

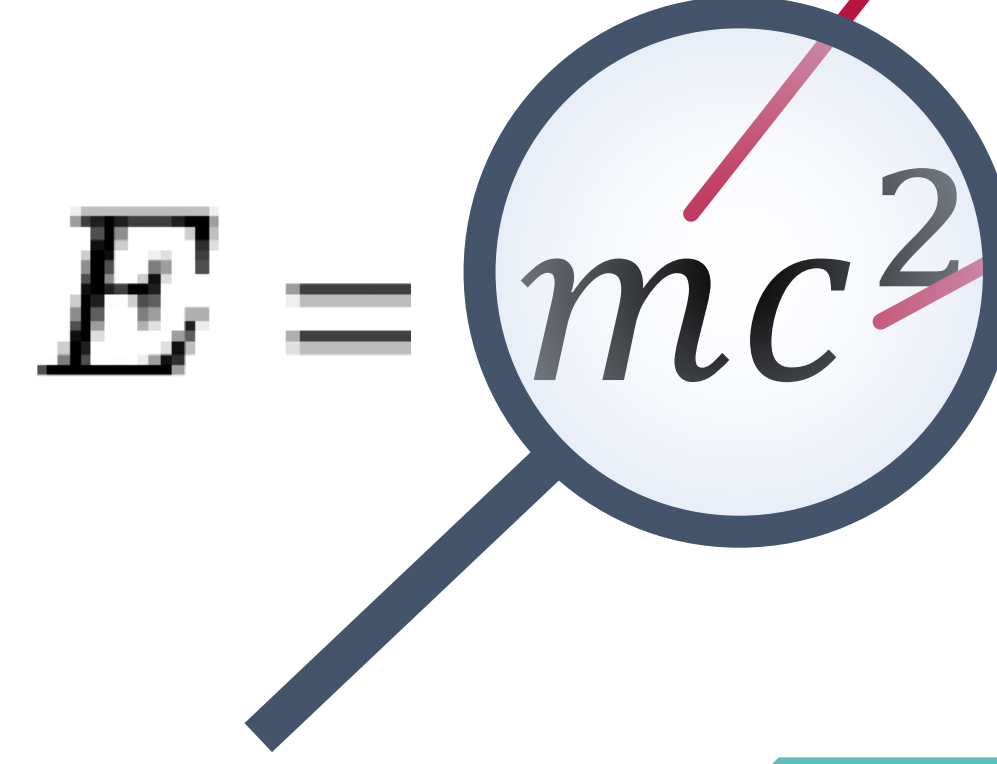
Mathematical Formulae in Wikimedia Projects 2020

Moritz Schubotz^{1,2}, André Greiner-Petter¹, Norman Meuschke^{1,3}, Olaf Teschke², Bela Gipp^{1,3}

ABSTRACT

This poster summarizes how mathematical formulae are processed in Wikimedia projects in 2020. We describe the transition from course-grained PNG images in 2001 to modern semantically enriched language-independent MathML formulae. Additionally, we provide a roadmap for further improving the accessibility and discoverability of mathematical knowledge in Wikimedia projects.

SEMANTICALLY ENHANCED MATH

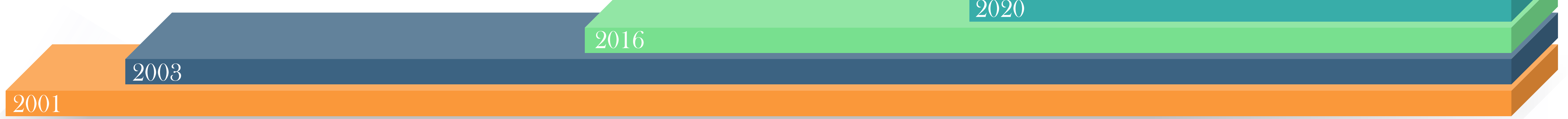


Mass
Property of matter to resist changes of the state of motion and to attract other bodies

