and Mathematics". [33]

2.1 Now Available (Although they improve with time)

2.1.1 Solve Math Problems

Many times in physics, we encounter equations which impede progress. The researchers who work on them can not find the solutions to them. These equations might have solutions (e.g. approximations, work-arounds, or more useful - mathematically equivalent - ways of expressing them) which are known to certain fields of math, or which have been used elsewhere in physics or the other sciences. Yet, finding them has so far been prevented by the use of jargon and the inefficiency of identifying collaborators who know something about the solution. Math-aware search engines help a researcher to identify systems, fields, and collaborators which would enable the system of interest to be better understood and/or applied. Good examples of this inter-disciplinary approach benefiting physics include the AdS/CFT correspondence and holographic duality theories. To expand upon this point, it has been said (in math) that “we don’t even know how much we know that we don’t know we know” [23]: finding what is applicable out of what we know is a good approach to make progress.

The development of emergent theories, fundamental physics, effective theories, applied physics, experiments, and phenomenology can all benefit from such improvements.

2.1.2 Design Models and Match them with Physical Systems

Finding mathematical models which are useful for a case can be done by first coming up with a model which is reasonable but hard to solve, then researching ways to make it more tractable. The first model which one comes up with need not, then, be the best model: searching the literature can locate refinements which make it more solvable. This might include avoiding the equation in question by changing the modeling setup, rewriting the equation into an equivalent but manageable format, or making approximations which other researchers have found to work well. Context will need to be considered, of course, but having this start at discovering better models is priceless. Keyword-based search engines can help with this, but they are inherently limited by varying terminologies between fields (i.e. jargon) and the need to know what kind of systems one wants to find. These equation-based search engines enable researchers to find mathematical models using the most appropriate language: that of math. They already know some of the “synonyms” of math (such as $a + b = b + a$), and will only improve with time.

2.1.3 Find Analogous Systems for Experimentation

Other times, we find it useful to experiment with analogous systems to ones which we can not directly observe. For instance, an investigation occurred into Black Holes and their particle production via an analogous system: dumb holes. [3] These sound-based analogues of black holes were experimented upon to better understand and predict black holes’ behaviours. [4,5] Another set of experiments has been pursued in order to explore additional fundamental predictions of quantum field theory, such as the Sauter-Schwinger effect; [34] discusses them in a colloquium.

2.1.4 Pattern-find within physics

Another application of finding analogous systems is to explore what physical theories might connect systems which are found to be analogous. For instance, one might find patterns for merging gravitation with quantum physics, by searching physics publications for quantum or gravity equations. While this application is limited, so far (because the math-aware search engines are not yet advanced enough to identify mathematically-equivalent notation), it still provides an excellent start.

2.1.5 Find a specific format of an equation

The form in which an equation is written can have significance. For example, writing the Schrodinger equation as a probability current indicates that it might be under study as a fluid flow. While this particular format has a name (“probability current”), not all significant formats of equations have a name. These un-named equations can be difficult to look up - as can some of the named ones - without knowing the jargon of the field. Math-language search engines let researchers get around the jargon, during the search phase.

2.1.6 Identify Special Functions

If a researcher can identify that a function of interest is a special function, then new resources become available with which to solve it.

However, this application is not well-developed yet. Although databases of special functions exist, their math-aware search engines do not recognize all possible ways of writing special functions. For instance, the definitions of the special function “Chebyshev polynomial of the first kind” are different on each of Wolfram Alpha, DLMF, and Wolfram Functions (from which the “primary definition” was taken); none of the special function databases’ search engines (i.e. Wolfram Alpha and DLMF) recognize the definitions which came from one of the other three websites. Yet, it can work to identify special functions via this route, given present capabilities: both Wolfram Alpha and the DLMF recognized the complementary error function’s ($\operatorname{erfc}(x)$) definitions which were drawn from the same three websites. These experiments with special functions are described in slightly more detail in section 3.4.

For now, using a search engine of this type to identify special functions is limited: it can provide a positive identification, but can not tell a researcher that a given function is *not* a special function. As math-aware search engines get better at translating between different math notations, this application will be more complete.

2.1.7 Look up partially-remembered theorems/laws

Long-ago learned equations might be remembered for their application, but the formula and name can be difficult to fully bring to mind. If enough of an formula is remembered, it can be looked up in a math-aware search engine to (hopefully) find the full version of that formula. [8]

As an example (adapted from [8]), suppose that a researcher remembers an integral inequality exists for expressions of the form $\int_D |f(x)g(x)|dx$. However, the researcher does not remember its name or full equation. The inequality which they half-remember is Holder’s inequality:

$$\int_D |f(x)g(x)|dx \leq \left(\int_D |f(x)|^p dx \right)^{\frac{1}{p}} \left(\int_D |g(x)|^q dx \right)^{\frac{1}{q}} \quad (2.1)$$

The researcher might try a keyword search for “integral inequality absolute value”, or “Integral |f(x)g(x)| approximate”, or similar phrases. However, for these search terms the first page of Google returns nothing which is obviously related to Holder’s inequality. Using math-aware search engines can return much more relevant results. For instance, the search engines included herein which are powered by MathWebSearch [8] - that is, zbMATH [35] or MathWebSearch’s online demo [36] - can locate Holder’s inequality given the input $\int_D |f(x)g(x)|dx$.

2.1.8 Look up the meaning of unfamiliar math notation

Especially when doing interdisciplinary research, it is easy for a researcher to come across notation which he or she is unfamiliar with. The context is not always explained adequately, in the text, for that researcher to easily find the notation’s meaning: field’s conventions and “common knowledge” are often not stated in publications. [37, 38] Math-aware search engines which use equations’ structure (i.e. “Presentation-based”) to search for results can be used to find more information about the unfamiliar notation. For example, the search engine “Tangent” proposes this application as a use of its software for both students and researchers [39]; for students, they give the example that the binomial coefficient’s notation $\binom{4}{2}$ could be drawn in m_{in} (section 5.3.1) and subsequently searched for [16, 40]. For researchers, an application might be to look up why $\pi(2) = 1$, which applies when discussing the “prime counting function”. [40]

2.1.9 Preparing LaTeX Equations for Publication

For complex or highly-formatted LaTeX equations, it can be useful to either copy and paste an existing equation from another researcher’s publication, or generate it via a drawing or template editor.

Copy and paste is now possible for a wider range of publications. It has been possible on some arXiv.org publications for some time: users can download the source code for a publication, which includes the LaTeX file [41]; this is somewhat tedious, but can be worth it. A simpler method applies, now, to some of the journal Springer’s publications. Using Springer’s LaTeXSearch website, researchers can bring up all equations in a paper, by searching for its DOI or title. Results will include an image and LaTeX code for each equation. Finding LaTeX code in this way, for direct copying, is one of the main purposes which drove LaTeXSearch to be made [42].

To generate LaTeX code for an equation, one can also draw it on the computer (such as in m_{in} [43]) or use a template editor such as those described in section 4.2.1.

2.2 Future Possibilities

2.2.1 Discover Alternate Interpretations via Different Equation Formats

Rewriting an equation can change the way a person understands it. For example the Schrodinger wave equation can seem less (or perhaps more) mysterious when it is written as a fluid flow in probability:

$$-\frac{\hbar}{2im}\nabla \cdot (\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) = \frac{d}{dt}(\Psi^*\Psi) \quad (2.2)$$

(This particular form of the Schrodinger equation is one whose structure arises in any theory within which a conservation law applies to an extensive quantity (such as mass, charge, or heat energy). However, let us pretend for a moment that we don't know that connection, yet, and are looking for ways in which certain branches of physics relate to one another - such as in quantum gravity.)

A math-aware search engine could find connections like this, wherein a standard equation is rewritten to a format which matches a different branch of physics (here, fluid mechanics). This discovery of how an equation fits into the current system of physics can seed new depth to understanding.

As math-aware search engines are able to recognize more of math's equivalencies, this application becomes better. However, a researcher using the present-day math-aware search engines would have a very difficult time finding these connections using them.

2.2.2 Discover new solution methods via Different Equation Formats

Writing an equation in a new way can also elucidate paths towards solving it. A simple example of this in everyday research is the use of substitutions in integrals, or integration by parts. Finding other researchers' writings of an equation gives the researcher an opportunity to see what other forms of writing other researchers might have found useful: it gives them direction towards finding a potential solution.

However, this application is only possible when the mathematical meaning of formulas can be recognized by math-aware search engines, to an extent which lets them match equivalent formulas which are written in different ways. Their present level of meaning-extraction is probably far too low to enable this use, although their in-development upgrade to parse meaning using *Mathematica*-type software will likely make this feasible.

Improvements to this application are possible with certain feature developments in the math-aware search engine. For instance, if the results could be *ranked according to solvability* - that is, whether an equation can and/or has been solved analytically (or numerically, in an efficient manner) - then researchers could find ways of writing their equation such that solutions fall out easily. This goes beyond *Mathematica*, since even *Mathematica* (and similar software) sometimes needs equations to be written in different representations in order to solve them. Furthermore, *Mathematica* can not change how the mathematical model is attached to the physical world (e.g. is x the linear distance from a point, or the distance

along a curved wire?), which a researcher who views equation results could do. When a researcher searches existing models, the searcher can also see what assumptions were useful to those models, and decide whether they can apply to the system at hand; *Mathematica* requires many such assumptions to be included as input to the equation. Finding information in the literature can therefore be of benefit beyond what *Mathematica*-like systems provide, for solving equations. If these search engines could determine whether a given equation format is solvable, or read that data from the document which includes it, then finding solvable mathematical models (or just equations) in the literature becomes easier.

2.2.3 Speed up computer programs' run-time

Some analytic equations run faster than others, either due to the number of steps required to compute them or to the run-time engine's specific architecture. As a simple example, rewriting

$$f(x) = \cos(a) \cos(b) \cos(c) - \sin(a) \sin(b) \cos(c) - \sin(a) \cos(b) \sin(c) - \cos(a) \sin(b) \sin(c) \quad (2.3)$$

to

$$f(x) = \cos(a + b + c) \quad (2.4)$$

will cause less computations to be performed when the code runs. Although the gain is small when solving this equation once, if it must be solved millions or billions of time during run-time then the small changes will add up.

Math-aware search engines can help to find *useful* re-writings of an equation, which result in faster run-time: there are many ways to rewrite equations, but the format which another researcher used may have been chosen because that researcher found it to work well for the application. Other researchers might have already put in the work to sort through some of the less efficient writings. Equivalent mathematical notation might be found in this way, in addition to approximations (which lose precision when used).

Using math-aware search engines for this task works best if the math-aware search engine "knows" a lot of mathematical equivalencies. Then, notations of the equation which are not obvious from its input format could be found. It is possible to do this application with current math-aware search engines, but it would be very limited due to their meaning-based comparisons limitations.

2.2.4 Custom "alerts" when new publications contain a specific equation

Routing of results - that is, updating users when new documents are available which includes their equation query - is one proposed use of math-aware search engines, by [44]. When journal articles comprise the database, this would mean that researchers can set up "alerts" to be sent to them when a new article with a certain equation is published (and is indexed into the search system's database).

Chapter 3

Extent of Engines' Abilities

Building a math-aware search engine is no easy task, but it has been accomplished to some degree. Over the past thirteen years [18], these search engines have gained the ability to interpret some basic math, to use equations' structure and appearance to find results, and to retrieve both exact and simply similar matches to a queried equation. They can take keywords alongside multiple equations in a single query, and include (named and un-named) wildcards. Their database coverage is impressive, and is expected to continue growing.

In the future, their developments are expected to include better extraction of math meaning from equations and their surrounding text [8,45], an improved ability to match equations using their mathematical meaning [22,26], a better understanding of what constitutes "similar" equations [19], improved database coverage [8,16,31,46–48], and a focus on user needs and interactions [45]. Some other neat tools - such as email alerts for new publications which contain a given equation - are also imagined for the future [44].

This section discusses the operational steps, current abilities, and future developments for these math-aware search engines. Beginning with an overview of the engines' generic operation and their database coverage, it then delves into the extent of each engine's abilities to extract and compare the mathematical meaning of equations. This information complements the usage-oriented presentation of each engine which is found in Chapter 5, by sketching the engines' limitations and rules for retrieving "similar" or "equivalent" equations. At the end of this chapter, future developments of this field are discussed, and a few physics-oriented developments are proposed.

3.1 Sequence of Events in an Engine's Operation

Both the query (input) equation and those located in the database's documents are processed in order to match equations in a math-aware way. This is done for presentation-based and meaning-based search.

Via various algorithms, variables' names and constants' values are either preserved or ignored. Preserving their values allows exact matches to the query to be found. Allowing one or both of them to be ignored broadens the search results while preserving the mathematical content of the equation. (although sometimes physical meaning is lost when constants' values are ignored). [16,26,49] To account for physical experiments' constants having varying

precision, one search engine (in-development) performs iterative rounding on each query’s constants. [26]

Interpreting operators and variable types requires some context about what the equation represents. For example, the variable \mathbf{R} might represent a matrix, a vector, or the real number line. A variable might be real, complex, or of another type. Math-search engines often extract context from an equation’s parent document, in order to decipher what the symbols represent. [26,45] However, extracting more context is an area of ongoing development. [8,45]

Equivalence of mathematical expressions must then be determined, between the query expression and those in the database. Currently, algorithms either base their comparison on the structural similarities between expressions, or the mathematical meaning behind the equations (regardless of their notational representation). [24] The former can be called “presentation-based” search (or “syntactic” search), and the latter “meaning-based search” (or, more properly, “semantic” search). [22,24] Meaning-extraction algorithms still contain errors [24], however, and meaning-based comparisons of equations are not yet perfect [8,45]; developments are underway to improve this [8,24].

An additional element which can be considered is the complexity of the equation, so that equations with similar complexity to the query are put nearer to the top of the results. [49]

3.2 Database coverage

Academic papers (such as journal articles), a curated encyclopedia, Wikipedia, and mathematics forums are all accessible via math-aware search engines. [9–11, 14, 17, 50]

The papers include many from physics, a large number from math, and some from other applied mathematics sciences. Their sources include the journal Springer (in LaTeXSearch), and the numerous journals which are accessed by the search services EuDML and zbMATH. [9–11] A portion of the arXiv is searchable by equation, but it is only for demonstration purposes; however, a developer has stated his team’s engine (MathWebSearch) to be ready to handle the entire arXiv. [31]

Encyclopedic and discussion-based sources cover some of physics, and a good portion of math. A curated encyclopedia is maintained by Wolfram Alpha, which brings up information about some physics topics when their appropriate equation is searched for. [14, 15] Wikipedia’s mathematical pages have been partially indexed by Tangent, although their content does not update when Wikipedia does [50]; hopefully this coverage extends soon. The math forums covered include planetmath.org, mathoverflow.net, and math.stackexchange.com, with searching provided by Uniquation. [17]

Scalability is still a challenge which many of the math-aware search engines have to overcome. [8, 16] However, the range of material covered by some of the engines so far is impressive. In particular, there are over 3 million publications in zbMATH [47], and over 250,000 on EuDML [10]; all of them are searchable using equations. The over 1.1 million arXiv.org pre-prints are also stated to be serviceable by MathWebSearch, provided that server space is provided for the search engine. [31] The developers of these search engines want to provide math-aware search over additional scientific resources (e.g. [31, 46–48]), including servicing databases on the order of current keyword-based search engines’ sizes [16].

