


Mathematical Entity Linking

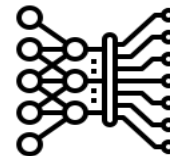
Methods and Applications

Doctoral Defense

Philipp Scharpf

$E = mc^2$









AnnoMathTeX

A Wikitext/LaTeX Document Annotation Recommender System

Search Wikipedia

$$x_{\mathrm{rms}} = \sqrt{\frac{1}{n} \left(x_1^2 + x_2^2 + \dots + x_n^2 \right)}.$$

 DEMO  SOURCE

☐ Local No Match

n

Source 0	Source 1	Source 2
rms	number	amount of substance
values	integer	number density

<https://annomathtex.wmflabs.org>

Introduction

Research problem

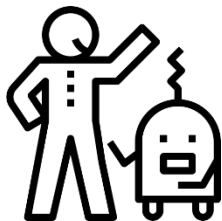
Research objective

Research questions





<https://wikidata.org>



Launched 2012 to connect multilingual Wikipedia articles
in language-independent concept items

Free, **open** access,
collaborative **semantic** knowledge-base

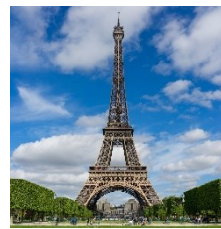
Humans or bots can
create, read, accept, decline or edit content



Wikidata Entity Linking



Paris Hilton stays in Hilton Paris Opera in Paris.



Paris Hilton (Q47899)

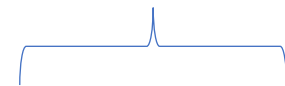
Hilton Paris Opera (Q2992084)

Paris (Q90)

<https://www.wikidata.org/wiki/Q47899>

<https://www.wikidata.org/wiki/Q2992084>

<https://www.wikidata.org/wiki/Q90>



帕丽斯-希尔顿

留在



希尔顿巴黎歌剧院

在



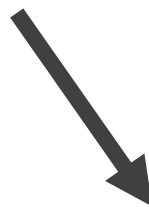
巴黎。

Problem: Classical Entity Linking (EL)



representations

- Washington
- President Washington
- George Washington
- G. Washington
- Father of the United States
- The American Fabius
- The First President



Item

George Washington (Q23)

concept



Problem: Mathematical EL (MathEL)



$\frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi + \left(\frac{m_0 c}{\hbar} \right)^2 \psi = 0$	$u_{tt} + Au + f(u) = 0$
$\partial_{ct}^2 h_n(z, t) - \partial_z^2 h_n(z, t) + \nu_n^2 h_n(z, t) = 0$	$\nabla^a \nabla_a \psi = \mu^2 \psi$
$\frac{\hbar^2}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \frac{\hbar^2 \partial^2 \psi}{\partial x^2} = -2i\hbar \frac{\partial \psi}{\partial \tau}$	$-\hbar^2 \frac{\partial^2 \psi}{\partial t^2} + c^2 \hbar^2 \nabla^2 \psi = m_0^2 c^4 \psi$
$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} - \frac{2\alpha + a}{c^2} \frac{\partial \phi}{\partial t} - \frac{\alpha^2 + a\alpha}{c^2} \phi = 0$	$u_{tt} - \Delta u + m^2 u + G'(u) = 0$
$\left(\eta^{\mu\nu} \frac{\partial}{\partial x^\mu} \frac{\partial}{\partial x^\nu} - \left(\frac{mc}{\hbar} \right)^2 \right) \phi = 0$	$u_{tt} - \Delta u + m u + \mathcal{P}'(u) = 0$
$\left(-\frac{1}{c^2} \frac{\partial^2}{\partial t^2} + \sum_{i=1}^p \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^i} - \left(\frac{mc}{\hbar} \right)^2 \right) \phi = 0$	

concept



Item

Klein–Gordon equation (Q868967)

representations

Different symbols for constants or variables are used

$$\frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi + \left(\frac{m_0 c}{\hbar} \right)^2 \psi = 0$$

Constants appear in **different terms**

$$-\hbar^2 \frac{\partial^2 \Psi}{\partial t^2} + c^2 \hbar^2 \nabla^2 \Psi = m_0^2 c^4 \Psi \quad u_{tt} + A u + f(u) = 0$$

Additional terms and functions are introduced

$$\frac{\hbar^2}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \frac{\hbar^2 \partial^2 \psi}{\partial^2 x^2} = -2i\hbar \frac{\partial \psi}{\partial \tau}$$

Additional (index or semantic) **sub- or superscripts** are introduced

$$\left(-\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \sum_{i=1}^p \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^i} - \left(\frac{mc}{\hbar} \right)^2 \right) \varphi = 0$$

Substitutions, i.e., identifiers are subsumed into others and then appear **implicitly**

