



CICM 2016, DML Track

A Smooth Transition to Modern mathoid-based Math Rendering in Wikipedia with Automatic Visual Regression Testing

Moritz Schubotz & Alan Sexton
(with material from Frédéric Wang)



Overview

1. Introduction to Wikipedia's new math rendering using mathoid
2. Improvements over the old math rendering using texvc
3. Comparing different math rendering engines using mathpipe
 1. Automation Framework
 2. Comparison of math images
4. Future Work



Overview

1. **Introduction to Wikipedia's new math rendering using mathoid**
2. Improvements over the old math rendering using texvc
3. Comparing different math rendering engines using mathpipe
 1. Automation Framework
 2. Comparison of math images
4. Future Work

👤 Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article **Talk** Read Edit More ▾ 🔍



- [Main page](#)
 - [Contents](#)
 - [Featured content](#)
 - [Current events](#)
 - [Random article](#)
 - [Donate to Wikipedia](#)
 - [Wikipedia store](#)
-
- Interaction
- [Help](#)
 - [About Wikipedia](#)

Bell's theorem

From Wikipedia, the free encyclopedia
(Redirected from [Bell inequality](#))

Bell's theorem is a ‘no-go theorem’ that draws an important distinction between quantum

Quantum mechanics

$$\hat{H}|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$

Schrödinger equation

[Introduction](#) · [Glossary](#) · [History](#)

Background [\[show\]](#)

Fundamentals [\[hide\]](#)

[Complementarity](#) · [Decoherence](#) ·

[Main page](#)[Contents](#)[Featured content](#)[Current events](#)[Random article](#)[Donate to Wikipedia](#)[Wikipedia store](#)

Bell's theorem is a ‘no-go theorem’ that draws an important

Quantum mechanics

mi.MJX-variant 12.0706 × 20.8

$$\hat{H}|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$

Schrödinger equation[Introduction](#) · [Glossary](#) · [History](#)[Background](#)[\[show\]](#)

Inspekt... Konso... Debugg... Stilbearbeit... Laufzeitana... Netzwerka...

[line.mwe-math-mat...](#) [math](#) [semantics](#) [mrow.MJX-TeXAtom-ORD](#) [mstyle](#) [mi.MJX-variant](#)

```
<mstyle displaystyle="true" scriptlevel="0">
  <mrow class="MJX-TeXAtom-ORD"></mrow>
  <mrow class="MJX-TeXAtom-ORD"></mrow>
  <mi></mi>
  <mo stretchy="false">(</mo>
  <mi>t</mi>
  <mo stretchy="false">)</mo>
  <mo fence="false" stretchy="false"></mo>
  <mo>=</mo>
  <mi>i</mi>
  <mi class="MJX-variant">
    ħ
    <!-- ħ -->
  </mi>
```



WIKIPEDIA

The Free Encyclopedia

[Main page](#)

[Contents](#)

[Featured content](#)

[Current events](#)

[Random article](#)

[Donate to Wikipedia](#)

[Wikipedia store](#)

Interaction

[Help](#)

[About Wikipedia](#)

[Community portal](#)

[Recent changes](#)

[Contact page](#)

[Article](#)

[Talk](#)

[Read](#)

[Edit](#)

[More](#) ▼



Bell's theorem

From Wikipedia, the free encyclopedia

(Redirected from [Bell inequality](#))

Bell's theorem is a ‘no-go theorem’ that draws an important distinction between quantum mechanics (QM) and the

Quantum mechanics

$$\hat{H}|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$

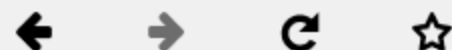
[Introduction](#)

[Complex](#)

[Entang](#)

[Meas](#)

[Quantum number](#) • [State](#)



Seite speichern unter...

