Mathematical Information Retrieval: Search and Question Answering

Other titles in Foundations and Trends® in Information Retrieval

Information Discovery in E-commerce
Zhaochun Ren, Xiangnan He, Dawei Yin and Maarten de Rijke
ISBN: 978-1-63828-462-8

Fairness in Search Systems
Yi Fang, Ashudeep Singh and Zhiqiang Tao
ISBN: 978-1-63828-498-7

User Simulation for Evaluating Information Access Systems
Krisztian Balog and ChengXiang Zhai
ISBN: 978-1-63828-378-2

Multi-hop Question Answering
Vaibhav Mavi, Anubhav Jangra and Adam Jatowt
ISBN: 978-1-63828-374-4

Conversational Information Seeking
Hamed Zamani, Johanne R. Trippas, Jeff Dalton and Filip Radlinski
ISBN: 978-1-63828-200-6

Perspectives of Neurodiverse Participants in Interactive Information Retrieval
Laurianne Sitbon, Gerd Berget and Margot Brereton
ISBN: 978-1-63828-202-0

Mathematical Information Retrieval: Search and Question Answering

Richard Zanibbi
Rochester Institute of Technology
rxzvcs@rit.edu

Behrooz Mansouri
University of Southern Maine
behrooz.mansouri@maine.edu

Anurag Agarwal
Rochester Institute of Technology
axasma@rit.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

R. Zanibbi *et al.*. *Mathematical Information Retrieval: Search and Question Answering*. Foundations and Trends® in Information Retrieval, vol. 19, no. 1-2, pp. 1–190, 2025.

ISBN: 978-1-63828-503-8

Foundations and Trends® in Information Retrieval
Volume 19, Issue 1-2, 2025

Editorial Scope

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2025, Volume 19, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Preface **2**

1 Sources and Information Tasks **6**
1.1 When and Where Do We Search? 7
1.2 Information Task Framework 10
1.3 Information Needs and Information Task Strategies 16
1.4 Retrieval Systems 22
1.5 User Interfaces and System Interactions 26

2 Annotating and Indexing Sources **37**
2.1 Consulting Sources for Mathematical Information 37
2.2 Annotating Formulas: Representations and Canonicalization 46
2.3 Indexing Formulas and Text 58

3 Math Retrieval Tasks and Evaluation Metrics **63**
3.1 Evaluation Overview 64
3.2 Retrieval Tasks: Search and Question Answering 67
3.3 Creating Test Collections 69
3.4 Evaluation Metrics 77
3.5 System Comparisons and Statistical Tests 84

4 Formula Search **86**
4.1 Test Collections for Formula Search 88
4.2 Formula Retrieval Models 93
4.3 Text-based and Tree-based Models 95
4.4 Dense Retrieval with Formula Tree Embeddings 101
4.5 Visual-spatial Models and Formula Autocompletion 105
4.6 Retrieval Effectiveness and Combining Representations . . 108

5 Math-Aware Search **111**
5.1 Test Collections for Math-aware Search 112
5.2 Searching with Formulas and Text 116
5.3 Federated Search: Combining Formula and Text Results . . 117
5.4 Unified Formula + Text Representations 121
5.5 Using LLMs for Math-aware Search 123

6 Math Question Answering **125**
6.1 Test Collections for Math QA 126
6.2 Solving Math Word Problems 130
6.3 LLMs and Mathematical Reasoning 137

7 Final Thoughts and Research Opportunities **145**
7.1 What is Next for MathIR? 145
7.2 Limitations: What MathIR Cannot Provide 148
7.3 Some Parting Thoughts 150

Acknowledgements **152**

Appendices **155**

References **161**

Mathematical Information Retrieval: Search and Question Answering

Richard Zanibbi[1], Behrooz Mansouri[2] and Anurag Agarwal[3]

[1] *Department of Computer Science, Rochester Institute of Technology, USA; rxzvcs@rit.edu*

[2] *Department of Computer Science, University of Southern Maine, USA; behrooz.mansouri@maine.edu*

[3] *School of Mathematics & Statistics, Rochester Institute of Technology, USA; axasma@rit.edu*

ABSTRACT

Mathematical information is essential for technical work, but its creation, interpretation, and search are challenging. To help address these challenges, researchers have developed multimodal search engines and mathematical question answering systems. This monograph begins with a simple framework characterizing the information tasks that people and systems perform as we work to answer math-related questions. The framework is used to organize and relate the other core topics of the monograph, including interactions between people and systems, representing math formulas in sources, and evaluation. We close by addressing some key questions and presenting directions for future work. This monograph is intended for students, instructors, and researchers interested in systems that help us find and use mathematical information.

Richard Zanibbi, Behrooz Mansouri and Anurag Agarwal (2025), "Mathematical Information Retrieval: Search and Question Answering", Foundations and Trends® in Information Retrieval: Vol. 19, No. 1-2, pp 1–190. DOI: 10.1561/1500000095.

Preface

This monograph provides an introduction to the foundations and current developments in *mathematical information retrieval* research, or *math IR*. In this area we focus on systems designed to assist with finding, collecting, and using mathematical information. With the advent of Large Language Models (LLMs), mathematical question answering is of particularly keen interest at the moment. Systems that combine LLMs with logic-based systems have been making news by solving problems from the International Mathematical Olympiad, for example.

The authors have spent more than a decade working on systems and interfaces for math-aware search engines in web pages, PDF documents, and even videos. We have more recently worked with interactive *conversational* systems for math IR, and looked into applications of LLMs. In this monograph we try to summarize what we have learned about systems for searching existing sources, and briefly introduce emerging math question answering systems that automatically generate responses based on usage patterns for text *and* formulas.

Our intended audience includes students, instructors, and researchers of information science, computer science, and mathematics.[1] Because the goal of this series is to introduce a topic from its fundamentals and then build up to the state-of-the-art, we needed to prioritize making

[1]With this said, we have often been 'unintended' readers of monographs, and warmly welcome anyone with even the slightest interest in what we have to say here.

the presentation as clear and concrete as possible. This means that we also had to make difficult decisions about what material to include. We have focused on covering what we understand to be core topics and concerns, rather than provide an exhaustive survey of techniques. For those interested in learning more than we could cover here, we refer you to other math IR surveys that are available (Zanibbi and Blostein, 2012; Guidi and Sacerdoti Coen, 2016; Dadure *et al.*, 2024).

Our approach in this work is searcher-centered, *i.e.,* focused on models and systems used directly by people. We acknowledge that there are *vast* bodies of literature concerned with searching and discovering mathematical information with little or no human interaction. Important examples include automated theorem proving, one of the oldest and most influential corners of artificial intelligence research, and work with mathematical knowledge databases in the Mathematical Knowledge Management community (*e.g.,* at the CICM[2] conferences), from which much of the early influential work in math IR also originates from. In Appendix B we present related work from the theorem proving space, but this is only a brief overview, and just the tip of the iceberg.

Structured representations for formulas and computations precede the electronic computer. The tree-based representations for formula structure presented in Section 2 were designed to be simple and capture key structural properties, *i.e.,* writing lines of symbols for formula appearance, and operation hierarchies for mathematical expressions represented by formulas. However, we make no assertion of their novelty; many analagous representations have been used in papers and systems. We've found that Symbol Layout Trees (SLTs) and Operator Trees (OPTs) generalize the graph types used for recognition and retrieval of formulas well.

We close here with a brief summary of the contents of this monograph. The sections of the monograph are summarized visually in Figure 1, and in more detail below.

Section 1 introduces math IR, and presents a framework that unifies information seeking activities performed by people *and* systems in (1) the real world, with (2) 'traditional' retrieval systems, and (3)

[2] https://cicm-conference.org/cicm.php

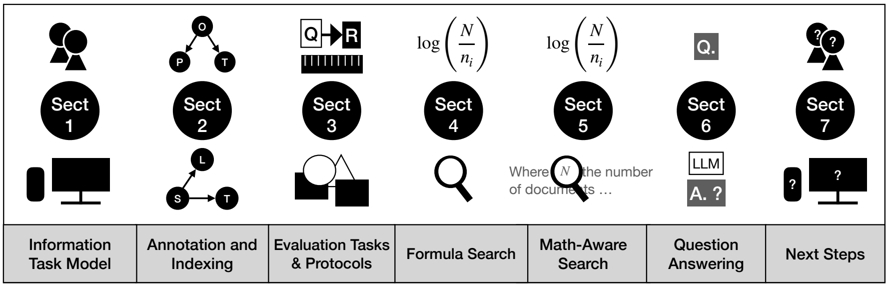

Figure 1: Visual summary of this monograph's contents

question answering systems. The framework is organized around information needs, sources, and tasks, with an informal 'source jar' model for human information seeking, and a structured task graph for systems. The systems-oriented model is used to organize material in Sections 2–6.

Section 2 considers the types of mathematical information present (and missing) in sources, and provides an overview of formula representations. The larger focus is annotating sources with additional information (*e.g.*, formula representations) and representing text and formulas in indexes for 'sparse' (*i.e.*, discrete pattern-based lookup) and 'dense' (*i.e.*, continuous vector space-based lookup) retrieval.

Section 3 presents the math IR tasks addressed in Sections 4–6, along with procedures for creating *test collections* for evaluation, and evaluation metrics used in benchmarks. These are presented together to emphasize similarities and differences between retrieval and question answering tasks.

