Generating OpenMath Content Dictionaries from Wikidata

Moritz Schubotz
Information Science Group
University of Konstanz
www.isg.uni.kn

Overview

1. Introduction
2. The MathML Benchmark MathMLBen
3. A Wikidata Content Dictonary
4. Wikidata as a `cdbase`
5. OpenMath vs. Wikidata
6. Conclusions and Future Works

Overview

1. **Introduction**
2. The MathML Benchmark MathMLBen
3. A Wikidata Content Dictonary
4. Wikidata as a `cdbase`
5. OpenMath vs. Wikidata
6. Conclusions and Future Works

Introduction

what symbol is ζ	Riemann zeta function, Damping ratio, Hurwitz zeta function, 1s Slater-type function, Jerk (physics), Oblate spheroidal coordinates, Routhian mechanics
define $\Leftrightarrow$ in A $\Leftrightarrow$ B	Monoidal t-norm logic, Logical equivalence, If and only if, Logical biconditional Contraposition, (Tautology (logic) or Exportation (logic))
definition $a \oplus b$	Direct sum, $\oplus$, Exclusive or
$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$	Determinant, Laplace expansion
define notation ${}_2F_1(a, b; c; z)$	Hypergeometric function

Introduction

what symbol is ζ	Riemann zeta function, Damping ratio, Hurwitz zeta function, 1s Slater-type function, Jerk (physics), Oblate spheroidal coordinates, Routhian mechanics
define $\Longleftrightarrow$ in A $\Longleftrightarrow$ B	Monoidal t-norm logic, Logical equivalence, If and only if, Logical biconditional Contraposition, (Tautology (logic) or Exportation (logic))
definition $a \oplus b$	Direct sum, $\oplus$, Exclusive or
$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$	Determinant, Laplace expansion
define notation ${}_2F_1(a, b; c; z)$	Hypergeometric function

- MathIR is complicated
- Evaluation of Information Needs involves human factors

Subproblem Information Extraction

$$\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} - \nabla^2\psi + \left(\frac{m_0 c}{\hbar}\right)^2 \psi = 0. \qquad (30)$$

$$\partial_{ct}^2 h_n(z,t) - \partial_z^2 h_n(z,t) + \nu_n^2 h_n(z,t) = 0 \qquad (19)$$

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

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} - \frac{2\alpha+a}{c^2}\frac{\partial\phi}{\partial t} - \frac{\alpha^2+a\alpha}{c^2}\phi = 0. \qquad (10)$$

$$-\hbar^2\frac{\partial^2\psi}{\partial t^2} + c^2\hbar^2\nabla^2\psi = m_0^2 c^4\psi \qquad (15)$$

$$u_{tt} + Au + f(u) = 0 \qquad (1)$$

$$u_{tt} - \Delta u + mu + P'(u) = 0 \quad (m > 0, P(u) \geqslant 0), \qquad (1)$$

$$\nabla^a\nabla_a\psi = \mu^2\psi\,, \qquad (4)$$

$$u_{tt} - \Delta u + m^2 u + G'(u) = 0, \qquad (1)$$

$$\left(\eta^{\mu\nu}\frac{\partial}{\partial x^\mu}\frac{\partial}{\partial x^\nu} - \left(\frac{mc}{\hbar}\right)^2\right)\phi = 0$$

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

Subproblem Information Extraction

- Can one create a benchmark for semantic extraction of „machine readable“ from human readable documents?