Seite bei Pocket speichern

Hintergrundgrafik anzeigen

Alles markieren

MathML-Quelle anzeigen

Seitenquelltext anzeigen

Seiteninformationen anzeigen

Element untersuchen (Q)

Copy MathML Formula

Copy Annotation



Dokument2 - Word Moritz Schubotz

Datei Start Einfügen Entwurf Layout Verweise Sendung Überprüf Ansicht MathPla Add-Ins Entwurf Sie...ünsce Freige...en

Einfügen

Calibri (Textkörp... 11

F **K** **U** **abc** **x₂** **x²** **A**

A **ab** **A** **Aa** **A** **A**

Schriftart

G **P** **F**

Y1 **N** **JN**

AaBbCcDd AaBbCcDd AaBbCcDd

Standard Kein Lee... Überschrif...

Formatvorlagen

Bearbeiten

Zwischenablage

$$H|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$

$$p = \frac{h}{\lambda}$$

De Broglie übertrug dies 1923 auf beliebige Teilchen:

$$\lambda = \frac{h}{p}$$

mit dem relativistischen Impuls:

$$p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

ergibt sich daraus die sogenannte De-Broglie-Wellenlänge:

$$\lambda = \frac{h \cdot \sqrt{1 - \frac{v^2}{c^2}}}{mv}$$

$$\text{In[8]:= } p == \frac{h}{\lambda}$$

$$\text{Out[8]= } p == \frac{h}{\lambda}$$

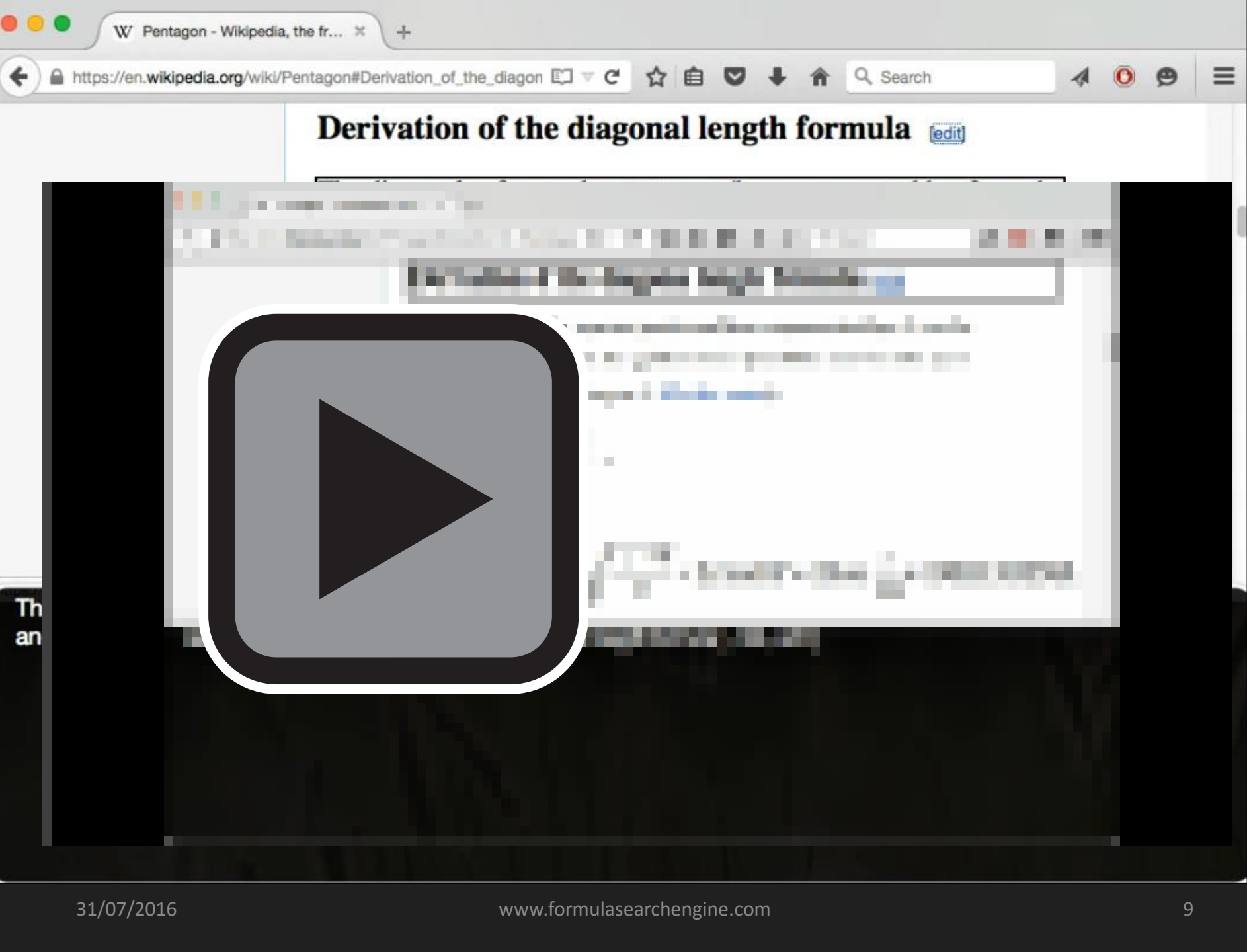
$$\text{In[9]:= } p = \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\text{Out[9]= } \frac{mv}{\sqrt{1 - \frac{v^2}{c^2}}}$$