Speed of Light

Speed at which all massless particles and associated fields travel in a vacuum

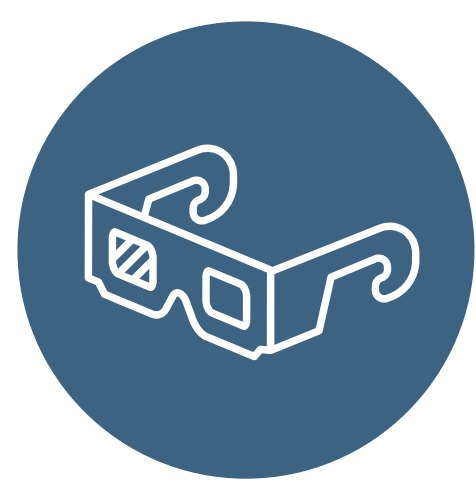
HISTORY OF MATH WITHIN WIKIPEDIA FROM 2001 UNTIL TODAY



- Wikipedia launched 2001
- Following the concept of free-as-in-freedom
- Community-driven online encyclopedia
- No TeX support for math expressions yet



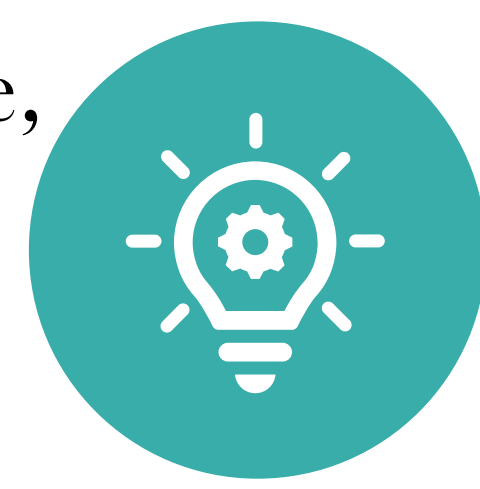
- TeX support introduced in 2003 [1]
- One can add math via math-tags using TeX directly in the Wikitext of an article
- Math still rendered as non-scalable PNG images



- First support of scalable math in 2016 [2]
- TeX input is converted to presentation MathML and SVG images
- Conversions are performed via JavaScript-based Mathoid that calls MathJax on server



- For the first time, math is linked with semantic concepts
- Math-tags can be annotated with Wikidata item IDs
- Annotations are language-independent since Wikidata is also multi-lingual



- Develop assistant tools for the Wikitext editor to improve the annotation process for editors
- Perform annotations automatically via mathematical information retrieval algorithms



MOTIVATION

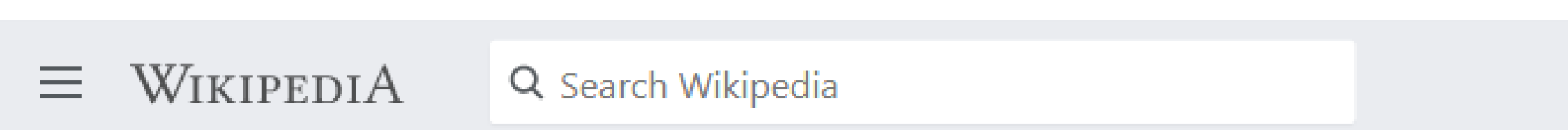
- Improve the overall quality of Wikipedia articles
- Ensure validity and quality of math
- Improve the accessibility and machine-readability of math, e.g., for search engines [3]
- Create a community-driven, multi-lingual, semantic dataset for mathematical expressions [4]



- Multi-lingual online encyclopedia
- No semantics in math expressions
- Additional information about math must be gathered from the surrounding text



- Collaboratively edited multi-lingual knowledge database
- Central structured database for Wikipedia projects
- Allows maintaining structured semantic and language-independent information about math
- Every item can be linked to Wikipedia pages



mass–energy equivalence

physical law

Math Formula Information

Formula: $E = mc^2$

Name: mass–energy equivalence

Type: physical law

Description: mass and energy are proportionate measures of the same underlying property of an object