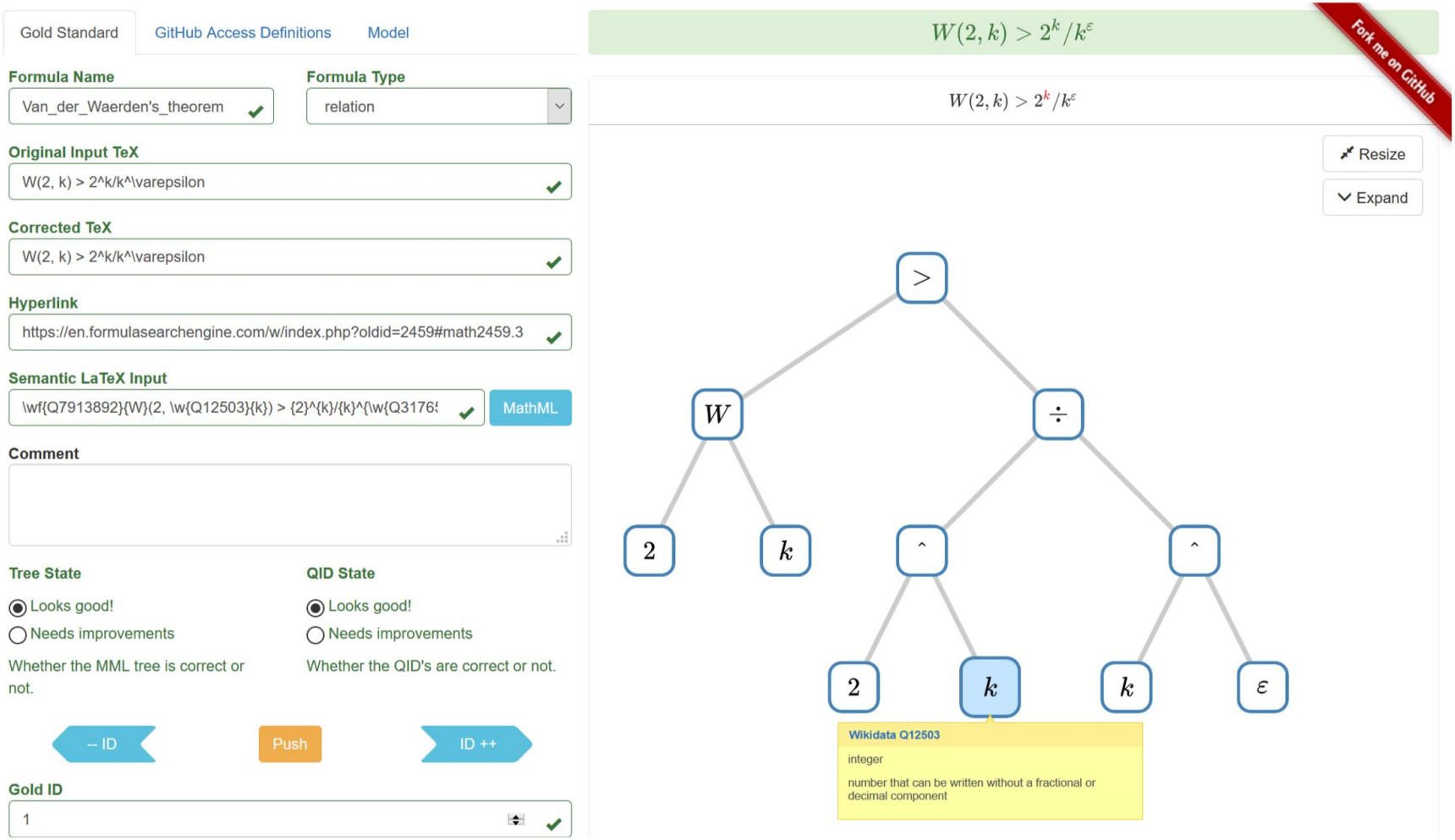
Overview

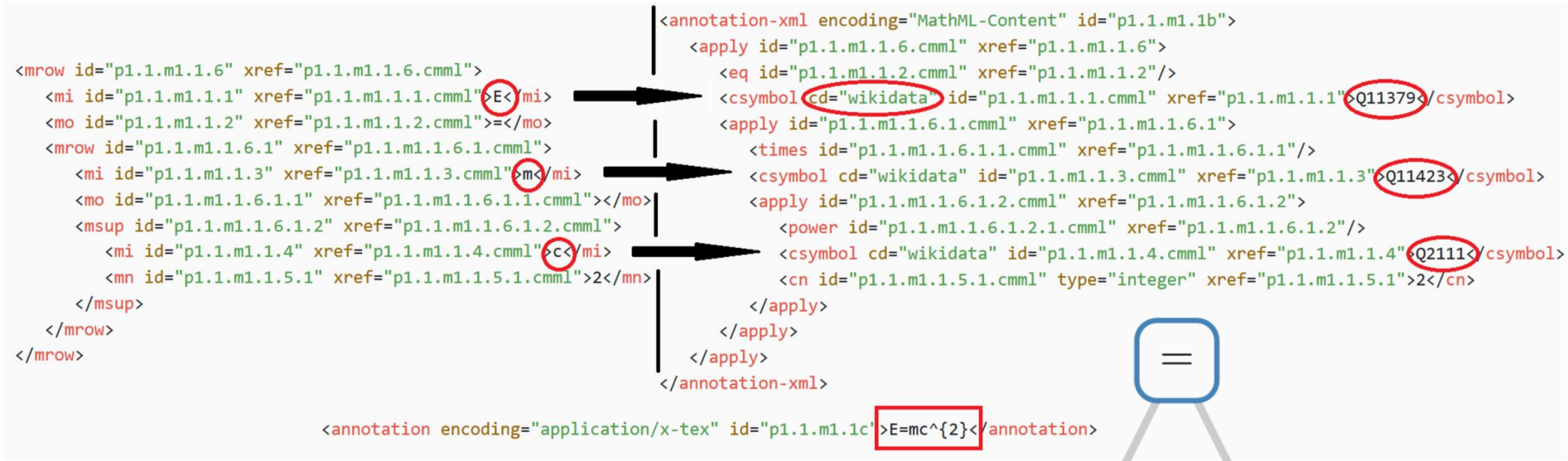
1. Introduction
2. **The MathML Benchmark MathMLBen**
3. A Wikidata Content Dictonary
4. Wikidata as a `cdbase`
5. OpenMath vs. Wikidata
6. Conclusions and Future Works

MathMLBen

- 300 formulae from
  - Wikipedia
  - arXiv
  - DLMF
- Includes formulae from NTCIR
- Open github project

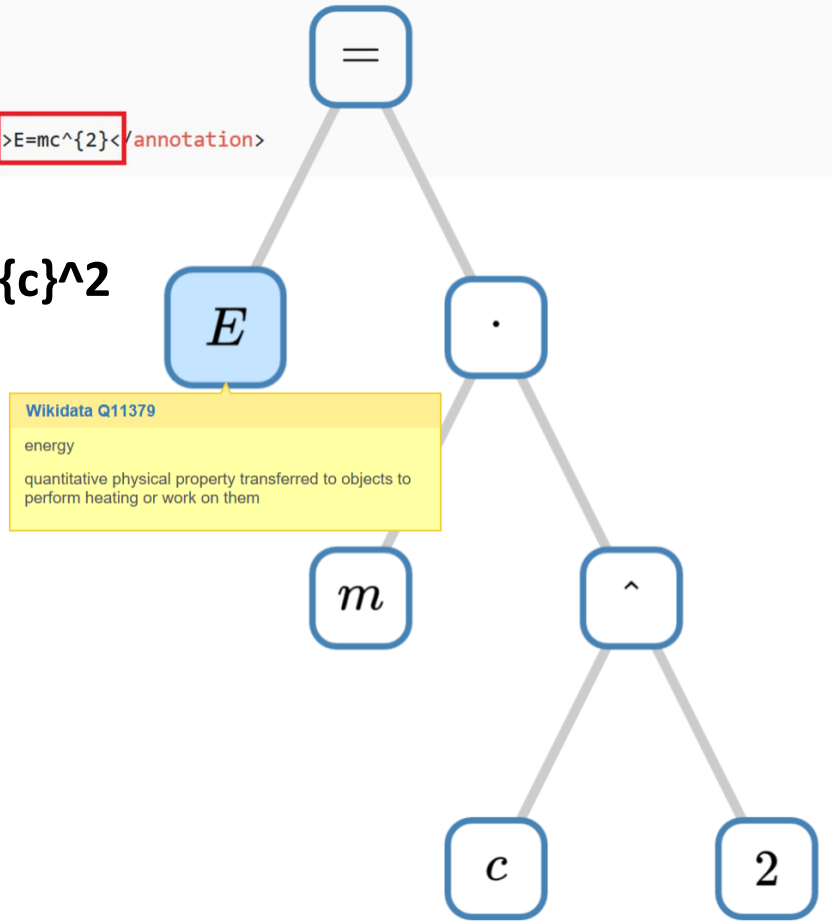
https://mathmlben.wmflabs.org





$$E = mc^2$$

\w{Q11379}{E} = \w{Q11423}{m} \w{Q2111}{c}^2



Paper and references are openly available from:

https://www.gipp.com/wp-content/papercite-data/pdf/scharpf2018.pdf

Overview

1. Introduction
2. The MathML Benchmark MathMLBen
3. **A Wikidata Content Dictonary**
4. Wikidata as a cdbase
5. OpenMath vs. Wikidata
6. Conclusions and Future Works