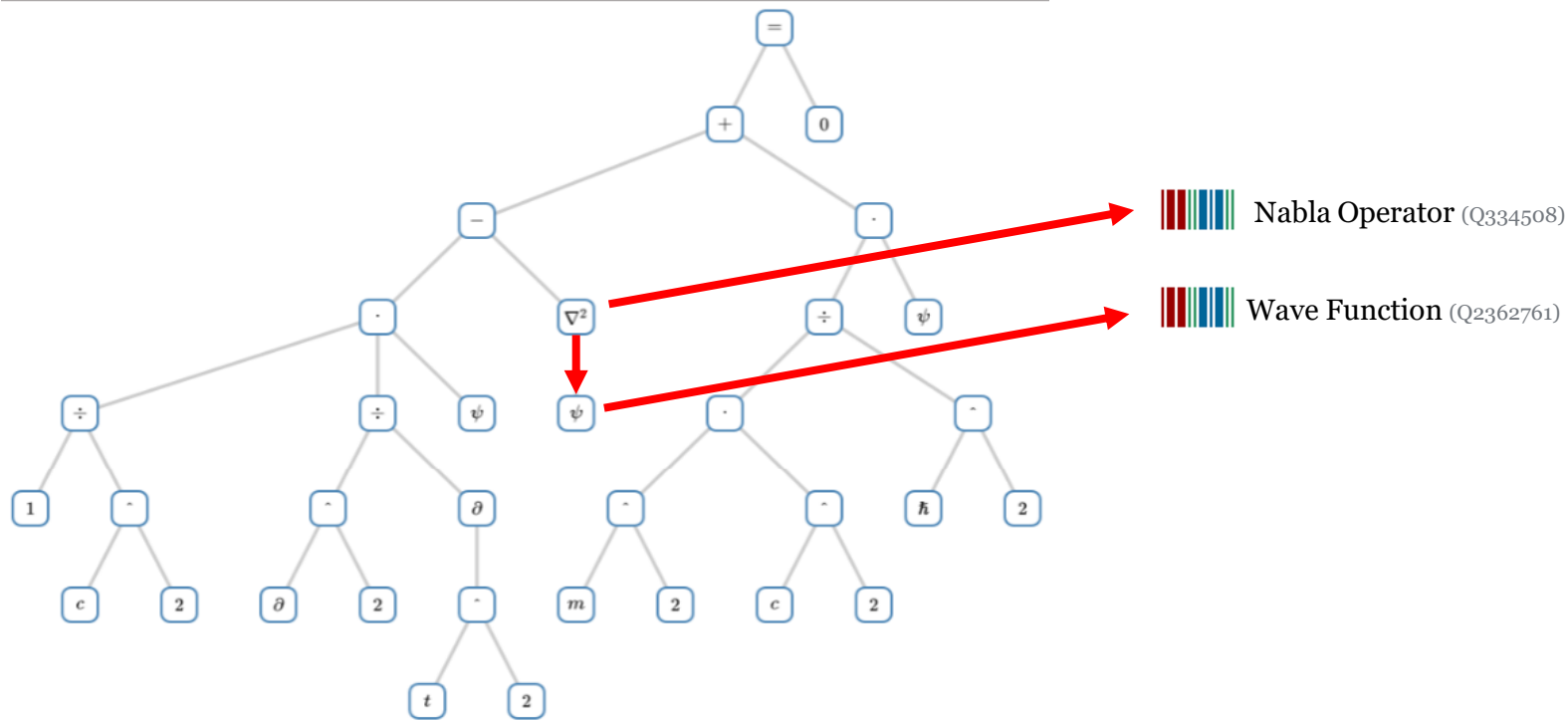


Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiteringer, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

Mathematical Entity Linking Challenges



$$\frac{1}{c^2} \frac{\partial^2}{\partial t^2} \psi - \nabla^2 \psi + \frac{m^2 c^2}{\hbar^2} \psi = 0$$



Resolving **mathematical** ambiguities is **difficult**

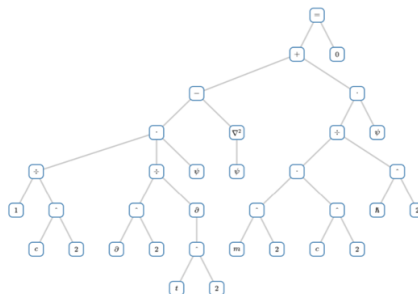
1. many **different** formula **representations** $\frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi + \left(\frac{m_0 c}{\hbar}\right)^2 \psi = 0$

$$u_{tt} + Au + f(u) = 0$$

- ## 2. many involved symbols

```
\nabla: "del" (Q334508), m: "mass" (Q11423) ,
\hbar: "Planck constant" (Q122894) , ...
```

- ### 3. very complex semantics





- How to **discover** a math entity?
- How to **recognize** a math entity?
- How to **represent** a math entity in a KG?
- How to **populate** the KG with math entities?
- How to **query** the KG for math entities?
- How to **apply** linking math entities?



Reviews („Where“)

- Entity Linking with Wikidata
([Scharpf24b?](#))
- Mathematics in Wikidata
([Scharpf21a](#))

Methods („How“)

- Formula Classification & Clustering
([Scharpf19a](#), [Scharpf23a](#))
- Formula Annotation Recommendation
([Scharpf19b](#), [Scharpf21b](#))
- Formula Benchmarking
([Schubotz18a](#), [Scharpf18](#))

Applications („Why“)

- Mathematical Question Answering
([Schubotz18b](#), [Scharpf20](#), [Scharpf22](#))
- Mathematical (Physics) Question Generation
([Scharpf22](#), [Scharpf23b](#))
- STEM Document Classification Explainability
([Schubotz20](#), [Scharpf20](#), [Scharpf24a?](#))

13 publications first author
+ 3 co-author



Reviews („Where“)

- Entity Linking with Wikidata
(*Scharpf24b?*)
- Mathematics in Wikidata
(*Scharpf21a*)

Methods („How“)

- Formula Classification & Clustering
(*Scharpf19a, Scharpf23a*)
- Formula Annotation Recommendation
(*Scharpf19b, Scharpf21b*)
- Formula Benchmarking
(*Schubotz18a, Scharpf18*)

Applications („Why“)

- Mathematical Question Answering
(*Schubotz18b, Scharpf20, Scharpf22*)
- Mathematical (Physics) Question Generation
(*Scharpf22, Scharpf23b*)
- STEM Document Classification Explainability
(*Schubotz20, Scharpf20, Scharpf24a?*)

Presented in depth in this talk

MathEL – Demonstrators

AnnoMathTeX
A Wikitex/LaTeX Document Annotation Recommender System

$$S x_{\left[\text{math}(rms) \right]} = \text{sqrt} \left(\frac{1}{n} \left(\left| x_{-1}^2 + x_{-2}^2 + \dots + x_{-n}^2 \right| \right) \right)$$

Token: n Annotated with: integer Type: Global

Annotation¹

MathMLben

Benchmark^{2,3}

mass–energy equivalence

Math Formula Information

Formula: $E = mc^2$

Name: mass–energy equivalence

Description: Physical law relating mass to energy

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>

Readability⁴

zbMATH^{Open}
THE FIRST RESOURCE FOR MATHEMATICS

Documents Authors Serials Classification Software Formulas

Structured Search

Search for documents

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

EM
EUROPEAN
MATHEMATICAL
SOCIETY

HEIDELBERGER AKADEMIE
DER WISSENSCHAFTEN
Akademie der Wissenschaften
des Landes Baden-Württemberg

Classification⁵

Mathematical Question Answering System (MathQA)

Language: English

What is the relationship between volume and temperature?