$$\text{In[10]:= } \text{Solve}[\%8, \lambda]$$

$$\text{Out[10]=}$$

$$\left\{ \left\{ \lambda \rightarrow \frac{h \sqrt{\frac{c^2 - v^2}{c^2}}}{mv} \right\} \right\}$$





Presentation MathML

- Enables Copy and Paste
- Good presentation
- Accessibility
- Better text integration
- Reduces bandwidth
 - Wikimedia 15 Billion visits
 - 5 pages per **person** and month (p/**pm**)
 - Central African Republic 0.1 p/**pm**
 - Ireland 19p/**pm**



- Lack of fonts
- No country of mathematicians
- Add-on required
- ... otherwise SVG

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} t^{z-1} e^{-t} dt. [4]$$

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} t^{z-1} e^{-t} dt. [4]$$

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} t^{z-1} e^{-t} dt. [4]$$

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} t^{z-1} e^{-t} dt. [4]$$

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{1}{z+n} + \int_1^{\infty} t^{z-1} e^{-t} dt. [4]$$



Overview

1. Introduction to Wikipedia's new math rendering using mathoid
- 2. Improvements over the old math rendering using texvc**
3. Comparing different math rendering engines using mathpipe
 1. Automation Framework
 2. Comparison of math images
4. Future Work