Elements of the Formula

energy	E	quantitative physical property transferred to objects to perform heating or work on them
mass	m	property of matter to resist changes of the state of motion and to attract other bodies
speed of light	c	speed at which all massless particles and associated fields travel in a vacuum

Data Source

<https://www.wikidata.org/wiki/Q35875>

WIKIPEDIA

Content is available under [CC BY-SA 3.0](#) unless otherwise noted.

[Terms of Use](#) • [Privacy policy](#) • [Desktop](#)

Figure 1: Screenshot of the special page for the mass-energy equivalence. The information is fetched from the Wikidata item Q35875. You can access this information by clicking on $E = mc^2$ in the English or German Wikipedia article about “Mass-energy equivalence”.

ANNOTATION WORKFLOW

- Find or generate Wikidata item that lists the elements of the formula in the “has part” section
 - Every element is a Wikidata item itself
 - If an element is not linked to an existing Wikipedia page, add a more generalized page to the “subclass of” attribute
 - The rendered math of an element is either given by the “defining formula” or the “quantity symbol” attribute
- Annotate the math-tag in the Wikipedia article with the corresponding QID of the Wikidata item that represents the entire formula, e.g., `$E=mc^2$`
- By clicking on the annotated formula in Wikipedia, the structured information is fetched from Wikidata (see Figure 1).

Grothendieck–Hirzebruch–Riemann–Roch Theorem Q1899432

$$\text{ch}(f_! \mathcal{F}^*) \text{td}(Y) = f_* (\text{ch}(\mathcal{F}^*) \text{td}(X))$$

$\text{ch}(\bullet)$ Chern Class Q1069818

$\text{td}(\bullet)$ Todd Class Q2438393

Y Quasiprojective Variety

X Smooth Quasiprojective Schemes Q85397895

f Proper Morphism Q7250176

$f_!$ Grothendieck Construction Push-Forward Q85416305

f_* Push-Forward (Chow Groups) Chow Ring Q85416010

PROBLEMS

- Manual annotations are required
- An entire formula must be annotated, partial annotation is not supported

OUTLOOK

- Develop a recommendation systems for annotations in the visual Wikitext editor
- Use MathIR tools to perform automatic annotations

REFERENCES

- 1) M. Schubotz and G. Wicke. “Mathoid: Robust, Scalable, Fast and Accessible Math Rendering for Wikipedia”. In: Proc. CISM. Vol. 8543. Springer, 2014, pp. 224-235
- 2) M. Schubotz and A. P. Sexton. “A Smooth Transition to Modern mathoid-based Math Rendering in Wikipedia with Automatic Visual Regression Testing”. In: WiP at CISM. Vol. 1785, CEUR-WS.org, 2016, pp. 132-145
- 3) M. Schubotz. “Making Math Searchable in Wikipedia”. In: CoRR (July 30, 2012). DOI: 10.14279/depositonce-5034. arXiv: 1304.5475
- 4) P. Scharpf et al. “AnnoMath TeX - A Formula Identifier Annotation Recommender System for STEM Documents”. In: Proc. RecSys, ACM, 2019, pp. 532-533. DOI: 10.1145/3298689.3347042



We thank Corinna Breiting for narrating the video for our poster and the Wikimedia Foundation, especially Marko Obrovac, for their continuous support. Our research was supported by the German Research Foundation DFG GI 1259/1.



¹University of Wuppertal
Wuppertal - Germany
{last}@uni-wuppertal.de
andre.greiner-petter@zbmath.org



²FIZ Karlsruhe / zbMATH
Berlin - Germany
{first.last}@fiz-karlsruhe.de



³University of Konstanz
Konstanz - Germany
{first.last}@uni-konstanz.de