$PV = nRT$

n (amount of substance)
Enter value: 0.3144588

R (molar gas constant)
Source: www.wikidata.org/wiki/Q211432

T (temperature)
Enter value:

MathQA

Question Answering⁶

PhysWikiQuiz
Physics Question Generation and Interrogation System

Enter Formula Concept Name (e.g., "speed"):

speed Generate

Formula Concept Question:

What is the distance s , given speed $v = 10 \text{ m s}^{-1}$, duration $t = 6 \text{ s}$?

Enter Formula Concept Question Answer:

60 m Answer

Value correct! Unit correct!

Solution from www.wikidata.org/wiki/Q3711325 formula $s = t \cdot v$ with $60 \text{ m} = 6 \text{ s} \cdot 10 \text{ m s}^{-1}$

PhysWikiQuiz

Question Generation⁷

Reviews
("Where?")



Entity Linking with Wikidata

Mathematics in Wikidata





Wikipedia



Wikidata

Classical
Entity
Linking
(EL)



EL with Wikipedia

EL with Wikidata

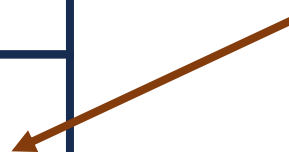
Mathematical
EL (MathEL)



MathEL with
Wikipedia

MathEL with
Wikidata

Research Gap

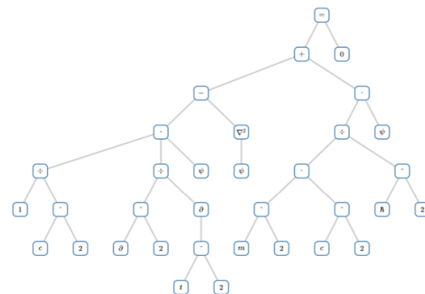


Why Wikidata?



Linking (mathematical) entities to **Wikidata** instead of **Wikipedia**, allows for

- modeling formula **expression substructure**

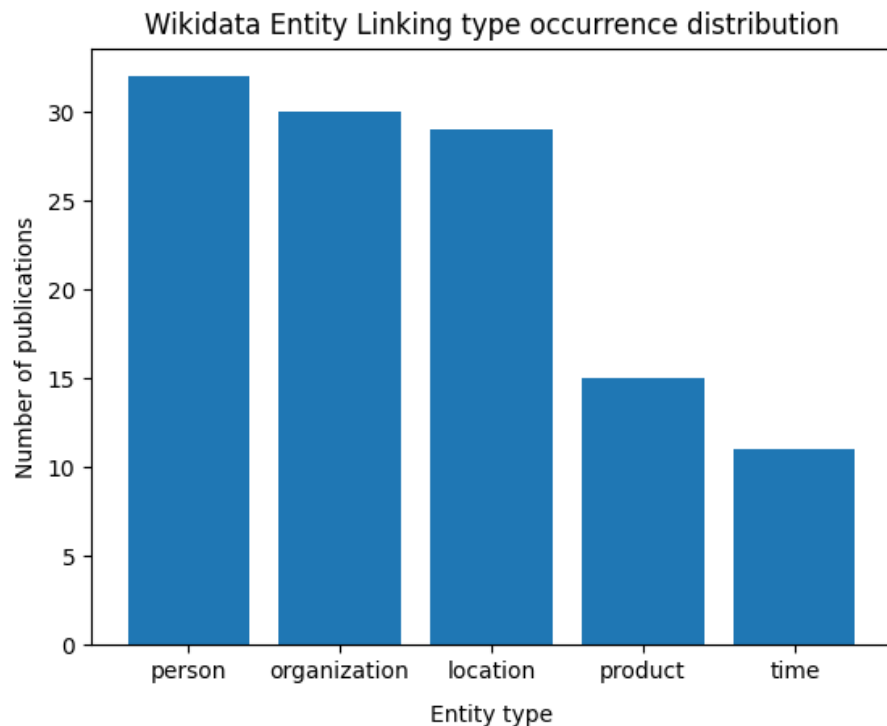


- structured queries



- multilingual entity representations





discovered
78 **entity types**
in literature



mathematical
formula entity
missing

...



Philipp Scharpf, Moritz Schubotz, Andreas Spitz, Norman Meuschke, André Greiner-Petter, and Bela Gipp. Entity linking with wikidata: A systematic literature review. In *CSUR*. Submitted to ACM Comput. Surv. (CSUR), Feb. 2024.

Methods
("How?")



Formula Concept Retrieval

Formula Annotation
Recommendation





How to **discover** mathematical concept entities?



