The MathDeck Formula Editor: Interactive Formula Entry Combining LaTeX, Structure Editing, and Search

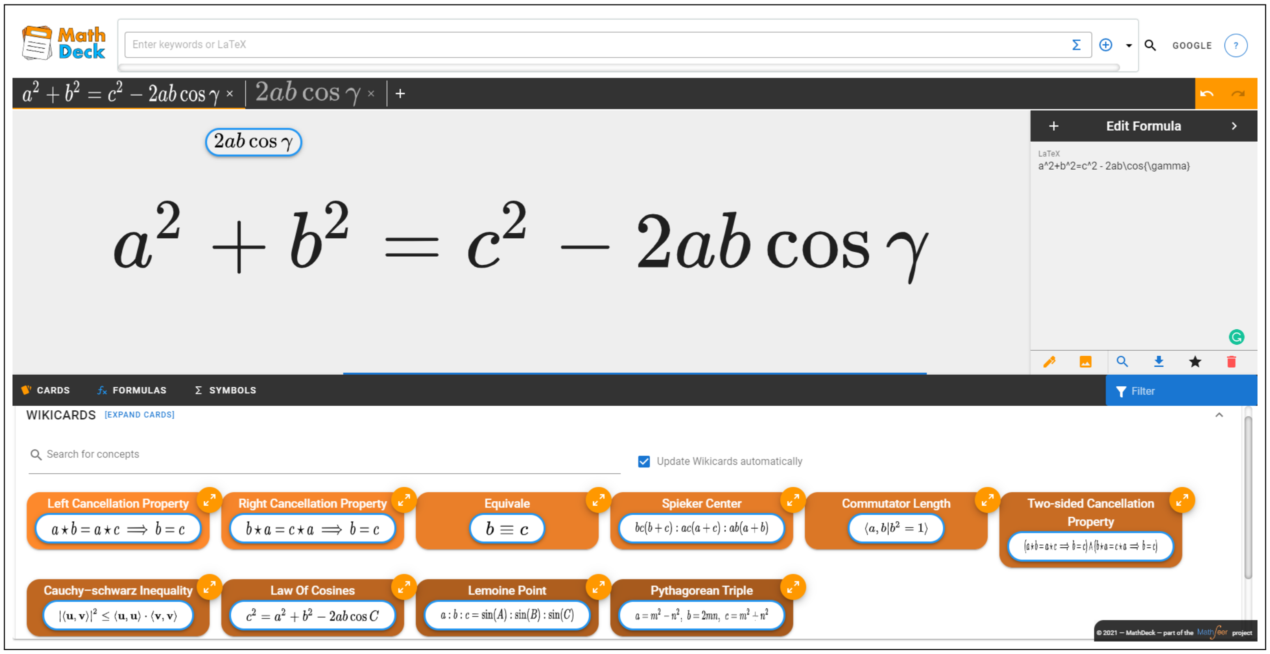
Yancarlos Diaz
yxd3549@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Gavin Nishizawa
ghn6069@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Behrooz Mansouri
bm3302@rit.edu
Rochester Institute of Technology
Rochester, NY, USA

Kenny Davila
kennydav@buffalo.edu
University at Buffalo
Buffalo, NY, USA

Richard Zanibbi
rxzvcs@rit.edu
Rochester Institute of Technology
Rochester, NY, USA


Math Deck
Enter keywords or LaTeX
Σ
GOOGLE
?
$a^2+b^2=c^2-2ab\cos\gamma$ ×
$2ab\cos\gamma$ ×
+
$2ab\cos\gamma$
$$a^2+b^2=c^2-2ab\cos\gamma$$
+
Edit Formula
›
LaTeX
CARDS
FORMULAS
Σ SYMBOLS
Filter
WIKICARDS
[EXPAND CARDS]
Search for concepts
Left Cancellation Property
$a\star b=a\star c\implies b=c$
Right Cancellation Property
$b\star a=c\star a\implies b=c$
Equivale
$b\equiv c$
Spieker Center
$bc(b+c):ac(a+c):ab(a+b)$
Commutator Length
$\langle a,b|b^2=1\rangle$
Two-sided Cancellation Property
$(a\star b=a\star c\implies b=c)\wedge(b\star a=c\star a\implies b=c)$
Cauchy–schwarz Inequality
$|\langle\mathbf{u},\mathbf{v}\rangle|^2\leq\langle\mathbf{u},\mathbf{u}\rangle\cdot\langle\mathbf{v},\mathbf{v}\rangle$
Law Of Cosines
$c^2=a^2+b^2-2ab\cos C$
Lemoine Point
$a:b:c=\sin(A):\sin(B):\sin(C)$
Pythagorean Triple
$a=m^2-n^2,\ b=2mn,\ c=m^2+n^2$
© 2021 – MathDeck – part of the MathSeer project