A wikidata content dictionary

- 49M Wikidata items
- MathMLBen uses 280 items
- https://cd.formulasearchengine.com/wikidata.ocd

Logistic function example entry

- wikidata.ocd provides access layer for standard OpenMath tools to Wikiata data

```
<CDDefinition>
  <Name>Q1052379</Name>
  <Role>application</Role>
  <Description>
    logistic function
    https://en.wikipedia.org/w/
    index.php?title=Logistic+function

    This description was generated from
    https://www.wikidata.org/w/
    index.php?oldid=648452086
  </Description>
</CDDefinition>
```

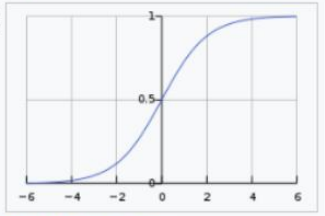
logistic function (Q1052379)

No description defined

Statements

subclass of
function
0 references

image



Logistic-curve.svg
600 × 400; 3 KB
0 references

defining formula
$f(x) = \frac{L}{1 + e^{-k(x - x_0)}}$
1 reference
imported from Wikimedia project English W

Identifiers

Freebase ID
/m/0lj5d
1 reference
stated in Freebase Data Dumps
publication date 28 October 2013

MathWorld identifier
LogisticEquation
1 reference
imported from Wikimedia project English Wikipedia

Quora topic ID
Logistic-Function
1 reference
imported from Wikimedia project Quora

PSH ID
11351
0 references

Sitelinks

Wikipedia (21 entries)

- arwiki دالة لوجستية
- cswiki Logistická funkce
- dewiki Logistische Funktion
- enwiki Logistic function
- eswiki Función logística
- frwiki Fonction logistique (Verhulst)
- hewiki פונקציה לוגיסטית
- itwiki Equazione logistica
- kowiki 로지스틱 방정식
- nlwiki Logistische functie
- ruwiki Логистическое уравнени
- skwiki Logistická funkcia
- svwiki Logistisk funktion
- trwiki Lojistik fonksiyon
- ukwiki Логістична рівність

Overview

1. Introduction
2. The MathML Benchmark MathMLBen
3. A Wikidata Content Dictonary
4. **Wikidata as a `cdbase`**
5. OpenMath vs. Wikidata
6. Conclusions and Future Works

Property talk:P5610

Documentation

OpenMath ID
OpenMath content dictionary symbol

Start a query – Current uses – Statistics by class – String len

Represents	OpenMath (Q1465300)
Data type	External identifier
Domain	Mathematical concepts (note: this should be moved to the property statements)
Allowed values	Any valid symbol from one of the official OpenMath content dictionaries in the form of CDName:Symbol https://www.openmath.org/standard/ (note: this information should be moved to a property format as a regular expression (P1793))
Example	absolute value (Q120812) → arith1#abs division (Q1226939) → arith1#divide greatest common divisor (Q131752) → arith1#gcd hyperbolic cosine (Q1253682) → transc1#cosh
Formatter URL	http://tools.wmflabs.org/wikidata-externalid-url/?p=5610&url_prefix=https://www.openmath.org/cd/&id=**$1**
Tracking: usage	no label (Q56039309)

Lists	• Count of items by number of statements (chart)
	  • Items with the most statements
	  • Items with the fewest statements
	• Count of items by number of sitelinks (chart)
	  • Items with the most sitelinks
	  • Items with the most sitelinks, but no link to one Wikipedia (sample: Welsh Wikipedia, cywiki)
	• Items with the most identifier properties
	• Items with no other external identifier
	• Items with no other statements
	• Items with novalue claims
	• Items with unknown value claims
	• Usage history
	• Database reports/Constraint violations/P5610

Proposal discussion	Property proposal/OpenMath ID
Current uses	52
Search for values	haswbstatement:P5610=arith1#abs **Search for 'arith1#abs'**

[create] Create a translatable help page (preferably in English) for this property to be included here

Single value: this property generally contains a single value. (Help)
Exceptions are possible as rare values may exist.
List of this constraint violations: Database reports/Constraint violations/P5610#Single value, SPARQL, SPARQL (new)