Transfer mathematical knowledge

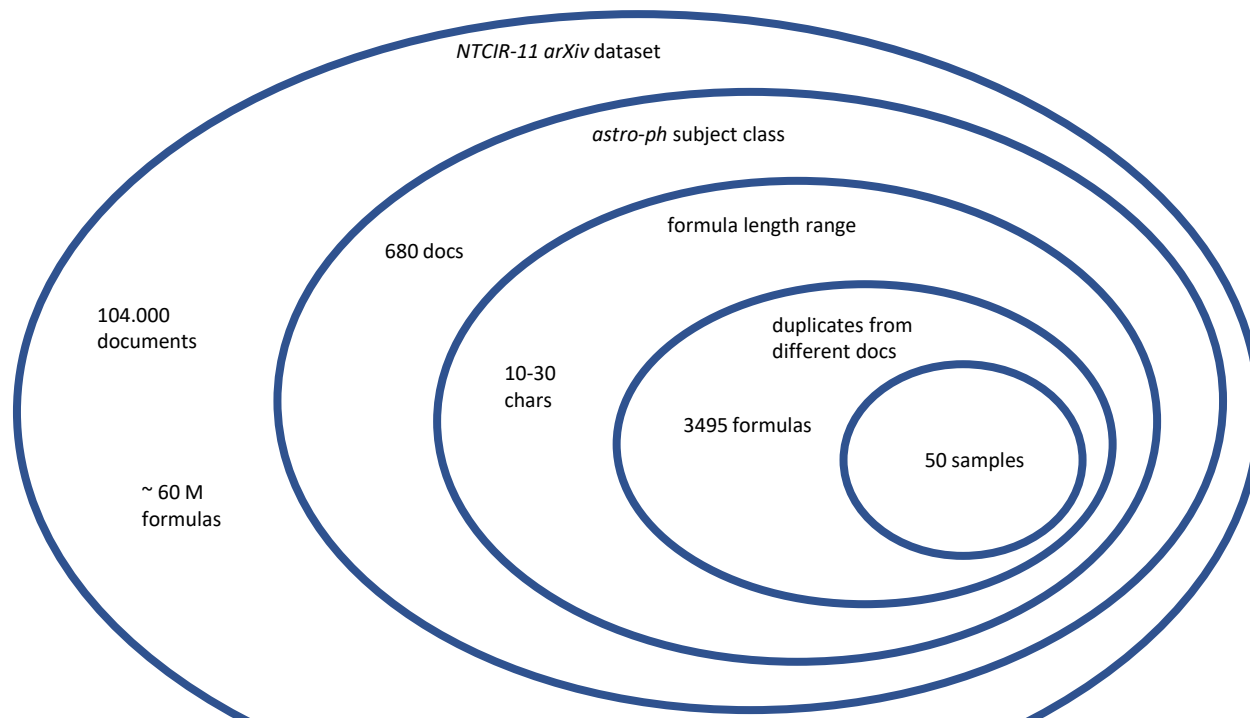
from *arXiv* (unstructured)

to *Wikidata* (structured)



arXiv.org





Aizawa, A., Kohlhase, M., Ounis, I., Schubotz, M.: NTCIR-11 math-2 task overview. In: NTCIR. National Institute of Informatics (NII) (2014)

Formula string candidates



Formula vectors
k-nearest neighbors

Hubble's law
$p = \omega \rho$
$p = \kappa \rho$
$\omega = p/\rho$
$p_d = \omega \rho_d$

...



Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiteringer, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

Formula name candidates



Surrounding text
Word window

Name (QID)
Hubble parameter (Q179916)
Equation of state (Q214967)
Accelerating universe (Q1049613)
Dark fluid (Q5223514)
Dark energy (Q18343)

Retrieving *equivalent representations* using the k-nearest neighbor algorithm

#	Formula	Name (QID)	d / \hat{d}	$s_{e_m}, s_{\hat{e}_m}, s_{e_s}, s_{\hat{e}_s}$	Encoding: sample formula
1	$H = \dot{a}/a$	hubble parameter (Q179916)	32 / 32	0.0, 0.1, 0.0, 0.9	$\hat{e}_s: H_i = \dot{R}/R$
2	$p = \omega \rho$	equation of state (Q214967)	6 / 5	0.3, 0.0, 0.1, 0.6	$e_s: p_d = w \rho_d$
3	$\omega = p/\rho$	accelerating universe (Q1049613)	4 / 3	0.7, 0.0, 0.0, 0.3	$e_m: p = \omega \rho$
4	$p = -A/\rho^\alpha$	dark fluid (Q5223514)	4 / 4	0.7, 0.0, 0.3, 0.0	$e_m: p = -\frac{A}{\rho^\alpha}$
5	$p_d = w \rho_d$	dark energy (Q18343)	4 / 3	0.3, 0.0, 0.3, 0.3	$e_s: p_X = \omega_X \rho_X$

...

rel. success

(e_m)	<i>math2vec</i>	0.70	} formulas text frequency sequence
(\hat{e}_m)	<i>semantics tf-idf</i>	0.15	
(e_s)	<i>semantics2vec</i>	0.11	
(\hat{e}_s)	<i>math tf-idf</i>	0.04	

We achieve best results for encoding the formula constituents with sequence (*math2vec*)



Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiterger, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

Retrieving *concept name candidates* from the surrounding text

#	Formula	Name (QID)	d / \hat{d}	$se_m, s_{\hat{e}_m}, se_s, s_{\hat{e}_s}$	Encoding: sample formula
1	$H = \dot{a}/a$	hubble parameter (Q179916)	32 / 32	0.0, 0.1, 0.0, 0.9	$\hat{e}_s: H_i = \dot{R}/R$
2	$p = \omega \rho$	equation of state (Q214967)	6 / 5	0.3, 0.0, 0.1, 0.6	$e_s: p_d = w \rho_d$
3	$\omega = p/\rho$	accelerating universe (Q1049613)	4 / 3	0.7, 0.0, 0.0, 0.3	$e_m: p = \omega \rho$
4	$p = -A/\rho^\alpha$	dark fluid (Q5223514)	4 / 4	0.7, 0.0, 0.3, 0.0	$e_m: p = -\frac{A}{\rho^\alpha}$
5	$p_d = w \rho_d$	dark energy (Q18343)	4 / 3	0.3, 0.0, 0.3, 0.3	$e_s: p_X = \omega_X \rho_X$

...

surrounding text window

For 72%, we retrieve useful formula name candidates (recommendations)

