Mathematics in Wikidata*

Philipp Scharpf[1], Moritz Schubotz[2,3], and Bela Gipp[3]

[1] University of Konstanz, Germany philipp.scharpf@uni-konstanz.de
[2] FIZ Karlsruhe, Germany moritz.schubotz@fiz-karlsruhe.de
[3] University of Wuppertal, Germany gipp@uni-wuppertal.de

Abstract. Documents from Science, Technology, Engineering, and Mathematics (STEM) disciplines usually contain a significant amount of mathematical formulae alongside text. Some Mathematical Information Retrieval (MathIR) systems, e.g., Mathematical Question Answering (MathQA), exploit knowledge from Wikidata. Therefore, the mathematical information needs to be stored in items. In the last years, there have been efforts to define several properties and seed formulae together with their constituting identifiers into Wikidata. This paper summarizes the current state, challenges, and discussions related to this endeavor. Furthermore, some data mining methods (supervised formula annotation and concept retrieval) and applications (question answering and classification explainability) of the mathematical information are outlined. Finally, we discuss community feedback and issues related to integrating Mathematical Entity Linking (MathEL) into Wikidata and Wikipedia, which was rejected in 33% and 12% of the test cases, for Wikidata and Wikipedia respectively. Our long-term goal is to populate Wikidata, such that it can serve a variety of automated math reasoning tasks and AI systems.

Keywords: Wikidata · Mathematical Information Retrieval · Mathematical Entity Linking · Mathematical Question Answering

1 Introduction

Mathematical Information Retrieval (MathIR) systems, such as Document Recommender (DocRec), Mathematical Question Answering (MathQA), and Automatic Document Classification (ADC), need to process and query mathematical formulae. Since Wikidata has been proven useful as a semantic grounding database for Natural Language Processing (NLP) approaches and applications, it was consequential to transfer and adapt classical IR and NLP methods to the special case of mathematical knowledge. In 2016, we implemented support for mathematical properties, such as 'defining formula' (P2534), which were pro-

 This work was supported by DFG grant GI-1259-1.

posed[4] and used[5]. Later, additional properties to include the semantics of the formula identifiers were added [6].

Our long-term goal is to build *math.wikipedia.org*, a large collaborative, semi-formal, machine-readable, language-independent mathematics encyclopedia. Its purpose will be to provide the backbone for automated reasoning tasks, concept entity linking, knowledge-graph population, question answering, and more.

Table 1. Different mathematical Wikidata properties with their occurrence frequencies (as of July 19, 2021) and an example.

Property	**Frequency**	**Example**
'defining formula' (P2534)	5166	E = m c^2
'in defining formula' (P7235)	703	E
'calculated from' (P4934)	780	mass
'has part' (P527)	179	energy

Table 1 shows the four most relevant and used properties for mathematical concept items. While the 'defining formula' property is employed to store an entire formula (e.g., `E=mc^2` in Q35875), 'in defining formula' (P7235), 'calculated from' (P4934), and 'has part' (P527) are used to denote the identifier information.

The occurrence frequency numbers were obtained by running Wikidata SPARQL queries[7]. For example, the number of items with 'defining formula' property can be retrieved using the following query snipped:

```
#Retrieve all items with 'defining formula' property P2534
SELECT ?formula WHERE {
  ?item wdt:P2534 ?formula.
  }
}
```

Figure 1 illustrates (using the 'mass-energy equivalence' Q35875 item as an example) how the defining formula is displayed in the Wikidata user interface.

Currently, as of July, 30th 2021, the usage frequency distribution is as shown in Table 1. The 'has part' property, which was historically used first was gradually replaced by 'calculated from', which is now more than four times as prominent. For a discussion of the differences between the two properties and their individual limitations, see Section 3.3.

In this resource paper, we discuss how the mathematical knowledge stored in Wikidata can be extended and employed for Mathematical Entity Linking (MathEL) and its applications, e.g., Mathematical Question Answering (MathQA) and Document Classification Explainability (DCE).

[4] https://www.wikidata.org/w/index.php?title=Property:P2534&oldid=303933381
[5] https://www.wikidata.org/w/index.php?title=Q35875&oldid=303968820
[6] https://www.wikidata.org/w/index.php?title=Property:P4934&oldid=646697942
[7] https://query.wikidata.org

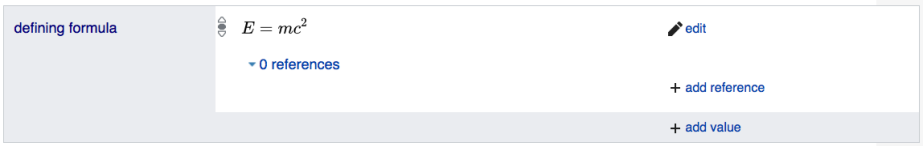

Fig. 1. Displaying the 'defining formula' property of the item 'energy-mass equivalence'. This excerpt from the Wikidata item page shows where the LaTeX formula string can be inserted.