3.3 Extent of abilities to: extract math in-context

Search engines which act upon journal articles and other web-based sources need to find and extract mathematical equations from their database documents. Because math is intended to be read by humans, rather than by computers, [37] this requires some algorithmic design [26, 38, 45].

There is a lot of symbol recognition which occurs before mathematical context is considered. For example, there are many ways to encode equations in LaTeX: a bold symbol could be encoded using “`\mathbf{symbol}`”, “`\pmb{symbol}`”, or a few other commands. [51] Individual symbols can also be written in many different ways; the Angstrom symbol in unicode can be written as a “latin capital letter A with ring above” (code 0xc5) or “Angstrom sign” (code 0x212b). [52] The math-aware search engines account for the different encodings of given symbols, but they do have troubles when those encodings are non-standard (for instance, see the example below concerning “`\dot`”).

To interpret math, these search engines pull context from the text and apply field-specific conventions whenever possible. For example, the text might specify that λ is a binding function, or that it is a variable [38]; or, f might be either a functional operator or a variable (such as frequency). An integral ($\int f(x)dx$) could be Riemann, Lebesgue, or any of the other dozen or so types of anti-derivation. [20] Vectors, matrices, and other symbols might also be defined in the text, which can resolve ambiguities such as whether \mathbf{R} is a matrix, vector, or the real number line. Extracting meaning from documents is difficult without this context; algorithms for context-extraction are known to improve the relevance of results from math-aware search engines. [45]

Examples of presently problematic notation include subscripts, incorrect notation, and conventions which are specific to geographical regions. For instance, the expression x_1^2 might indicate the square of the variable x_1 ($(x_1)^2$), or might be the first entry in a sequence of x^2 terms ($(x^2)_1$). [38] Notation which is encoded in a human-readable but formally incorrect manner is also problematic. The times operator is a good example of this: it can be properly encoded as “`\dot`”, but can also be written as “.”. The latter appears as a decimal point, which is problematic when it appears between numbers. For example, the the historical formula

$$\sqrt{2} = 1 + \frac{1}{3} + \frac{1}{34} - \frac{1}{3434} \tag{3.1}$$

is here encoded using “`\dot`”, but when the encoding uses “.”, then the last denominator is interpreted as 3.4.34 by MIA.S. Aside from this being a senseless number - three point four point thirty-four - it in no way matches the intention of the formula. [25] A good example of a geographic difficulty with conventions is the notation for the binomial coefficient: the Russian convention is C_n^k , which is the opposite of the C_k^n French standard. [53]

To summarize, it is difficult to parse the human-oriented mathematics equations properly. A lot of progress has been made, and developments continue. [24, 25, 54] Search engines which use context to extract meaning from equations’ surrounding text have been noted to retrieve more relevant results. [45].

3.4 Extent of abilities to: match formulae

These search engines take a queried equation, and find equivalent and similar equations to it. The main algorithms used, in this task, are presentation-based search and meaning-based search. The applied uses of each are overviewed in section 4.1.1. Within these two categories, the abilities of the available math-aware search engines vary considerably.

In general, most of the meaning-based search engines can understand commutativity ($a + b = b + a$), allow variables' names and constants' values to change, and allow variations in the equations' structure (i.e. adding or dropping terms, allowing entire expressions to take the place of a variable, etc.). The grand majority of them can not do associativity ($ab = ba$) or distributivity ($a + (b + c) = (a + b) + c$), except by allowing variables' names to vary and terms to be re-arranged. More complex math equivalencies are generally not included; however, progress is being made to incorporate *Mathematica*-like Computer Algebra Systems into the engines' algorithms, so that all the mathematical rules, simplifications, and other tools of *Mathematica* can be used in identifying which equations are equivalent, mathematically. [22, 26]

In this section, the abilities of each meaning-based and presentation-based search engine is reviewed. The meaning-based engines' discussion focuses on the extent to which their algorithms are able to recognize mathematically equivalent expressions which differ visually. In contrast, the presentation-based engines' discourse considers which criteria the engines define to say which equations are "structurally similar".

3.4.1 Mathematical meaning-based comparisons

Meaning-based search algorithms attempt to find equations which math the mathematical meaning of the searched-for ("queried") equation. [8, 22, 24] Each uses a different approach, and all fall short of being able to fully identify mathematical equivalencies given different writings of the math (including Wolfram Alpha). Nevertheless, they are able to identify mathematical equivalencies to varying degrees.

Wolfram Alpha

Wolfram Alpha, which both searches its internal encyclopedia and computes results, uses all the mathematical rules and simplifications which *Mathematica* knows. [14, 15] As such, it is very powerful. However, it still can not recognize all formats of a special function. To experiment upon this, the definition of the Chebyshev polynomial of the first kind (hereafter Cheb-I) was found on the Wolfram Functions website to be

$$T_n(z) = \frac{\delta_{n,0}}{2} + \frac{n}{2} \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{(-1)^k (n-k-1)! (2z)^{n-2k}}{k!(n-2k)!} + 2^{n-1} z^n; n \in \mathbb{N} \quad (3.2)$$

(this was the first "primary definition" on Wolfram Functions). [27] Yet, when this equation was entered into Wolfram Alpha's search query - using a simple copy-paste of the *Mathematica* text from Wolfram Functions, with a check for errors - Wolfram Alpha failed to parse it. Various formats of the equation were tried, and eventually Wolfram Alpha recognized

the formula - but, it did not identify it as the Cheb-I polynomial. Wolfram Alpha has its own definition for Cheb-I, which is found by searching by keyword for the function:

$$T_n(x) = \cos(n \cos^{-1}(x)) \quad (3.3)$$

[28]. When this equation is entered into the Wolfram Alpha search box, it identifies that it is Cheb-I. On the other hand, another experiment which was performed with the same method revealed Wolfram Alpha to correctly identify two other websites' definitions of the complementary error function ($\operatorname{erfc}(z)$). In that experiment, definitions were pulled from the Wolfram Functions website,

$$\operatorname{erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \sum_0^{\infty} \frac{(-1)^k z^{2k+1}}{k!(2k+1)} \quad (3.4)$$

[55] and the DLMF ¹,

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-t^2} dt = 1 - \operatorname{erf}(z) \quad (3.5)$$

[57]. Wolfram Alpha defined the $\operatorname{erfc}(z)$ to be

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) \quad (3.6)$$

[58]. When each of the DLMF and Wolfram Functions $\operatorname{erfc}(z)$ definitions were input to Wolfram Alpha, the results identified that the equation was the complementary error function.

MathWebSearch

MathWebSearch is the only working math-aware search engine for journal articles which uses meaning-based search algorithms. Determining the limits of its abilities, as with Wolfram Alpha, would require a full understanding of its algorithms, which is beyond the scope of this paper. However, many of its features and their limits are explicitly stated in the publications which discuss it.

Variable names can either be preserved (for “non-query” variables) or ignored (for “query”, or “wildcard” variables). [29] This means that a user can enforce commutativity to hold (although it may be encoded into MathWebSearch already - the literature does not mention commutativity) simply by writing their equation using query variables. It supports the use of “named” query variables, so that a user can specify both that certain variables should have the same name (as in $ab + a$) and that the actual variable used is irrelevant (so $xy + x$ would also match) [59, 60]. Based upon experimentation with zbMATH [35], query variables in MWS are allowed to be replaced with subscripted variables, non-subscript

¹The DLMF was not included for the Cheb-I experiment because it defined Cheb-I in terms of a few other functions, which were defined across many of its webpages, in such a way that no formula for Cheb-I was actually written in its database. [56]

variables, or numbers: $(?x)^2 + (?y)^2$ can retrieve $a^2 + b^2$, $x_1^2 + x_2^2$, and $8^1 + 1^2$. Furthermore, when non-query variables are used, the variable names are not allowed to vary at all ².

Another basic math function, associativity, is also included in MathWebSearch [30].

To better recognize math operators' changeable names, MathWebSearch also generalizes elementary functions such as $\sin(x)$ to generic operators like $f(x)$. [8] When searching for an equation, it also lets only part of the searched equation match the results, or only part of the result to match the searched-for equation. [8] These two items combine to make it good for retrieving half-forgotten formulas and finding approximations. For example, consider Holder's inequality:

$$\int_D |f(x)g(x)|dx \leq \left(\int_D |f(x)|^p dx \right)^{\frac{1}{p}} \left(\int_D |g(x)|^q dx \right)^{\frac{1}{q}} \quad (3.7)$$

Searching for $\int_D |f(x)g(x)|dx$ could bring up equation 3.7, if it was in the database. [8] In addition, a search for $\int_R |\sin(x)\cos(x)|dx$ could also find equation 3.7. [8] However, searching for $\int_R |\sin(x)\cos(2x)|dx$ will not identify equation 3.7 as a match: the algorithms of MWS interpret the x and $2x$ to be the arguments of the function, causing it to parse this equation's generalization as $\int_D |f(x)g(y)|dx$ and not $\int_D |f(x)h(x)|dx$. [8]

3.4.2 Presentation-based (Structural) Comparisons

These engines find equations which are structurally similar to the query. The results retrieved might differ from the query by changed variable names; added, dropped, or re-arranged terms; subexpressions substituted for variables; or other visual-based criteria. [13, 21, 22]

DLMF

Searching its own database of special functions, the DLMF allows queried equations to bring up results whose structure differs slightly from that of the input equation. For example, it will add terms which it thinks appropriate: \sin^2 will be appended with and (x) to make it $\sin^2(x)$. [61] If a query returns no exact matches, then the DLMF's engine allows the symbols in the equation to be re-arranged (thus approximating commutativity and associativity) and to be separated by other math terms; this is a form of "query relaxation". [13, 25]

Unique to DLMF is a "metadata" search feature. It allows a search to contain keywords for certain branches of math (e.g. "trigonometric"), such that those keywords represent any function from that category (e.g. $\sin(x)$, $\cos(x)$, etc.). The keyword does not have to appear on the search results page.

For example, a user could search for "trigonometric^2 + trigonometric^2", and retrieve identities such as $\sin^2(x) + \cos^2(x) = 1$. Due to the DLMF's query relaxation, such a search

²To test whether non-query variables were allowed to vary, the expression " $c^2 + d^2$ " was searched on zbMATH's formula-oriented search page, [35]. Only two pages of results were found, all of which used exactly the formula $c^2 + d^2$. However, searching for the query variable form of this equation, " $(?c)^2 + (?d)^2$ ", brought up 497 results, which included those using variables other than "c" and "d"; examples included $x_1^2 + x_2^2$, $x^2 + y^2$, and $8^2 + 1^2$.

also returns $\sec^2(z) = 1 + \tan^2(z)$, into which terms were inserted between the “trigonometric²” terms; $\sin(2z) = (2 \tan(z))/(1 + \tan^2(z))$, in which one “trigonometric²” term was dropped; and many other functions.

Tangent

Visual aspects of equations are highly emphasized in Tangent’s search over Wikipedia articles. [21,62] For instance, $1 + x^2$ is considered to be different from $x^2 + 1$, although each may come up in the results of the other as one of the “similar” equations. [63] The engine matches equations based on the relative positions, and on the types, of the symbols therein. [21] Wildcards are allowed in the web-available version (which is not the latest development), but each wildcard can only represent a single symbol. [62]

Based on experimentation with the online version, equations which are considered “similar” to one another can have interchanged variables and constants, differently named variables, constants of differing values, and terms added or removed. [63] For example, searching for “ $g(z) = 0$ ” returns results which include the following: $g(z) = 0$, $g(E) = 0$, $g(0) = 0$, and $h(z) = 0$. [16,63] The latter three of these results all share the same “score” - that is, they are considered to have the same degree of similarity to the input equation. [63]

The web-available demo version of Tangent is an old version of the software (v0.1). Developers have now released version 0.3.1, which is available for download and use; that version was used for a recent competition in math-aware information retrieval. [50,64] Its upgrades include better discovery of “similar” equations, broader wildcard support (including allowing a single wildcard to represent a subexpression, instead of just one symbol), keywords alongside equations in a query, and scalability. [21,62,64,65] It allows subexpressions to be substituted for other subexpressions, regardless of their mathematical meaning or size - for example, $\sin(a + b)$ has the subexpression $(a + b)$, which can be substituted for any other expression inside parenthesis when retrieving results. [64] Notationally, this updated engine can also query matrices, prefix subscripts (e.g. ${}_nC$), and prefix superscripts (e.g. nC). [65] Possible upcoming features indicated by the developers include incorporating textual context into equation matching [62], improving keyword-vs-equation weighting when ranking results [64], and adjusting the weight which is given to matching variables (versus allowing them to vary) when retrieving and ranking results [64].

Math Indexer and Searcher (MIaS)

The developers of MIaS give a fair bit of detail about what their algorithms can do. Their engine - currently used to search over journal articles - uses presentation-based search but is moving towards incorporating meaning-based algorithms. [22]

When finding similar equations, the engine allows constants’ values and variables’ names to change. [22] However, it distinguishes between variables with and without subscripts (e.g. x_1 vs x), and does not allow them to be interchanged. For example, the query $x_1 + y_1$ would *not* retrieve $x + y$. [22] This treatment of subscripts can be problematic for physics applications, because subscripts are often used to denote which physical characteristics a variable is attached to (such as x_i for the position at a certain time, or d_s for the distance to Saturn); these subscripts have no mathematical significance, except to distinguish one variable from

another. Another equation characteristic which is allowed to vary is that subexpressions in the query are searched for: for example, the query $a + b^{c+2}$ has the subexpressions b^{c+2} and $c + 2$. [22] Furthermore, the structure of equations is recently able to be considered in matching formulae: $x + y/z$ would structurally match $a + \sqrt{b}/c$, having the form “item1 + item2 / item3”. Finally, the mathematical equivalency of commutativity is accounted for, so that $a + 5$ is considered equal to $5 + a$. [22]

Some problematic queries, for MIaS, are elucidated in [25]. These were queries which were sent with knowledge of what should be retrieved, based on a human assessment, and which the algorithm failed at. Some situations where the query did not succeed included: (a) a query used “=”, whereas the target equation used “ \equiv ”, (b) the query wrote text in math-mode using “ $\text{\text{}}$ ”, whereas the target did so with “ $\text{\rm{}}$ ”, (c) the target equation was broken up into multiple lines, whereas the input was on one line. [25]

Developments are underway to expand MIaS’s ability to identify similar equations and design algorithms which understand a much broader range of math. [22,66] Query relaxation - that is, allowing a query to have terms added, dropped, or re-arranged - has been experimented with and is being optimized. [22,66] Another under-experimentation development is operator grouping: the addition and subtraction operators can be classified as one “type” of operator, and allowed to be substituted for one another freely; for example, this would let $x + y$ “exactly” match $x - y$, $-x + y$, and $-x - y$. [66] Wildcard inclusion is also planned to be investigated [22], and machine learning algorithms will be looked at for use in ranking and disambiguating equations [66].

To include drastically more math understanding in the engine’s procedures, the developers also plan to investigate the inclusion of *Mathematica*-like systems (a.k.a. “Computer Algebra Systems”) in their mathematics parsing stages. [22] Their successful incorporation would transform MIaS into an engine which can translate math - and find matching and similar equations based on that - on par with Wolfram Alpha. This approach has already been partially implemented in an under-development search engine (geared towards physicists), as is discussed in section 5.6.1. [26]

LaTeXSearch

Little information is available about LaTeXSearch’s algorithms, for its search over some of Springer’s journal articles. [42] Independent experiments by this author and [22] indicated that it performs presentation-based searching, retrieving both exact matches and similar formulae to a queried equation.