Sections 4–6 present past and current systems for formula search, text + formula search ('math-aware' search), and question answering. These systems depend upon indexing and evaluation techniques covered in Sections 2 and 3. Each section begins with a summary of test collections for evaluation, followed by a presentation and comparison of methods.

Section 7 provides a closing summary, a discussion of what math IR *cannot* provide, and directions for future work organized around the information task framework from Section 1.

Richard Zanibbi, Rochester, NY, USA
Behrooz Mansouri, Portland, ME, USA
Anurag Agarwal, Rochester, NY, USA
December 2024

Appendices

A

Online Resources

Links for some of the resources mentioned in this monograph may be found below in Table A.1.

Table A.1: Resources for Math Search and Question Answering

Name	Type	Link	Year
Tools			
LaTeXML	LaTeX-MathML Convertor	https://github.com/brucemiller/LaTeXML	2004
SnuggleTeX	LaTeX-MathML Convertor	https://github.com/davemckain/snuggletex	2008
LeanDojo	Theorem provers	https://github.com/lean-dojo/LeanDojo	2023
Test Collections			
NTCIR	Search Test Collection	https://ntcir-math.nii.ac.jp/data/	2013-16
ARQMath	Search Test Collection	https://www.cs.rit.edu/~dprl/ARQMath/	2020-22
AQUA-RAT	MWP Dataset	https://github.com/google-deepmind/AQuA	2017
MathQA	MWP Dataset	https://math-qa.github.io/	2019
ASDiv	MWP Dataset	https://github.com/chao-chun/nlu-asdiv-dataset.	2020
GSM8K	MWP Dataset	https://huggingface.co/datasets/openai/gsm8k	2021
MATH	MWP Dataset	https://github.com/hendrycks/math	2021
GeoQA	Geometry QA	https://github.com/chen-judge/GeoQA	2021
UniGeo	Geometry QA	https://github.com/chen-judge/UniGeo	2022
MathVista	Visual Math Problem	https://mathvista.github.io/	2024
MATH-Vision	Visual Math Problem	https://github.com/mathllm/MATH-V	2024
Systems			
MIAS	Formula Search	https://github.com/MIR-MU/MIaS	2011
Tangent-S	Formula Search	https://github.com/MattLangsenkamp/tangent-s	2017
Tangent-V	Formula Search	https://www.cs.rit.edu/~dprl/files/TangentV-source.zip	2019
Tangent-CFT	Formula Search	https://github.com/BehroozMansouri/TangentCFT	2019
NTFEM	Formula Search	https://github.com/NTFEM/Formulae-Embedding	2020
MathEmb	Formula Search	https://github.com/Franknewchen/MathEmb	2021
WikiMirs	Math-Aware Search	https://github.com/huxuan/WikiMirs	2016
Tangent-L	Math-Aware Search	https://github.com/fras2560/Tangent-L	2018
Approach0	Math-Aware Search	https://github.com/approach0/search-engine	2019
MathDowsers	Math-Aware Search	https://github.com/kiking0501/MathDowsers-ARQMath	2021
ALBERT-for-Math-AR	Math-Aware Search	https://github.com/AnReu/ALBERT-for-Math-AR	2021
MABOWDOR	Math-Aware Search	https://github.com/approach0/pya0	2023
CrossMath	Multilingual Math-Aware S.	https://github.com/jgore077/CrossMath	2024
LILA	QA Dataset/LLM	https://lila.apps.allenai.org/	2022
MAmmoTH	Math-LLM	https://github.com/TIGER-AI-Lab/MAmmoTH	2023
LLEMMA	Math-LLM	https://github.com/EleutherAI/math-lm	2024
ToRA	Math-LLM	https://github.com/microsoft/ToRA	2024
MetaMath	Math-LLM	https://github.com/meta-math/MetaMath	2024

B

Search and QA for Theorem Proving

In Sections 5 and 6 we discussed math-aware search and math question-answering. A topic closely related to these is theorem proving. Automated theorem proving aims to explore the application of computers in proving mathematical theorems. This is actually one of the earliest topics of interest in computer science. In the early 1960s, scientists started exploring the use of machines for theorem proving for the quantification theory (Davis and Putnam, 1960; Davis *et al.*, 1962) using deduction rules to prove assertions.

Informal theorem proving refers to the way that humans approach proving theorems using reasoning through notation and natural language, likely with some missing details (*e.g.,* assumed definitions) and skipped computational steps, for example. In contrast, formal theorem proving represents theorems in a machine-readable format, making verification by logical rules possible, at the cost of more information being explicitly stated (*e.g.,* all variable types, definitions for all operators, etc.).

Converting informal to formal proof steps is referred to as *Autoformalization*. The Mizar[1] language is commonly used by mathemati-

[1] https://mizar.uwb.edu.pl/.

cians as a formal language for writing definitions and proofs. Using Mizar, Wang *et al.* (2018) explored converting (translation) of informal LaTeX-written text into formal Mizar language. This work explored different seq2seq architectures, with LSTM and attention providing the highest BLEU score for this translation.

Researchers have also explored other formalization languages, including applying large language models to generate statements in Codex (Chen *et al.*, 2021b) has been studied recently (Wu *et al.*, 2022). This translates statements in natural text into formalized theorems for the interactive proof assistant Isabelle (Wenzel *et al.*, 2008). As an example, the system translated "Prove that there is no function f from the set of non-negative integers into itself such that $f(f(n)) = n + 1987$ for every n" perfectly to Codex as:

```
theorem
    fixes f :: "nat \<Rightarrow> nat"
    assumes "\<forall> n. f (f n) = n + 1987"
    shows False
```

Another task for theorem proving is retrieving useful lemmas that will help with proving steps, known as *premise selection*. This is a form of search problem, where relevance is defined in terms of suitability for proving a specific conjecture. The problem is defined as (Alama *et al.*, 2014):

> *Given an Automated Theorem Provers and a large number of premises, find premises that are useful to the prover for constructing a proof for a new conjecture.*

DeepMath (Irving *et al.*, 2016) is one of the earliest works to apply deep learning for this task. Conjecture and axiom sequences are embedded separately, concatenated, and then passed to a fully connected neural network for predicting the usefulness of the axiom. Embeddings are at character level for formulas, and word-level for statements defining symbols. The convolutional network model FormulaNet (Wang *et al.*, 2017a) used a similar idea and applied graph neural networks. Using formulas in higher-order logic (Church, 1940), each formula is first parsed into an OPT: internal nodes represent a quantifier or a constant or variable function, and leaf nodes represent variable or constant values.

Edges connect a quantifier to all instances of its quantified variables. After creating the tree, a merging step is applied to merge leaf nodes representing the same constant/variable. Finally, a unification technique is used to replace variable names with 'VAR' and function names with 'VARFUNC'. After building the graph, convolution or message passing is applied to get node embeddings. These embeddings are used with max-pooling to form an embedding for the graph.

As theorems are built upon existing mathematical knowledge, a graph representation of mathematical concept statements such as lemma, and definitions is a common approach for this task. One way of building this graph is to use statements as nodes and ordered edges from node s_1 to s_2 if there is statement 1 is a premise of statement 2 (Ferreira and Freitas, 2020a). With this definition of a graph, the problem can be viewed as link prediction, for which a Deep Graph Convolutional Neural Network (DGCNN) architecture was applied (Zhang *et al.*, 2018). The textual content embedding of each node in this work is encoded using Doc2Vec (Le and Mikolov, 2014) model with mathematical concepts being encoded as linearized trees, with every sub-expression represented as a token.

For example, the formula $(x + y) \times c$ is represented as a sequence of tokens for subexpressions {'x', 'y', '$(x + y)$', '$(x + y) \times c$'} (similar to what is used to generate subexpression tokens for the WikiMirs search model discussed earlier). The authors later introduced STAtement Representation (STAR) cross-modal representation (Ferreira and Freitas, 2021), treating mathematical formulas as natural language. Each statement is viewed as a combination of words and formulas. However, for embedding, they proposed two separate self-attention layers, one for formulas and one for words. The output of the self-attention layers is then concatenated and passed to a Bi-LSTM to get the final representation of the statement. To find the relatedness score between conjecture and premises, a Siamese neural network was applied.

Generative and large language models have recently been applied for premise selection. LeanDojo (Yang *et al.*, 2023) is an open-source toolkit based on the Lean[2] programming language, that introduces

[2]https://lean-lang.org/

ReProver (Retrieval-Augmented Prover). ReProver is a language model-based prover, augmented with retrieval for selecting premises. Given the initial state of proof, it retrieves a set of useful premises (set at 100) using a Dense Passage Retriever. These premises are concatenated to the initial state and passed to a fine-tuned ByT5 (Xue *et al.*, 2022) model to generate steps toward the proof.

As the application of LLMs for theorem proving is gaining attention, new datasets are being introduced for theorem proving. NATURAL-PROOFS (Welleck *et al.*, 2021) for example, is a multi-domain corpus of mathematical statements and their proofs, written in natural mathematical language. The main tasks in this dataset are: 1) *mathematical reference retrieval*: given a theorem, retrieve a set of references that occur in the theorem proof, and 2) *mathematical theorem generation*: generate the sequence of references that occur in a given theorem's proof. Table B.1 summarizes some of the existing resources for this task.