The remainder of this paper is structured as follows. In Section 2, we describe how the knowledge can be distilled by annotating mathematical documents (papers, articles, etc.). We show how this can be accelerated using an annotation recommender system. In Section 3, we present standards and systems for benchmarking the knowledge for Mathematical Information Retrieval (MathIR) experiments. Mathematical Entity Retrieval and Linking methods are introduced, and community feedback on incorporating MathEL data into Wikidata and Wikipedia is discussed. Section 4 outlines MathQA and DCE as two example applications of MathEL and concludes with an outlook to challenges and future work.

2 Mathematical Entity Annotation

The process of Mathematical Entity Linking can be comprised of 1) Mathematical Entity Annotation and 2) Mathematical Entity Retrieval. In this chapter, we start with 1) by presenting approaches for document annotation and its acceleration by annotation recommendation.

2.1 Document Annotation

Document annotations are generally employed to provide additional information about a resource (e.g., comments) or to link resources (e.g., to URLs). The Web Annotation Data Model[8] specifies the annotation model structure (id, type, property, relationship) in JSON format. Moreover, RDF classes and ontologies should be defined and serialized according to the Web Annotation Vocabulary[9]. Several annotation tools and recommender systems for linked data have been developed so far. Tietz et al. present a system for Wordpress [24] that recommends DBpedia resources and visualizes the annotation process. Users can explore background information and relationships between named entities. Vagliano et al. provide a technical report [25] on semantic annotation of user reviews using DBpedia and Wikidata. Purwitasari et al. introduce an ontology-based annotation

[8] https://www.w3.org/TR/annotation-model
[9] https://www.w3.org/TR/annotation-vocab

recommender for learning material [10] using Latent Semantic Analysis (LSA) and WordNet to determine the context of content categories, which are then structured into an ontology model. Lastly, Wiesing et al. developed an RDF annotation tool (KAT) specific for STEM documents in XHTML format [26].

2.2 Annotation Recommendation

To disambiguate and match mathematical expressions in Wikipedia articles to Wikidata items [12], the 'AnnoMathTeX' formula and identifier annotation recommender system[10] was developed. The system is designed to suggest Wikidata item name and QID candidates provided from several sources, such as the arXiv[11], Wikipedia, Wikidata, or the text that surrounds the formula. In the first evaluation, it could be shown that 78% of the identifier name recommendations were accepted by the user. In additional experiments, the community acceptance of the Wikipedia article link and Wikidata item seed edits was assessed [15]. For 88% of the edited Wikipedia articles and 67% of the Wikidata items, the contributions were accepted. Moreover, the annotation could be accelerated by a speedup of factor 1.4 for formulae and 2.4 for identifiers. The 'AnnoMathTeX' system is ready to be integrated seamlessly into the Wikimedia user interfaces via a 'MathWikiLink' API.

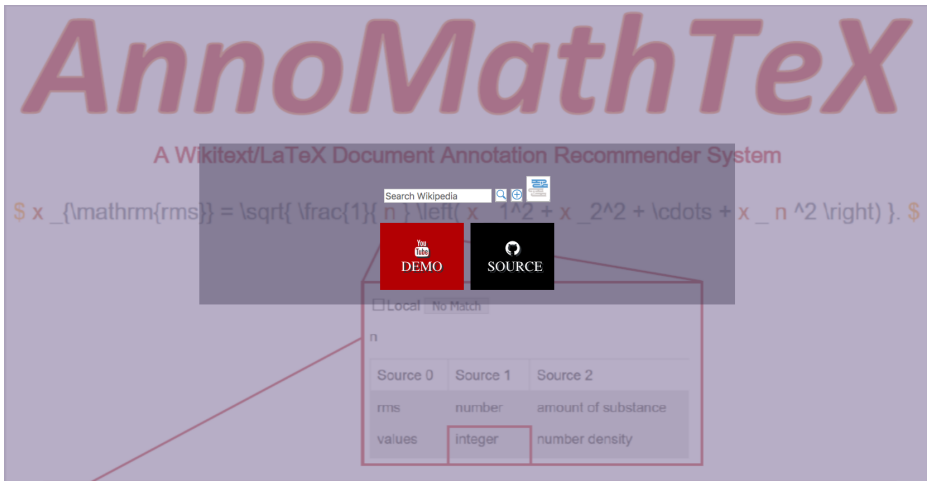

Fig. 2. The start screen of AnnoMathTeX, where the user can start or continue annotating selected Wikipedia articles.