a^2+b^2=c^2 - 2ab\cos{\gamma}

☑ Update Wikicards automatically

Figure 1: MathDeck Search Interface. Formulas are created using the rendered formula, the LaTeX string at right, formulas in tabs, and 'chips' on the canvas and in the 'deck' at bottom. Cards in the deck update for the current formula, and may be expanded to reveal descriptions. Formulas and keywords may be added to the query bar, and sent to a search engine.

ABSTRACT

Writing formulas in LaTeX can be difficult, especially for complex formulas. MathDeck simplifies LaTeX formula entry by: 1) allowing rendered formulas to be edited directly alongside their associated LaTeX strings, 2) helping build formulas from smaller ones, and 3) providing searchable formula cards with associated names and descriptions. Cards are searchable by formula and title.

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8095-9/21/05.
https://doi.org/10.1145/3411763.3451564

CCS CONCEPTS

• **Information systems → Search interfaces**; • **Mathematics of computing → Mathematical software**.

KEYWORDS

equation editor, structure editor, LaTeX, Mathematical Information Retrieval (MIR)

ACM Reference Format:
Yancarlos Diaz, Gavin Nishizawa, Behrooz Mansouri, Kenny Davila, and Richard Zanibbi. 2021. The MathDeck Formula Editor: Interactive Formula Entry Combining LaTeX, Structure Editing, and Search. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3411763.3451564

1 INTRODUCTION

Writing formulas in LaTeX is sometimes difficult, especially for novice users unfamiliar with the various commands to create special symbols (e.g., Greek letters) and structures such as fractions and square roots. Writing complex formulas in LaTeX can be tedious even for expert users, as one's attention needs to be divided between editing the LaTeX string and looking at the appearance of the rendered formula to locate where changes have been made, e.g., when trying to replace a small subexpression within a large formula, or moving part of the expression inside a fraction or other structure. On the other hand, for very small formulas, working with LaTeX strings directly may be easy, even for non-experts. Further, for well-known formulas it would be desirable to retrieve their LaTeX representation, rather than recreate the formula.

Rather than replace LaTeX, what we seek is a useful combination of the strengths of LaTeX editing with directly manipulating rendered formulas. To manage complexity and reduce effort, we also want the ability to work on formulas in smaller pieces, and support for math formula search in collections of LaTeX formulas.

To allow edits to be more directly viewed where they are made, we introduce a new LaTeX formula editor as part of the MathDeck search interface [6] (see Fig. 1). MathDeck is a *math-aware* search engine that allows queries containing keywords and formulas to be sent to various search engines. To allow multiple formulas to be edited concurrently and combined to construct larger formulas, we introduce a tabbing system to the formula editing canvas in MathDeck. The tabs behave similarly to those in web browsers, but our tabs hold subexpressions that may be used directly in editing formulas. This complements and extends the formula 'chips' used to hold and reuse math formulas in editing from the original MathDeck (in Fig. 1, there is a chip on the formula editing canvas holding '$2ab\cos\gamma$'). To better support retrieval of LaTeX formulas, we have reworked the interface elements and engine supporting formula search, including the addition of text search for formula names/concepts.

In the remainder of this paper we outline our design goals, and then discuss key features of the new MathDeck formula editor as they relate to addressing the problems described above. We conclude by summarizing our demonstration at CHI 2021.[1]

2 DESIGN GOALS

We had three main design goals in producing the new formula editor for MathDeck, which we describe below.

2.1 Building Large Formulas from Smaller Ones

As formulas get larger and more complex, it becomes harder to edit them. Users can become lost in a long LaTeX string when they are perhaps just looking to edit a small part of a formula. A possible solution to this problem is helping users build formulas in pieces. This requires an interface that allows the user to easily define, use, and combine those pieces. We attempt to address this in MathDeck with formula 'chips' and tabs (see Fig. 1).