Table B.1: Resources for Math Search and Question Answering

Name	**Type**	**Link**	**Year**
NaturalProof	Proof Dataset	https://github.com/wellecks/naturalproofs	2021
FormulaNet	Theorem Proving	https://github.com/princeton-vl/FormulaNet	2017
STAR	Theorem Proving	https://github.com/ai-systems/crossmodal_embedding	2021

References

Ahmed, S., K. Davila, S. Setlur, and V. Govindaraju. (2021). "Equation Attention Relationship Network (EARN) : A Geometric Deep Metric Framework for Learning Similar Math Expression Embedding". In: *2020 25th International Conference on Pattern Recognition (ICPR)*. Milan, Italy: IEEE. 6282–6289. DOI: 10.1109/ICPR48806.2021.9412619.

Aizawa, A., M. Kohlhase, and I. Ounis. (2013). "NTCIR-10 Math Pilot Task Overview". In: *Proc. NTCIR-10*. 1–8.

Aizawa, A., M. Kohlhase, and I. Ounis. (2014). "NTCIR-11 Math-2 Task Overview". In: *Proc. NTCIR-11*. 88–98.

Alama, J., T. Heskes, D. Kühlwein, E. Tsivtsivadze, and J. Urban. (2014). "Premise Selection for Mathematics by Corpus Analysis and Kernel Methods". *Journal of Automated Reasoning*. 52(2): 191–213. DOI: 10.1007/s10817-013-9286-5.

Alexeeva, M., R. Sharp, M. A. Valenzuela-Escárcega, J. Kadowaki, A. Pyarelal, and C. Morrison. (2020). "MathAlign: Linking Formula Identifiers to their Contextual Natural Language Descriptions". English. In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association. 2204–2212. URL: https://aclanthology.org/2020.lrec-1.269.

Almazán, J., A. Gordo, A. Fornés, and E. Valveny. (2014). "Word Spotting and Recognition with Embedded Attributes". *IEEE Trans. Pattern Anal. Mach. Intell.* 36(12): 2552–2566. DOI: 10.1109/TPAMI.2014.2339814.

Amador, B., M. Langsenkamp, A. Dey, A. K. Shah, and R. Zanibbi. (2023). "Searching the ACL Anthology with Math Formulas and Text". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*. ACM. 3110–3114. DOI: 10.1145/3539618.3591803.

Amini, A., S. Gabriel, S. Lin, R. Koncel-Kedziorski, Y. Choi, and H. Hajishirzi. (2019). "MathQA: Towards Interpretable Math Word Problem Solving with Operation-Based Formalisms". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics. 2357–2367. DOI: 10.18653/v1/N19-1245.

Asakura, T., Y. Miyao, and A. Aizawa. (2022). "Building Dataset for Grounding of Formulae — Annotating Coreference Relations Among Math Identifiers". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association. 4851–4858. URL: https://aclanthology.org/2022.lrec-1.519.

Asakura, T., Y. Miyao, A. Aizawa, and M. Kohlhase. (EasyChair, 2021). *MioGatto: A Math Identifier-Oriented Grounding Annotation Tool*. EasyChair Preprint 6209.

Avenoso, R. (2021). "Spatial vs. Graph-Based Formula Retrieval". *Master's thesis*. Rochester, NY, USA: Rochester Institute of Technology.

Avenoso, R., B. Mansouri, and R. Zanibbi. (2021). "XY-PHOC Symbol Location Embeddings for Math Formula Retrieval and Autocompletion". In: *Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021*. Vol. 2936. *CEUR Workshop Proceedings*. CEUR-WS.org. 25–35. URL: https://ceur-ws.org/Vol-2936/paper-02.pdf.

Azerbayev, Z., H. Schoelkopf, K. Paster, M. D. Santos, S. McAleer, A. Q. Jiang, J. Deng, S. Biderman, and S. Welleck. (2024). "Llemma: An Open Language Model For Mathematics". URL: https://arxiv.org/abs/2310.10631.

Baker, J. B., A. P. Sexton, and V. Sorge. (2010). "Faithful mathematical formula recognition from PDF documents". In: *The Ninth IAPR International Workshop on Document Analysis Systems, DAS 2010, June 9-11, 2010, Boston, Massachusetts, USA. ACM International Conference Proceeding Series*. ACM. 485–492. DOI: 10.1145/1815330.1815393.

Banarescu, L., C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. (2013). "Abstract Meaning Representation for Sembanking". In: *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*. 178–186.

Bates, M. J. (1989). "The Design of Browsing and Berrypicking Techniques for the Online Search Interface". *Online review*.

Belkin, N. J. (1980). "Anomalous States of Knowledge as a Basis for Information Retrieval". *Canadian journal of information science*. 5(1): 133–143.

Bender, E. M., T. Gebru, A. McMillan-Major, and S. Shmitchell. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*. ACM. 610–623. DOI: 10.1145/3442188.3445922.

Blei, D. M., A. Y. Ng, and M. I. Jordan. (2003). "Latent Dirichlet Allocation". *J. Mach. Learn. Res.* 3: 993–1022. URL: http://jmlr.org/papers/v3/blei03a.html.

Bobrow, D. G. (1964). "A question-answering system for high school algebra word problems". In: *Proceedings of the October 27-29, 1964, Fall Joint Computer Conference, Part I. AFIPS '64 (Fall, part I)*. San Francisco, California: Association for Computing Machinery. 591–614. DOI: 10.1145/1464052.1464108.

Bojanowski, P., E. Grave, A. Joulin, and T. Mikolov. (2017). "Enriching Word Vectors with Subword Information". *Transactions of the Association for Computational Linguistics*. 5(Dec.): 135–146. DOI: 10.1162/tacl_a_00051.

Borlund, P. (2003). "The IIR Evaluation Model: A Framework For Evaluation of Interactive Information Retrieval Systems". *Inf. Res.* 8(3). URL: http://www.informationr.net/ir/8-3/paper152.html.

Bozzon, A. and P. Fraternali. (2010). "Chapter 8: Multimedia and Multimodal Information Retrieval". In: *Search Computing: Challenges and Directions*. Berlin, Heidelberg: Springer Berlin Heidelberg. 135–155. DOI: 10.1007/978-3-642-12310-8_8.

Broder, A. Z. (2002). "A Taxonomy of Web Search". *SIGIR Forum*. 36(2): 3–10. DOI: 10.1145/792550.792552.

Buckley, C. and E. M. Voorhees. (2004). "Retrieval Evaluation with Incomplete Information". In: *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*. Sheffield, United Kingdom: ACM Press. 25. DOI: 10.1145/1008992.1009000.

Chen, J., T. Li, J. Qin, P. Lu, L. Lin, C. Chen, and X. Liang. (2022a). "UniGeo: Unifying Geometry Logical Reasoning via Reformulating Mathematical Expression". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 3313–3323. DOI: 10.18653/v1/2022.emnlp-main.218.

Chen, J., J. Tang, J. Qin, X. Liang, L. Liu, E. Xing, and L. Lin. (2021a). "GeoQA: A Geometric Question Answering Benchmark Towards Multimodal Numerical Reasoning". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Online: Association for Computational Linguistics. 513–523. DOI: 10.18653/v1/2021.findings-acl.46.

Chen, M., J. Tworek, H. Jun, Q. Yuan, H. P. D. O. Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, *et al.* (2021b). "Evaluating large language models trained on code". *arXiv preprint arXiv:2107.03374*.

Chen, W., X. Ma, X. Wang, and W. W. Cohen. (2022b). “Program of Thoughts Prompting: Disentangling Computation from Reasoning for Numerical Reasoning Tasks”. *arXiv preprint arXiv:2211.12588*.

Cheung, A. C. and R. E. Slavin. (2013). “The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis”. *Educational Research Review*. 9(June): 88–113. DOI: 10.1016/j.edurev.2013.01.001.

Church, A. (1940). “A Formulation of the Simple Theory of Types”. *The journal of symbolic logic*. 5(2): 56–68.

Cobbe, K., V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano, C. Hesse, and J. Schulman. (2021). “Training Verifiers to Solve Math Word Problems”. URL: https://arxiv.org/abs/2110.14168.

Croft, W. B., D. Metzler, and T. Strohman. (2009). *Search Engines - Information Retrieval in Practice*. Pearson Education. URL: http://www.search-engines-book.com/.

Dadure, P., P. Pakray, and S. Bandyopadhyay. (2024). “Mathematical Information Retrieval: A Review”. *ACM Comput. Surv.* 57(3). DOI: 10.1145/3699953.

Dai, Y., L. Chen, and Z. Zhang. (2020). “An N-ary Tree-based Model for Similarity Evaluation on Mathematical Formulae”. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Toronto, ON, Canada: IEEE. 2578–2584. DOI: 10.1109/SMC42975.2020.9283495.

Davila, K., R. Joshi, S. Setlur, V. Govindaraju, and R. Zanibbi. (2019). “Tangent-V: Math Formula Image Search Using Line-of-Sight Graphs”. In: *Advances in Information Retrieval*. Vol. 11437. Cham: Springer International Publishing. 681–695. DOI: 10.1007/978-3-030-15712-8_44.

Davila, K., S. Ludi, and R. Zanibbi. (2014). “Using Off-line Features and Synthetic Data for On-line Handwritten Math Symbol Recognition”. In: *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE. 323–328.

Davila, K., F. Xu, S. Setlur, and V. Govindaraju. (2021). "FCN-LectureNet: Extractive Summarization of Whiteboard and Chalkboard Lecture Videos". *IEEE Access*. 9: 104469–104484. DOI: 10.1109/ACCESS.2021.3099427.

Davila, K. and R. Zanibbi. (2017a). "Layout and Semantics: Combining Representations for Mathematical Formula Search". In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Shinjuku Tokyo Japan: ACM. 1165–1168. DOI: 10.1145/3077136.3080748.