Uniquation

Very little of Uniquation’s workings is published, and that only on Twitter and the Uniquation web-interface. Its search over math forums appears to make use of presentation-based search, and finds both similar equations and exact matches. [17] Basic abstract algebra is supported, such as $\mathbb{Z}[t][\sqrt{t^2 - 1}]$. [67] Modular arithmetic is also accessible, including $\sum_{n=1}^m n^m \equiv 1 \pmod{p}$. [67] (These two search queries are presented in section 6.7 as screenshots of Uniquation’s search results. Their LaTeX encodings are also included therein.)

3.5 Future Developments

Developers in the field of math-language search engines have identified a few directions for their development, and a few neat ways in which to expand upon their abilities.

Ideas of this author’s, which are specific to physics applications, are also included for discussion and/or development.

3.5.1 Field’s identified directions

Algorithms for Search

Field experts have indicated that more of the search engines need to focus on extracting mathematical meaning (including context and symbols’ meanings) from documents. [45] Algorithms with which to extract theorems from publications, and to restrict searches to formulas which are consistent with a query’s mathematical assumptions, are also desired. [8] It is being explored to allow the user to define which parts of a formula should be kept fixed, when searching for results, and which are allowed to bring up “similar” results (see “simto” regions, [68]). There is also work being done to improve the identification of equations which share the same mathematical meaning, but are written differently. [22,26] Precision of results and recall speeds are other areas of improvement; the recall speed is only a few seconds at most, but that is still slower than standard text-based search engines. [18] Unbiased evaluation metrics are also recommended, to allow rigorous comparison of the math-aware search engines. [18] Finally, scalability is a target area of advancement, with the eventual goal of covering the entire web. [16]

Database expansion

Future databases are planned to include the arXiv [31,69], the digital math library DML-CZ [46], and other added journal articles [46,47]. The DLMF is also exploring ways to improve or enlarge their special function content (for which they are soliciting community feedback - see “news” section of [12]). Formalized math libraries (e.g. Mizar) and theorem proving libraries (e.g. Thousands of Problems for Theorem Proving) are also having math-aware search added to them. [8]

Less formal media which are in-development for math-aware search include lecture videos [70], Wikipedia [31,69], and additional math forum websites (such as the question-and-answer forum Math Overflow) [71]. In addition, an idea has been put forward to use math-aware search on input code for systems such as *Mathematica*, MatLab, and R. [8]

Eventually, coverage of all mathematical science material is desired [48], with database sizes expected to rival present-day keyword-based search websites [16].

User experience

The user-experience is another focus in development. This includes re-visiting the subjective rules for which results are presented at the top of the results list, as well as the appearance of the results. [45] Other considerations include ease of use and learning curve. [45]

3.5.2 New tools proposed by developers

Developers have proposed a few tools which use the math-aware search engines, or similar algorithms.

Alerts for new publications containing a specific equation

Routing of results - that is, updating users when new documents are available which includes their equation query - is one proposed use of math-aware search engines, by [44]. When journal articles comprise the database, this would mean that researchers can set up “alerts” to be sent to them when a new article with a certain equation is published (and is indexed into the search system’s database).

New kind of search: Use Diagrams

The group [16] want to enable search queries to contain diagrams, as well as text and equations.

3.5.3 For Physics: Ideas for Useful Directions

Rank according to Solvability

When using these search engines to solve equations, it would be a big help if results were ranked according to whether the equation therein had been solved (exactly or approximately), and how quick it is to solve (computationally or by a skilled mathematician).

Doing so would require first obtaining this information, likely by adding a step to the pre-processing of documents in these databases. One possible, but probably time-intensive, way to do so would simply be to attempt to solve every one of the database documents’ equations in Mathematica, or a similar program. However, the choice of input parameters would need to be decided upon. Even an approximate measure of how solvable an equation is would - in general - be better than having a list of unranked equation search results. Another possible method is to try to extract the information about an equation’s solvability from the text or equations of any document which contains that exact written form of the equation; this includes searching mathematical encyclopedias, educational resources, and any other available documents. This, too, would require a lot of development to be fully implemented.

Downloading results for post-processing

None of the search engines allow users to download search results, so that post-processing of equations may occur.

Reasons a researcher might want to post-process equations include:

- Input retrieved equations to Mathematica, to see if their written format is easier to solve than the query’s written form.
- Pattern-search, algorithmically, within the results. For example, this could help researchers to identify phenomenological patterns, similarities, and analogies. To implement this, machine learning and data science algorithms may be useful: experimental results and theoretical predictions can both count as “data” for these algorithms.

- Computational speedup of code: see next proposed application.

Computational speedup: replace slow equations as they occur

Some computations can be done analytically, using Mathematica or similar software. (Which ones can be done this way is limited by the computation time and ability of the software platform.) Computations which are done this way often produce secondary equations during run-time: that is, they combine or solve a set of equations, then try to combine it with a second set of equations. One of the secondary equations can slow down the program's run-time enough that it becomes impractical. If that equation can be replaced, during run-time, with a faster-solving rewriting of it, then the problem of its slow-down can be fixed. That replacement could be done by making a run-time call to a "search engine" for math equations, which obtains a faster-solving mathematical representation for the problematic equation.

To implement this application, one could place a maximum time on computations. If the time is exceeded, the math-language search engine is called from the computer code itself. With no human interaction, the search engine would return a list of equations which match the mathematical meaning of the problematic equation. A (library) function could sort through the equation results, finding one which lets the code run within a reasonable amount of time. Preferably, these equations would be ranked by their solvability (as proposed above); however, if they were not sorted that way, the code could iteratively try various of the equivalent equations until either one works or a maximum time limit is reached.

Note that this application requires one or both of the above proposed applications to be done.

Algorithmically search for patterns, to help direct new research

Already done in Mathematics, this application involves algorithmically predicting "probable links" between known concepts using their mathematical relations, axioms, etc. These links can then be investigated, and used to help direct new cutting-edge research. [72]

In physics, we could algorithmically find phenomenological patterns, similarities, and analogies. Considerations would need to include the experimental uncertainty in each experiment, context of equations (for example, does x represent a bead's position on a wire, or a straight distance from an object?), and assumptions behind each experiment and theory which we apply our pattern-discovery algorithms to. Although these challenges may be considerable to overcome, performing pattern-discovery research without having yet solved them can still provide useful pointers as to where beneficial areas of research might be.

Math-aware search engines' algorithms for matching equations based on their mathematical meaning - both now and as they continue to develop - can provide a good basis for porting this application to physics.

Chapter 4

Guide to Use

Just as searching Google is a bit of an art, searching by equation requires some training. The tips given herein may be useful to researchers, when learning the use of these math-aware search engines. Of course, simply picking an engine at random and inputting an equation will often provide good results, without any of the knowledge elucidated here.

Troubleshooting, wildcards use, tips for successful queries, and algorithm choice are all discussed here. Also included is a short list of tools which help generate LaTeX or MathML equation encodings, as well as tricks for re-using (and navigating) results lists.

4.1 Choosing which Engine to Use

Equation-based search engines vary a lot in their search algorithms [26], database coverage, and results' presentation style. Yet, no one engine is “clearly” better than the rest, according to [18].

Features of each web-available search engine are summarized in tables 5.1 and 5.2. Discussing what they mean, and how they might impact a search's goal, are the subjects of this section.

Important considerations for a purpose-driven (or exploratory) search may include: whether the search algorithm matches the meaning or the presentation of the math; what fields and types of documents it searches over; and how well you can query given the interface's allowed options.

4.1.1 Presentation vs. Meaning-Based Search

Both meaning-based and presentation-based searches have their uses in search, according to developers and this author's analysis.

All meaning-based search algorithms are incomplete, so far: none of them can extract full mathematical meaning from the text around the equation in order to fully use the equation's proper meaning. [8, 24] This is a targeted area of development for the field. [45] However, some very good progress into meaning-based searching has been made. [8, 45] Presentation-based searching also has developments underway, surrounding how best to find “similar” equations (the meaning of which is naturally subjective) to an input query. [19]

Definitions

Presentation-based searching attempts to match the visual structure of equations [13, 21, 22], whereas meaning-based algorithms work to account for equivalencies of mathematical meaning between differently-structured equations [8, 22, 24]. For example, $(x^2 + y^2)/x^2$ has the same meaning as does $1 + (y^2/x^2)$, but is not visually equivalent. The two equations presented here would match in a meaning-based search, but not in a presentation-based search.

Uses of Each

Presentation-based search enables equations which are found to be written in similar notation and/or format to the one which was input. Two studies [16, 73, 74] have suggested that these properties - notation and format - play a role in how individuals reason about math.

If the true physical phenomena are sought after, however, meaning-based math matching is required, for it recognizes that math is just a representation of physical or abstract systems. To describe a system, one can use many different math notations: the system does not care which one you use, it will behave the same either way.

As such, presentation-based search is good when users seek familiar notation. For example, it would be useful to look up information related to a formula which is fairly standardized, to understand an unfamiliar notation of a formula from a document [16, 75], and to find out about specific questions or formulae in a given branch of math [40]. When a person does not know the name of a type of math, or of a specific formula, then presentation-based search also provides great usefulness.

Math-meaning searches complement presentation-based search. Formulas found from meaning-based searches can be found in presentation-based search to enable better understanding of the math therein. On the other side of the coin, presentation-based searching can provide better context and refine a search, if that is desired, before using meaning-based search. In some cases, of course, only one or the other will be needed.

Meaning-based searches, alone, can hit directly upon the deep connections which math notation hides. In this sense, they are far more useful for branching between fields of physics and/or math, and for (eventually) algorithmically searching for patterns between the phenomena which math represents. They help researchers to gain better understanding of what the system is actually doing, by changing the human construct imposed upon it by the math: this can help the researcher to separate the system from the mathematical formulations.

Complexity of equation: A consideration

A small experiment [24] found that presentation-based search engines provided better results than did meaning-based engines for the following types of equations: elementary functions (such as logarithm), and equations which have only one way of being written (excluding, for example, certain trigonometric expressions). Special functions which have a standardized representation, such as the Legendre-Q function, were suggested to be equally well-matched by presentation-based and meaning-based search engines. Equations which have multiple written forms were found to be better-matched by meaning-based search engines;

examples include trigonometric expressions and certain special functions (e.g. Hermite-H, Poly-Gamma).

4.1.2 Documents Searched: Field and Document Type

Different databases are searched by each search engine. Documents include journal articles, mathematics' special functions databases, and Wikipedia. Fields of research covered include physics, astronomy, applied mathematics, statistics, pure math, and others; each of the ones named here is especially applicable to physics research.

4.1.3 Query Features

Wildcards

Some engines allow wildcard searches. Which ones do, and their wildcard characters, is given in table 5.2.

Named wildcards are allowed by some searches. They can be made, for example, by preceding a character with a "?", such as "?a". If the same named wildcard occurs more than once in a query, then it enforces that results should have the same variable in that wildcard's place; for example, "?a + ?b^{?a}" could return $x + y^x$ but not $x + y^z$. Expressions, single variables, or constants can be allowed to take the place of a named wildcard. [35, 76, 77]

In the literature, "query variables" is another term which is used to describe wildcards. [22, 76]

Mutli-equation queries: Initial conditions inclusion

Some search engines allow the user to search for multiple equations at once. [12, 16, 49, 78] This allows initial conditions to be added to a search, among other things. The searched equations do not have to occur in the same portion of the paper. As with Google, these equation-based search engines may relax the query if the equations can not be found to occur in the same document: as does MIaS [49], they might require only one or more of the multiple query equations to be found in the paper.

Author, DOI, & other searches alongside equations

In many of these equation-based search interfaces, queries can include both equations and keywords from a specified portion of the publication. Author, year, journal, DOI, and title are common options.

4.2 Successful Queries

Successful and efficient querying is the goal of this section.

4.2.1 Making the query

Copy and paste an equation from a publication

If the equation query is a complex equation from an existing publication, then it is sometimes possible to access its LaTeX code. Springer articles can be found via LaTeXSearch, with their equations listed for easy review; however, this database contains a limited number of Springer’s articles, so finding the equation is hit-and-miss. [11, 32, 42] Another route is to download the LaTeX source code from arXiv.org, if the publication is available there.

Download code for a special function

If a special function is your query, it might be found in the DLMF [12] or Wolfram Functions website [79]. Both are searchable by keyword (and the DLMF is also searchable by portions of equations).

To download an equation’s code in DLMF, click the equation’s chapter/section link in the results list. This will lead to a page with more information, and many equations. Next to each, there is an encircled “i”: click it (do not just mouse-over it). Select “TeX” to download a TeX file with the equation in LaTeX code. For MathML, click “pMML” (“presentation MathML”) to open a new webpage with the code. [12]

Wolfram Functions allows code to be downloaded in MathML, by going to a specific formula’s webpage therein. For instance, one can search for “gamma function”, click the results titled “gamma function”, and in the list of its (193) formulas, click on “primary definition”. On the page which comes up, click the equation. The resulting webpage has MathML code for the formula, as well as *Mathematica* code. [79]

Generate equation code

Input of an equation query is usually done in LaTeX or MathML. Generating code in either language is simple with the aid of online template-based editors. Drawing an equation or uploading its image are both possibilities [80], although image upload is currently being fixed. [81] Translators also exist between the two languages.

Examples of tools are given here; they create or translate LaTeX or MathML code for equations.

Draw equation The online editor m_{in} allows users to draw equations, then send them as queries to various search engines or download their code as TeX or MathML. [80] To copy the code for a drawn expression, right-click on the expression’s image in the small window near the top right corner of the screen; a drop-down menu with “save math as” will appear. [82]

Upload equation’s image LaTeX code can be generated from an image or PDF file of an equation. This is not only useful for grabbing equations from academic papers, but also for extracting equations from online services such as the Wolfram (special) Functions website.

A desktop software exists to do this, which translates TIF, GIF, PNG, BMP, or PDF files to LaTeX or MathML. [83] It can translate entire documents, or single-equation images. [84]

The free demo version is fully featured, and allows up to 5 pages of documents per day. [85] The software is called InftyReader. [84]

The web-hosted tool m_{in} , which has been discussed elsewhere in this document, also allows images to be uploaded for translation to LaTeX; however, that feature is under repair at the moment, and is expected to be fixed in late 2016 [81]. Equations which are uploaded in this way enter into m_{in} 's drawing board, where they can be edited. The final equation can then be sent to various math-aware search engines as a query, or its LaTeX code can be downloaded. [80, 82]

Template editors LaTeX: At least a few template-based editors are available which generate LaTeX code from a template editor. One online version is Springer's LaTeX Sandbox. [86] **MathML:** A good list of equation editors and viewers ("renderers"), for MathML, is provided at [87]. Equation editors include downloadable interfaces (such as fMath and MathMagic) and plugins (such as FireMath, a Firefox plugin). [87] A good renderer for MathML equations, which generates a math equation image from the code, is Wolfram Research's MathML renderer. [88] **Both LaTeX and MathML:** For licensed or subscription users, *Mathematica* and Wolfram Alpha Pro each allow export of formulas to MathML or LaTeX code. In *Mathematica*, exporting is done by right-clicking on an output formula and selecting "Copy As". In Wolfram Alpha, a mouse-over on the formula's image brings up a menu; clicking "data" brings up a menu from which Pro users can download the formula's code.