2.2 Direct Manipulation of Rendered Formulas

In existing LaTeX formula editors, there is a disconnect between the LaTeX formula in text and the formula appearance, which is presented separately from the LaTeX string. If users see they have an error in the rendered version of the formula, they must search the LaTeX string for the location of the error. It would be easier if they could change the rendered formula image directly.

We introduce a novel formula structure editor in MathDeck [5], which allows editing operations to be applied directly to rendered formulas (see Fig. 2), and the user may alternate between using the structure editor and editing the full LaTeX string for an expression shown in the TeX panel (at right in Fig. 1). New operations include the ability to insert formula chips into the current formula, and to replace part of a formula image using an arbitrarily complex LaTeX string or formula chip. The structure editor handles cases where replacements or deletions break a formula into multiple subexpressions, in which case chips are produced for disconnected subexpressions, as illustrated in Fig. 3c.

2.3 Interactive Formula Search and Autocomplete

Entering complex formulas can be time-consuming even for expert users. Sometimes users wish to enter well-known formulas that they know a name for, and/or is tedious to type in LaTeX. With plaintext, this burden is often eased with the help of autocomplete. As the user types a word or sentence, they are presented with common words that match their input.

We provide both formula autocompletion and search-by-name in MathDeck. As the user enters formulas, cards containing formulas along with their names and descriptions (*formula concept cards*) are searched by formula, and returned in the card 'deck' at the bottom of the interface (see Figs. 1 and 4). Formula cards may also be searched by title, allowing a formula to be selected without typing LaTeX at all. More broadly, the cards provide a formula collection that can be searched and browsed using both formulas and text as queries.

3 FORMULA EDITING

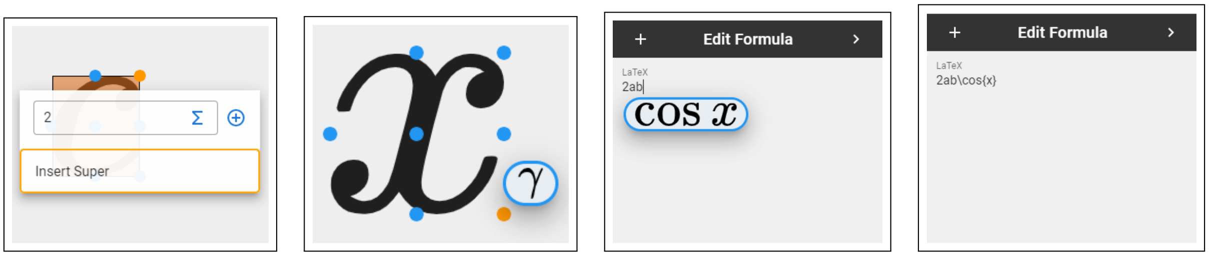
MathDeck introduces a novel formula structure editor that allows both a LaTeX formula string and its rendered image to be manipulated while editing. We summarize some key aspects below.

3.1 Formula Chips and Formula Tabs

MathDeck provides two ways for users to work in small pieces and then combine them. The first is through formula 'chips.' As the user enters formulas, the interface saves them as chips that can be moved around the interface. Multiple chips may sit on top of the formula editing canvas for later use (see Fig. 1), and appear in the deck at the bottom of the interface. Combined with the direct formula image manipulation the interface provides, chips may be used to build big expressions by dragging the chip onto control points located around symbols on the formula canvas, or into the editing panel for the full LaTeX string.

Users can favorite formula chips they use often, to make them more readily accessible (see Fig. 4), and chips can be downloaded as jpeg images for later use, and to share with others. Dragging a chip image file onto MathDeck restores the formula chip, as all chip

[1] Online MathDeck system: https://mathdeck.cs.rit.edu



(a) LaTeX insertion (b) Chip insertion (c) Chip over LaTeX panel (d) After dropping chip in (c)

Figure 2: Inserting Formulas. (a) Inserting '2' as an exponent to 'c' using a LaTeX string. (b) Inserting 'γ' in a chip as subscript of x. (c) A chip containing '$\cos x$' hovering over the LaTeX panel (at right in Fig. 1) - the cursor is at the end of the LaTeX string. (d) After dropping '$\cos x$' on the canvas, appending '$\cos x$' at the cursor location.

information is serialized in metadata fields of the formula jpeg file, including the associated LaTeX string. All chips have a context menu, allowing them to be duplicated, inserted in the LaTeX panel, copied as a chip onto the canvas, and deleted. Additionally, all formula chips may be searched using the currently selected search engine.