Davila, K. and R. Zanibbi. (2017b). "Whiteboard Video Summarization via Spatio-Temporal Conflict Minimization". In: *14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017*. IEEE. 355–362. DOI: 10.1109/ICDAR.2017.66.

Davila, K. and R. Zanibbi. (2018). "Visual Search Engine for Handwritten and Typeset Math in Lecture Videos and LATEX Notes". In: *16th International Conference on Frontiers in Handwriting Recognition, ICFHR 2018, Niagara Falls, NY, USA, August 5-8, 2018*. IEEE Computer Society. 50–55. DOI: 10.1109/ICFHR-2018.2018.00018.

Davis, M., G. Logemann, and D. Loveland. (1962). "A Machine Program for Theorem-proving". *Commun. ACM*. 5(7): 394–397. DOI: 10.1145/368273.368557.

Davis, M. and H. Putnam. (1960). "A Computing Procedure for Quantification Theory". *J. ACM*. 7(3): 201–215. DOI: 10.1145/321033.321034.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics. 4171–4186. DOI: 10.18653/v1/N19-1423.

Diaz, Y. (2021). "Introducing Handwriting into a Multimodal LaTeX Formula Editor". *Master's thesis*. Rochester Institute of Technology.

Diaz, Y., G. Nishizawa, B. Mansouri, K. Davila, and R. Zanibbi. (2021). "The MathDeck Formula Editor: Interactive Formula Entry Combining LaTeX , Structure Editing, and Search". In: *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan: ACM. 1–5. DOI: 10.1145/3411763.3451564.

Dmello, A. (2019). "Representing Mathematical Concepts Associated With Formulas Using Math Entity Cards". *Master's thesis*. Rochester Institute of Technology.

Dua, D., Y. Wang, P. Dasigi, G. Stanovsky, S. Singh, and M. Gardner. (2019). "DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs". URL: https://arxiv.org/abs/1903.00161.

Duncan, D. (2021). *Index, a History of the*. New York: Norton.

Durgin, S., J. Gore, and B. Mansouri. (2024). "MathMex: Search Engine for Math Definitions". In: *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part V*. Vol. 14612. *Lecture Notes in Computer Science*. Springer. 194–199. DOI: 10.1007/978-3-031-56069-9_17.

Ethayarajh, K., W. Xu, N. Muennighoff, D. Jurafsky, and D. Kiela. (2024). "KTO: Model Alignment as Prospect Theoretic Optimization". URL: https://arxiv.org/abs/2402.01306.

Faggioli, G., L. Dietz, C. L. A. Clarke, G. Demartini, M. Hagen, C. Hauff, N. Kando, E. Kanoulas, M. Potthast, B. Stein, and H. Wachsmuth. (2024). "Who Determines What Is Relevant? Humans or AI? Why Not Both?" *Commun. ACM*. 67(4): 31–34. DOI: 10.1145/3624730.

Ferreira, D. and A. Freitas. (2020a). "Premise Selection in Natural Language Mathematical Texts". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics. 7365–7374. DOI: 10.18653/v1/2020.acl-main.657.

Ferreira, D. and A. Freitas. (2021). "STAR: Cross-modal [STA]tement [R]epresentation for selecting relevant mathematical premises". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Online: Association for Computational Linguistics. 3234–3243. DOI: 10.18653/v1/2021.eacl-main.282.

Ferreira, D. and A. Freitas. (2020b). "Natural Language Premise Selection: Finding Supporting Statements for Mathematical Text". URL: https://arxiv.org/abs/2004.14959.

Formal, T., B. Piwowarski, and S. Clinchant. (2021). "SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking". In: *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM. 2288–2292. DOI: 10.1145/3404835.3463098.

Fraser, D., A. Kane, and F. W. Tompa. (2018). "Choosing Math Features for BM25 Ranking with Tangent-L". In: *Proceedings of the ACM Symposium on Document Engineering 2018*. Halifax NS Canada: ACM. 1–10. DOI: 10.1145/3209280.3209527.

Freund, Y., R. Iyer, R. E. Schapire, and Y. Singer. (2003). "An Efficient Boosting Algorithm for Combining Preferences". *Journal of machine learning research*. 4(Nov): 933–969.

Fuhr, N. (2017). "Some Common Mistakes In IR Evaluation, And How They Can Be Avoided". *SIGIR Forum*. 51(3): 32–41. DOI: 10.1145/3190580.3190586.

Gao, L., Z. Jiang, Y. Yin, K. Yuan, Z. Yan, and Z. Tang. (2017). "Preliminary Exploration of Formula Embedding for Mathematical Information Retrieval: can mathematical formulae be embedded like a natural language?" *arXiv:1707.05154 [cs]*. Aug. URL: http://arxiv.org/abs/1707.05154.

Gao, L., K. Yuan, Y. Wang, Z. Jiang, and Z. Tang. (2016). "The Math Retrieval System of ICST for NTCIR-12 MathIR Task". In: *NTCIR*.

Gao, L., A. Madaan, S. Zhou, U. Alon, P. Liu, Y. Yang, J. Callan, and G. Neubig. (2023). "PAL: Program-aided Language Models". In: *Proceedings of the 40th International Conference on Machine Learning*. Vol. 202. *Proceedings of Machine Learning Research*. PMLR. 10764–10799. URL: https://proceedings.mlr.press/v202/gao23f.html.

Gao, S. and Y.-K. D. Ng. (2023). “Recommending Answers to Math Questions Based on KL-Divergence and Approximate XML Tree Matching”. In: *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region. SIGIR-AP ’23.* Beijing, China: Association for Computing Machinery. 21–31. DOI: 10.1145/3624918.3625337.

Giacalone, B., G. Paiement, Q. Tucker, and R. Zanibbi. (2024). “Beneath the [MASK]: An Analysis of Structural Query Tokens in ColBERT”. In: *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part III.* Vol. 14610. *Lecture Notes in Computer Science.* Springer. 431–439. DOI: 10.1007/978-3-031-56063-7_35.

Gore, J., J. Polletta, and B. Mansouri. (2024). “CrossMath: Towards Cross-lingual Math Information Retrieval”. In: *The 10th ACM SIGIR / The 14th International Conference on the Theory of Information Retrieval.* URL: https://openreview.net/forum?id=IYcJxgnJ9Q.

Gou, Z., Z. Shao, Y. Gong, Y. Shen, Y. Yang, M. Huang, N. Duan, and W. Chen. (2024). “ToRA: A Tool-Integrated Reasoning Agent for Mathematical Problem Solving”. URL: https://arxiv.org/abs/2309.17452.

Graf, P. (1995). “Substitution Tree Indexing”. In: *Rewriting Techniques and Applications.* Springer Berlin Heidelberg. 117–131.

Greiner-Petter, A., M. Schubotz, C. Breitinger, P. Scharpf, A. Aizawa, and B. Gipp. (2023). “Do the Math: Making Mathematics in Wikipedia Computable”. *IEEE Trans. Pattern Anal. Mach. Intell.* 45(4): 4384–4395. DOI: 10.1109/TPAMI.2022.3195261.

Grossman, J., Z. Lin, H. Sheng, J. T.-Z. Wei, J. J. Williams, and S. Goel. (2019). “MathBot: Transforming Online Resources for Learning Math into Conversational Interactions”. *AAAI 2019 Story-Enabled Intelligence.*

Grover, A. and J. Leskovec. (2016). “node2vec: Scalable Feature Learning for Networks”. In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining.* 855–864.

Guidi, F. and C. Sacerdoti Coen. (2016). "A Survey on Retrieval of Mathematical Knowledge". *Mathematics in Computer Science.* 10(4): 409–427. DOI: 10.1007/s11786-016-0274-0.

Ha, J., R. Haralick, and I. Phillips. (1995). "Recursive XY Cut using Bounding Boxes of Connected Components". In: *Proceedings of 3rd International Conference on Document Analysis and Recognition.* Vol. 2. 952–955 vol.2. DOI: 10.1109/ICDAR.1995.602059.

Hallinen, N. R., L. N. Sprague, K. P. Blair, R. M. Adler, and N. S. Newcombe. (2021). "Finding Formulas: Does Active Search Facilitate Appropriate Generalization?" *Cognitive Research: Principles and Implications.* 6(1): 50. DOI: 10.1186/s41235-021-00316-y.

Hambasan, R., M. Kohlhase, and C. Prodescu. (2014). "MathWebSearch at NTCIR-11". In: *Proc. NTCIR-11.* 114–119.

Hamilton, W., Z. Ying, and J. Leskovec. (2017). "Inductive Representation Learning on Large Graphs". In: *Advances in Neural Information Processing Systems.* Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7ebea9-Paper.pdf.

He, K., X. Zhang, S. Ren, and J. Sun. (2016). "Deep Residual Learning for Image Recognition". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Head, A., K. Lo, D. Kang, R. Fok, S. Skjonsberg, D. S. Weld, and M. A. Hearst. (2021). "Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* Yokohama Japan: ACM. 1–18. DOI: 10.1145/3411764.3445648.

Hearst, M. (2009). *Search User Interfaces.* Cambridge university press.

Hendrycks, D., C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, and J. Steinhardt. (2021). "Measuring Mathematical Problem Solving With the MATH Dataset". URL: https://arxiv.org/abs/2103.03874.