Translators Translating from MathML to LaTeX can be useful, for example when copying special functions from the Wolfram Functions website. However, given the limited options available for this, it might be easier to copy the relevant equations' pictures and upload them to a software which translates them into LaTeX. Although many LaTeX to MathML translators exist online (e.g. MathMLCentral [89], MathToWeb [90], LaTeX2MathML [91], TeXZilla [92]), no operational web-hosted services for MathML to LaTeX could be found (one example of an inoperative one is MathML to TeX online Translator [93]). There are a couple of desktop solutions, although there seems to be no working desktop software. One solution is to use the github-hosted program web-XSLT [94]: there are no usage instructions included, but the developer specified on a forum that it should be used with Saxon (or a different xslt2 engine) and run via the command "java -jar saxon9.jar myfile.xml pmml2tex.xsl" [95]. Alternatively, the XSLT MathML Library can be used with Saxon, via "saxon -o output.tex input.mml xsltml.2.1.2/mmltex.xsl". [96]

4.2.2 Wildcards

Some search engines allow wildcards. This is discussed in section 4.1.3. Each search engine's treatment of wildcards is summarized in table 5.2.

4.2.3 Troubleshooting

If an equation query returns no results, or unsatisfactory ones, then the following might help target results better.

- Check whether this math-aware search engine requires equations to be bracketed by TeX $\$$ s.
 - For instance, it might require the equation to be entered as “ $\$equation\$$ ”, “ $\$equation\$$ ”, or “equation” (without the quotation marks).
- Change the notation: [18] point out that non-standard math notation can make a query hard to locate.
 - For example, some authors use “.” (which appears as “.”) or “\bullet” (“•”) to indicate the dot product, instead of “\cdot” (\cdot). Different LaTeX encodings are accounted for to a certain extent by the math-aware search engines, but this type of non-standard notation can provide a problem; this is especially pertinent for “.”, which is interpreted as a decimal point. Therefore, it can be useful to try using both standard and non-standard LaTeX notation in the query. [25]
- Simplify or otherwise re-write the formula.
 - For instance, the formula $\sinh(\log(\cos(x)))$ simplifies to $(-1/2)\sin(x)\tan(x)$ (and a few other forms) [15], which would likely provide different search results. As a simpler example, searching for $a\sin(a) + a\cos(a)$ can provide different results to a query of $a(\sin(a) + \cos(a))$. These rewritings are only necessary because the math-aware search engines don’t yet have the ability to compare equations based on their meaning in math; the next generation of engines should “know” all of the same math rules as Computer Algebra Systems (such as *Mathematica*), and are under development already [22, 26].
- Generalize the formula.
 - For example, one can transform elementary functions into simple variables (e.g. $\sin(2x)$ into $f(x)$) or exchange an expression in a query for a variable (e.g. rewrite $\sin(\log(\sqrt{x+2}))$ to $\sin(\log(a))$).
- Add a keyword to the search.
 - For example, to find information about how the “prime counting function” implies that $\pi(2) = 1$, it can be useful to search for both $\pi(2) = 1$ and “prime counting function” in the same query. [40]

4.3 Results

Search-easing features of some search engines, effect of different algorithms for ranking results, and expectations for display of results are included in this section.

4.3.1 Easing the search, in results stage

Result as a new query

The LaTeX text of results can be accessed in LaTeXSearch, making re-use of a result easy. [42] In addition, m_{in} allows images of equations to be used as search queries [97], making it much easier to re-use results from any search engine’s query; however, this feature is currently in-operational, and is expected to be fixed shortly [81]. Tangent provides a direct link from

each result to open its equation in m_{in} (section 5.3.1), where users can edit it via drawing or text, then generate LaTeX or MathML code. [82]

Filter results

Results can be filtered or sorted by publication year, author, journal, and other subjects, in some search interfaces.

4.3.2 Ranking Algorithms' Effects

Results rankings differ between engines. Since different rankings are appropriate to certain fields of study [38, 49] and goals of search, it may be beneficial to both try a few different engines and skim a few pages of results.

4.3.3 Presentation Expectations

Most of the search engines display results very well, including a display of the matching equation, text for context, a link to the article or webpage, DOI, and articles' titles and authors. Engines which do not display results well have that noted in their summary of section 5.4.

Chapter 5

Available Math-Aware Search Engines

The math-aware search engines which are available include many which are intended for use online, and a few which are downloadable for local creation of a math-searchable database.

Included here are summaries of each engines' abilities, including only those which work on the web or (for downloadable ones) have a working implementation in an existing database. Their features are summarized, for comparison and quick-reference in tables 5.1 to 5.4.

Because algorithms and databases differ, it can be useful to search using many math-aware search engines. To make the search more efficient, a multi-engine interface exists (m_{in}). [43] In addition, it is simple to copy and paste equations from one search engine to another, since they all accept LaTeX input.

5.1 Engines Included: Working, Web-Available

Many prototypes are in development, and many old architectures are no longer available (or, operational) on the web. [18] Because this document is tailored towards use of the search engines, only working and web-available search engines are included. Some of these have back-end software which is available for download, for the creation of equation-searchable databases from local or web-crawled documents; these features are also summarized.

One exception to the web-available rule is that a physics-catered search engine for equations is included in this summary. Its developers (from CERN) have worked to include physics-specific considerations in the search algorithms. Physics, after all, is not purely math. The features of this up-and-coming equation-search engine are overviewed.

Architectures which were discovered, but not included due to absent web-availability, are listed here in case the reader comes across them elsewhere. They may be in early stages of development, or old enough to be discontinued. They include: EgoMath2, Ego-Math, Physik-erwelt, WikiMirs, Formula Search Engine, MathDex (formerly MathFind), LeActiveMath, ActiveMath, Math Go!, and Math Aware Search Engine (MASE). [22, 24–26, 49, 75, 98, 99]

5.2 Keeping Up-To-Date

Keywords to help researchers keep abreast of newly-developed math-aware search engines include:

- Math Information Retrieval (MIR)
- Mathematics Retrieval
- Mathematical Knowledge Management (MKM)
- Digital Mathematics Libraries (DML)
- Math Search Interfaces
- Math-Aware Search Engines
- Formula Search

Terminology in this field can vary, as conventions do not seem to be yet set. For example, the following types of search can vary in their detailed definitions (while matching broad ideas) between documents: “similarity”, “presentation”-based, “syntactic”, “semantic”, “content-based”, and “meaning-based”.

5.3 Multi-Engine Interface

5.3.1 m_{in}

Four equation-based search services, plus Wikipedia and Google, are searchable from the single input interface m_{in} . [97] The math-aware search engines available are:

- Tangent (searches an old version of Wikipedia)
- NIST DLMF (provides special functions)
- LaTeXSearch (searches science and math journals)
- Wolfram Alpha (generates information, sometimes including special functions).

Each selected search is opened in a new tab, so that the original input is preserved for use in another search engine. [43]

Drawing or typing equations is allowed. Image recognition of png-format equations is also a feature, but is under repair at the moment. [81,97]

5.4 Web-Available Engines

The engines summarized here have a working online interface, at the time of this writing. Their features are summarized in tables 5.1 and 5.2, which are contained on a single page for quick reference. Example searches can be found in section 3.4, which outlines the extent of each engines’s mathematical search ability, and chapter 6, which contains screenshot of searches with each engine.

5.4.1 LaTeXSearch

Covering physics and math journal articles, ranging from quantum theory to energy, Springer’s LaTeXSearch enables presentation-based equation retrieval. A subset of their articles are included, but Springer does not specify which subset ¹. [32,42]

¹Springer specifies that 120,000 peer-reviewed articles are covered by LaTeXSearch. However, Springer’s main webpage states that the size of their entire collection is over 5 million articles. [11,32,42]

This platform is proprietary, and provides very little information about its internal search algorithms. Both exact matches and similar equations are found. [42] A “Speed-search” feature searches for only exact matches to the input LaTeX code (which is different from searching for exact matches to the equation, as a given equation can have multiple LaTeX encodings). It runs much faster than does the full LaTeXSearch. [100] Based on this author’s experimentation, this engines’ “similar equations” include those which have different variable names, different constant values, or variables in the place of constants (and visa-versa).

Presentation-based search appears to be used, based upon independent experiments by: this author, [49] and [24].

The language of query is LaTeX. Equation terms which are entered into the query box are all considered to be part of a single equation: multiple equations can not be searched for². Advanced search options allow subject, publication, and year to be added to the equation search. [100] Alternatively, a single article’s equations can all be recalled by entering the DOI or title keywords of that publication. [11]

LaTeX code for the discovered equations is easy to download, and is located below the equation’s image in the results. [100]

The developers want to expand the database, including possibly indexing the arXiv. They ask people who are interested in helping to contact them. [42]

5.4.2 zbMATH

Natural sciences (including physics), computer science, and other applied math publications comprise a significant portion of the zbMATH database. Pure math publications make up the rest. They include articles from around 3000 journals, and 170,000 books, comprising over 3 million publications in total. Many old texts have been digitized, as well, for inclusion in the database. [9, 47]

Using MathWebSearch as its engine (detailed in section 5.7.2), zbMATH provides meaning-based math search. Exact matches to an equation are found, unless named wildcards are used, in which case similar equations are found. Named wildcards are made, in this case, by preceding a character with a “?”, such as “?a”. [101]

A query can include a single “LaTeX-style” equation, along with several optional keywords. [35]

Results are displayed as a list of titles with author and year. Clicking on one of the results expands it, showing a highlighted form of the located equation, along with the article’s abstract, DOI, and other information. [35]

A subscription service of zbMATH is also available, and is not to be confused with its free equation-search platform. The subscription service adds other query options, such as author, publication year, and mathematical classification, to an equation-based search. It can be previewed, but only three results are available in the free version. Access to it is found by clicking the “zbMATH” logo at the top of the formula-search (free) platform, [35].

²This was found from experimentation. Specifically, a paper was found which contained at least two equations, and the two equations were searched for independently (to make sure they were found) and then together, in the format (without the $\langle \rangle$ brackets): \langle “equation1” AND “equation2” \rangle When they were searched for together, no results were found. Searching for pieces of a single equation, in the format \langle “piece1” AND “piece2” \rangle , does work to bring up individual equations which contain both pieces.

5.4.3 EuDML

Mathematical Physics journal articles, and other math journal articles, are included in EuDML. Over 250,000 documents are searchable, including a few books. [10] The developers are negotiating to include more content, but they note that it is difficult to convince “big commercial publishers” (such as Elsevier and Springer) to agree to their article control policies. Other publishers with whom they have been under negotiations with include Euclid, Math-Ru.Net, and EMS Publishing House. [47]

EuDML allows a search query to include a single equation and any number of keywords. Note that, while its engine (MIaS) does allow multiple equations to be included in one query [49], experimentation reveals that the EuDML interface does not seem to do so. [10]. Other search categories can be specified in EuDML, as well, including author, year, and language of writing. [10]

Results’ presentation could be improved, but is sufficient: the equation from the result is not highlighted, and is often not displayed in the text of the result. Finding the equation, and discovering whether it was an exact vs. similar match to the query, requires accessing and downloading the article (or book) which was found to contain a match. Included in each result is the title, author, year, and usually some text. [10]

Presentation-based search is performed; both similar and exact formulae are found. The back-end engine used, MIaS [22], is further described in sections 3.4.2 and 5.7.3.

5.4.4 NIST Digital Library of Mathematical Functions (DLMF)

Special functions from the textbook “Handbook of Mathematical Functions” have been digitized and expanded upon to provide an equation-searchable library. Highly curated, this library consists of formulae, tables, 3D visualization, and graphs. [12]

Formulae can be downloaded for their LaTeX or (presentation) MathML versions. [12]

Search types possible include text, phrases (in double-quotes), LaTeX-like math expressions, boolean operators, and proximity operators (which denote how close a term should be to another term). Categories of math, such as “trigonometric” can be used in the search. [102] Pieces of many equations, such as $\langle (1-z^2)^{(1/2)} \text{ and } e^{\nu(1-x)} \rangle$, can be searched for by connecting expressions with $\langle \text{ and } \rangle$; the query’s exact equations, or their similar counterparts (see next paragraph) must all appear in the same section of the online textbook in order for it to be retrieved as a result. [12]

Presentation-based search is used by DLMF. First, the search engine attempts to find an exact match to the input equation. After those have been found, the symbols of the equation are allowed to be re-ordered or separated by other math terms. [13,24] As a consequence, this search engine can not identify special functions which are written in a way which significantly differs from that of its internal database. However, it can still be used to look up some representations of special functions. Software by which to solve special functions is also listed in the DLMF, including a table specifying which software solves specific special functions. [103]

5.4.5 Tangent

A section of Wikipedia is the database for this search engine. The content is frozen to a certain date: it does not update as Wikipedia evolves. [50]

Only ten results are allowed to be viewed, in this demo. [63] It is not clear why; however, this may often be enough to find the desired results on Wikipedia.

Presentation-based search is used, which the developers say helps non-specialists find understandable information. [16] Visual aspects, including left-to-right structure, are considered when ranking results. [21, 62, 63] Exact matches are listed first; similar matches might have variables' names changed, constants in place of variables (or visa-versa), constants' values changed, or changes in the structure of the equation. [16, 63]

Users can search for multiple equations at once. [63] While keywords can be added to the query, the results may be unexpected since they are not supported in this web-available version. [21, 63] Input is allowed via LaTeX or a drawing interface (named m_{in}). [16, 78] Wildcard use is supported, with the “?” symbol representing any single symbol. [16]

A downloadable, more advanced version of Tangent is available for enabling math-search over custom databases. It is accompanied by clear instructions as to how to run a webserver which searches a website for equations. (See section 5.7.1 for more details.)

5.4.6 Uniquation

This engine searches over math “help”, or “question and answer” websites, such as planet-math.org, mathoverflow.net, and math.stackexchange.com.

Not much information is available about it. No scientific papers can be found which focus on it, although it is mentioned in a few papers (such as [26]).

It appears to use presentation-based search: it finds structurally similar and equivalent expressions. Variable names are ignored, when searching.

TeX is used as the language of query. Keyword-searching does not appear to be supported.

Results are presented grouped together by formula; the actual webpages are accessed using a link above each formula. [17]

5.4.7 Wolfram Alpha

Wolfram Alpha both retrieves mathematical (and other) information from an encyclopediac database, and calculates results to queries. It has roughly 1000 curated data sets, and draws from real-time data sources. [14] Many special functions can be found, although it does not recognize all forms of the special functions (see 3.4.1); definitions, information, and plots about functions are also presented. [15] It can process math based on its mathematical meaning, due to its use of the math processing power of *Mathematica*. [14]

Wolfram Alpha does not bring up instances of formulas in journal articles or webpages, but rather brings up information from its own internal curated database; in this sense, it is more like an encyclopedia which is accessed solely via search queries. [15]

Input can be given in natural language [14] (English [104]), TeX [105], and (the majority of) small pieces of *Mathematica* code [104].

5.5 Ready for ArXiv deployment: Needs Server Space

The developer of the equation-based search engine MathWebSearch, Michael Kohlhase, states on his webpage that MathWebSearch is ready for deployment on arXiv.org. To do so, he needs 128GB of additional server space, and that their restriction comes especially from the RAM; donations of server space are invited. [31]

MathWebSearch currently powers zbMATH, and is available for download to build local and web-crawled searchable databases. [31] It also has a demo interface at [36], but that demo is only partially operational at this time.