For 82% of the retrieved name candidates, there is a *Wikidata QID* available

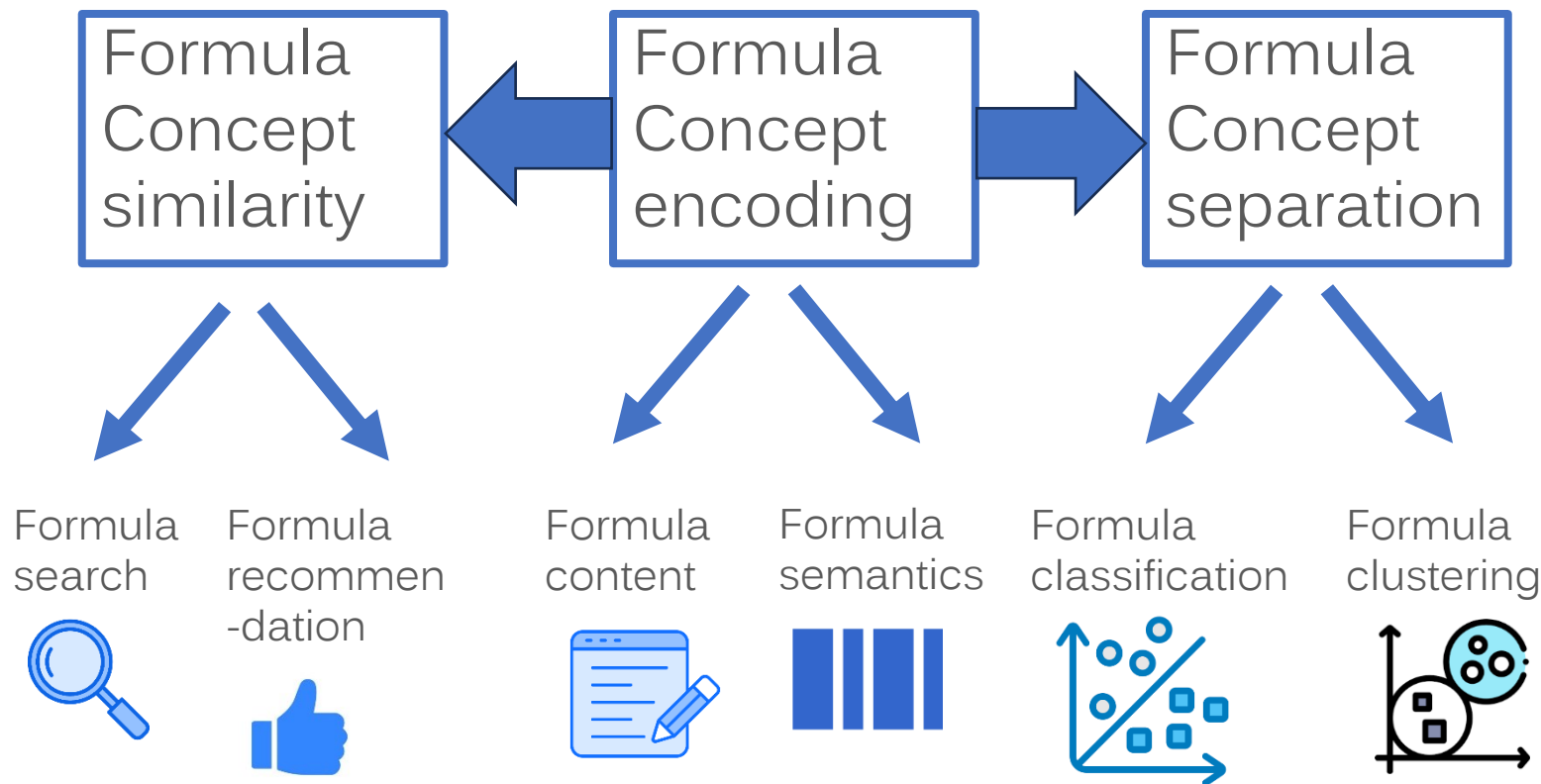


Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiterger, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

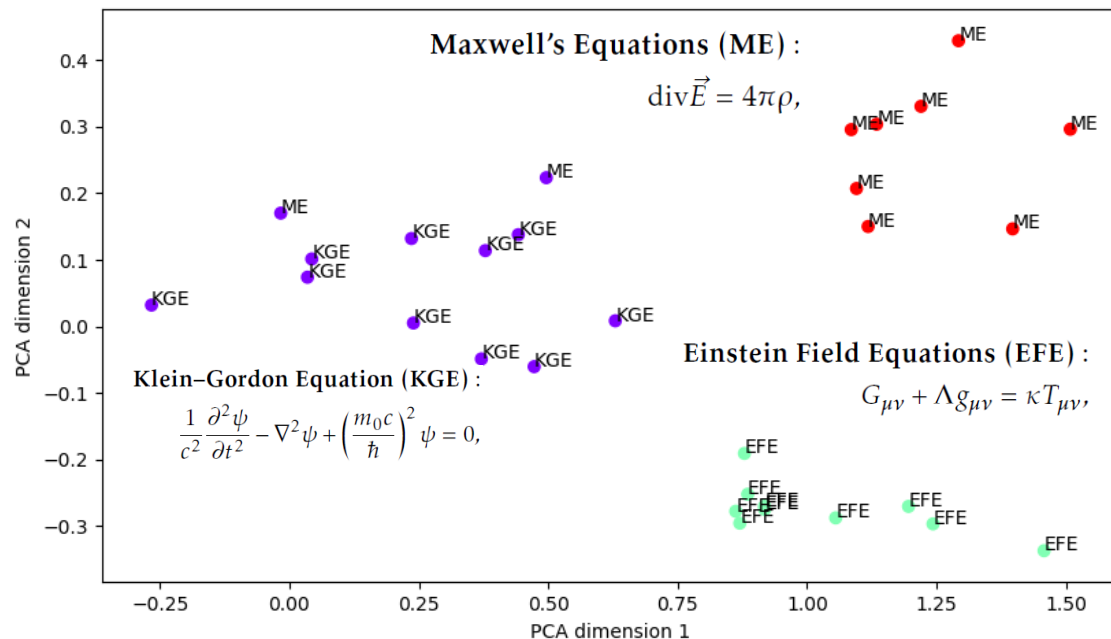


How to **recognize** mathematical concept entities?