Comparison to old rendering

B I   

▼ Advanced ▶ Special characters ▶ Help

Heading ▼ Format      

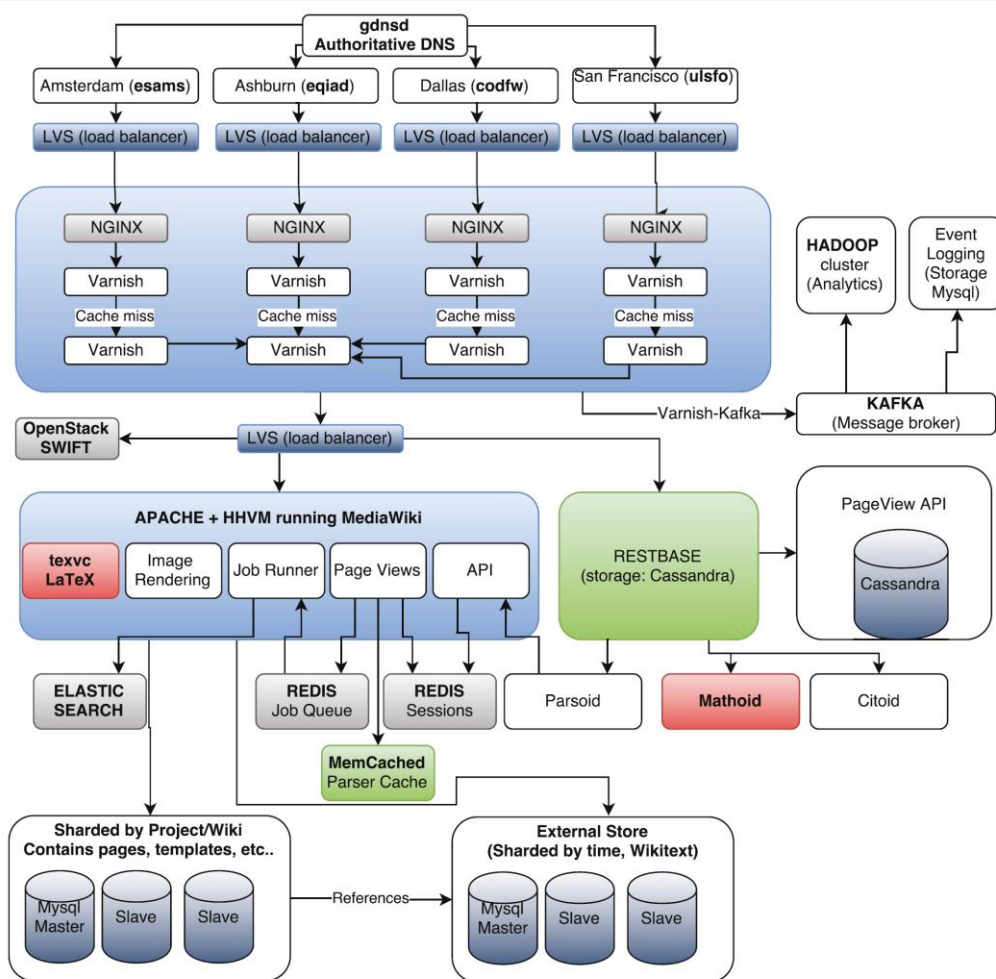
A⁺ A⁻ A[^] A₋ Insert    

```
| bodystyle = width:
| {{width|19.0em}};
| title = [[Quantum mechanics]]
| image = <math>\hat{H} | \psi(t)
\angle = i \hbar \frac{\partial}{\partial t} | \psi(t) \angle
</math>
| captionstyle =
font-size:90%;padding-top:0.4em;font-
style:italic;
| caption = [[Schrödinger equation]]
| listtitlestyle = text-align:center;
| liststyle = border-top:1px solid
#aaa;border-bottom:1px solid #aaa;
```

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$
$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle$$