5.6 In development for physics

5.6.1 $5e^{x+y}$

CERN has recently been developing a physics-catered search engine for equations. It is meant to be used on journal articles and other scientific documents, and is currently catered to the CERN Document Server (CDS). [26]

Meaning-based search is used: equations are parsed for comparisons using *Mathematica*. The “Simplify” and “FullSimplify” commands, in Mathematica, are used to normalize every equation in the database, and every query, so that they can be compared based on their mathematical meaning rather than their syntax (or “presentation”). [26, 52]. These commands in *Mathematica* can re-arrange the equations to a standardized form, using special function definitions, elementary function definitions, and of course simple rules including associativity and distributivity (which the other search engines only approximate).³

Physics considerations are also in the forefront. Query constants’ values are included in the search in a way which accounts for different experiments of the same phenomenon giving different results: their value is iteratively rounded, for matching purposes. Mathematical structure is emphasized in the search. Specific fields of study have common patterns, such as $\langle S = \int_{t_1}^{t_2} \mathcal{L} dt \rangle$ or $\langle \text{originalParticle} \rightarrow \text{subparticle1, subparticle2, subparticle3} \rangle$; many of these were manually added to the search algorithms, although the list is not meant to be exhaustive. [26] The developers also tried to account for the fact that variables’ names follow certain conventions in some cases, and are irrelevant in others. [52]

More development is needed before it is ready for full-scale deployment on CDS and similar systems. [26] However, a working demo version was available on Invenio’s main webpage [52]; it has been taken down now [106]. Hopefully work will continue on this physics-catered project.

5.7 Build a Custom Equation-Searchable Database

Universities, institutions, and individuals can enable math-equation searching on custom databases of MathML (using MWS, Tangent, MIaS) and LaTeX (using Tangent). Web-

³Another, currently-available math-aware search engine is investigating including *Mathematica*-like systems for performing the mathematical comparisons: MIaS, which has its present state discussed in sections 3.4.2 and 5.7.3. [22]

crawling (using MWS) or local databases (using Tangent, MIaS, MWS) can be used.

The software for this also powers the web-available engines. Other software might exist for this, which has no working web demo or interface; those are not included here.

Features of the included engines are summarized in tables 5.3 and 5.4. More details about the mathematical processing abilities of these engines is given in section 3.4.2.

5.7.1 Tangent

More advanced than its web demo, which is v0.1, Tangent v0.3.1 is available for download from the research group’s webpage: [50].

This updated version is the latest release under development by [50]. It finds similar equations to a query, based on their symbols and layout. [62] Keywords are supported in the same query as equations, and wildcards can represent subexpressions or single symbols. [64] This scalable [62] engine can process local documents in TeX and MathML [39,107], and has been used on parts of the arXiv and Wikipedia [64].

5.7.2 MathWebSearch (MWS)

Providing meaning-based search [8,22,24,49], this scalable engine can work independently [8] or with other “wrapper” software which adapt it to different database types, user interfaces, and ranking schemes [108]. A version for the arXiv has been tested [8], and needs only server space to fully implement [31].

MWS uses its own algorithms to provide meaning-based search, and is based on Lucene for text-based searching. [49] It supports named wildcards, so that a character can be replaced by any expression or single variable in results; they are denoted by preceding a character with a “?”. [35,76,77] A description of named wildcards is given in section 4.1.3.

A working search interface which uses MWS is zbMATH. [35] Another demo, called Sentido-search ([36]), showcases an MWS input panel which both has a template and allows for and coding in many languages, including LaTeX; it has a search feature, but it is currently not fully functional.

ArXiv.org deployment of MWS is stated to need only an additional 128GB of RAM; donations of server space are invited. [31] MWS developers tested its performance on about 65% of the arXiv (as of 2012). Query times averaged 25 - 75 ms for most queries, and went just past 300 ms for a more difficult class of query. The system uses distributed computing, so that its estimated memory requirements for the entire arXiv (170GB) can be accommodated. A user interface for arXiv retrieval was designed, and appears to use LaTeX for its query language. [8]

The github version of MWS allows indexing of MathML formulae from a website or a local repository of XHTML documents. [109] (To use it on arXiv.org documents, the developers use an XML-translation of the arXiv.org files. [8,110]) Instructions for installation of MWS are included on the github page ([109]).

5.7.3 Math Indexer and Searcher (MIaS)

MIaS uses presentation-based searching; upgrades to meaning-based search, with the help of *Mathematica*-like systems, is being explored. [22, 24, 25, 49] First locating queries' exact matches, in the database, it then moves on to allow variable names to be ignored (unless the query consists of solely a single variable), constants' values to not matter, and only a portion of the formulae (i.e. their subformulae) to match. New developments also allow it to match equations which have similar visual structure. [66] The complexity of both the query and database equations are considered when ordering the retrieved formulae: in general, more complex formulae are ranked higher in the list of results. [49]

Queries allowed include multiple equations, alongside textual keywords. When multiple equations are searched for, only one needs to appear in the results: the equations are connected internally by the OR operator. [49] Input is done, at present, in LaTeX or MathML; a visual rendering of the input is provided. [49] Wildcards are currently being explored for inclusion. [22]

Already used in EuDML, this back-end engine is designed to support large-scale libraries such as arXiv. [22, 49] It runs on the platform Apache Lucene [49], which is already in use for the CERN internal document server (CDS) [26].

A demo of the downloadable interface, WebMIaS, is provided at [111] and described in [46]. The demo searches around 100,000 arXiv.org articles (the “NTCIR-12” dataset). [46, 112, 113] Input is either TeX or MathML, and keyword-search in the “title” or “content” can accompany an equation in a query. Presentation- and/or content- MathML language can be used: the former can be better for visually-matching equations, and the latter for matching their meaning. [46] The results display well, with formulae matches highlighted. Unfortunately, at present, the documents from which results are drawn are inaccessible from the webpage; however, arXiv numbers can often be recognized in their titles, so the results can be found by the determined user. [111]

The github versions of MIaS [114] and WebMIaS [46, 115] use Lucene to index and search a local database of files; math in the files must be written in MathML. ⁴ A plug-in to Solr and Lucene is also available, called MIASMath. [116] Instructions for use are provided on the github webpages ([114, 116]).

⁴For arXiv-interested users, note that a partial version of the arXiv.org documents is available in MathML: it is called arXMLiv. [110]

Table 5.1: Web-Available Search Engines for Math

Search Engine	Discipline ^A	Database	Search type	Back-end	Web address
LaTeXSearch	Math-Sci Math	Journal articles	Presentation	Custom	http://latexsearch.com/home.do
zbMATH	Math-Sci Math	Journal articles, Books	Meaning	MWS	http://search.mathweb.org/zbl
EuDML	Math-Sci Math	Journal articles, Books	Presentation	MIaS	https://eudml.org/
DLMF	Math	Handbook of Mathematical Functions	Presentation	Custom	http://dlmf.nist.gov/
Tangent	All	Wikipedia (old subset)	Presentation	Custom	http://saskatoon.cs.rit.edu/tangent/
Uniquation	Math	Forums, collaborative encyclopedias	Presentation	Unknown	http://uniquation.com/en/
Wolfram Alpha	Math-Sci Math	Encyclopedia, curated. (Also generates results.)	Meaning	<i>Mathematica</i>	https://www.wolframalpha.com/

^A “Math-Sci” indicates “Mathematical Sciences”; for each engine listed here, that includes physics content.

Table 5.2: Querying in Web-Available Engines

Search Engine	Language	\$ ¹	>1 eqn. ²	Keywords ³	Loc. ⁴	Wildcards (un-named ⁵)
LaTeXSearch	LaTeX	No	No	No	Yes	No ⁶
zbMATH	LaTeX MathML	No	No	Yes	No	Preface symbol with “?”. (Named wildcard.)
EuDML	TeX MathML	\$ or \$\$	No	Yes	Yes	No ⁷
DLMF	LaTeX-like	No	Yes	Yes	N/A	“\$” for many characters, “?” for 0 or 1 characters.
Tangent	TeX	No	Yes	No	No	“?” for single symbol.
Uniquation	TeX	No	No	No	No	Natively ⁷
Wolfram Alpha	Natural Language, <i>Mathematica</i> , LaTeX	No	No ⁸	Yes	No	No ⁶

¹ Bracket text with \$’s, as for TeX math mode?

² Query can have >1 equation.

³ Query can have keywords and equation(s).

⁴ Specify location (e.g. in title) of query terms.

⁵ Unless specified.

⁶ Not specified; experimentation excluded “\$” and “?”.

⁷ Ignores variable names.

⁸ Information not available; experimentation indicates “no”. This applies to encyclopediac information searches using equations, and not to solving equations directly with this tool.

Table 5.3: Downloadable Versions (& one Physics-Catered Upcoming Engine)

Search Engine	Application	Doc. Encoding ^A	Crawler/Local	Search type	Web-server instructions	Download	Demo
Tangent	Custom database	TeX	Local MathML	Presentation	Yes	[50, 117]	[63]
MathWebSearch	Custom database, Ready for arXiv	MathML	Web, MathML	Meaning	Interface-dependant Local	[109]	[35, 36]
MIaS	Custom database	MathML	Local	Presentation ^B	Yes (JVM instructions)	[114, 115]	[111]
$5e^{x+y}$	In-development,	LaTeX	Unknown	Meaning	N/A	N/A	N/A

^A Documents whose math is encoded using these languages can be indexed for math-aware search.

^B Work is proposed to investigate including *Mathematica*-like software in MIaS's meaning-recognition algorithms.

Table 5.4: Querying in Downloadable Versions

Search Engine	Language	$\$$ ¹	>1 eqn. ²	Keywords ³	Wildcards ⁴
Tangent	TeX	No	Yes	Yes	"?" for a symbol
MathWebSearch	Varies ⁵	Varies ⁵	No	Yes	Named: "?" before a symbol
MIaS	LaTeX MathML	Yes, $\$$	Yes	Yes	"?" for a symbol

¹ Bracket text with $\$$'s or $\$\$$'s, as for TeX math mode?

² Query can have >1 equation.

³ Query can have keywords and equation(s).

⁴ Wildcards are un-named unless specified.

⁵ Depends on the interface which this engine is paired with.

Chapter 6

Examples

A range of simple and neat equation queries are shown in this chapter, including at least one from each of the working web-based search engines. A few features of the search interfaces are also pointed out.

6.1 m_{in}

The query in figure 6.1 is in the process of being drawn, using m_{in} . Instructions for m_{in} 's use are located at [97].

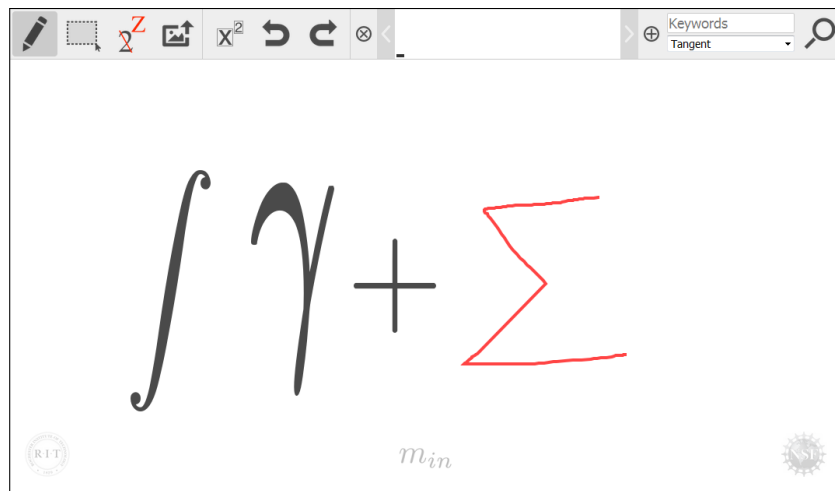


Figure 6.1: Drawing in-progress, in m_{in} .

6.2 LaTeXSearch

In the simple search shown in figure 6.2, note the post-search filters located in three locations: within the tabs, just below the tabs, and in the left vertical side-bar. In the display, equations from each paper are grouped together; the number of equations in each paper is listed to the left of the publication's name. LaTeX code can be downloaded via the green "View Latex Code" links.

The screenshot shows the LaTeXSearch interface. At the top, the search query is "\epsilon" AND "2\pi". The search results are displayed in a list format. The first result is titled "Delamination of thin bonded cement-based overlays: analytical analysis" and is from the journal "Materials and Structures" (2011) 44:43-51, January 01, 2011. It has 3 results. The second result is titled "An analysis of the flow of a power law fluid due to ciliary motion in an infinite channel" and is from the journal "Journal of Biorheology" (2010) 24:56-69, December 01, 2010. It has 4 results. The interface includes a left sidebar with filters for SOURCE and PUBLICATION YEAR, and a top navigation bar with links to Home, About, Contact Us, and LaTeX Sandbox.

SOURCE

- Quantum Theory of Conducting Matter 7
- Encyclopedia of Complexity and Systems Science 4
- Nanoscale Research Letters 4
- Deconvolution Problems in Nonparametric Statistics 3

PUBLICATION YEAR

1995 2012

SEARCH RESULT

Exact Results (94) Similar Results (106) << 1 2 3 4 5 >>

Sort by: Relevance Results per page 10 View Latex Code

3 results

Delamination of thin bonded cement-based overlays: analytical analysis
Materials and Structures (2011) 44:43-51, January 01, 2011

EXACT $\epsilon = \frac{1}{2\pi} \ln \frac{1-\beta}{1+\beta}$

View Latex Code

EXACT $\sigma_{22} = Re \left(\frac{K}{\sqrt{2\pi r}} r^{i\epsilon} \right)$

View Latex Code

EXACT $\sigma_{22} + i\sigma_{12} = \frac{K}{\sqrt{2\pi r}} r^{i\epsilon}$

View Latex Code

4 results


An analysis of the flow of a power law fluid due to ciliary motion in an infinite channel
Journal of Biorheology (2010) 24:56-69, December 01, 2010

Figure 6.2: A simple search on LaTeXSearch: “\epsilon” AND “2\pi”.

6.3 zbMATH

A neat combined keyword and equation search in zbMATH is shown in figure 6.3.

When generating this screenshot, it was necessary to click on the result's grey title box to show its full entry. The equation matched, $(\sin^2(x))$ does appear in the results, but is too low down to include in the screenshot.


 the first resource for mathematics

probability Search Example

$\sin^2(x)$

$\sin^2(x)$

« 1 »

Nagai, Katsumi (2006): *The integrated density of states of one-dimensional random Schrödinger operator with white noise potential and background.*

<http://zbmath.org/?q=an:1126.34053>
Title: The integrated density of states of one-dimensional random Schrödinger operator with white noise potential and background.
Author(s): Nagai, Katsumi
Published: 2006
Class: 34L40 47B80 60H25
Doctype: serial article
Keywords: random Schrödinger operator; integrated density of states; white noise potential
Language: EN

The author discusses the integrated density of states $N(\lambda)$, $\lambda \in \mathbb{R}$, of the random Schrödinger operator

$$H = -\frac{1}{r(t)} \frac{d}{dt} \left(\frac{1}{p(t)} \frac{d}{dt} \right) + \frac{q(t)}{r(t)} + \frac{cB'(t)}{r(t)}, 0 \leq t < \infty.$$

Here, we suppose that the coefficients of H satisfy the following conditions: $(p_\omega(t))_{t \geq 0}$, $(q_\omega(t))_{t \geq 0}$, $(r_\omega(t))_{t \geq 0}$ are continuous semi-martingales on a **probability** space (Ω, \mathcal{F}, P) with a filtration $(\mathcal{F}_t)_{t \geq 0}$.

Figure 6.3: A keyword (“probability”) and equation $(\sin^2(x))$ search on the zbMATH platform. The formula searched for is also present in the results’ display, but it does not fit into this screen capture.