We presented the system with its applications at the Wikiworkshop21 (WWW21 conference) [15]. Figure 2 shows the User Interface of the 'AnnoMath-

[10] https://annomathtex.wmflabs.org
[11] https://arxiv.org

TeX' system at the start, where Wikipedia Wikitext or arXiv LaTeX articles can be selected, loaded, and deleted.

If the user clicks on a formula or identifier in the loaded document , recommendations are displayed as shown in Figure 3 for the example formula $F = m \cdot a$, which is seeded into Wikidata as the item 'Newton's second law of motion' (Q3268014).

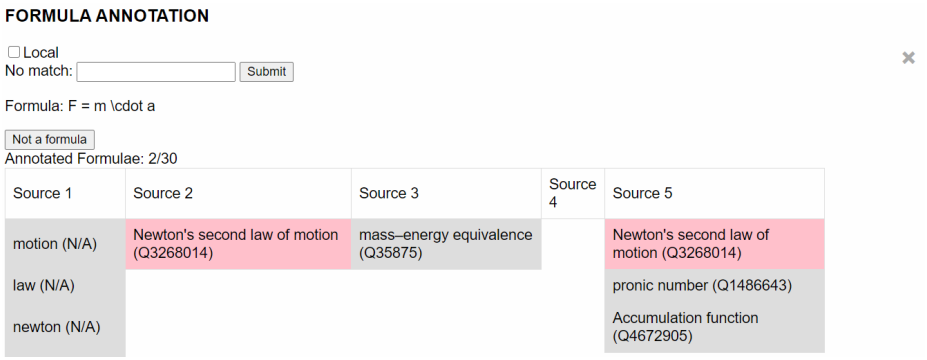

Fig. 3. Popup table containing recommendations for the annotation of the formula '$F = m \cdot a$', provided from different sources (cut off after third ranked).

Figure 4 shows an example where the formula name recommendation is very specific within a concept hierarchy.

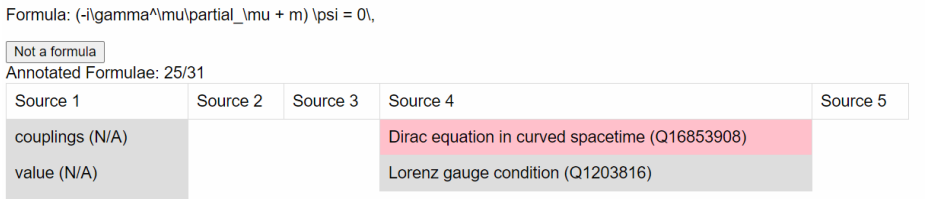

Fig. 4. Example popup table providing a very specific applicable recommendation, which is selected (highlighted in red).

2.3 Annotation Guidelines and Issues

The purpose of the first testing phase of the system was to elaborate on how the mathematics knowledge contained in Wikipedia articles can be transferred to Wikidata statements. For the annotation, we developed the following annotation rules or guidelines:

- Annotate identifiers first, such that the formula name recommendation retrieval from Wikidata via the 'has part' properties is enabled;
- Do not annotate identifier describing objects, such as 'gas', 'solid', 'line' instead of quantities or constants;
- Ignore derivative d characters, such as in d/dt and all indices (superscript or subscript);
- Locally different meanings of the same identifier within an article should be avoided (appeal to editors);
- Ignore block-level formulae that are not relations (equations, inequations, etc.) or do not have a single identifier right-hand side, e.g., $\sum_i I_i = \sum_i r_i^2 m_i$, $0 = ...$, $dE = ...$ Also, ignore formulae in tables, and derivations;
- Proper names (e.g., 'Planck constant') must be capitalized according to the conventions from 'Content dictionary description' (DRMF) [3].

During the annotation process, we discovered the following issues:

- It is not possible to parse equations with no spaces between identifiers, e.g., in the right-hand side of the LaTeX string '`L = rmv`';
- There are different common practices to denote vectors in LaTeX, e.g., `\vec` vs. `\mathbf`;
- There are different common practices for properties in Wikidata to include the semantics of the formula elements or identifier, e.g., 'has part' (P527) 'calculated from' (P4934) - see the discussion in Section 3.3;
- Sometimes two names are both commonly used to denote the same Formula Concept, e.g., 'M-sigma relation' (Q3424023) and 'Faber–Jackson relation' (Q1390162);
- In case the Wikidata item for a Formula Concept was missing, and we had to create it, we needed to reinsert the new QID into the annotations table manually.