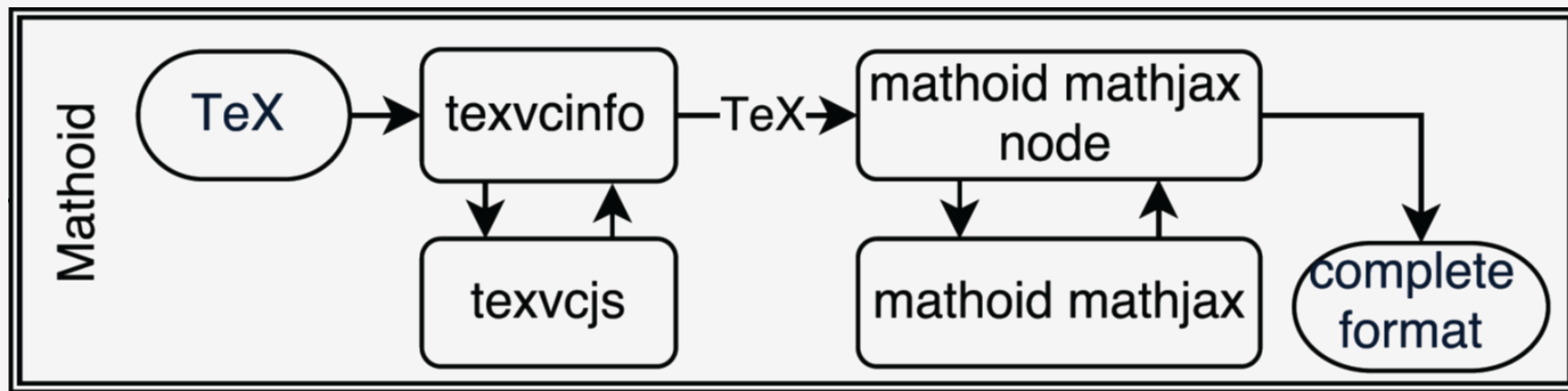


<https://github.com/wikimedia/mathoid>

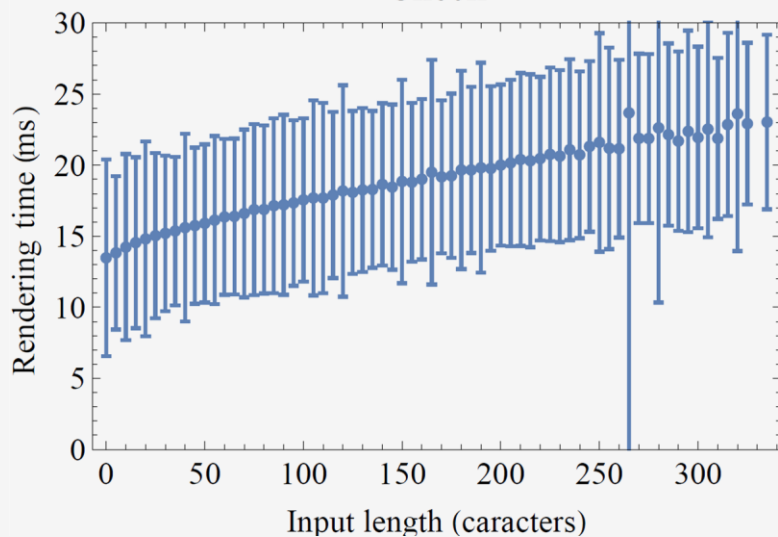




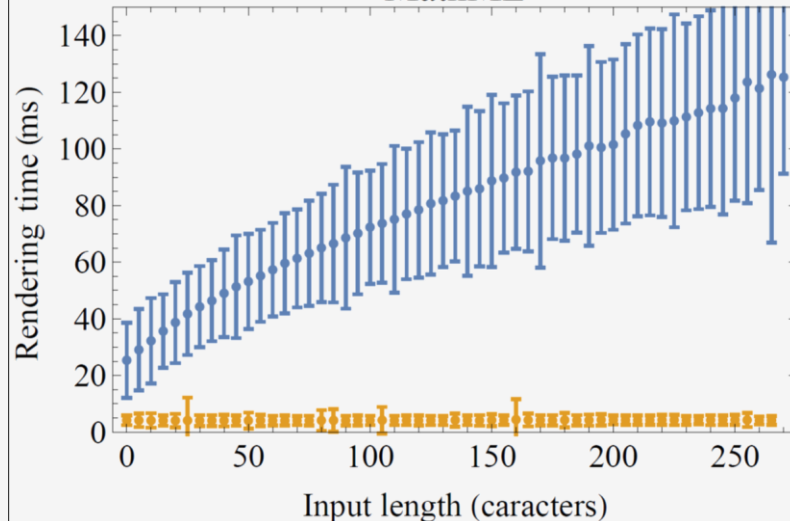
Performance



Check



MathML



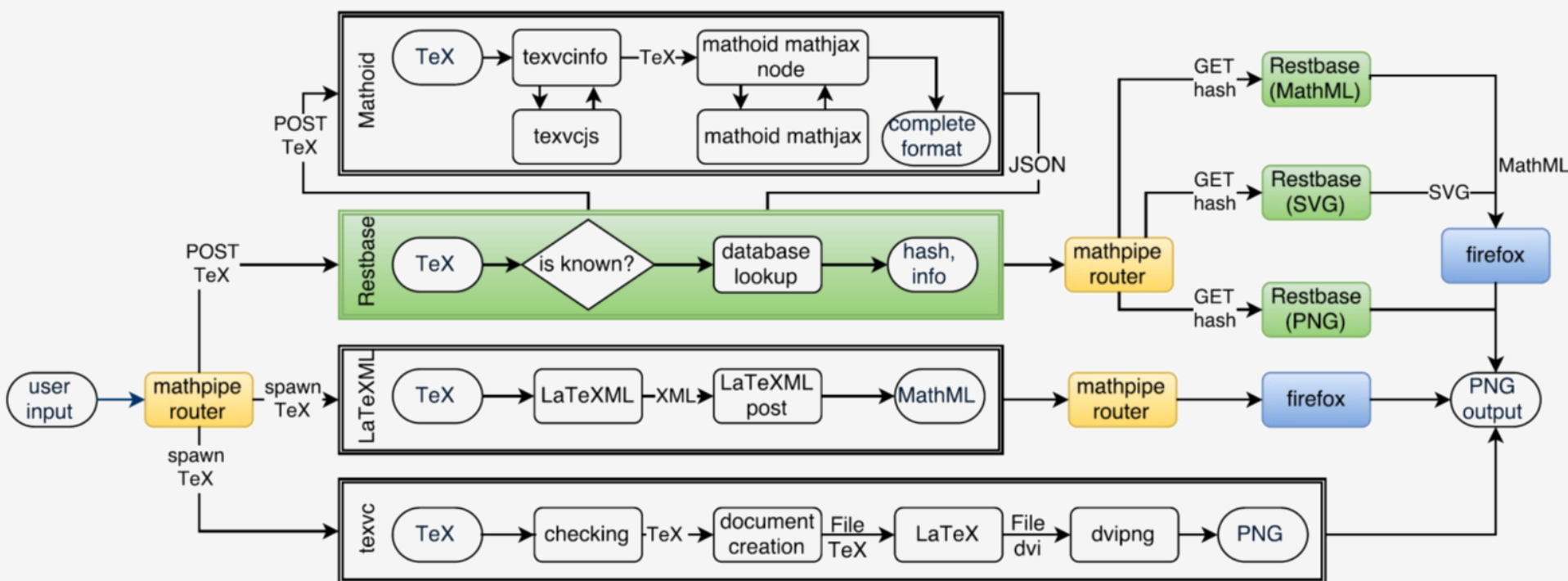


Overview

1. Introduction to Wikipedia's new math rendering using mathoid
2. Improvements over the old math rendering using texvc
- 3. Comparing different math rendering engines using mathpipe**
 - 1. Automation Framework**
 2. Comparison of math images
4. Future Work



mathpipe Automation





Overview

1. Introduction to Wikipedia's new math rendering using mathoid
2. Improvements over the old math rendering using texvc
- 3. Comparing different math rendering engines using mathpipe**
 1. Automation Framework
 - 2. Comparison of math images**
4. Future Work



Comparison of math images

$$P = 3B_0 \left(\frac{1 - \eta}{\eta^2} \right) e^{\frac{3}{2}(\textcolor{teal}{B}'_0 - 1)(1 - \eta)}$$

$$P = 3B_0 \left(\frac{1 - \eta}{\eta^2} \right) e^{\frac{3}{2}(\textcolor{violet}{B}'_0 - 1)(1 - \eta)}$$

$$P = \textcolor{brown}{3}B_{\textcolor{brown}{0}} \left(\frac{1 = \eta}{\eta^2} \right) e^{\frac{3}{2}((\textcolor{teal}{B}'_0 - 1))(1 = \eta)}$$



Comparison of math images

Significant changes

- Missing characters
- Different shapes
- Space changes
- Spatial misalignments
- Touching and breaking of symbols