6.4 EuDML

Using the advanced search function for EuDML, the equation $E = mc^2$ was queried. Putting post-search filters of “English” and “Articles” produced the result in Figure 6.4.

The screenshot shows the EuDML website interface. At the top, there is a search bar with the text "Title, Author, Keyword, Citation, Date..." and a "Search" button. The language is set to "English (en)". There are links for "Login" and "Register". Below the search bar, there are navigation tabs: "Home", "Advanced Search" (selected), "Browse by Subject", "Browse by Journals", and "Refs Lookup".

The "Advanced Search" section is active. It shows a search rule: "Match All of the following rules". The first rule is "Any field contains" followed by a text input field containing the formula $E = mc^2$. There are buttons for "Add Sub-clause" and "Add Another Rule". Below the input field, there is a preview of the formula: $E = mc^2$. There is also a checkbox for "Only documents with accessible full-text".

At the bottom of the search results, there is a summary: "Currently displaying 1 – 20 of 48376". It shows "Showing 20 per page" and "Order by Relevance | Title | Year of publication".

There are two main columns of results. The left column shows the title of the first result: "Pseudoautomorphisms of Bruck loops and their generalizations" by Mark Greer and Michael Kinvon, published in 2012 in Commentationes Mathematicae Universitatis Carolinae. The right column shows "Used filters" (Document types: Article, Languages: en) and "Journals" (Fundamenta Mathematicae: 2750, Studia Mathematica: 2421, Commentationes Mathematicae Universitatis Carolinae: 2239, Czechoslovak Mathematical Journal: 2230, Acta Arithmetica: 1987). There is also a "Years" section showing 2015 with 113 results.

Figure 6.4: A sample search in EuDML: $E = mc^2$. Other interesting result titles, lower on the page, included “The equivalence principle meets the uncertainty principle” (1988), and “On the validity of Wilson’s approach to general relativity” (1997), and “Recovering Asymptotics at Infinity of Perturbations of Stratified Media” (2000). Whether these are exact or similar formula matches is unknown: given the search setup, determining this requires accessing and downloading the article.

6.5 DLMF

The queries used here may not work in other search engines, since they were drawn from the DLMF help webpage. [102]

In figure 6.5's upper panel, the search is for the sum of the squares of any two trigonometric functions. Its lower panel shows a search for equations with Γ and $=$ both present and separated by no more than 5 terms.

The figure consists of two vertically stacked screenshots of the DLMF search interface. Both screenshots feature a sidebar on the left with navigation links: Index, Notations, trigonometri (partially visible), Search, Need Help?, How to Cite, and Customize. Below these is a 'What's New' section with 'About the Project' and the NIST logo (National Institute of Standards and Technology).

Top Screenshot:
 Search query: `trigonometric^2+trigonometric^2`
 Results: 1—10 of 91 matching pages
 1: [4.21 Identities](#)
 ▶ ...
 4.21.12 $\sin^2 z + \cos^2 z = 1,$
 4.21.13 $\sec^2 z = 1 + \tan^2 z,$
 4.21.14 $\csc^2 z = 1 + \cot^2 z.$
 ▶ ...
 4.21.27 $\sin(2z) = 2\sin z \cos z = \frac{2\tan z}{1 + \tan^2 z}$

Bottom Screenshot:
 Search query: `Gamma near/5 =`
 Results: 1—10 of 131 matching pages
 1: [5.2 Definitions](#)
 ▶ ...
 5.2.1 $\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt,$
 $\Re z > 0.$
 ▶ When $\Re z \leq 0$, $\Gamma(z)$ is defined by analytic continuation. ...
 5.2.2 $\psi(z) = \Gamma'(z)/\Gamma(z),$
 $z \neq 0, -1, -2, \dots$
 ▶ ...
 5.2.5 $(a)_n = \Gamma(a+n)/\Gamma(a).$

Figure 6.5: Tricky search examples, using language which is specific to the DLMF. The top panel query is “trigonometric^2 + trigonometric^2”; the bottom is “Gamma near/5 =”.

6.6 Tangent

Simple queries are shown in figure 6.6, on the Tangent search engine platform.

Tangent	Tangent
<p data-bbox="240 409 737 443">\forall a, b</p> <p data-bbox="269 485 743 508">5852 results found in 1104 ms (750 ms parsing, 353 ms searching).</p> <p data-bbox="269 569 331 596">$\forall a, b,$</p> <p data-bbox="269 600 570 644">Document: Wikipedia - Happened-before Score: 0.750 - Edit query - Search for this</p> <p data-bbox="269 659 310 686">a, b</p> <p data-bbox="269 690 610 735">Document: Wikipedia - Equivalence partitioning Score: 0.667 - Edit query - Search for this</p> <p data-bbox="269 749 347 777">$\forall a, b, c$</p> <p data-bbox="269 781 570 825">Document: Wikipedia - Happened-before Score: 0.571 - Edit query - Search for this</p> <p data-bbox="269 840 318 867">$P_{a,b}$</p> <p data-bbox="269 871 719 915">Document: Wikipedia - Power dividers and directional couplers Score: 0.500 - Edit query - Search for this</p>	<p data-bbox="831 409 1328 443">\epsilon / 2\pi</p> <p data-bbox="860 485 1334 508">8152 results found in 908 ms (640 ms parsing, 268 ms searching).</p> <p data-bbox="860 569 922 596">$\Gamma / 2\pi$</p> <p data-bbox="860 600 1247 644">Document: Wikipedia - Electrophoretic light scattering Score: 0.500 - Edit query - Search for this</p> <p data-bbox="860 659 919 686">$1 / 2\pi$</p> <p data-bbox="860 690 1162 735">Document: Wikipedia - Hartley transform Score: 0.500 - Edit query - Search for this</p> <p data-bbox="860 749 922 777">$\omega / 2\pi$</p> <p data-bbox="860 781 1198 825">Document: Wikipedia - SeibergWitten invariant Score: 0.500 - Edit query - Search for this</p> <p data-bbox="860 840 928 867">$= \epsilon / 2$</p> <p data-bbox="860 871 1162 915">Document: Wikipedia - Machine epsilon Score: 0.500 - Edit query - Search for this</p>

Figure 6.6: Neat and simple search query examples, and a showcase of the Tangent platform. The left panel's query is " $\forall a, b$ ", and the right is " $\epsilon / 2\pi$ ".

6.7 Uniquation

Uniquation’s inclusion of abstract algebra (top panel) and a general example of how sums might be searched for (bottom panel) are both shown in figure 6.7. Note that results are grouped by the formula which they contain; this is evident in the lower panel of figure 6.7, where two results are shown under formula “1”.


(uni)quation ^{alpha}

[Description of query language \(TeX\)](#)

2 formulae were found.

- [Finding the Units in the Ring - Mathematics - Stack Exchange](#)

$$\mathbb{Z}[t][\sqrt{t^2 - 1}]$$

 Finding the Units in the Ring $\mathbb{Z}[t][\sqrt{t^2 - 1}]$ - Mathematics - Stack Exchange
- [Finding the Units in the Ring - Mathematics - Stack Exchange](#)

$$\mathbb{Z}[t][\sqrt{t^2 - 1}] \subset \mathbb{Z}\left[\frac{1}{2}\right][x, x^{-1}]$$


(uni)quation ^{alpha}

[Description of query language \(TeX\)](#)

One formula was found.

- [Show all links to this formula \(3\)](#)

$$\sum_{n \in [1, m]} n^m = 1 \pmod{p}$$

 modular arithmetic - Why are the only numbers m for which $n^{m+1} \equiv n \pmod{m}$ is true also unique for $\sum_{n=1}^m n^m \equiv 1 \pmod{m}$? - Mathematics - Stack Exchange


 Why are the only numbers m for which $n^{m+1} \equiv n \pmod{m}$ is true also unique for $\sum_{n=1}^m n^m \equiv 1 \pmod{m}$? - Mathematics - Stack Exchange

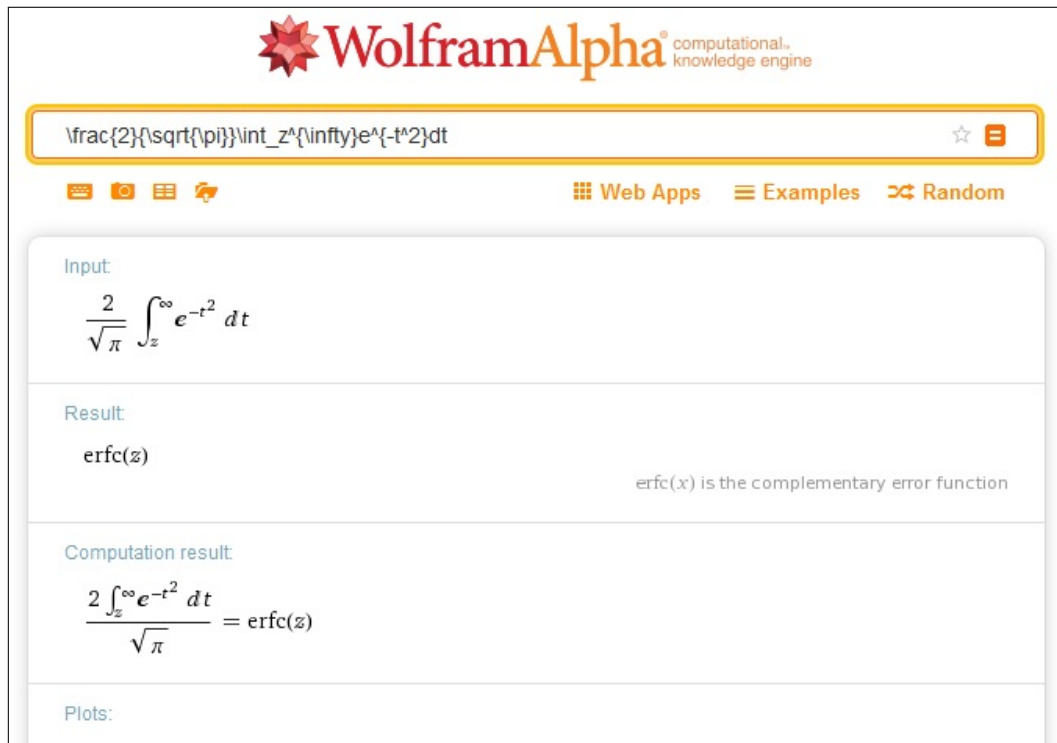
Figure 6.7: Examples of neat searches, using Uniquation. The query in the top panel is “ $\mathbb{Z}[t][\sqrt{t^2 - 1}]$ ”, and that in the bottom panel is “ $\sum_{n=1}^m n^m \equiv 1 \pmod{p}$ ”. Examples are drawn from [67].

6.8 Wolfram Alpha

A search for the complementary error function's integral representation,

$$\frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-t^2} dt \quad (6.1)$$

is shown in figure 6.8. Wolfram Alpha correctly identifies this as $\operatorname{erfc}(z)$.



The screenshot shows the Wolfram Alpha interface. At the top is the logo "WolframAlpha" with the tagline "computational knowledge engine". Below the logo is a search bar containing the LaTeX query $\frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-t^2} dt$. Below the search bar are navigation icons for "Web Apps", "Examples", and "Random". The main content area is divided into sections: "Input" showing the original integral, "Result" showing $\operatorname{erfc}(z)$ with a note "erfc(x) is the complementary error function", "Computation result" showing the equation $\frac{2 \int_z^{\infty} e^{-t^2} dt}{\sqrt{\pi}} = \operatorname{erfc}(z)$, and "Plots:" which is currently empty.

Figure 6.8: A Wolfram Alpha search, using a LaTeX-encoded query, for the integral representation of the $\operatorname{erfc}(z)$.

Chapter 7

Conclusion

New research methods are enabled by equation-based search engines. From the simple benefit of looking up a concept by its equation - whether half-remembered or fully - to the advanced application of helping researchers discover new interdisciplinary connections, the possible applications elucidated so far have likely barely scratched the surface.

Meaning-based equation searches can open new doors to math solutions, perspectives, collaborations, and faster-running computer code. Presentation-based searches can help with the same, except that they require equations' structures to be similar in order for a match to be found; they can also help with finding out what unfamiliar math means.

These uses of math-aware search engines can lead to better understanding, predictability, and applications of physical systems. Fields which may see progress accelerated by these engines range from fundamental physics (e.g. gravitation and quantum physics) to effective theories (e.g. within condensed matter and fluid mechanics) to technological applications (e.g. quantum computing and energy extraction).

Databases which are presently searchable in this way include journal articles in physics, math, and other mathematical disciplines (LaTeXSearch, EuDML, zbMATH); Wikipedia (Tangent); a curated math encyclopedia which includes some physics (Wolfram Alpha); and math forums (Uniquation). Expansion to the entire arXiv.org website is feasible with MathWebSearch, a meaning-based math-aware search engine. Other expansions are also planned. A physics-specific engine is also in development, at CERN, which would search CERN's document server using math-aware equations; it takes into account some specific desires of physicists, during the math matching stage.

Other developments planned include improving the: ranking algorithms for results, mathematical-meaning extraction and comparison, and speed of retrieval (which is passable but not as fast as Google). A different sort of search which is also being considered is search by diagram.

The search tools and discussion in this document provide a resource for physicists (and other researchers) to encompass the language of math in their searches. Whether these math-aware search engines are used for targetted information searches, exploration, or other inspiration, they are sure to prove useful in physics.

Appendix A

Glossary

Terms used in this discussion of math-aware search engines are defined here. Many of these terms are not yet set into conventions, in this field. Some of the definitions may seem very straightforward, but can have varying definitions between and within research fields.

Math-Aware Search Engine / Math-language search engine A software which searches a database of mathematical texts, using a mathematical equation as at least part of its search query. It accounts, at least partially, for the mathematical equivalencies between different writings of equations.

Query The set of items which the user inputs as search terms. It can consist, for example, of equations, keywords, phrases, DOIs, and publication date range.

Database A collection of data. For equation-based search engines, it might consist of journal articles, wikis, forums, encyclopedias, or other documents.

Equation match An equation which matches another equation, within the search engines criteria. For example, the following equations might match: $ax + 5 = b$, and $sx + 5 = t$.

Result The items from the database (e.g. equations, pieces of text) which match the search query given the algorithms of the search engine.

Rank of a result The relevance of a database equation, relative to the query. It is described by a number. Results with a higher rank are located higher in the list of results. [16, 18, 38, 49]

Subformula Pieces of a formula. For example, $a + b^{2+c}$ has two subformulae: b^{2+c} and $2 + c$. The expression $a + b^{2+c}$ is itself a subformula of $2/(a + b^{2+c})$. [26, 49]

Meaning-based search Searches based on the true mathematical meaning of an equation, with the goal of identifying all equations which have equivalent or similar meanings to the query equation. For example, the meaning of $\sinh(x)$ and $(1/2)(e^x - e^{-x})$ are equal. [18] Meaning-based search engines which are now web-ready obtain only part of this goal, but they do account for some mathematical meaning. [8, 24, 45]

Presentation-based search Bases search results on the visual appearance of an equation. For example, $1 + x^2$ is visually different from $x^2 + 1$. The equations $1 + x^2/2$ and $1 + y^2/2$ match visually, so would count as similar equations in a presentation-based search. Some presentation-based search engines account for algebraic rearrangement of terms (e.g., $1 + x^2$ versus $x^2 + 1$), and others do not. [21, 24, 25, 63]

Appendix B

Nifty Math Tools

There is a lot of scientific software which is publicly available, but it can be hard to find unless a researcher knows exactly what to look for.