In the future, the process of discovering new issues and requirements to improve the system and extend the annotation guidelines will be continued. Wikimedia users can collaboratively contribute to this joint endeavor.

3 Mathematical Entity Linking

3.1 Mathematical Entity Benchmarking

The open-source and open access formula benchmark system *MathMLben*[12] was introduced to facilitate the conversion between different mathematical formats such as LaTeX variations and Computer Algebra Systems (CAS) [19]. Figure 5 shows the Graphical User Interface (GUI) of the system, displaying the expression tree of an example formula. Each formula identifier can be annotated with Wikidata QID macros. The annotation functionality was motivated by the potential to define semantic relatedness for formulae by counting Wikidata links

[12] https://mathmlben.wmflabs.org

between them [14]. The MathMLben database contains 375 expressions or formulae (GoldIDs) from Wikipedia, the arXiv, and the Digital Library of Mathematical Functions (DLMF). The content is ranging from individual symbols to complex multi-line formulae. It additionally contains meta-information, such as the source URL or document page it is retrieved from. Expressions 1 to 100 are random samples taken from the *National Institute of Informatics Testbeds and Community for Information access Research Project* (NTCIR) 11/12 Math Wikipedia Task [1]. Expressions 101 to 200 are random samples taken from the *NIST Digital Library of Mathematical Functions* (DLMF) [6] available on the website https://dlmf.nist.gov containing around 10.000 labeled LaTeX formulae with semantic markup classified in 36 categories [2, 4]. Expressions 201 to 305 were selected from the NTCIR arXiv and NTCIR-12 Wikipedia dataset retrieval. 70 % of these formulae were taken from the arXiv and 30 % from a Wikipedia dump. The remaining formulae were extracted from an annotation of 25 selected Wikipedia articles from physics (classical mechanics) [15].

For each Gold ID entry or formula, there is an input field for the *Formula Name*, *Formula Type* (definition, equation, relation or general formula), **Original Input TeX** and manually **Corrected TeX** together with a **Hyperlink** to the source. The **Semantic LaTeX Input** field is used for the semantic annotations, as a grounding for the generation of Content MathML with Wikidata annotations by LaTeXML [9, 5]. The corrected TeX is rendered in real time by Mathoid [23]. Moreover, an expression tree is displayed, rendered by our visualization tool VMEXT [20]. For each symbol in the tree, the assigned annotation is shown as a yellow mouse-over infobox containing the Wikidata QID, name, and description (if available). The system includes a user guide on how to access raw data or contribute by extending or correcting the expression tree or (Wikidata) annotations.

3.2 Formula Concept Seeding and Retrieval

In 2018, we first introduced linking mathematical formula content to Wikidata, both in MathML and LaTeX markup [19, 14]. In 2019, we called out for a 'Formula Concept Discovery (FCD) and Formula Concept Recognition (FCR) challenge' to elaborate automated Mathematical Entity Linking. For our FCD approach, we could achieve a recall of 68% for retrieving equivalent representations of frequent formulae and 72% for extracting the formula name (assigned to a Wikidata item) from the surrounding text on the NTCIR arXiv dataset [1]. We defined a 'Formula Concept' as a 'labeled collection of mathematical formulae that are equivalent but have different representations through notation, e.g., the use of different identifier symbols or commutations' [13]. For example, the formula $E = mc^2$ can be regarded as being one representation of the Formula Concept labeled 'mass-energy equivalence'. A different representation of this same concept with different notation and rearrangement could be $\mu = \epsilon/c^2$.

The following snipped exemplifies how Einstein's famous formula $E = mc^2$, the item 'mass-energy equivalence' (Q35875) can be found via a SPARQL query. Based on the snippet, a *formula search engine* on Wikidata can be implemented.