Holmes, S., A. Moorhead, R. Bond, H. Zheng, V. Coates, and M. McTear. (2019). "Usability Testing of a Healthcare Chatbot: Can we use Conventional Methods to Assess Conversational user Interfaces?" In: *Proc. 31st European Conference on Cognitive Ergonomics*. 207–214.

Hopkins, M., R. Le Bras, C. Petrescu-Prahova, G. Stanovsky, H. Hajishirzi, and R. Koncel-Kedziorski. (2019). "SemEval-2019 Task 10: Math Question Answering". In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. Minneapolis, Minnesota, USA: Association for Computational Linguistics. 893–899. DOI: 10.18653/v1/S19-2153.

Hosseini, M. J., H. Hajishirzi, O. Etzioni, and N. Kushman. (2014). "Learning to Solve Arithmetic Word Problems with Verb Categorization". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics. 523–533. DOI: 10.3115/v1/D14-1058.

Hu, X., L. Gao, X. Lin, Z. Tang, X. Lin, and J. B. Baker. (2013). "WikiMirs: A Mathematical Information Retrieval System for Wikipedia". In: *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries - JCDL '13*. Indianapolis, Indiana, USA: ACM Press. 11. DOI: 10.1145/2467696.2467699.

Huang, D., S. Shi, C.-Y. Lin, J. Yin, and W.-Y. Ma. (2016). "How well do Computers Solve Math Word Problems? Large-Scale Dataset Construction and Evaluation". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Berlin, Germany: Association for Computational Linguistics. 887–896. DOI: 10.18653/v1/P16-1084.

Irving, G., C. Szegedy, A. A. Alemi, N. Een, F. Chollet, and J. Urban. (2016). "DeepMath - Deep Sequence Models for Premise Selection". In: *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/f197002b9a0853eca5e046d9ca4663d5-Paper.pdf.

Ito, T., T. Matsuzaki, and S. Sato. (2017). “Coreference Resolution on Math Problem Text in Japanese”. In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Taipei, Taiwan: Asian Federation of Natural Language Processing. 373–377. URL: https://aclanthology.org/I17-2063.

Jiang, A. Q., A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. l. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, *et al.* (2023). “Mistral 7B”. *arXiv preprint arXiv:2310.06825*.

Jo, H., D. Kang, A. Head, and M. A. Hearst. (2021). “Modeling Mathematical Notation Semantics in Academic Papers”. In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. Punta Cana, Dominican Republic: Association for Computational Linguistics. 3102–3115. DOI: 10.18653/v1/2021.findings-emnlp.266.

Joachims, T. (2006). “Training Linear SVMs in Linear Time”. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 217–226.

Jones, K. S. (1972). “A Statistical Interpretation of Term Specificity and its Application in Retrieval”. *J. Documentation*. 28(1): 11–21. URL: https://api.semanticscholar.org/CorpusID:2996187.

Kamali, S. and F. W. Tompa. (2013). “Retrieving Documents with Mathematical Content”. In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. Dublin Ireland: ACM. 353–362. DOI: 10.1145/2484028.2484083.

Kane, A., Y. K. Ng, and F. W. Tompa. (2022). “Dowsing for Answers to Math Questions: Doing Better with Less”. In: *CLEF (Working Notes)*. 40–62.

Karpukhin, V., B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih. (2020). “Dense Passage Retrieval for Open-Domain Question Answering”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics. 6769–6781. DOI: 10.18653/v1/2020.emnlp-main.550.

Khattab, O. and M. Zaharia. (2020). “ColBERT: Efficient and Effective Passage Search via Contextualized Late Interaction over BERT”. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR ’20.* Virtual Event, China: Association for Computing Machinery. 39–48. DOI: 10.1145/3397271.3401075.

Kirsch, D. (2010). “Detexify: Recognition of Hand-drawn LaTeX Symbols”. *Master’s thesis*. University of Münster (in German).

Kohlhase, A. and M. Kohlhase. (2007). “Reexamining the MKM Value Proposition: From Math Web Search to Math Web ReSearch”. In: *Towards Mechanized Mathematical Assistants, 14th Symposium, Calculemus 2007, 6th International Conference, MKM 2007, Hagenberg, Austria, June 27-30, 2007, Proceedings*. Vol. 4573. *Lecture Notes in Computer Science*. Springer. 313–326. DOI: 10.1007/978-3-540-73086-6_25.

Kohlhase, M. and C. Prodescu. (2013). “MathWebSearch at NTCIR-10”. In: *Proc. NTCIR-10*. 675–679.

Kohlhase, M. and I. Sucan. (2006). “A Search Engine for Mathematical Formulae”. In: *Artificial Intelligence and Symbolic Computation*. Vol. 4120. Berlin, Heidelberg: Springer Berlin Heidelberg. 241–253. DOI: 10.1007/11856290_21.

Koncel-Kedziorski, R., H. Hajishirzi, A. Sabharwal, O. Etzioni, and S. D. Ang. (2015). “Parsing Algebraic Word Problems into Equations”. *Transactions of the Association for Computational Linguistics*. 3: 585–597.

Kristianto, G. Y. and A. Aizawa. (2014). “Extracting Textual Descriptions of Mathematical Expressions in Scientific Papers”. *D-Lib Magazine*. 20(11): 9.

Kristianto, G. Y., G. Topic, and A. Aizawa. (2016a). “MCAT Math Retrieval System for NTCIR-12 MathIR Task”: 323–330.

Kristianto, G. Y., G. Topic, and A. Aizawa. (2017). “Utilizing dependency relationships between math expressions in math IR”. *Inf. Retr. J.* 20(2): 132–167. DOI: 10.1007/S10791-017-9296-8.

Kristianto, G. Y., G. Topić, and A. Aizawa. (2016b). "Entity Linking for Mathematical Expressions in Scientific Documents". In: *Digital Libraries: Knowledge, Information, and Data in an Open Access Society*. Cham: Springer International Publishing. 144–149.

Krstovski, K. and D. M. Blei. (2018). "Equation Embeddings". *arXiv:1803.09123 [cs, stat]*. Mar. URL: http://arxiv.org/abs/1803.09123.

Lai, V., A. Pouran Ben Veyseh, F. Dernoncourt, and T. Nguyen. (2022). "SemEval 2022 Task 12: Symlink - Linking Mathematical Symbols to their Descriptions". In: *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*. Seattle, United States: Association for Computational Linguistics. 1671–1678. DOI: 10.18653/v1/2022.semeval-1.230.

Langkilde, I. and K. Knight. (1998). "Generation that Exploits Corpus-Based Statistical Knowledge". In: *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL '98, August 10-14, 1998, Université de Montréal, Montréal, Quebec, Canada. Proceedings of the Conference*. Morgan Kaufmann Publishers / ACL. 704–710. DOI: 10.3115/980845.980963.

Langsenkamp, M., B. Amador, and R. Zanibbi. (2024). "A Study of PHOC Spatial Region Configurations for Math Formula Retrieval". *Tech. rep.* DOI: 10.48550/ARXIV.2408.09283. arXiv: 2408.09283.

Langsenkamp, M., B. Mansouri, and R. Zanibbi. (2022). "Expanding Spatial Regions and Incorporating IDF for PHOC-Based Math Formula Retrieval at ARQMath-3". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*. Vol. 3180. *CEUR Workshop Proceedings*. CEUR-WS.org. 63–82. URL: https://ceur-ws.org/Vol-3180/paper-04.pdf.

Le, Q. and T. Mikolov. (2014). "Distributed Representations of Sentences and Documents". In: *Proceedings of the 31st International Conference on Machine Learning*. Vol. 32. *Proceedings of Machine Learning Research*. No. 2. Bejing, China: PMLR. 1188–1196. URL: https://proceedings.mlr.press/v32/le14.html.

Lewkowycz, A., A. Andreassen, D. Dohan, E. Dyer, H. Michalewski, V. Ramasesh, A. Slone, C. Anil, I. Schlag, T. Gutman-Solo, Y. Wu, B. Neyshabur, G. Gur-Ari, and V. Misra. (2022). "Solving Quantitative Reasoning Problems with Language Models". In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc. 3843–3857. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/18abbeef8cfe9203fdf9053c9c4fe191-Paper-Conference.pdf.

Liang, Z., J. Zhang, L. Wang, W. Qin, Y. Lan, J. Shao, and X. Zhang. (2022). "MWP-BERT: Numeracy-Augmented Pre-training for Math Word Problem Solving". In: *Findings of the Association for Computational Linguistics: NAACL 2022*. Seattle, United States: Association for Computational Linguistics. 997–1009. DOI: 10.18653/v1/2022.findings-naacl.74.

Libbrecht, P. and E. Melis. (2006). "Methods to Access and Retrieve Mathematical Content in ActiveMath". In: *Mathematical Software - ICMS 2006*. Vol. 4151. Berlin, Heidelberg: Springer Berlin Heidelberg. 331–342. DOI: 10.1007/11832225_33.

Ling, W., D. Yogatama, C. Dyer, and P. Blunsom. (2017). "Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics. 158–167. DOI: 10.18653/v1/P17-1015.

Líška, M., P. Sojka, M. Ruzicka, and P. Mravec. (2011). "Web Interface and Collection for Mathematical Retrieval : WebMIaS and MREC". In: *Towards a Digital Mathematics Library*. Bertinoro, Bertinoro, Italy: Masaryk University Press (Brno, Czech Republic). 77–84.

Liu, Y., K. Ding, and Y. Zhou. (2019). "AiFu at SemEval-2019 Task 10: A Symbolic and Sub-symbolic Integrated System for SAT Math Question Answering". In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. Minneapolis, Minnesota, USA: Association for Computational Linguistics. 900–906. DOI: 10.18653/v1/S19-2154.