Distinct values: this property likely contains a value that is different from all other items. (Help)
Exceptions are possible as rare values may exist.
List of this constraint violations: Database reports/Constraint violations/P5610#Unique value, SPARQL (every item), SPARQL (by value), SPARQL (new)

Format "([a-z]+[0-9]*)\#([a-z_]+)": value must be formatted using this pattern (PCRE syntax). (Help)
Exceptions are possible as rare values may exist.
List of this constraint violations: Database reports/Constraint violations/P5610#Format, SPARQL, SPARQL (new)

Categories: All Properties | Properties with external-id-datatype | Properties used on 10+ items | Properties with single value constraints | Properties with unique value constraints | Properties with format constraints

URI = cdbase + ’/’ + cd-name + ’#’ + symbol-name.

```
SELECT ?x WHERE { ?x wdt:P5610 "arith1#plus".}
```

```
<CDDefinition>
  <Name>Q32043</Name>
  <Role>application</Role>
  <Description>
    addition
    arithmetic operation of adding (augend+addend=summand+summand=sum,
       total). (Add, Sum, Plus, Increase, Total)
    https://en.wikipedia.org/w/index.php?title=Addition
    See also
    https://www.openmath.org/cd/arith1#plus

    This description was generated from
    https://www.wikidata.org/w/index.php?oldid=720254931
  </Description>
</CDDefinition>
```

Overview

1. Introduction
2. The MathML Benchmark MathMLBen
3. A Wikidata Content Dictonary
4. Wikidata as a `cdbase`
5. **OpenMath vs. Wikidata**
6. Conclusions and Future Works

gamma

Description:

Euler's gamma function

Commented Mathematical property (CMP):

gamma(z)=\int_0^{+\infty} t^{z-1} e^{-z} dt (Re(z)>0)

Formal Mathematical property (FMP):

OpenMath XML (source)
Strict Content MathML
Prefix
Popcorn
Rendered Presentation MathML

$$\text{real}(z) > 0 \Rightarrow \text{gamma}(z) = \int_0^{\infty} t^{(z-1)} e^{(-z)} dt$$

Example:

gamma(n) = (n-1)! (n \in N)

OpenMath XML (source)
Strict Content MathML
Prefix
Popcorn
Rendered Presentation MathML

$$n \in \mathbb{N} \Rightarrow \text{gamma}(n) = (n-1)!$$

Signatures:

sts

Gamma function

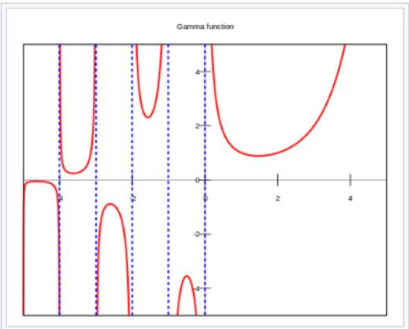
From Wikipedia, the free encyclopedia

In mathematics, the **gamma function** (represented by Γ, the capital Greek alphabet letter gamma) is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers. If n is a positive integer,

$$\Gamma(n) = (n-1)!$$

The gamma function is defined for all complex numbers except the non-positive integers. For complex numbers with a positive real part, it is defined via a convergent improper integral:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$$



The gamma function along part of the real axis

This integral function is extended by analytic continuation to all complex numbers except the non-positive integers (where the function has simple poles), yielding the meromorphic function we call the gamma function. It has no zeroes, so the reciprocal gamma function $1/\Gamma(z)$ is a holomorphic function. In fact the gamma function corresponds to the Mellin transform of the negative exponential function:

$$\Gamma(z) = \{\mathcal{M} e^{-x}\}(z)$$

The gamma function is a component in various probability-distribution functions, and as such it is applicable in the fields of probability and statistics, as well as combinatorics.

arith1#plus (OpenMath vs. Wikidata)

```
<CDDefinition>
<Name>plus</Name>
<Role>application</Role>
<Description>
The symbol representing an n-ary commutative
function plus.
</Description>
<CMP> for all a,b | a + b = b + a </CMP>
<FMP>
<OMOBJ version="2.0">
  <OMBIND><OMS cd="quant1" name="forall"/>
    <OMBVAR> <OMV name="a"/> <OMV name="b"/>
    </OMBVAR>
    <OMA> <OMS cd="relation1" name="eq"/>
    <OMA><OMS cd="arith1" name="plus"/>
      <OMV name="a"/>
      <OMV name="b"/></OMA>
    <OMA><OMS cd="arith1" name="plus"/>
      <OMV name="b"/>
      <OMV name="a"/></OMA></OMA>
  </OMBIND>
</OMOBJ>
</FMP>
</CDDefinition>
```