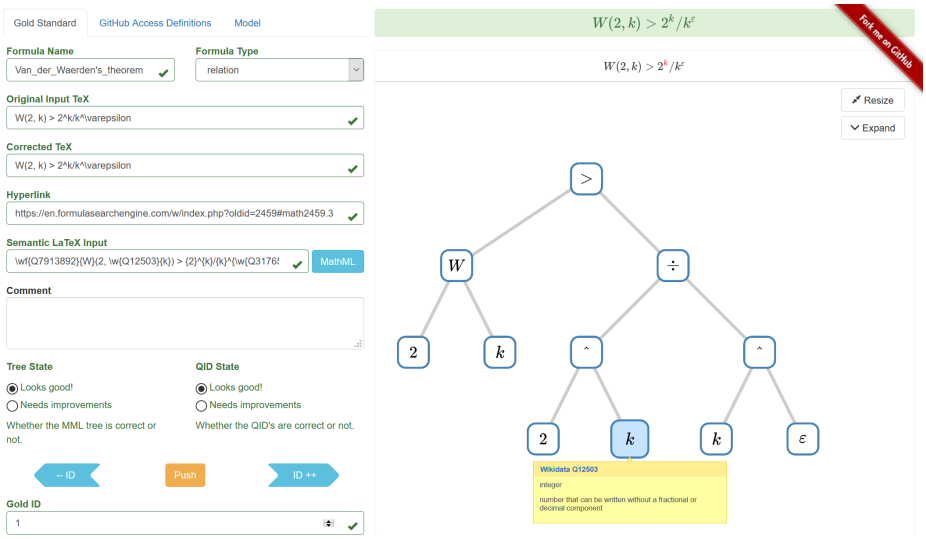

Fig. 5. Graphical User Interface of *MathMLben* providing several TeX input fields (left) and a mathematical expression tree rendered by the VMEXT visualization tool (right) [19].

```
#Retrieve all items with label, description,
and formula, whose defining formula property (P2534)
contains the string 'E=mc^2'
    SELECT ?item ?itemLabel ?itemDescription
    ?definingFormula
    WHERE {
    ?item wdt:P2534 ?definingFormula;
    FILTER( contains(?definingFormula, 'E=mc^2'@en))
    SERVICE wikibase:label
    }
}
```

Figure 6 illustrates how to make use of the 'has part' (P527) property to get all items with formula whose identifiers are annotated as 'energy' (Q11379) and 'speed of light' (Q2111). Based on the snippet, a *semantic formula search engine* on Wikidata can be implemented.

3.3 Community Feedback on Wikidata and Wikipedia

In [15] we presented the evaluation of our *AnnoMathTeX* formula and identifier annotation recommender system on a selection of 25 Wikipedia articles from physics. The linked formula concepts were seeded into Wikidata and persisted in our formula benchmark system *MathMLben* (see Section 3.1).

```
SELECT ?item ?itemLabel ?itemDescription
WHERE {
?item wdt:P527 wd:Q11379.
?item wdt:P527 wd:Q2111
SERVICE wikibase:label
{ bd:serviceParam wikibase:language "en" .} }
```

Fig. 6. SPARQL query making use of the 'has part' property. It returns all items that that are connected to 'energy' (Q11379) and 'speed of light' (Q2111) through the 'has part' property (P527).

The formula linkings from the annotated Wikipedia articles were included in the Wikitext via `qid` attribute of the `<math>` tag. After uploading the edited articles to Wikipedia, the following issues were pointed out by the community:

- One community member responded that Wikidata should be usable independently of Wikipedia, not having to comply with the technical requirements for the special page display.
- It was pointed out that currently, for the Wikdata items that have a 'defining formula' (P2534), the use of the 'calculated from' (P4934) property is much higher than 'has part' (P527). The claim was that 'has part' is only a relict from the past, which will not be used anymore for newly populated items.
- Studying some sample equations with 'calculated from' properties, another user found that for 'Gauss's law for magnetism' (Q1195250) with the formula $\nabla \cdot \mathbf{B} = 0$ calculated from 'magnetic field' B and 'divergence' $\nabla\cdot$ does not make sense. On the other hand, it was asked if 'length' or 'time' was indeed a 'part of' 'acceleration'? In summary, concerns about the general validity of both properties were expressed.
- Furthermore, the coexistence and different benefits of the properties 'quantity symbol (string)' (P416), 'quantity symbol (LaTeX)' (P7973), and 'defining formula' (P2534) for the subexpression strings were discussed.
- One user argued that using the latter, only two properties (P2234 and P527) would be needed to develop applications, such as the special page, which is planned to be displayed as popup in the future[13].
- It was argued that the property P2234 would be general enough to be suitable for longer expressions, such as `f(x)` or `\exp`, which are not just symbols.
- Another user replied that one should distinguish between a definition (typically using an '=' sign) and citation (possibly using different symbols, such as m or μ for 'mass') of a quantity.