The second way MathDeck helps users work in small pieces is the new tab system. Tabs in MathDeck are analogous to browser tabs - each formula currently being edited has a different tab, and the formula in the currently selected tab appears on the editing canvas (see Fig. 1). Users can create as many tabs as they like, and they can move between them to quickly change context and work on a different formula. Additionally, users can directly drag a tab onto the canvas or onto the editing panel to insert the contents of the tab at the specified location, like they can do with formula chips. Fig. 1 show the MathDeck interface with two tabs. The first tab contains the formula '$a^2 + b^2 = c^2 - 2ab\cos\gamma$' and the second one contains the formula '$2ab\cos\gamma$.' The first tab is actually the result of dragging the second tab onto the canvas when it contained only the formula '$a^2 + b^2 = c^2 -$.'

3.2 Formula Structure Editing

We created operations to directly edit the rendered formula in-place, reducing the disconnect between the LaTeX string for a formula and the formula's appearance [5]. Operations include the ability to select subexpressions of formulas and view, edit, delete, move and copy into a new formula chip ('lift'), or replace the selection with a formula chip or a formula entered as LaTeX.

Rendered formula images may be edited directly, by using control points around symbols and selections to insert formulas provided as LaTeX strings or in formula chips and tabs (see Fig. 2). Chips and tabs can also be used to insert LaTeX in the LaTeX panel of the interface. In Fig. 3 we see LaTeX being entered after clicking a blue dot on the canvas (i.e., in-place). As the user types the formula, they can preview the rendered version. If LaTeX is invalid, this is indicated during typing by turning the text box border red. As shown in Fig. 3 part (d), users can drag a formula chip to replace a subexpression - inserting a chip into the center control point replaces the x with the contents of the chip, resulting in $c^2 - 2ab\cos\gamma$. Inserting at other locations will insert the formula at the position indicated, as seen in Fig. 2.

Additionally, for deletions and replacements where the formula tree is split into multiple subexpressions, subexpressions disconnected from the main expression after the operation are lifted up in-place into floating chips. This means that not only can symbols replace symbols, and expressions of the same shape be replaced, but also that selections and replacements may have different shapes. Parts (b) and (c) of Fig. 3 show this behavior. In this example, the user selects $a + b$ from $a^2 + b^2$ and replaces it with x, resulting in x^2 but also having the extra '2' appear floating as a formula chip.

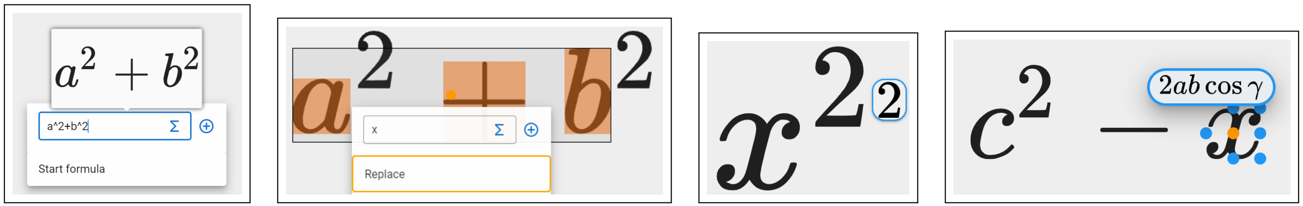
4 FORMULA CONCEPT CARDS AND INTERACTIVE FORMULA SEARCH

MathDeck provides the user with formula concept cards [2] in the deck area of the interface, which allows users to search and browse formulas and their associated information while editing. The cards contain a formula chip, a title, and a short, focused description (see Fig. 4). Card data is currently taken from Wikipedia (via Wikidata). The cards reduce entry effort when formulas exist in the collection, and enrich the formula entry process with contextual information that allows users to learn more about associated concepts. The importance of concept search for mathematical information has been reported previously [4]. This study showed that in a general search engine, 400,000 queries out of 27 million query records contained at least one distinctive mathematical term.