Lu, P., H. Bansal, T. Xia, J. Liu, C. Li, H. Hajishirzi, H. Cheng, K.-W. Chang, M. Galley, and J. Gao. (2024). "MathVista: Evaluating Mathematical Reasoning of Foundation Models in Visual Contexts". URL: https://arxiv.org/abs/2310.02255.

Lu, P., L. Qiu, K.-W. Chang, Y. N. Wu, S.-C. Zhu, T. Rajpurohit, P. Clark, and A. Kalyan. (2023). "Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning". URL: https://arxiv.org/abs/2209.14610.

Lv, Y. and C. Zhai. (2011). "Lower-Bounding Term Frequency Normalization". In: *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. 7–16.

Macina, J., N. Daheim, S. Chowdhury, T. Sinha, M. Kapur, I. Gurevych, and M. Sachan. (2023). "MathDial: A Dialogue Tutoring Dataset with Rich Pedagogical Properties Grounded in Math Reasoning Problems". In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Singapore: Association for Computational Linguistics. 5602–5621. DOI: 10.18653/v1/2023.findings-emnlp.372.

Manning, C. D., P. Raghavan, and H. Schütze. (2008). *Introduction to information retrieval*. Cambridge University Press. DOI: 10.1017/CBO9780511809071.

Mansouri, B. and Z. Jahedibashiz. (2023). "Clarifying Questions in Math Information Retrieval". In: *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval. ICTIR '23*. Taipei, Taiwan: Association for Computing Machinery. 149–158. DOI: 10.1145/3578337.3605123.

Mansouri, B. and R. Maarefdoust. (2024). "Using Large Language Models for Math Information Retrieval". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '24*. Washington DC, USA: Association for Computing Machinery. 2693–2697. DOI: 10.1145/3626772.3657907.

Mansouri, B., V. Novotný, A. Agarwal, D. W. Oard, and R. Zanibbi. (2022a). "Overview of ARQMath-3 (2022): Third CLEF Lab on Answer Retrieval for Questions on Math". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Cham: Springer International Publishing. 286–310.

Mansouri, B., V. Novotný, A. Agarwal, D. W. Oard, and R. Zanibbi. (2022b). "Overview of ARQMath-3 (2022): Third CLEF Lab on Answer Retrieval for Questions on Math (Working Notes Version)". In: *Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022*. Vol. 3180. *CEUR Workshop Proceedings*. CEUR-WS.org. 1–27. URL: https://ceur-ws.org/Vol-3180/paper-01.pdf.

Mansouri, B., D. W. Oard, A. Agarwal, and R. Zanibbi. (2021a). "Effects of Context, Complexity, and Clustering on Evaluation for Math Formula Retrieval". *CoRR*. abs/2111.10504. arXiv: 2111.10504. URL: https://arxiv.org/abs/2111.10504.

Mansouri, B., D. W. Oard, and R. Zanibbi. (2020). "DPRL Systems in the CLEF 2020 ARQMath Lab". CEUR Workshop Proceedings 2696. Ed. by L. Cappellato, C. Eickhoff, N. Ferro, and A. Névéol. URL: https://ceur-ws.org/Vol-2696/paper%5C_223.pdf.

Mansouri, B., D. W. Oard, and R. Zanibbi. (2022c). "Contextualized Formula Search Using Math Abstract Meaning Representation". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*. ACM. 4329–4333. DOI: 10.1145/3511808.3557567.

Mansouri, B., D. W. Oard, and R. Zanibbi. (2022d). "Contextualized Formula Search Using Math Abstract Meaning Representation". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management. CIKM '22*. Atlanta, GA, USA: Association for Computing Machinery. 4329–4333. DOI: 10.1145/3511808.3557567.

Mansouri, B., S. Rohatgi, D. W. Oard, J. Wu, C. L. Giles, and R. Zanibbi. (2019a). "Tangent-CFT: An Embedding Model for Mathematical Formulas". In: *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. Santa Clara CA USA: ACM. 11–18. DOI: 10.1145/3341981.3344235.

Mansouri, B., R. Zanibbi, and D. W. Oard. (2019b). "Characterizing Searches for Mathematical Concepts". In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. Champaign, IL, USA: IEEE. 57–66. DOI: 10.1109/JCDL.2019.00019.

Mansouri, B., R. Zanibbi, and D. W. Oard. (2021b). "Learning to Rank for Mathematical Formula Retrieval". In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Virtual Event Canada: ACM. 952–961. DOI: 10.1145/3404835.3462956.

Miao, S.-y., C.-C. Liang, and K.-Y. Su. (2020). "A Diverse Corpus for Evaluating and Developing English Math Word Problem Solvers". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics. 975–984. DOI: 10.18653/v1/2020.acl-main.92.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. (2013). "Efficient Estimation of Word Representations in Vector Space". *arXiv preprint arXiv:1301.3781*.

Miller, B. R. and A. Youssef. (2003). "Technical Aspects of the Digital Library of Mathematical Functions": 19.

Miller, B. R. (2013). "Three Years of DLMF: Web, Math and Search". In: *Intelligent Computer Mathematics*. Berlin, Heidelberg: Springer Berlin Heidelberg. 288–295.

Miner, R. and R. Munavalli. (2007). "An Approach to Mathematical Search Through Query Formulation and Data Normalization". In: *International Conference on Mathematical Knowledge Management*. Springer. 342–355.

Mishra, S., M. Finlayson, P. Lu, L. Tang, S. Welleck, C. Baral, T. Rajpurohit, O. Tafjord, A. Sabharwal, P. Clark, and A. Kalyan. (2022). "LILA: A Unified Benchmark for Mathematical Reasoning". In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. 5807–5832. DOI: 10.18653/v1/2022.emnlp-main.392.

Mišutka, J. and L. Galamboš. (2008). "Extending Full Text Search Engine for Mathematical Content". In: *Proc. Digital Mathematical Libraries 2008*. 55–67. Birmingham, United Kingdom.

Mitra, A., H. Khanpour, C. Rosset, and A. Awadallah. (2024). "Orca-Math: Unlocking the potential of SLMs in Grade School Math". URL: https://arxiv.org/abs/2402.14830.

Nagy, G. and S. Seth. (1984). “Hierarchical Representation of Optically Scanned Documents”. In: *Proc. International Conference on Pattern Recognition*. Vol. 1. Montréal, Canada. 347–349.

National Council of Teachers of Mathematics. (2011). “Strategic Use of Technology in Teaching and Learning Mathematics - National Council of Teachers of Mathematics”. URL: https://www.nctm.org/Standards-and-Positions/Position-Statements/Strategic-Use-of-Technology-in-Teaching-and-Learning-Mathematics/.

Ng, Y. K., D. J. Fraser, B. Kassaie, and F. W. Tompa. (2021). “Dowsing for Answers to Math Questions: Ongoing Viability of Traditional MathIR”. In: *CLEF (Working Notes)*. 63–81.

Ng, Y. K., D. J. Fraser, B. Kassaie, G. Labahn, M. S. Marzouk, F. W. Tompa, and K. Wang. (2020). “Dowsing for Math Answers with Tangent-L”. CEUR Workshop Proceedings 2696. Ed. by L. Cappellato, C. Eickhoff, N. Ferro, and A. Névéol. URL: https://ceur-ws.org/Vol-2696/paper%5C_167.pdf.

Nishizawa, G. (2020). “Visual Structure Editing of Math Formulas”. *Master's thesis*. Rochester Institute of Technology.

Nogueira, R., Z. Jiang, and J. Lin. (2021). “Investigating the Limitations of Transformers with Simple Arithmetic Tasks”. URL: https://arxiv.org/abs/2102.13019.

Norman, D. A. (1988). *The Psychology of Everyday Things*. Basic books.

Ohri, A. and T. Schmah. (2021). “Machine Translation of Mathematical Text”. *IEEE Access*. 9: 38078–38086. DOI: 10.1109/ACCESS.2021.3063715.

Okamoto, M. and B. Miao. (1991). “Recognition of Mathematical Expressions by Using the Layout Structures of Symbols”. In: *Prof. International Conference on Document Analysis and Recognition*. Vol. 1. St. Malo, France. 242–250.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. (2002). “Bleu: a Method for Automatic Evaluation of Machine Translation”. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics. 311–318. DOI: 10.3115/1073083.1073135.

Pavan Kumar, P., A. Agarwal, and C. Bhagvati. (2012). "A Structure Based Approach for Mathematical Expression Retrieval". In: *Multi-disciplinary Trends in Artificial Intelligence*. Vol. 7694. Berlin, Heidelberg: Springer Berlin Heidelberg. 23–34. DOI: 10.1007/978-3-642-35455-7_3.

Peng, S., K. Yuan, L. Gao, and Z. Tang. (2021a). "MathBERT: A Pre-Trained Model for Mathematical Formula Understanding". *arXiv:2105.00377 [cs]*. May. URL: http://arxiv.org/abs/2105.00377.

Peng, S., K. Yuan, L. Gao, and Z. Tang. (2021b). "Mathbert: A pre-trained model for mathematical formula understanding". *arXiv preprint arXiv:2105.00377*.

Perozzi, B., R. Al-Rfou, and S. Skiena. (2014). "Deepwalk: Online Learning of Social Representations". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

Petersen, F., M. Schubotz, A. Greiner-Petter, and B. Gipp. (2023). "Neural Machine Translation for Mathematical Formulae". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics. 11534–11550. DOI: 10.18653/v1/2023.acl-long.645.