This appendix covers a sample of interesting math tools, and gives references to a few good databases and lists of tools. These were discovered along the course of this research, and the author thought them nifty: the list is by no means exhaustive.

B.1 Math tools: A Sample of Interesting Tools

Below is a subset of Mathematical software tools: it includes a few which I thought might be of interest to physicists, but is intended to be a sample only. In the next section (section B.2), some resources for finding other math software are given.

B.1.1 Number Knowledge Lookup

Inverse Symbolic Calculator

Enter a decimal number. The tool tries to find combinations of mathematical constants which could make up that number. It ignores factors of 10; why this is so is not stated on the webpage. ¹ [118]

The On-Line Encyclopedia of Integer Sequences

Enter the first few numbers of an integer sequence; this tool will try to identify whether it is part of a formal sequence, and bring up references for it. It is recommended, on the “hints” page, that about 6 terms in the sequence be used, starting with the second or third term. [119]

¹It is not clear to me why having 10 of something should be special, as opposed to having, say 8 of it: why are factors of 10 more important than are factors of 8, or 4, or 3? However, perhaps the factor of 10 is special, given our base-10 number system, from number theory.

Encyclopedia of Combinatorial Structures

This site is like the Online Encyclopedia of Integer Sequences, except that it focuses on sequences which are relevant to decomposable combinatorial structures. Search options include: the first few terms from a sequence, keyword, generating function, or closed form. [120]

B.1.2 Exploration and/or Lookup

The KnotPlot Site

Contains a collection of knots and links images, and a link to a software by which to visualize and manipulate three- and four-dimensional mathematical knots. [121]

Interactive Geometry Software: Cinderella

Designed for university-level math, this free downloadable software helps users explore geometry visually. [122]

B.1.3 Proof Assistants

Mathematical proof assistants include the semi-interactive theorem prover Coq [123] and the coding language Mizar [124]. [18]

B.1.4 Special and Elementary Function databases

Aside from the equation-searchable special function databases, which were mentioned in the main paper (DLMF, Wolfram Alpha), there are some keyword-searchable special function databases. One example is the Wolfram Functions website, where a special function's entry can contain many different definitions and written forms. A keyword by which to search for more such sites is "Digital Mathematics Library" (DML).

There is also an in-development upgrade to searching special functions and orthogonal polynomials, which includes the content of the DLMF and a number of special functions textbooks. It is planned to be easily searchable [125]; since it aims to improve upon the DLMF, hopefully this will include a math-aware search engine which can parse equation queries. Community comments on each formula's page will also be possible. A demo of it is available at [126] or [127].

B.1.5 Modeling tools

Many modeling tools are listed on swMath.org. A few examples are the modeling-oriented coding language Modelica (thank you to Erik Schnetter for pointing out this tool) [128], and Wolfram's Systems Modeler platform [129].

B.2 Resources to find other tools

B.2.1 Database of Custom Scientific & Mathematical Software

swMath.org [130] collects subject-browsable software which has been made for researchers. It can also be searched by keyword; software name, authors, and description; programming language; and category of research (“MSC Classification”). Covered subjects include:

- Quantum theory
- Astronomy and astrophysics
- Approximation and expansions
- Numerical analysis
- Other math and physics subjects

References are included to publications which discuss or use each software; the number of such “citations” of a software is presented in search results.

Some of the information in swMath.org is outdated, in that the software is no longer available and/or operational. But, it is a great place to start for researchers who seek knowledge about whether software solutions exist for their particular problem. For general browsing, it is also very informative.

B.2.2 Lists of Math Tools

The following places have good lists of references to other tools, and/or to other lists. They are ones which I happened upon; many other good lists undoubtedly exist.

- [131]: a resource list by the developer of the Inverse Symbolic Calculator (see section B.1.1).
- [132]: the section “Specialized Mathematical Databases”, and the subsequent couple of sections, starting at page 120.

Bibliography

- [1] J. Maldacena, “The large- n limit of superconformal field theories and supergravity,” *International Journal of Theoretical Physics*, vol. 38, no. 4, pp. 1113–1133, 1999. [Online]. Available: <http://dx.doi.org/10.1023/A:1026654312961>
- [2] V. E. Hubeny, “The AdS/CFT Correspondence,” *Class. Quant. Grav.*, vol. 32, no. 12, p. 124010, 2015.
- [3] W. Unruh, “Dumb holes: analogues for black holes,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1877, pp. 2905–2913, 2008. [Online]. Available: <http://rsta.royalsocietypublishing.org/content/366/1877/2905>
- [4] O. Lahav, A. Itah, A. Blumkin, C. Gordon, S. Rinott, A. Zayats, and J. Steinhauer, “Realization of a sonic black hole analog in a bose-einstein condensate,” *Physical Review Letters*, vol. 105, p. 240401, Dec 2010. [Online]. Available: <http://link.aps.org/doi/10.1103/PhysRevLett.105.240401>
- [5] S. Carroll, “Holes of silence,” Blog, Jun. 2009, accessed Aug 2016. [Online]. Available: <http://www.preposterousuniverse.com/blog/2009/06/12/holes-of-silence/>
- [6] J. D. Wells, *Effective Theories in Physics: From Planetary Orbits to Elementary Particle Masses*. Springer-Verlag Berlin Heidelberg, 2012, iSBN 978-3-642-34891-4.
- [7] E. Seiler and I.-O. Stamatescu, *Approaches to Fundamental Physics: An assessment of current theoretical ideas*, ser. Lecture Notes in Physics, E. Seiler and I.-O. Stamatescu, Eds. Springer-Verlag Berlin Heidelberg, 2007, vol. 721, iSBN: 978-3-540-71115-5 or 978-3-642-09021-9.
- [8] M. Kohlhase, B. A. Matican, and C.-C. Prodescu, *MathWebSearch 0.5: Scaling an Open Formula Search Engine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 342–357. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-31374-5_23
- [9] FIZ Karlsruhe Leibniz Institute for Information Infrastructure GmbH (FIZ Karlsruhe), European Mathematical Society (EMS), FIZ Karlsruhe, the Heidelberg Academy of Sciences and Humanities, and Springer, “About zbMATH,” 2016, accessed Mar 2016. [Online]. Available: <https://zbmath.org/about/>
- [10] “The European Digital Mathematics Library (EuDML),” v2.1.5. Accessed Mar 2016. [Online]. Available: <https://eudml.org/>
- [11] Springer, “Springer Innovations: LaTeXSearch.com,” accessed Aug 8, 2016. [Online]. Available: <http://www.springer.com/gp/partners/society-zone-issues/springer-innovations--latexsearch-com/4516>
- [12] “NIST Digital Library of Mathematical Functions,” online companion to [133]. Release 1.0.11 of 2016-06-08. Accessed Aug 9 2016. [Online]. Available: <http://dlmf.nist.gov/>
- [13] R. F. Boisvert, C. W. Clark, D. W. Lozier, and F. W. Olver, “A special functions handbook for the digital age,” *Notices of the American Mathematical Society*, vol. 58, pp. 905–911, Aug. 2011. [Online]. Available: www.ams.org/notices/201107/rtx110700905p.pdf
- [14] E. Weisstein, *Computable Data, Mathematics, and Digital Libraries in Mathematica and Wolfram—Alpha*. Cham: Springer International Publishing, 2014, pp. 26–29. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08434-3_3

- [15] Wolfram Research, “Wolfram alpha,” accessed August 2016. [Online]. Available: <https://www.wolframalpha.com/>
- [16] R. Zanibbi and A. Orakwue, *Intelligent Computer Mathematics: International Conference, CICM 2015, Washington, DC, USA, July 13-17, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. Math Search for the Masses: Multimodal Search Interfaces and Appearance-Based Retrieval, pp. 18–36. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-20615-8_2
- [17] “(Uni)quation alpha,” accessed Mar 2016. [Online]. Available: <http://uniquation.com/en/>
- [18] F. Guidi and C. Sacerdoti Coen, *Intelligent Computer Mathematics: International Conference, CICM 2015, Washington, DC, USA, July 13-17, 2015, Proceedings*. Cham: Springer International Publishing, 2015, ch. A Survey on Retrieval of Mathematical Knowledge, pp. 296–315. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-20615-8_20
- [19] Q. Zhang and A. Youssef, *An Approach to Math-Similarity Search*. Cham: Springer International Publishing, 2014, pp. 404–418. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08434-3_29
- [20] M. Kohlhase and I. Sukan, *A Search Engine for Mathematical Formulae*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 241–253. [Online]. Available: http://dx.doi.org/10.1007/11856290_21
- [21] R. Zanibbi, K. Davila, A. Kane, and F. W. Tompa, “The tangent search engine: Improved similarity metrics and scalability for math formula search,” *CoRR*, vol. abs/1507.06235, 2015. [Online]. Available: <http://arxiv.org/abs/1507.06235>
- [22] M. Líška, “Building the ultimate math search engine,” Brno, 2015. [Online]. Available: http://is.muni.cz/th/255768/fi_r/?lang=en
- [23] M. Hazewinkel, “Mathematical knowledge management is needed,” *eprint arXiv:cs/0410055*, Oct. 2004, keynote speech at the November, 2003 MKM meeting at Herriott-Watt, Edinburg, UK.
- [24] M.-Q. Nghiem, G. Y. Kristianto, G. Topić, and A. Aizawa, *Which One Is Better: Presentation-Based or Content-Based Math Search?* Cham: Springer International Publishing, 2014, pp. 200–212. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08434-3_15
- [25] M. Líška, “Evaluation of mathematics retrieval,” Master’s thesis, Masaryk University, Faculty of Informatics, Brno, 2013. [Online]. Available: http://is.muni.cz/th/255768/fi_m/
- [26] O. Urazmetov, “ $5e^{x+y}$: A Math Aware Search Engine for CDS,” Ph.D. dissertation, Ecole Polytechnique, Lausanne, Mar 2014, presented 09 Apr 2014. [Online]. Available: <https://cds.cern.ch/record/1670010>
- [27] Wolfram Research, “ChebyshevT,” Wolfram Functions Website, Aug. 2001, accessed Aug 2016. [Online]. Available: <http://functions.wolfram.com/Polynomials/ChebyshevT/02/0001/>
- [28] “Wolfram alpha chebyshev polynomial,” Wolfram Research, Wolfram Alpha, accessed Aug 2016. [Online]. Available: <https://www.wolframalpha.com/input/?i=chebyshev+polynomial>
- [29] D. Hasegan, “Sorted unification in mathwebsearch,” Jacobs University - School of Engineering and Science, Mar. 2014. [Online]. Available: <https://svn.eecs.jacobs-university.de/svn/eecs/archive/bsc-2014/hasegan.pdf>
- [30] M. Iancu, M. Kohlhase, and C. Prodescu, *Representing, Archiving, and Searching the Space of Mathematical Knowledge*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 26–30. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-44199-2_5
- [31] KWARC group at Jacobs University, “MathWebSearch: Searching math on the web,” accessed Mar 2016. [Online]. Available: <http://search.mathweb.org/>
- [32] Springer, “Springer Link,” accessed Mar 2016. [Online]. Available: <http://link.springer.com/>
- [33] Foundational Questions Institute (FQXi), “Trick or truth: the mysterious connection between physics and mathematics,” accessed July 2016. [Online]. Available: <http://fqxi.org/community/essay/winners/2015.1>

- [34] R. Schuetzhold, “Fundamental quantum effects in the laboratory,” Colloquium, Aug 2012, accessed Aug 2016. [Online]. Available: <http://pirsa.org/12080045/>
- [35] “zbMATH,” accessed April 2016.
- [36] “Math WebSearch: a semantic search engine,” accessed Aug 2016. Powered by: Sentido Labs.com, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Jacobs University. [Online]. Available: <http://search.mathweb.org/sentido/>
- [37] J. H. Davenport, *MKM from Book to Computer: A Case Study*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 17–29. [Online]. Available: http://dx.doi.org/10.1007/3-540-36469-2_2
- [38] R. Zanibbi and D. Blostein, “Recognition and retrieval of mathematical expressions,” *International Journal on Document Analysis and Recognition (IJДАР)*, vol. 15, no. 4, pp. 331–357, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10032-011-0174-4>
- [39] D. Stalnaker, “Math expression retrieval using symbol pairs in layout trees,” Thesis, David Stalnaker, <http://scholarworks.rit.edu/theses/5533>, Aug. 2013.
- [40] K. D. V. Wangari, R. Zanibbi, and A. Agarwal, “Discovering real-world use cases for a multimodal math search interface,” in *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, ser. SIGIR ’14. New York, NY, USA: ACM, 2014, pp. 947–950. [Online]. Available: <http://doi.acm.org/10.1145/2600428.2609481>
- [41] Cornell University Library, “Unpacking retrieved papers,” accessed Aug 2016. [Online]. Available: <https://arxiv.org/help/unpack>
- [42] Springer, “About LaTeXSearch,” accessed Mar 2016. [Online]. Available: <http://latexsearch.com/static/about.jsp>
- [43] “Min,” accessed Aug 2016. [Online]. Available: <http://saskatoon.cs.rit.edu/min/>
- [44] A. Youssef, *Mathematical Knowledge Management: 5th International Conference, MKM 2006, Wokingham, UK, August 11-12, 2006. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, ch. Roles of Math Search in Mathematics, pp. 2–16. [Online]. Available: http://dx.doi.org/10.1007/11812289_2
- [45] A. Aizawa, M. Kohlhase, I. Ounis, and M. Schubotz, “NTCIR-11 math-2 task overview,” in *NTCIR*, 2014.
- [46] M. Líška, P. Sojka, and M. Růžička, *Intelligent Computer Mathematics: International Conference, CICM 2014, Coimbra, Portugal, July 7-11, 2014. Proceedings*. Cham: Springer International Publishing, 2014, ch. Math Indexer and Searcher Web Interface, pp. 444–448. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08434-3_36
- [47] J. Rákosník, P. Stanchev, and R. Pavlov, “Recent developments in digital mathematics libraries,” in *Proceedings of the Digital Presentation and Preservation of Cultural and Scientific Heritage, Veliko Tarnovo, Bulgaria, 18.9.2014 - 21.9.2014*, vol. IV. Institute of Mathematics and Informatics - BAS, 2014, pp. 59–66.
- [48] M. Kohlhase, B. A. Matican, and C. C. Prodescu, “Mathwebsearch 0.5: Scaling an open formula search engine,” Slides on web, Jul. 2012, accessed Aug 2016. [Online]. Available: <http://kwarc.info/kohlhase>
- [49] P. Sojka and M. Líška, “Indexing and searching mathematics in digital libraries – architecture, design and scalability issues,” in *Intelligent Computer Mathematics Lecture Notes in Computer Science, 2011, Volume 6824/2011*, J. U. F. R. James H. Davenport, William M. Farmer, Ed. Berlin / Heidelberg: Springer, 2011, pp. 228–243. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-22673-1_16
- [50] Document and Pattern Recognition Lab (DPRL) at Rochester Institute of Technology, “Software: Demos and source code,” 2016, accessed Aug 2016. [Online]. Available: <https://www.cs.rit.edu/~dprl/Software.html>
- [51] S. Pakin, *The comprehensive LaTeX Symbol List*, Nov. 2015, page 220, Table 517.