The visible concept cards are automatically updated as the user edits a formula. If the user enters '$a^2 + b^2$' for example, one of the concept cards they will see in the deck is that of the Pythagorean theorem ($a^2 + b^2 = c^2$). In this way, the concept cards act as a formula autocomplete. The card collection is searched interactively in real-time using a modified version of Tangent-CFT [3] that uses both formula appearance and syntax for retrieval.

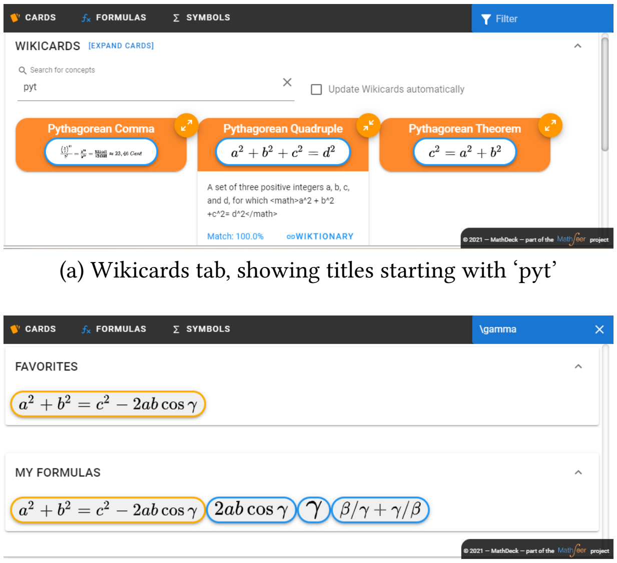
Users can also search concept cards by the name of the formula or concept they are looking for. As shown in Fig. 4, to find the Pythagorean Theorem, a user may enter 'pyt' in the formula deck search box to find the card. This avoids the need to enter LaTeX at all, if the desired formula is already in the card collection.

Users can edit and save cards for later use, and are provided with a list of their own formula cards that they have created or edited. We imagine this could be expanded and customized by users in the future, e.g., having different groups of formulas for different domains of mathematics, or grouping formulas for a specific math class (e.g., as prepared by the instructor). These groups might be



(a) Initial LaTeX (b) Replace 'a+b' by 'x' (c) Result from (b) (d) Replace 'x' by a formula

Figure 3: Replacing Subexpressions. (a) through (c) show a LaTeX formula entered on the canvas, after which a subexpression is replaced by a smaller one producing a chip for the disconnected exponent ('2'). (d) shows how to replace a selection with a chip: dragging the chip over the central dot. This operation produces '$c^2 - 2ab\cos\gamma$.' Chips may also be inserted at blue dots on the right, left, above, below, superscripted, and subscripted relative to a selected symbol or subexpression.


CARDS FORMULAS Σ SYMBOLS Filter

WIKICARDS [EXPAND CARDS]

Search for concepts
pyt

□ Update Wikicards automatically

Pythagorean Comma
Pythagorean Quadruple $a^2 + b^2 + c^2 = d^2$
Pythagorean Theorem $c^2 = a^2 + b^2$

A set of three positive integers a, b, c, and d, for which <math>a^2 + b^2 +c^2= d^2</math>

Match: 100.0% WIKTIONARY

© 2021 – MathDeck – part of the MathSeer project

(a) Wikicards tab, showing titles starting with 'pyt'

CARDS FORMULAS Σ SYMBOLS \gamma

FAVORITES

$a^2 + b^2 = c^2 - 2ab\cos\gamma$

MY FORMULAS

$a^2 + b^2 = c^2 - 2ab\cos\gamma$ $2ab\cos\gamma$ γ $\beta/\gamma + \gamma/\beta$

© 2021 – MathDeck – part of the MathSeer project

(b) Formulas tab, showing chips containing gamma

Figure 4: The Deck. (a) Shows the Wikicards tab, while (b) shows the tab containing formulas entered and favorited by the user. A third tab provides a symbol palette with chips for symbols and common subexpressions. Users may filter tabs: in (a) cards with a title beginning with 'pyt' are shown, and in (b) formulas containing 'γ' are shown.

represented as different card 'decks.' A user tagging system for cards is another interesting future direction.