Pfahler, L. and K. Morik. (2020). "Semantic Search in Millions of Equations". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Virtual Event CA USA: ACM. 135–143. DOI: 10.1145/3394486.3403056.

Piękos, P., M. Malinowski, and H. Michalewski. (2021). "Measuring and Improving BERT's Mathematical Abilities by Predicting the Order of Reasoning." In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Online: Association for Computational Linguistics. 383–394. DOI: 10.18653/v1/2021.acl-short.49.

Pirolli, P. and S. Card. (1999). "Information Foraging". *Psychological Review*. 106(4): 643–675. DOI: 10.1037/0033-295x.106.4.643.

Rafailov, R., A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. Finn. (2023). "Direct Preference Optimization: Your Language Model is Secretly a Reward Model". In: *Advances in Neural Information Processing Systems*. Vol. 36. Curran Associates, Inc. 53728–53741. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/a85b405ed65c6477a4fe8302b5e06ce7-Paper-Conference.pdf.

Reichenbach, M. S., A. Agarwal, and R. Zanibbi. (2014). "Rendering Expressions to Improve Accuracy of Relevance Assessment for Math Search". In: *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 851–854.

Reusch, A., J. Gonsior, C. Hartmann, and W. Lehner. (2024). "Investigating the Usage of Formulae in Mathematical Answer Retrieval". In: *Advances in Information Retrieval*. Cham: Springer Nature Switzerland. 247–261.

Reusch, A., M. Thiele, and W. Lehner. (2022). "Transformer-Encoder-Based Mathematical Information Retrieval". In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Cham: Springer International Publishing. 175–189.

Robertson, S. (2004). "Understanding Inverse Document Frequency: On theoretical arguments for IDF". *J. Documentation*. 60(5): 503–520. DOI: 10.1108/00220410410560582.

Robertson, S. E. and S. Walker. (1994). "Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval". In: *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 3-6 July 1994 (Special Issue of the SIGIR Forum)*. ACM/Springer. 232–241. DOI: 10.1007/978-1-4471-2099-5_24.

Robertson, S. E. and H. Zaragoza. (2009). "The Probabilistic Relevance Framework: BM25 and Beyond". *Found. Trends Inf. Retr.* 3(4): 333–389. DOI: 10.1561/1500000019.

Rohatgi, S., W. Zhong, R. Zanibbi, J. Wu, and C. L. Giles. (2019). "Query Auto Completion for Math Formula Search". arXiv: 1912.04115 [cs.IR].

Roy, S. and D. Roth. (2015). "Solving General Arithmetic Word Problems". In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics. 1743–1752. DOI: 10.18653/v1/D15-1202.

Ruzicka, M., P. Sojka, and M. Líška. (2016). "Math Indexer and Searcher under the Hood: Fine-tuning Query Expansion and Unification Strategies". In: *Proc. of the 12th NTCIR Conference on Evaluation of Information Access Technologies. Noriko Kando, Tetsuya Sakai, and Mark Sanderson,(Eds.) NII Tokyo*. 331–337.

Sakai, T. (2007). "Alternatives to Bpref". In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*. Amsterdam, The Netherlands: ACM Press. 71. DOI: 10.1145/1277741.1277756.

Satpute, A., N. Gießing, A. Greiner-Petter, M. Schubotz, O. Teschke, A. Aizawa, and B. Gipp. (2024). "Can LLMs Master Math? Investigating Large Language Models on Math Stack Exchange". In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '24*. Washington DC, USA: Association for Computing Machinery. 2316–2320. DOI: 10.1145/3626772.3657945.

Schellenberg, T., B. Yuan, and R. Zanibbi. (2012). "Layout-Based Substitution Tree Indexing and Retrieval for Mathematical Expressions". In: *Document Recognition and Retrieval XIX, part of the IS&T-SPIE Electronic Imaging Symposium, Burlingame, California, USA, January 25-26, 2012, Proceedings*. Vol. 8297. *SPIE Proceedings*. SPIE. 82970I. DOI: 10.1117/12.912502.

Schubotz, M., A. Grigorev, M. Leich, H. S. Cohl, N. Meuschke, B. Gipp, A. S. Youssef, and V. Markl. (2016). "Semantification of Identifiers in Mathematics for Better Math Information Retrieval". In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '16*. Pisa, Italy: Association for Computing Machinery. 135–144. DOI: 10.1145/2911451.2911503.

Schubotz, M., P. Scharpf, K. Dudhat, Y. Nagar, F. Hamborg, and B. Gipp. (2018). "Introducing MathQA: a Math-Aware Question Answering System". *Information Discovery and Delivery*. 46(4): 214–224. DOI: 10.1108/IDD-06-2018-0022.

Shah, A. K., A. Dey, and R. Zanibbi. (2021). "A Math Formula Extraction and Evaluation Framework for PDF Documents". In: *16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part II*. Vol. 12822. *Lecture Notes in Computer Science*. Springer. 19–34. DOI: 10.1007/978-3-030-86331-9_2.

Shen, Y. and C. Jin. (2020). "Solving Math Word Problems with Multi-Encoders and Multi-Decoders". In: *Proceedings of the 28th International Conference on Computational Linguistics*. Barcelona, Spain (Online): International Committee on Computational Linguistics. 2924–2934. DOI: 10.18653/v1/2020.coling-main.262.

Shirahama, K. and M. Grzegorzek. (2016). "Towards Large-Scale Multimedia Retrieval Enriched by Knowledge about Human Interpretation". *Multim. Tools Appl.* 75(1): 297–331. DOI: 10.1007/S11042-014-2292-8.

Shumailov, I., Z. Shumaylov, Y. Zhao, Y. Gal, N. Papernot, and R. Anderson. (2024a). "The Curse of Recursion: Training on Generated Data Makes Models Forget". URL: https://arxiv.org/abs/2305.17493.

Shumailov, I., Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal. (2024b). "AI Models Collapse When Trained on Recursively Generated Data". *Nature*. July. DOI: 10.1038/s41586-024-07566-y.

Sojka, P. and M. Líška. (2011). "Indexing and Searching Mathematics in Digital Libraries". In: *Intelligent Computer Mathematics*. Vol. 6824. Berlin, Heidelberg: Springer Berlin Heidelberg. 228–243. DOI: 10.1007/978-3-642-22673-1_16.

Sojka, P., M. Růžička, and V. Novotný. (2018). "MIaS: Math-Aware Retrieval in Digital Mathematical Libraries". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. Torino Italy: ACM. 1923–1926. DOI: 10.1145/3269206.3269233.

Song, Y. and X. Chen. (2021). “Searching for Mathematical Formulas Based on Graph Representation Learning”. In: *Intelligent Computer Mathematics*. Vol. 12833. Cham: Springer International Publishing. 137–152. DOI: 10.1007/978-3-030-81097-9_11.

Stathopoulos, Y., S. Baker, M. Rei, and S. Teufel. (2018). “Variable Typing: Assigning Meaning to Variables in Mathematical Text”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics. 303–312. DOI: 10.18653/v1/N18-1028.

Stathopoulos, Y. and S. Teufel. (2015). “Retrieval of Research-level Mathematical Information Needs: A Test Collection and Technical Terminology Experiment”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 334–340.

Sudholt, S. and G. A. Fink. (2016). “PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents”. In: *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE Computer Society. 277–282.

Tan, M., L. Wang, L. Jiang, and J. Jiang. (2022). “Investigating Math Word Problems using Pretrained Multilingual Language Models”. In: *Proceedings of the 1st Workshop on Mathematical Natural Language Processing (MathNLP)*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics. 7–16. DOI: 10.18653/v1/2022.mathnlp-1.2.

Tausczik, Y. R., A. Kittur, and R. E. Kraut. (2014). “Collaborative problem solving: a study of MathOverflow”. In: *Computer Supported Cooperative Work, CSCW '14, Baltimore, MD, USA, February 15-19, 2014*. ACM. 355–367. DOI: 10.1145/2531602.2531690.

Thanda, A., A. Agarwal, K. Singla, A. Prakash, and A. Gupta. (2016). “A Document Retrieval System for Math Queries”. In: *Proc. NTCIR-12*. 346–353. Tokyo, Japan.

Truong, T.-N., C. Nguyen, R. Zanibbi, H. Mouchére, and M. Nakagawa. (2024). "A Survey on Handwritten Mathematical Expression Recognition: The Rise of Encoder-Decoder and GNN Models". *Pattern Recognition*: (to appear).

Tucker, Q. (2024). "Improving Histogram-Based Mathematical Formula Search with Local Matching and Learned Embeddings". *Tech. rep.* Rochester Institute of Technology. URL: https://www.cs.rit.edu/~rlaz/projects_theses/Tucker_Local_PHOC_Project_May2024.pdf.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. (2017). "Attention Is All You Need". *CoRR*. abs/1706.03762. URL: http://arxiv.org/abs/1706.03762.

Wallace, E., Y. Wang, S. Li, S. Singh, and M. Gardner. (2019). "Do NLP Models Know Numbers? Probing Numeracy in Embeddings". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics. 5307–5315. DOI: 10.18653/v1/D19-1534.

Wang, K., J. Pan, W. Shi, Z. Lu, M. Zhan, and H. Li. (2024). "Measuring Multimodal Mathematical Reasoning with MATH-Vision Dataset". arXiv: 2402.14804 [cs.CV]. URL: https://arxiv.org/abs/2402.14804.