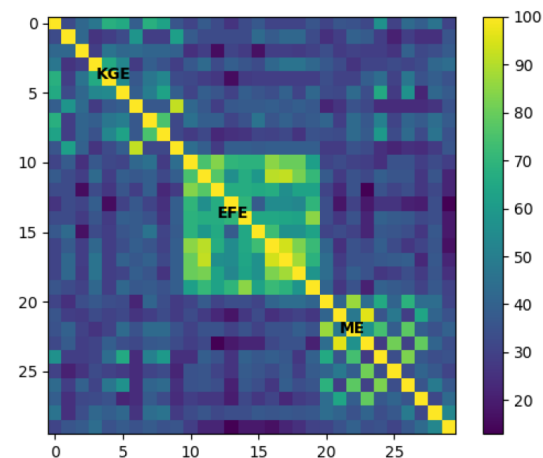




Formula Content Space



- Classification accuracy up to 0.95
- Cluster purity up to 0.87



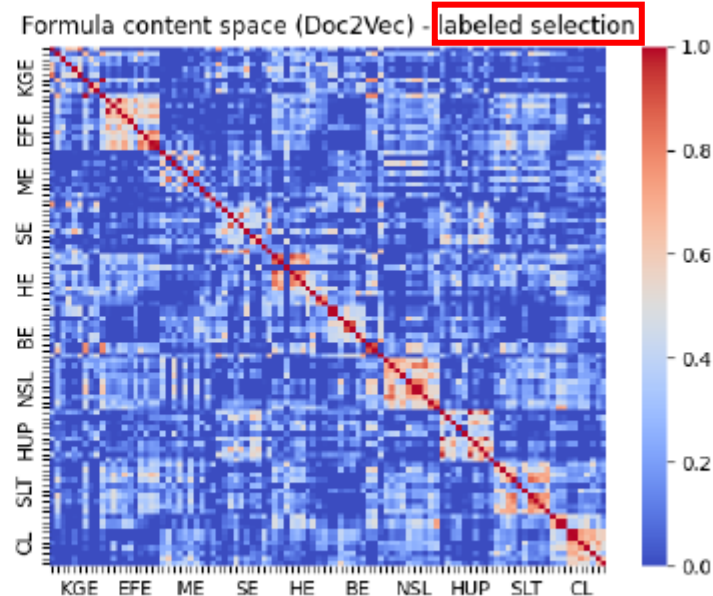
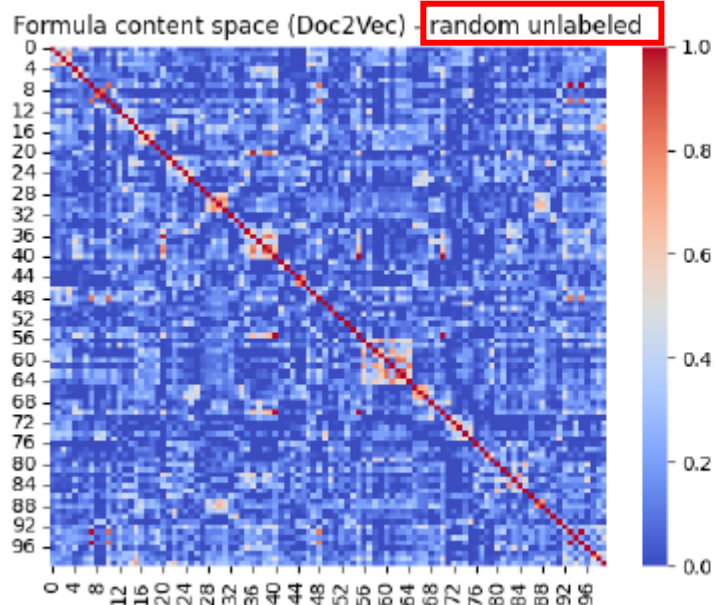
Fuzzy string similarity percentages

heatmap



Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiterger, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

Formula Concept Similarity



High intra-class similarity / coherence (red) and low inter-class similarity / separability (blue)



Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Corinna Breiteringer, and Bela Gipp. Discovery and recognition of formula concepts using machine learning. Springer Scientometrics Journal, Feb. 2023.

Methods
("How?")



Formula Concept Retrieval

Formula Annotation
Recommendation



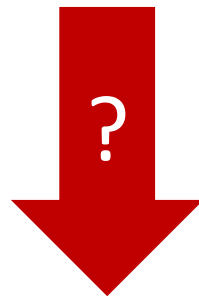
How to **populate** the (Wikidata) Knowledge Graph
with mathematical concept entities?



Transfer mathematical knowledge

from *Wikipedia* (semi-structured)

to *Wikidata* (structured)



WIKIPEDIA
The Free Encyclopedia





WIKIDATA
SPARQL



AnnoMathTeX

A Wikitext/LaTeX Document Annotation Recommender System

Search Wikipedia

 DEMO  SOURCE

☐ Local No Match

n

Source 0	Source 1	Source 2
rms	number	amount of substance
values	integer	number density

<https://annomathtex.wmflabs.org>



FORMULA ANNOTATION

☐ Local

No match:

$E = mc^2$



Formula: $E=mc^2$

Wikidata1	Wikidata2	WordWindow	FormulaConceptDB	Manual
mass–energy equivalence (Q35875)	mass–energy equivalence (Q35875)	formula (N/A)	mass–energy equivalence (Q35875)	
	time-independent Schrödinger equation (Q25829357)	einstein (N/A)		



Philipp Scharpf, Moritz Schubotz, and Bela Gipp. Fast linking of mathematical wikidata entities in wikipedia articles using annotation recommendation. In *Proceedings of the Web Conference (WWW) 2021*. ACM / IW3C2, April 2021.



IDENTIFIER ANNOTATION

☐ Local

No match:

Submit

Identifier: m

Not an identifier

Annotated Identifiers: 61/96

Source 1	Source 2	Source 3	Source 4	Source 5
mass (Q11423)	hypothesis tests (N/A)	N/A (N/A)	motion (N/A)	exhaust gas mass flow (Q320176)
field (Q185674)	edges (N/A)		law (N/A)	mass (Q11423)
integer (Q12503)	mass (Q11423)		newton (N/A)	
number (N/A)	message (N/A)		acceleration (Q11376)	
particle (N/A)	rest mass (N/A)		mass (Q11423)	



Philipp Scharpf, Moritz Schubotz, and Bela Gipp. Fast linking of mathematical wikidata entities in wikipedia articles using annotation recommendation. In *Proceedings of the Web Conference (WWW) 2021*. ACM / IW3C2, April 2021.