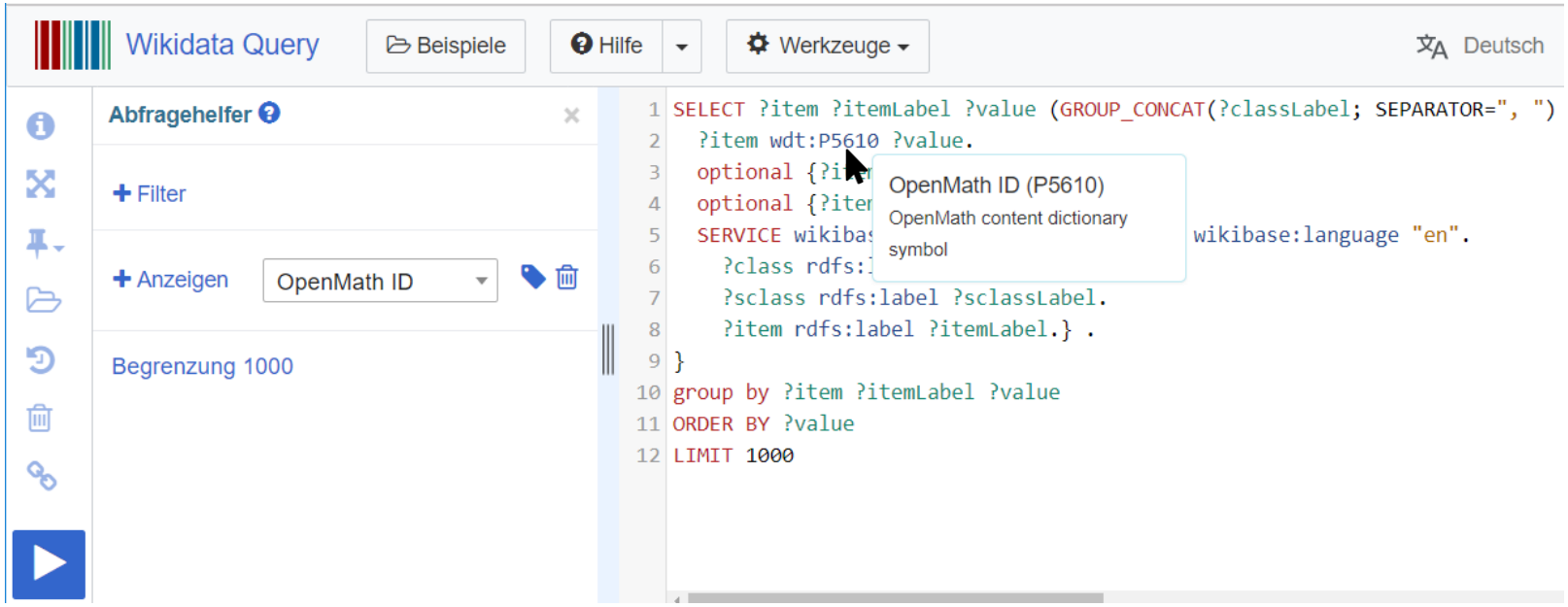
```
<CDDefinition>
  <Name>Q32043</Name>
  <Role>application</Role>
  <Description>
addition
arithmetic operation of adding (
augend+addend=summand+summand=sum, total). (
Add, Sum, Plus, Increase, Total)
https://en.wikipedia.org/w/
index.php?title=Addition
See also
https://www.openmath.org/cd/arith1#plus

This description was generated from
https://www.wikidata.org/w/
index.php?oldid=720254931
  </Description>
</CDDefinition>
```

Language support (arith1#plus)