- [52] A. Oviedo, N. Kasioumis, and K. Aberer, “ $5e^{x+y}$: Searching over Mathematical Content in Digital Libraries,” in *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '15. New York, NY, USA: ACM, 2015, pp. 283–284. [Online]. Available: <http://doi.acm.org/10.1145/2756406.2756953>
- [53] M. Kohlhase and I. Sukan, *A Search Engine for Mathematical Formulae*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 241–253. [Online]. Available: http://dx.doi.org/10.1007/11856290_21
- [54] M.-Q. Nghiem, G. Y. Kristianto, and A. Aizawa, “Using MathML parallel markup corpora for semantic enrichment of mathematical expressions,” *IEICE TRANSACTIONS on Information and Systems*, vol. E96-D, no. 8, pp. 1707–1715, Aug. 2013.
- [55] Wolfram Research, “Erfc - complementary error function,” Wolfram Functions Website, accessed Aug 2016. [Online]. Available: <http://functions.wolfram.com/GammaBetaErf/Erfc/02/0001/>
- [56] DLMF, “18 orthogonal polynomials: Classical orthogonal polynomials,” Jun. 2016, accessed Aug 2016. [Online]. Available: <http://dlmf.nist.gov/18.3#T1.t1>
- [57] “7 error functions, dawsons and fresnel integrals: Properties,” DLMF, Jun. 2016, accessed Aug 2016. [Online]. Available: <http://dlmf.nist.gov/7.2#i>
- [58] Wolfram Research, “Wolfram alpha - erfc(x),” Oct. 2001, accessed Aug 2016. [Online]. Available: [https://www.wolframalpha.com/input/?i=erfc\(z\)](https://www.wolframalpha.com/input/?i=erfc(z))
- [59] M. Kohlhase, B. A. Matican, and C.-C. Prodescu, *MathWebSearch 0.5: Scaling an Open Formula Search Engine*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 342–357. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-31374-5_23
- [60] “zbMATH: MathWebSearch search page,” accessed Mar 2016. [Online]. Available: <http://search.mathweb.org/zbl>
- [61] A. Youssef, “Advances in math search,” *PAMM*, vol. 7, no. 1, pp. 1 010 501–1 010 502, 2007. [Online]. Available: <http://dx.doi.org/10.1002/pamm.200700111>
- [62] R. Zanibbi, K. Davila, A. Kane, and F. W. Tompa, “Multi-stage math formula search: using appearance-based similarity metrics at scale,” in *SIGIR 2016 July 18-20, Pisa, Italy*, no. ISBN 978-1-4503-4069-4/16/07. ACM, 2016.
- [63] “Tangent,” accessed Aug 2016. [Online]. Available: <http://saskatoon.cs.rit.edu/tangent/>
- [64] K. Davila, A. Kane, R. Zanibbi, and F. W. Tompa, “Tangent-3 at the NTCIR-12 MathIR task,” in *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, June 7 - 10, Tokyo Japan, 2016*.
- [65] N. Pattaniyil and R. Zanibbi, “Combining TF-IDF text retrieval with an inverted index over symbol pairs in math expressions: The Tangent math search engine at NTCIR 2014,” in *NTCIR-11, 2014*.
- [66] M. Ružička, P. Sojka, and M. Liška, “Math indexer and searcher under the hood: Fine-tuning query expansion and unification strategies,” in *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, June 7 - 10, Tokyo Japan, 2016*.
- [67] “Uniquation: News,” accessed Mar 2016. [Online]. Available: <https://twitter.com/uniquation>
- [68] R. Zanibbi, A. Aizawa, M. Kohlhase, I. Ounis, G. Topić, and K. Davila, “NTCIR-12 MathIR task overview,” in *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, June 7 - 10 2016, Tokyo Japan, 2016*.
- [69] M. Liška, P. Sojka, and M. Ružička, “Combining text and formula queries in math information retrieval: Evaluation of query results merging strategies,” *CoRR*, vol. abs/1508.01929, 2015. [Online]. Available: <http://arxiv.org/abs/1508.01929>

- [70] K. Davila, “Appearance-based retrieval of mathematical notation in documents and lecture videos,” in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '16. New York, NY, USA: ACM, 2016, pp. 1165–1165. [Online]. Available: <http://doi.acm.org/10.1145/2911451.2911477>
- [71] T. T. Nguyen, K. Change, and S. C. Hui, “A math-aware search engine for math question answering system,” in *CIKM'12, October 29 - November 2, 2012, Maui, HI USA*, ACM 978-1-4503-1156-4/12/10, 2012.
- [72] O. A. Nevzorova, N. Zhiltsov, A. Kirillovich, and E. Lipachev, *OntoMath PRO Ontology: A Linked Data Hub for Mathematics*. Cham: Springer International Publishing, 2014, pp. 105–119. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-11716-4_9
- [73] D. Landy and R. L. Goldstone, “How abstract is symbolic thought?” *Journal of experimental psychology: learning, memory, and cognition*, vol. 33, no. 4, pp. 720–733, Jul. 2007.
- [74] M. S. Reichenbach, A. Agarwal, and R. Zanibbi, “Rendering expressions to improve accuracy of relevance assessment for math search,” in *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2014, pp. 851–854.
- [75] K. D. V. Wangari, “Discovering real-world usage scenarios for a multimodal math search interface,” Thesis, Rochester Institute of Technology, 2013. [Online]. Available: <http://scholarworks.rit.edu/theses/7808>
- [76] A. Aizawa, M. Kohlhase, and I. Ounis, “NTCIR-10 math pilot task overview,” in *NTCIR*, 2013.
- [77] M. Kohlhase and C.-C. Prodescu, “MathWebSearch at NTCIR-10,” in *NTCIR*, 2013.
- [78] S. Bureau, “RIT graduates create award-winning math search engine,” University News, Feb 2015, accessed Mar 2016. [Online]. Available: <https://www.rit.edu/news/story.php?id=51200>
- [79] Wolfram Research, “The Wolfram Functions Site,” 2016, accessed Apr 2016. [Online]. Available: <http://functions.wolfram.com/>
- [80] C. Sasarak, K. Hart, R. Pospesel, D. Stalnaker, L. Hu, R. Livolsi, S. Zhu, and R. Zanibbi, “min: A multimodal web interface for math search,” hCIR 2012, Oct 4 - 5, 2012, Cambridge, Massachusetts, USA.
- [81] R. Zanibbi, Aug 2016, personal Communication.
- [82] RIT lab, “Multimodal math input with min,” accessed Aug 2016. [Online]. Available: <https://www.cs.rit.edu/~rlaz/files/minDemo.mp4>
- [83] M. Suzuki, “InftyReader version 3.0 - outline of the software,” web, 2015, accessed Aug 2016. [Online]. Available: <http://www.inftyproject.org/download/AboutInftyReaderE.txt>
- [84] Science Accessibility Net, “InftyReader version 3.0 - outline of the software,” web, Jul 2016, accessed Aug 2016. [Online]. Available: <http://www.sciaccess.net/en/InftyReader/>
- [85] “Infty Project - software,” web, accessed Aug 2016. [Online]. Available: <http://www.inftyproject.org/en/software.html#InftyReader>
- [86] Springer, “LaTeX Sandbox,” accessed Apr 2016. [Online]. Available: <http://www.latexsearch.com/sandbox.do>
- [87] W3C Math Working Group, “Tools,” Jan 2016, accessed Apr 2016. [Online]. Available: <https://www.w3.org/Math/wiki/Tools>
- [88] Wolfram Research, “Render MathML from Text,” 2016, accessed Apr 2016. [Online]. Available: <http://www.mathmlcentral.com/Tools/FromMathML.jsp>
- [89] “MathML Central - convert expression to MathML,” Wolfram Research, accessed Aug 2016. [Online]. Available: <http://www.mathmlcentral.com/Tools/ToMathML.jsp>

- [90] P. Hunter, “MathToWeb,” accessed Aug 2016. [Online]. Available: <https://www.mathtowebonline.com/>
- [91] J. Oden, “LaTeX2MathML - Online demonstration,” accessed Aug 2016. [Online]. Available: <http://latex2mathml.freewebmaster.fr/demo.php>
- [92] “TeXZilla - Live Demo,” accessed Aug 2016. [Online]. Available: <https://fred-wang.github.io/TeXZilla/>
- [93] “MathML to TeX Online Translator,” accessed Aug 2016. [Online]. Available: <http://www.orcca.on.ca/MathML/texmml/mmltotex.html>
- [94] D. Carlisle, “web-xslt - pmml2tex,” 2010, accessed Aug 2016. [Online]. Available: <https://github.com/davidcarlisle/web-xslt/tree/master/pmml2tex>
- [95] David Carlisle, “Thread: I want to convert MathML to LaTeX,” Comment on StackExchange, Dec 2012, accessed Aug 2016. [Online]. Available: <http://tex.stackexchange.com/questions/85639/i-want-to-convert-mathml-to-latex>
- [96] user Zelo, “Thread: MathML to LaTeX conversion,” Comment on StackOverflow, Jan 2014, accessed Aug 2016. [Online]. Available: <http://stackoverflow.com/questions/6587864/mathml-to-latex-conversion>
- [97] R. Pospesel, K. Hart, L. Hu, S. Zhu, D. Stalnakar, C. Sasarak, R. LiVolsi, A. Orakwue, and R. Zanibbi, “Overview of the min interface,” Feb 2014, accessed Mar, 2016. [Online]. Available: http://saskatoon.cs.rit.edu/min_instructions/
- [98] X. Hu, L. Gao, X. Lin, Z. Tang, X. Lin, and J. B. Baker, “Wikimirs: A mathematical information retrieval system for wikipedia,” in *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '13. New York, NY, USA: ACM, 2013, pp. 11–20. [Online]. Available: <http://doi.acm.org/10.1145/2467696.2467699>
- [99] P. Libbrecht and E. Melis, *Mathematical Software - ICMS 2006: Second International Congress on Mathematical Software, Castro Urdiales, Spain, September 1-3, 2006. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, ch. Methods to Access and Retrieve Mathematical Content in ActiveMath, pp. 331–342. [Online]. Available: http://dx.doi.org/10.1007/11832225_33
- [100] “Latex search beta,” accessed Apr 2016. [Online]. Available: <http://latexsearch.com/>
- [101] R. Hambasan, M. Kohlhase, and C.-C. Prodescu, “MathWebSearch at NTCIR-11,” in *NTCIR*, 2014.
- [102] NIST DLMF, “Guide to Searching the DLMF,” Aug 2015, v1.0.10. Accessed April 2016. [Online]. Available: <http://dlmf.nist.gov/help/search>
- [103] “Software index: NIST Digital Library of Mathematical Functions,” online companion to [133]. Release 1.0.11 of 2016-06-08. Accessed Aug 9 2016. [Online]. Available: <http://dlmf.nist.gov/software/>
- [104] “Frequently asked questions,” Wolfram Alpha, Wolfram Research, accessed Aug 2016. [Online]. Available: <http://www.wolframalpha.com/faqs7.html>
- [105] The Wolfram—Alpha Math Team, “Talk to Wolfram—Alpha in TeX,” Sep 2010, accessed Aug 2016. [Online]. Available: <http://blog.wolframalpha.com/2010/09/30/talk-to-wolframalpha-in-tex/>
- [106] A. Oviedo, Personal Communication, Mar 2016.
- [107] “DPRL,” github, Apr 2015, accessed Apr 2016. [Online]. Available: <https://github.com/DPRL>
- [108] “MathWebSearch project,” Apr. 2013, accessed April 2016. [Online]. Available: <https://trac.mathweb.org/MWS/wiki>
- [109] The KWARC group at Jacobs University, “MWS,” github, Nov 2015, accessed April 2016. [Online]. Available: <https://github.com/KWARC/mws>

- [110] H. Stamerjohanns and M. Kohlhase, *Transforming the arXiv to XML*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 574–582. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-85110-3_46
- [111] “WebMIaS,” accessed: Apr. 2016. [Online]. Available: <https://mir.fi.muni.cz/webmiasdemo/>
- [112] “WebMIaS,” github, Feb. 2016, accessed April 2016. [Online]. Available: <https://mir.fi.muni.cz/webmias/>
- [113] NTCIR-12, “Data - NTCIR-12 MathIR Task,” accessed Feb 2016. [Online]. Available: <http://ntcir-math.nii.ac.jp/data/>
- [114] “MIaS,” github, Mar. 2016, accessed April 2016. [Online]. Available: <https://github.com/martinliska/MIaS>
- [115] “WebMIaS,” github, Mar. 2016, accessed April 2016. [Online]. Available: <https://github.com/martinliska/WebMIaS>
- [116] “MIASMath,” github, Feb. 2016, accessed April 2016. [Online]. Available: <https://github.com/martinliska/MIASMath>
- [117] Document and Pattern Recognition Lab (DPRL) at Rochester Institute of Technology, “Tangent,” github, Mar 2014, accessed Mar 2016. [Online]. Available: <https://github.com/DPRL/tangent>
- [118] N. Singer, A. Shouldice, L. Ye, T. Daske, P. Dobcsanyi, D. Manna, O.-Y. Chan, and J. Borwein, “isc+: Inverse symbolic calculator,” accessed August 2016. [Online]. Available: <https://isc.carma.newcastle.edu.au/>
- [119] The OEIS Foundation Inc., “The on-line encyclopedia of integer sequences,” Aug. 2016, accessed August 2016. [Online]. Available: <http://oeis.org/>
- [120] S. Petit, A. Darrasse, and F. Chyzak, “Encyclopedia of combinatorial structures (introduction),” Mar. 2011, accessed April 2016. [Online]. Available: <http://algo.inria.fr/encyclopedia/>
- [121] R. G. Scharein, “The KnotPlot Site,” Jan. 2014, accessed April 2016. [Online]. Available: <http://www.knotplot.com/>
- [122] “The interactive geometry software cinderella,” Sep. 2013, accessed April 2016. [Online]. Available: <http://www.cinderella.de/tiki-index.php>
- [123] “The coq proof assistant,” accessed April 7, 2016. [Online]. Available: <https://coq.inria.fr/>
- [124] “Mizar home page,” 2016, accessed April 2016. [Online]. Available: <http://www.mizar.org/>
- [125] H. S. Cohl, M. A. McClain, B. V. Saunders, M. Schubotz, and J. C. Williams, *Digital Repository of Mathematical Formulae*. Springer International Publishing, 2014, pp. 419–422. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-08434-3_30
- [126] “Digital repository of mathematical formulae,” Apr. 2014, accessed July 2016. [Online]. Available: http://gw32.iu.xsede.org/index.php/Main_Page
- [127] “Main page - nist digital repository of mathematical formulae,” Nov. 2015, accessed July 2016. [Online]. Available: http://drmf.wmflabs.org/wiki/Main_Page
- [128] The Modelica Association, “Modelica,” 2016, accessed April 2016. [Online]. Available: <https://www.modelica.org/>
- [129] Wolfram Research, “Wolfram system modeler,” 2016, accessed April 2016. [Online]. Available: <http://www.wolfram.com/system-modeler/>
- [130] “swMATH,” accessed April 2016. [Online]. Available: <http://www.swmath.org/>
- [131] J. Borwein, “Jon borwein’s mathematics portal,” university of Newcastle. Accessed April 2016. [Online]. Available: <https://carma.newcastle.edu.au/jon/portal.html>

- [132] National Research Council, *Developing a 21st Century Global Library for Mathematics Research*. Washington, D.C.: The National Academies Press, 2014.
- [133] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, Eds., *NIST Handbook of Mathematical Functions*. New York, NY: Cambridge University Press, 2010, print companion to [12].