Identifiers	Time (seconds)
Recommendation	2.6
Manual	6.3
Formulae	Time (seconds)
Recommendation	2.8
Manual	4.0

Speedup
by a factor of
2.4 and 1.4
compared to
manual labeling



Philipp Scharpf, Moritz Schubotz, and Bela Gipp. Fast linking of mathematical wikidata entities in wikipedia articles using annotation recommendation. In *Proceedings of the Web Conference (WWW) 2021*. ACM / IW3C2, April 2021.

Applications ("Why?")



Mathematical
Question Answering

Physics
Question Generation



How to **apply** and **evaluate**
linking mathematical entities?

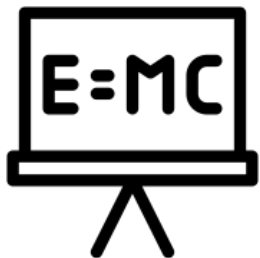


Applications ("Why?")



Mathematical
Question Answering

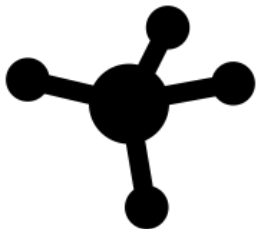
Physics
Question Generation



Concept Questions



Geometry Questions



Relationship Questions



Formula Input



WIKIPEDIA
The Free Encyclopedia



CoreNLP

arXiv.org



[Home](#)[Examples](#)[Video](#)[Github](#)

Mathematical Question Answering System (MathQA)

Language

English



Search

<https://mathqa.wmflabs.org>



P. Scharpf, M. Schubotz, and B. Gipp. "Mining Mathematical Documents for Question Answering via Unsupervised Formula Labeling". In: *Proc. ACM/IEEE JCDL*. ACM, June 2022, pp. 1–11. doi: 10.1145/3529372.3530925.



What is the formula for an ideal gas?

Search

$$PV = nRT$$

T (temperature)

293

R (gas constant)

8.3144598

n (amount of
substance)

2

Submit

Solution

$$P \cdot V = 4.87e+03$$

<https://mathqa.wmflabs.org>



P. Scharpf, M. Schubotz, and B. Gipp. "Mining Mathematical Documents for Question Answering via Unsupervised Formula Labeling". In: *Proc. ACM/IEEE JCDL*. ACM, June 2022, pp. 1–11. doi: 10.1145/3529372.3530925.



Search

m_i
 m

Submit





$$E = mc^2$$

c (speed of light)

m (mass)

Source: www.wikidata.org/wiki/Q35875



MathQA – Direct Formula Input



Gipp Lab
Scientific Information Analytics

$$T = \frac{\hbar c^3}{8\pi k G M}$$

G

pi

c

k

M

hbar

<https://mathqa.wmflabs.org>



P. Scharpf, M. Schubotz, and B. Gipp. "Mining Mathematical Documents for Question Answering via Unsupervised Formula Labeling". In: *Proc. ACM/IEEE JCDL*. ACM, June 2022, pp. 1–11. doi: 10.1145/3529372.3530925.

MathQA – Competitors



Gipp Lab
Scientific Information Analytics

MathQA

formula for speed Search

Mathematical Question Answering System (MathQA)

$v = s/t$

S (distance) t (duration)

Submit

<https://mathqa.wmflabs.org>



Equation: Calculate: speed

$v = a t$	>> time:
v speed	<input type="text" value="1 s"/>
t time	>> acceleration:
a acceleration	<input type="text" value="1 m/s^2"/>

Compute

<https://www.wolframalpha.com>



$s = \text{speed}$ $s = \frac{d}{t}$

$d = \text{distance traveled}$

$t = \text{time elapsed}$

<https://www.google.com>



P. Scharpf, M. Schubotz, and B. Gipp. "Mining Mathematical Documents for Question Answering via Unsupervised Formula Labeling". In: *Proc. ACM/IEEE JCDL*. ACM, June 2022, pp. 1–11. doi: 10.1145/3529372.3530925.



Query Concept Name	MathQA Formula	Wolfram Alpha Formula	Google Formula
acceleration	$a = dv/dt$	$v = at$	$\bar{a} = \Delta v / \Delta t$
angular acceleration	$\alpha = d\omega/dt$	$\omega = \alpha t$	$\alpha = \Delta\omega / \Delta t$
angular frequency	$\omega = 2\pi f$	$\nu = \omega / (2\pi)$	$\omega = 2\pi / t$
angular momentum	$L = r \times p$	$L = I\omega, \omega = 2\pi n$	$L = mvr$
angular velocity	$\omega = d\varphi/dt \cdot u$	$\omega = \alpha t$	$\omega = \Delta\theta / \Delta t$

Relevant	52%	48%	68% (58% boxed)
Calculation	80%	81%	2%

<https://mathqa.wmflabs.org>



P. Scharpf, M. Schubotz, and B. Gipp. "Mining Mathematical Documents for Question Answering via Unsupervised Formula Labeling". In: *Proc. ACM/IEEE JCDL. ACM*, June 2022, pp. 1–11. doi: 10.1145/3529372.3530925.

Applications ("Why?")



Mathematical
Question Answering

Physics
Question Generation



PhysWikiQuiz

<https://physwikiquiz.wmflabs.org>

Physics Question Generation and Test System

Enter Formula Concept Name (e.g., 'speed'):

Generate

Formula Concept Question:

What is the distance s , given speed $v = 10 \text{ m s}^{-1}$, duration $t = 6 \text{ s}$?

Enter Formula Concept Question Answer:

Answer

Value correct! Unit correct!