Less than half a day after the execution of the script, two Wikipedia editors started a discussion on our user's talk page[14], pointing out the following issues:

[13] https://phabricator.wikimedia.org/T208758
[14] https://en.wikipedia.org/wiki/User_talk:PhilMINT

- The 'defining formula' property of the corresponding Wikidata item is edited and evolving independently from the formula strings linked in the Wikipedia articles;
- The Wikidata items need to be very specific to account for a particular formula. See for example 'kinetic energy' (Q46276): $T = \frac{1}{2}mv^2$ vs. 'kinetic energy of rotating body' (Q104145205): $E_\mathrm{r} = 12I\omega^2$;
- If a Wikidata item has two or more 'defining formula' properties, only the first is displayed in the 'Special page' (even if the 'has part' identifier annotations refer to another). This is problematic, e.g., with 'Hooke's law' (Q170282), which currently has both $F = kX$ and $\sigma = E\varepsilon$ as 'defining formula';
- Sometimes disambiguation is needed to distinguish the physics terms from other word meanings, e.g., 'work' (Q42213) with description: 'energy transferred to an object via the application of force on it through a displacement' vs. 'work' (Q6958747): 'particular form of activity, sold by many people to sustain themselves';
- In the 'Equations of motion' article, a one-line formula includes three sub formulae with different meanings that would need three different QIDs;
- It should be possible for a Wikipedia reader to edit the formula and identifiers directly on the special page, such that the changes get transferred to Wikidata;
- The special pages and corresponding Wikidata items should have a 'what links here' link, providing a list of every page with a Wikilink to this page. This way, dependencies can be analyzed, and editors warned.

In summary, issues with the formula data representation in both Wikipedia and Wikidata as well as their exchange communication were identified. A subsequent discussion on the issue to find an appropriate property for the identifier semantics was started[15].

4 Applications and Outlook

In this final section, we outline some applications of Mathematical Entity Linking (with Wikidata) and conclude with an outlook to future work potential.

4.1 Mathematical Question Answering

Motivated by an increasing number of questions on Question Answering (QA) platforms like Math Stack Exchange (MSE) [17], signifying a growing information need to answer math-related questions, we developed a Mathematical Question Answering (MathQA) system [21]. Our open source and open data approach retrieves mathematical formulae using their concept names or querying formula identifier relationships from the Wikidata knowledge graph. Furthermore, we

[15] https://www.wikidata.org/wiki/Wikidata:Property_proposal/symbol_represents

developed Unsupervised Formula Labeling (UFL) for semantic formula search and question answering on the arXiv preprint repository and Wikipedia.

Figure 7 shows the MathQA UI with an example relationship question. A *Question Parsing Module* transforms natural language questions into a triple representation. Subsequently, a *Formula Retrieval Module* queries the Wikidata knowledge-base for the requested formula and presents the result to the user. The user can subsequently choose values for the occurring variables and order a calculation that is done by a *Calculation Module*. For the retrieved formula match, the Wikidata source link is displayed and identifier names and constant values (e.g., for the speed of light) are fetched from the respective items. The system employs the formula properties that are discussed in Section 1.

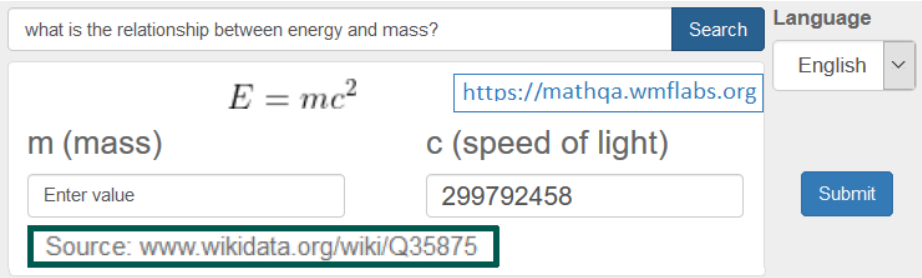

Fig. 7. MathQA semantic search example relationship question with identifier name and value retrieval and calculation. Demovideo at purl.org/mathqa.

4.2 Document Classification Explainability

Another application of MathEL can is document subject classification [18, 22], which is essential for structuring (digital) libraries and allowing readers to search for literature within a specific field. Supervised (automatic) or semi-supervised (semi-automatic) Machine Learning algorithms can support human domain expert classifiers by predicting subject classes for unclassified documents using classification information of documents that are already labeled. While humans can, in principle, explain their decisions, for machines, it is more difficult. This shortcoming motivated explainable AI research to address the problem of Machine Learning decisions being a black box [8]. In 2016, 'LIME' - a method for 'Local Interpretable Model-agnostic Explanations' was introduced to improve human trust in a classifier [11]. In 2019, 'SHAP' - an approach to improve the interpretability of tree-based models using game-theoretic Shapley values was presented to enhance understanding global model structure based on combining many local explanations [7]. Both LIME and SHAP models are available open-source and heavily employed. For the classification of natural language texts, such as legal or medical documents, explainer approaches have already successfully been applied. However, documents from Science, Technology, Engineering,

and Mathematics (STEM) disciplines are more difficult to tackle since they contain a significant amount of mathematical formulae alongside text [14, 12].