Wang, M., Y. Tang, J. Wang, and J. Deng. (2017a). "Premise Selection for Theorem Proving by Deep Graph Embedding". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/18d10dc6e666eab6de9215ae5b3d54df-Paper.pdf.

Wang, Q., C. Kaliszyk, and J. Urban. (2018). "First Experiments with Neural Translation of Informal to Formal Mathematics". In: *Intelligent Computer Mathematics*. Cham: Springer International Publishing. 255–270.

Wang, X., C. Macdonald, N. Tonellotto, and I. Ounis. (2023). "Reproducibility, Replicability, and Insights into Dense Multi-Representation Retrieval Models: from ColBERT to Col". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*. ACM. 2552–2561. DOI: 10.1145/3539618.3591916.

Wang, Y., X. Liu, and S. Shi. (2017b). "Deep Neural Solver for Math Word Problems". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics. 845–854. DOI: 10.18653/v1/D17-1088.

Wang, Z., M. Zhang, R. G. Baraniuk, and A. S. Lan. (2021). "Scientific Formula Retrieval via Tree Embeddings". In: *2021 IEEE International Conference on Big Data (Big Data)*. 1493–1503. DOI: 10.1109/BigData52589.2021.9671942.

Wangari, K. D. V., R. Zanibbi, and A. Agarwal. (2014). "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*. ACM. DOI: 10.1145/2600428.2609481.

Wei, J., X. Wang, D. Schuurmans, M. Bosma, b. ichter brian, F. Xia, E. Chi, Q. V. Le, and D. Zhou. (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc. 24824–24837. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/9d5609613524ecf4f15af0f7b31abca4-Paper-Conference.pdf.

Welleck, S., J. Liu, R. L. Bras, H. Hajishirzi, Y. Choi, and K. Cho. (2021). "NaturalProofs: Mathematical Theorem Proving in Natural Language". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)*. URL: https://openreview.net/forum?id=Jvxa8adr3iY.

Wenzel, M., L. C. Paulson, and T. Nipkow. (2008). "The Isabelle Framework". In: *Theorem Proving in Higher Order Logics*. Berlin, Heidelberg: Springer Berlin Heidelberg. 33–38.

White, R. W. (2016). *Interactions with Search Systems*. Cambridge University Press.

Wu, Y., A. Q. Jiang, W. Li, M. Rabe, C. Staats, M. Jamnik, and C. Szegedy. (2022). "Autoformalization with Large Language Models". In: *Advances in Neural Information Processing Systems*. Vol. 35. Curran Associates, Inc. 32353–32368. URL: https://proceedings.neurips.cc/paper_files/paper/2022/file/d0c6bc641a56bebee9d985b937307367-Paper-Conference.pdf.

Xue, L., A. Barua, N. Constant, R. Al-Rfou, S. Narang, M. Kale, A. Roberts, and C. Raffel. (2022). "ByT5: Towards a Token-Free Future with Pre-trained Byte-to-Byte Models". *Transactions of the Association for Computational Linguistics*. 10: 291–306. DOI: 10.1162/tacl_a_00461.

Yang, K., A. Swope, A. Gu, R. Chalamala, P. Song, S. Yu, S. Godil, R. J. Prenger, and A. Anandkumar. (2023). "LeanDojo: Theorem Proving with Retrieval-Augmented Language Models". In: *Advances in Neural Information Processing Systems*. Vol. 36. Curran Associates, Inc. 21573–21612. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/4441469427094f8873d0fecb0c4e1cee-Paper-Datasets_and_Benchmarks.pdf.

Yasunaga, M. and J. D. Lafferty. (2019). "TopicEq: A Joint Topic and Mathematical Equation Model for Scientific Texts". *Proceedings of the AAAI Conference on Artificial Intelligence*. 33(July): 7394–7401. DOI: 10.1609/aaai.v33i01.33017394.

Yu, L., W. Jiang, H. Shi, J. Yu, Z. Liu, Y. Zhang, J. T. Kwok, Z. Li, A. Weller, and W. Liu. (2024). "MetaMath: Bootstrap Your Own Mathematical Questions for Large Language Models". URL: https://arxiv.org/abs/2309.12284.

Yu, M., G. Li, D. Deng, and J. Feng. (2016). "String Similarity Search and Join: A Survey". *Frontiers Comput. Sci.* 10(3): 399–417. DOI: 10.1007/S11704-015-5900-5.

Yuan, K., D. He, Z. Jiang, L. Gao, Z. Tang, and C. L. Giles. (2020). "Automatic Generation of Headlines for Online Math Questions". *Proceedings of the AAAI Conference on Artificial Intelligence*. 34(05): 9490–9497. DOI: 10.1609/aaai.v34i05.6493.

Yue, X., X. Qu, G. Zhang, Y. Fu, W. Huang, H. Sun, Y. Su, and W. Chen. (2023). "MAmmoTH: Building Math Generalist Models through Hybrid Instruction Tuning". URL: https://arxiv.org/abs/2309.05653.

Zanibbi, R., A. Aizawa, and M. Kohlhase. (2016a). "NTCIR-12 MathIR Task Overview". In: *Proc. NTCIR-12*. 299–308.

Zanibbi, R. and D. Blostein. (2012). "Recognition and Retrieval of Mathematical Expressions". *International Journal on Document Analysis and Recognition (IJDAR)*. 15(4): 331–357. DOI: 10.1007/s10032-011-0174-4.

Zanibbi, R., K. Davila, A. Kane, and F. W. Tompa. (2016b). "Multi-Stage Math Formula Search: Using Appearance-Based Similarity Metrics at Scale". In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. Pisa Italy: ACM. 145–154. DOI: 10.1145/2911451.2911512.

Zanibbi, R. and L. Yu. (2011). "Math Spotting: Retrieving Math in Technical Documents Using Handwritten Query Images". In: *2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, September 18-21, 2011*. IEEE Computer Society. 446–451. DOI: 10.1109/ICDAR.2011.96.

Zeng, H., C. Luo, B. Jin, S. M. Sarwar, T. Wei, and H. Zamani. (2024). "Scalable and Effective Generative Information Retrieval". In: *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*. ACM. 1441–1452. DOI: 10.1145/3589334.3645477.

Zhang, A., A. Goyal, W. Kong, H. Deng, A. Dong, Y. Chang, C. A. Gunter, and J. Han. (2015). "adaQAC: Adaptive Query Auto-Completion via Implicit Negative Feedback". In: *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 143–152.

Zhang, D., L. Wang, L. Zhang, B. T. Dai, and H. T. Shen. (2020a). "The Gap of Semantic Parsing: A Survey on Automatic Math Word Problem Solvers". *IEEE Trans. Pattern Anal. Mach. Intell.* 42(9): 2287–2305. DOI: 10.1109/TPAMI.2019.2914054.

Zhang, M., S. Baral, N. Heffernan, and A. Lan. (2022). "Automatic Short Math Answer Grading via In-context Meta-learning". In: *Proceedings from the 15th International Conference on Educational Data Mining*.

Zhang, M., Z. Cui, M. Neumann, and Y. Chen. (2018). "An End-to-End Deep Learning Architecture for Graph Classification". *Proceedings of the AAAI Conference on Artificial Intelligence*. 32(1). DOI: 10.1609/aaai.v32i1.11782.

Zhang, T., V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. (2020b). "BERTScore: Evaluating Text Generation with BERT". URL: https://arxiv.org/abs/1904.09675.

Zhao, J., M.-Y. Kan, and Y. L. Theng. (2008). "Math Information Retrieval: User Requirements and Prototype Implementation". In: *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries - JCDL '08*. Pittsburgh PA, PA, USA: ACM Press. 187. DOI: 10.1145/1378889.1378921.

Zhong, L., J. Wu, Q. Li, H. Peng, and X. Wu. (2024). "A Comprehensive Survey on Automatic Knowledge Graph Construction". *ACM Comput. Surv.* 56(4): 94:1–94:62. DOI: 10.1145/3618295.

Zhong, W., S.-C. Lin, J.-H. Yang, and J. Lin. (2023). "One Blade for One Purpose: Advancing Math Information Retrieval using Hybrid Search". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '23*. Taipei, Taiwan: Association for Computing Machinery. 141–151. DOI: 10.1145/3539618.3591746.

Zhong, W. and R. Zanibbi. (2019). "Structural Similarity Search for Formulas Using Leaf-Root Paths in Operator Subtrees". In: *Advances in Information Retrieval*. Vol. 11437. Cham: Springer International Publishing. 116–129. DOI: 10.1007/978-3-030-15712-8_8.

Zhong, W., X. Zhang, J. Xin, R. Zanibbi, and J. Lin. (2021). "Approach Zero and Anserini at the CLEF-2021 ARQMath Track: Applying Substructure Search and BM25 on Operator Tree Path Tokens". *Proc. CLEF 2021 (CEUR Working Notes)*.

Zhu, L., C. Zheng, W. Guan, J. Li, Y. Yang, and H. T. Shen. (2024). "Multi-Modal Hashing for Efficient Multimedia Retrieval: A Survey". *IEEE Trans. Knowl. Data Eng.* 36(1): 239–260. DOI: 10.1109/TKDE.2023.3282921.

Zobel, J. and A. Moffat. (2006). "Inverted Files for Text Search Engines". *ACM Comput. Surv.* 38(2): 6–es. DOI: 10.1145/1132956.1132959.