Solution from www.wikidata.org/wiki/Q3711325 formula $s = t \cdot v$ with $60 \text{ m} = 6 \text{ s} \cdot 10 \text{ m s}^{-1}$.

<https://physwikiquiz.wmflabs.org>



Philipp Scharpf, Moritz Schubotz, Andreas Spitz, André Greiner-Petter, and Bela Gipp. Collaborative and ai-aided exam question generation using wikidata in education. In *Wikidata@ISWC*

Conclusion & Outlook

Conclusion

Contributions

Future Work

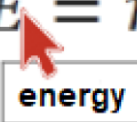


- MathEL *improves* the **representation, retrieval, and application** of math entities in a KG

- MathEL **methods and applications** for

- *defining* mathematical concepts
- *recognizing* entities in documents
- *structuring* data in the KG
- *populating* the KG
- *querying* math entities
- *applying* MathEL in various IR domains

$$E = mc^2$$



energy

retrieval & **analysis** (examples)

classification, clustering, matrices

high community **agreement**

annotation recommendation **speedup**

formula **search & question** answering

e.g., **reliable** question generation



1. **Formula classification** with up to **95%** accuracy
2. **Formula search** outperforming the search-engine *Google* by **24%** in MRR
3. **Formula annotation** acceleration by a factor of up to **2.4 x** compared to human
4. **Formula question answering** outperforming the knowledge-engine *Wolfram Alpha* with **8%** more relevant answers
5. **Formula question generation** (reliable AND scalable) using a KG and a CAS with **13 x** more coverage than its best-performing competitor



Mass–energy equivalence

From Wikipedia, the free encyclopedia

In **physics**, **mass–energy equivalence** is the principle that **mass** is a form of **energy** and that in the **rest frame**, mass and energy are equivalent and differ only by a constant.^{[1][2]} The principle is fundamental to many fields of physics, including **nuclear** and **particle physics** and is described by **Albert Einstein**'s famous formula:^[3]

Mass–energy relation

$$E = mc^2$$

Math Formula Information

Formula: $E = mc^2$

Name: mass–energy equivalence

Description: Physical law relating mass to energy

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>

ARTICLE

FORMULA DETAIL PAGE

https://en.wikipedia.org/wiki/Mass%E2%80%93energy_equivalence

Fine-Grained MSC Prediction API

zbMATH  Open

THE FIRST RESOURCE FOR MATHEMATICS

Example Queries

[Text to Keywords](#) [Text to MSCs](#) [Keywords to MSCs](#)

Public API (<https://automscpredictor.wmcloud.org>) to be used in production

by the mathematical library **zbMATH Open** (currently accessed around 40 Million times per year*)

*<https://www.fiz-karlsruhe.de/sites/default/files/FIZ/Dokumente/Jahresberichte/\Jahresbericht-2022.pdf>



FORMULA ANNOTATION

☐ Local

No match:

Formula: $E=mc^2$

\$ $E = mc^2$ \$

+ unsupervised
auto-annotation of
full document



(Open)LLaMA	Falcon	BERT & Co.	GPT-J/NeoX	Manual
mass–energy equivalence (Q35875)	mass–energy equivalence (Q35875)	formula (N/A)	mass–energy equivalence (Q35875)	
	time-independent Schrödinger equation (Q25829357)	einstein (N/A)		



Philipp Scharpf, Moritz Schubotz, and Bela Gipp. Fast linking of mathematical wikidata entities in wikipedia articles using annotation recommendation. In *Proceedings of the Web Conference (WWW) 2021*. ACM / IW3C2, April 2021.



PhysWikiQuiz (rule-based)

- Knowledge Graph (Wikidata)
- Computer Algebra System (SymPy)

+

GPT (learning-based)

- Large Language Model

= verified creative question

PhysWikiQuiz

<https://physwikiquiz.wmflabs.org>

Physics Question Generation and Test System

Enter Formula Concept Name (e.g., 'speed'):

Generate

Formula Concept Question:

What is the distance s , given speed $v = 10 \text{ m s}^{-1}$, duration $t = 6 \text{ s}$?

Enter Formula Concept Question Answer:

Answer

Value correct! Unit correct!

Solution from www.wikidata.org/wiki/Q3711325 formula $s = t \cdot v$ with $60 \text{ m} = 6 \text{ s} \cdot 10 \text{ m s}^{-1}$.

<https://physwikiquiz.wmflabs.org>




Philipp Scharpf, Moritz Schubotz, Andreas Spitz, André Greiner-Petter, and Bela Gipp. Collaborative and ai-aided exam question generation using wikidata in education. In *Wikidata@ISWC*

Future Work (PhysWikiQuiz)



Gipp Lab
Scientific Information Analytics

PhysWikiQuiz API

 physwikiquiz.wmflabs.org/api/v1?name=speed

```
{"explanation_text": "Solution from www.wikidata.org/wiki/Q3711325 formula  $v = s/t$  with  $2.5 \text{ m s}^{-1} = 10 \text{ m} / 4 \text{ s}$  .", "formula_unit_dimension": "m s^{-1}", "identifier_values": [2.5, 10, 4], "question_text": "What is the speed  $v$ , given distance  $s = 10 \text{ m}$ , duration  $t = 4 \text{ s}$  ?"}
```

GPT-4 (API)



ME

create a story question from this: What is the distance s , given speed $v = 4 \text{ m s}^{-1}$, duration $t = 9 \text{ s}$?



In the adventure story "Race Against Time," the protagonist has a magical pair of boots that allow him to run at a constant speed of 4 m/s without getting tired. At a crucial point in the story, he must reach a portal to save his friends, but the portal is only open for 9 seconds. How far can the protagonist travel during this limited time?

Enter Answer:

36 m

Answer

Value correct! Unit correct!

GPT-4 provided
wrong answers
in almost 40%
of our test cases!

<https://physwikiquiz.wmflabs.org>



Philipp Scharpf, Moritz Schubotz, Andreas Spitz, André Greiner-Petter, and Bela Gipp. Collaborative and ai-aided exam question generation using wikidata in education. In *Wikidata@ISWC*