- Label in 88 languages
- Descriptions in 14 languages

निरपेक्ष मान	|a| उस संख्या के चिह्न के बिना उसके आंकिक मान के बराबर होता है। उदाहरण
Absolute waarde	absolute waarde van 'n reële getal die nie-negatiewe waarde van die getal sonder inagneming van die getal se teken
valeur absolue	Distance à 0, valeur numérique d'un nombre réel sans tenir compte de son signe
wartość bezwzględna	funkcja matematyczna, wartość liczbowa nieuwzględniająca znaku danej liczby
Valore assoluto	Funzione dove per valori negativi della x si ottiene lo stesso risultato.
absolute value	magnitude of the number on the real number line; (of a real number x) non-negative value of x without regard to its sign
Betragsfunktion	mathematische Funktion
Mütləq qiymət	riyaziyyatda bir həqiqi ədədin işarəsiz qiyməti
Valor absoluto	ven dado pola seguinte expresión: Podemos notar que o valor absoluto dun número sempre tomará valores non-negativos, é dicir:
Модуль ліка	гэта адлегласць гэтага ліку ад нуля
модуль числа	математична функція і термін
ערך מוחלט	ערך מוחלט הוא פונקציה המודדת את גודלם של איברים בשדה, במתמטיקה
قيمة مطلقة	إذا كان يساو :هي دالة رياضية تخضع للمواصفات الثلاثة التالية

Plus vs Q32043

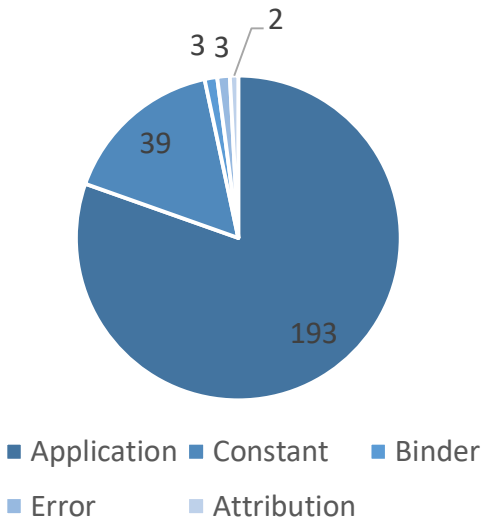


Wikidata Label	OpenMath ID	Instance of
absolute value	arith1#abs	piecewise function, even function, idempotent function
division	arith1#divide	binary operation
greatest common divisor	arith1#gcd	function
subtraction	arith1#minus	binary operation, operation
addition	arith1#plus	binary operation
exponentiation	arith1#power	operation
nth root	arith1#root	type of mathematical function, algebraic function
sum	arith1#sum	mathematical expression
multiplication	arith1#times	binary operation
opposite number	arith1#unary_minus	
derivative	calculus1#diff	unary operation, mathematical concept
Lambda expression	fns1#lambda	Wikimedia disambiguation page

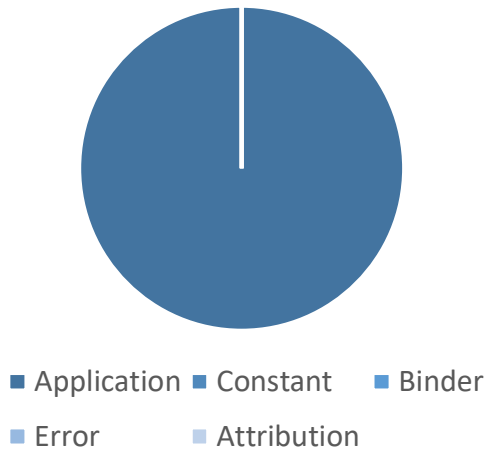
Role

www.isg.uni.kn

- OpenMath



- Wikidata



- OpenMath
- 289 official symbols
- 38 content dictionaries
- 5 roles
- 149 examples
- 180 FMP
- 179 CMP
- 131 average description length
- English names

- Wikidata
- 330 symbols
- 1 content dictionary
- 1 role
- 0 examples
- 0 FMP
- 0 CMP
- 198 average description length
- QIds

Overview

1. Introduction
2. The MathML Benchmark MathMLBen
3. A Wikidata Content Dictonary
4. Wikidata as a `cdbase`
5. OpenMath vs. Wikidata
6. **Conclusions and Future Works**

Conclusion and Future works

- https://cd.formulasearchengine.com/wikidata.ocd
  - 330 Wikidata based CDSymbols
  - 50 traditional OpenMath symbols
- P5610 OpenMath ID (Property in Wikidata)
- Can we derive more formal entries from Wikidata to match the quality of OpenMath entries?
- How can we use alignments?