Mathematical Entity Linking can help to foster explainability for STEM documents. We are currently working on first approaches for *STEM document classification explainability* using classical and mathematical Entity Linking [16]. Mining a collection of documents from the arXiv preprint repository (NTCIR and zbMATH dataset), we could show that mathematical entities have the potential to provide high explainability as they are a crucial part of a STEM document. Our full paper contribution has been submitted very recently and is currently under review.

4.3 Conclusion and Future Work

In this resource paper, we showed how Wikidata can be employed for the semantic grounding of mathematical entities in documents. We introduced the data model of mathematical statements and discussed the use of different properties for the semantics of formula identifier parts. We presented some example SPARQL queries to access the mathematical knowledge in Wikidata. Next, we discussed how to obtain entity linking data via document annotation using systems such as our 'AnnoMathTeX' formula and identifier annotation recommender system. We introduced guidelines for formula annotation and reported issues occurring in the annotation process. Moreover, we presented possibilities to benchmark Mathematical Entity Linking (MathEL) with Wikidata using our 'MathMLben' gold-standard UI. Next, we discussed community feedback on Wikidata and Wikipedia on the usage of MathEL and the different properties involved. Finally, we introduced two applications of MathEL: Mathematical Question Answering (MathQA) and document classification explainability.

Since mathematical Formula Concepts [13] link both mathematical and natural language, MathEL can bridge two worlds and is thus a very valuable method to foster the methodological understanding of mathematical knowledge. Wikidata can help to achieve this by storing and linking both the concept name (with QID) and the formula string, e.g., linking 'mass-energy equivalence' (Q35875) with `E=mc^2`. However, there are still numerous challenges to tackle. Different symbols are used for constants and variables, such as m or μ for 'mass'. Different unit systems and substitutions are rendering some identifiers or terms implicit. And other types of notational freedom (e.g., differential vs. integral forms or derivative notation) are making automated MathEL more difficult.

Summarizing, our key findings are 1) Wikidata is a valuable resource for storing and retrieving mathematical knowledge; 2) The data model and representation of mathematical statements still suffers from various issues, such as disagreement on the use of specific properties; 3) By creating a formula semantics benchmark (MathMLben) and annotation recommender system (AnnoMathTeX), we are able to foster reproducibility of MathIR experiments and speed up the process of Wikidata knowledge-base and knowledge-graph population; 4) The machine-interpretable mathematical knowledge can be exploited

by various MathIR AI systems, such as question answering or automated mathematical reasoning.

MathEL research is currently at an early stage, and there is still a lot of questions and potential to explore. In the future, we will continue our research on the discovery and recognition of Formula Concept as a basis for MathEL. Moreover, we will promote applications, such as MathQA and STEM document classification explainability, by extending our existing systems and approaches. Furthermore, we will introduce new applications, such as our currently developing 'PhysWikiQuiz' physics question generation and interrogation system using Wikidata. Finally, we will create and publish large MathEL benchmark datasets and integrate annotation entity linking recommendations into the editing view of Wikipedia articles.

References

1. Aizawa, A., Kohlhase, M., Ounis, I., Schubotz, M.: NTCIR-11 math-2 task overview. In: NTCIR. National Institute of Informatics (NII) (2014)
2. Cohl, H.S., McClain, M.A., Saunders, B.V., Schubotz, M., Williams, J.C.: Digital repository of mathematical formulae. In: Conference on Intelligent Computer Mathematics (CICM), Coimbra, Portugal. pp. 419–422 (2014). https://doi.org/10.1007/978-3-319-08434-3_30
3. Cohl, H.S., Schubotz, M.: Content dictionary description: select symbols from chapter 9 of the kls dataset in the drmf (2017)
4. Cohl, H.S., Schubotz, M., McClain, M.A., Saunders, B.V., Zou, C.Y., Mohammed, A.S., Danoff, A.A.: Growing the digital repository of mathematical formulae with generic sources. pp. 280–287 (2015). https://doi.org/10.1007/978-3-319-20615-8_18
5. Ginev, D., Stamerjohanns, H., Kohlhase, M.: The latexml daemon: Editable math on the collaborative web. In: LWA 2011, Magdeburg, Germany. pp. 255–256 (2011)
6. Lozier, D.W.: NIST digital library of mathematical functions. Ann. Math. Artif. Intell. **38**(1-3), 105–119 (2003)
7. Lundberg, S.M., Erion, G.G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.: Explainable AI for trees: From local explanations to global understanding. CoRR (2019)
8. Mahoney, C.J., Zhang, J., Huber-Fliflet, N., Gronvall, P., Zhao, H.: A framework for explainable text classification in legal document review. In: IEEE BigData. pp. 1858–1867. IEEE (2019)
9. Miller, B.: `LaTeXML`: A LaTeX to XML converter. http://dlmf.nist.gov/LaTeXML/, http://dlmf.nist.gov/LaTeXML/, accessed: 2018-05-09
10. Purwitasari, D., Yuniar, E., Yuhana, U.L., Siahaan, D.O.: Ontology-based annotation recommender for learning material using contextual analysis. In: Proceedings of the IETEC'11 Conference, Kuala Lumpur, Malaysia. (2011)
11. Ribeiro, M.T., Singh, S., Guestrin, C.: "why should I trust you?": Explaining the predictions of any classifier. In: KDD. pp. 1135–1144. ACM (2016)
12. Scharpf, P., Mackerracher, I., Schubotz, M., Beel, J., Breitinger, C., Gipp, B.: *AnnoMath TeX* - a formula identifier annotation recommender system for STEM documents. In: RecSys. pp. 532–533. ACM (2019)
13. Scharpf, P., Schubotz, M., Cohl, H.S., Gipp, B.: Towards formula concept discovery and recognition. In: BIRNDL@SIGIR. CEUR Workshop Proceedings, vol. 2414, pp. 108–115. CEUR-WS.org (2019)