5 IMPLEMENTATION

The interface is developed using Vue.js, and makes use of a customized MathJax library for rendering math. Tree transformations are naturally expressed in functional languages, and so the visual structure operations are implemented with ReasonML, which compiles OCaml code to Javascript. Web services used for the concept cards and session storage used by the interface are written in Python 3. LaTeX to MathML conversions are produced using LaTeXML[2].

Formula search for the Wikicards is provided by Tangent-CFT [3]. Tangent-CFT computes embeddings for mathematical formulas represented as both Symbol Layout Trees (SLTs) and Operator Trees (OPTs). SLTs capture formula appearance while OPTs represent operation syntax. We use three representations for each formula: SLT, OPT and SLT-Type. SLT-Type supports unification of variables, numbers, and other types by relabeling symbols with their associated type in the SLT. Tangent-CFT has been updated for MathDeck: instead of summing three vector representations and using a single vector representation for retrieval, we instead combine retrieval results from each individual representation with Reciprocal Rank Fusion (RRF) [1]. Cosine similarity is used to compute embedding vector rank scores.

As a user edits a formula, the SLT and OPT representations are extracted from the formula's LaTeX representation using LaTeXML, and the card collection is searched interactively using the modified version of Tangent-CFT described above. This allows users to view suggested and relevant formulae as they enter a formula.

6 EXPERIENCE

The demonstration will be performed in a virtual presentation session online, sharing a screencast of a running MathDeck session along with an image of the presenter. There will also be a chat channel monitored by the authors. The online MathDeck system will also be available for use[1] (Note: currently the program runs best in the Chrome browser). Participants will be encouraged to try the system during the session, and post questions and comments to the chat window.

We will have multiple 2-minute examples prepared, with each example illustrating a formula editing task. We will repeatedly present an example, and then field questions for 2-3 minutes. This will include an illustration of how our formula editing and search operations are integrated within the larger MathDeck search interface, which allows queries containing keywords and formulas to be submitted to search engines including Wolfram Alpha and Google.

ACKNOWLEDGMENTS

We wish to acknowledge Abishai Dmello, who developed the formula concept cards for his Master's thesis. This material is based upon work supported by the Alfred P. Sloan Foundation under Grant No. G-2017-9827 and the National Science Foundation (USA) under Grant No. IIS-1717997.

[2] https://dlmf.nist.gov/LaTeXML

REFERENCES

[1] Gordon V Cormack, Charles LA Clarke, and Stefan Buettcher. 2009. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. Association for Computing Machinery, New York, NY, USA, 758–759.

[2] Abishai Dmello. 2019. *Representing mathematical concepts associated with formulas using math entity cards*. Master’s thesis. Rochester Institute of Technology, USA.

[3] Behrooz Mansouri, Shaurya Rohatgi, Douglas W. Oard, Jian Wu, C. Lee Giles, and Richard Zanibbi. 2019. Tangent-CFT: An Embedding Model for Mathematical Formulas. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2019, Santa Clara, CA, USA, October 2-5, 2019*, Yi Fang, Yi Zhang, James Allan, Krisztian Balog, Ben Carterette, and Jiafeng Guo (Eds.). ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/3341981.3344235

[4] Behrooz Mansouri, Richard Zanibbi, and Douglas W. Oard. 2019. Characterizing Searches for Mathematical Concepts. In *19th ACM/IEEE Joint Conference on Digital Libraries, JCDL 2019, Champaign, IL, USA, June 2-6, 2019*, Maria Bonn, Dan Wu, J. Stephen Downie, and Alain Martaus (Eds.). IEEE, New York, NY, USA, 57–66. https://doi.org/10.1109/JCDL.2019.00019

[5] Gavin Nishizawa. 2020. *Visual Structure Editing of Math Formulas*. Master’s thesis. Rochester Institute of Technology, USA.

[6] Gavin Nishizawa, Jennifer Liu, Yancarlos Diaz, Abishai Dmello, Wei Zhong, and Richard Zanibbi. 2020. MathSeer: A Math-Aware Search Interface with Intuitive Formula Editing, Reuse, and Lookup. In *Advances in Information Retrieval - 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 12036)*, Joemon M. Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J. Silva, and Flávio Martins (Eds.). Springer, New York, NY, USA, 470–475. https://doi.org/10.1007/978-3-030-45442-5_60