14. Scharpf, P., Schubotz, M., Gipp, B.: Representing mathematical formulae in content mathml using wikidata. In: BIRNDL@SIGIR. CEUR Workshop Proceedings, vol. 2132, pp. 46–59. CEUR-WS.org (2018)
15. Scharpf, P., Schubotz, M., Gipp, B.: Fast linking of mathematical wikidata entities in wikipedia articles using annotation recommendation. In: Proceedings of the Web Conference (WWW) 2021. ACM / IW3C2 (April 2021). https://doi.org/10.1145/3442442.3452348
16. Scharpf, P., Schubotz, M., Gipp, B.: Towards explaining stem document classification using mathematical entity linking. arXiv preprint arXiv:2109.00954 (2021)
17. Scharpf, P., Schubotz, M., Greiner-Petter, A., Ostendorff, M., Teschke, O., Gipp, B.: Arqmath lab: An incubator for semantic formula search in zbmath open? In: CLEF (Working Notes). CEUR Workshop Proceedings, vol. 2696. CEUR-WS.org (2020)
18. Scharpf, P., Schubotz, M., Youssef, A., Hamborg, F., Meuschke, N., Gipp, B.: Classification and clustering of arxiv documents, sections, and abstracts, comparing encodings of natural and mathematical language. In: JCDL. pp. 137–146. ACM (2020)
19. Schubotz, M., Greiner-Petter, A., Scharpf, P., Meuschke, N., Cohl, H.S., Gipp, B.: Improving the representation and conversion of mathematical formulae by considering their textual context. In: JCDL. pp. 233–242. ACM (2018)
20. Schubotz, M., Meuschke, N., Hepp, T., Cohl, H.S., Gipp, B.: VMEXT: A visualization tool for mathematical expression trees. In: CICM, Edinburgh, UK. pp. 340–355 (2017). https://doi.org/10.1007/978-3-319-62075-6_24
21. Schubotz, M., Scharpf, P., Dudhat, K., Nagar, Y., Hamborg, F., Gipp, B.: Introducing mathqa - A math-aware question answering system. Information Discovery and Delivery **42 - No. 4**, 214–224 (2019). https://doi.org/10.1108/IDD-06-2018-0022
22. Schubotz, M., Scharpf, P., Teschke, O., Kühnemund, A., Breitinger, C., Gipp, B.: Automsc: Automatic assignment of mathematics subject classification labels. In: CICM. Lecture Notes in Computer Science, vol. 12236, pp. 237–250. Springer (2020)
23. Schubotz, M., Wicke, G.: Mathoid: Robust, scalable, fast and accessible math rendering for wikipedia. In: CICM. Lecture Notes in Computer Science, vol. 8543, pp. 224–235. Springer (2014)
24. Tietz, T., Waitelonis, J., Jäger, J., Sack, H.: refer: a linked data based text annotation and recommender system for wordpress. In: International Semantic Web Conference (Posters & Demos). CEUR Workshop Proceedings, vol. 1690. CEUR-WS.org (2016)
25. Vagliano, I., Monti, D., Scherp, A., Morisio, M.: Content recommendation through semantic annotation of user reviews and linked data. In: K-CAP. pp. 32:1–32:4. ACM (2017)
26. Wiesing, T., Schmoll, F.: KAT: an annotation tool for STEM documents. In: FM4M/MathUI/ThEdu/DP/WIP@CIKM. CEUR Workshop Proceedings, vol. 1785, pp. 66–72. CEUR-WS.org (2016)