Insignificant changes

- Font changes
- Space changes
- Touching and breaking of symbols

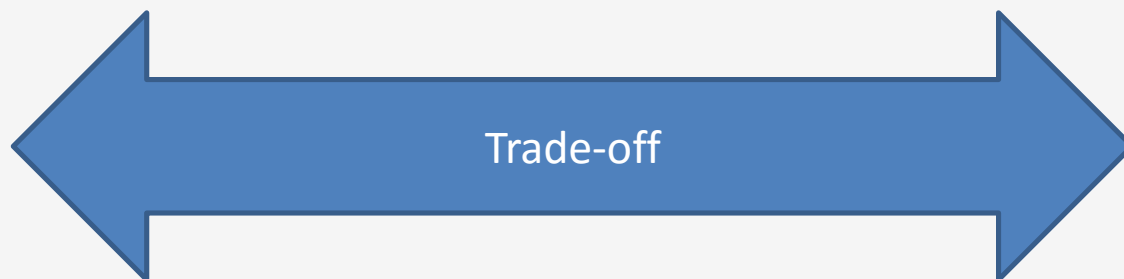




Image comparison method

Binarisation & Connected component analysis

Cropping & Scaling

Simple component pairing

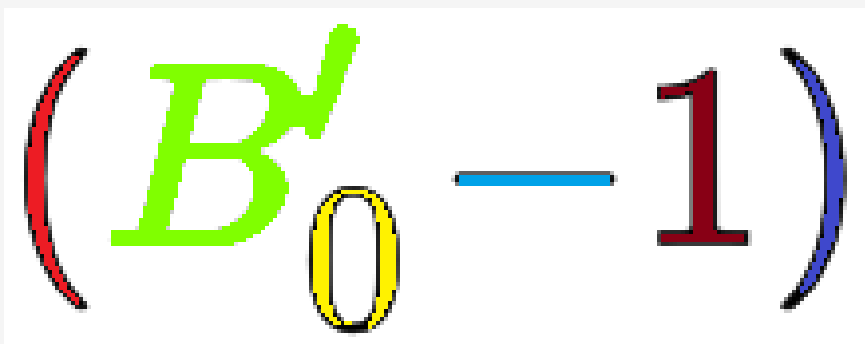
Touching component analysis

Result reporting



Binarisation & Connected component analysis

- Binarisation based on threshold
 - Different encodings
 - Grey scale vs. Transparency
- Connected components
 - Standard methods [He et al.]





Cropping & Scaling

- Remove superfluous background padding
- Avoid resize artefacts
- Virtual scaling using floats

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$



Cropping & Scaling

- Remove superfluous background padding
- Avoid resize artefacts
- Virtual scaling using floats

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$



Cropping & Scaling

- Remove superfluous background padding
- Avoid resize artefacts
- Virtual scaling using floats

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$



Cropping & Scaling

- Remove superfluous background padding
- Avoid resize artefacts
- Virtual scaling using floats

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$

$$\lim_{k \rightarrow \infty} ||\mathbf{T}^k|| = 0,$$



Simple component pairing

Iterative algorithm

1. Tight constraints on

- position
- size
- aspect ratio

2. Verification based on metric shape similarity

3. Relaxing parameters

$$P \rightleftharpoons 3B_0 \left(\frac{1-\eta}{\eta^2} \right) e^{\frac{3}{2}(\mathbf{B}'_0-1)(1-\eta)}$$
$$P \rightleftharpoons 3B_0 \left(\frac{1-\eta}{\eta^2} \right) e^{\frac{3}{2}(\mathbf{B}'_0-1)(1-\eta)}$$

Possible results

a) No unpaired components

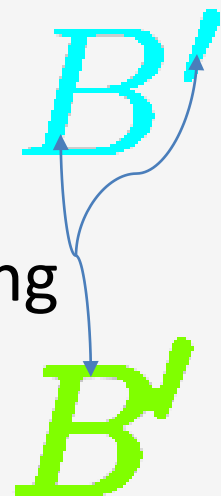
b) Relaxing parameters exceeds limit

c) Multi match



Touching component analysis

- Iterative algorithm
 - Start at largest unpaired component
 1. Try all combinations based on neighbouring unmatched components
 2. Shape similarity test
 3. If success → remove element from set
 - Finish on no remaining elements == finished processing the smallest element





Result reporting

- Components grouped into four groups
 1. Simple matches
 2. Touching matches
 3. First image unpaired
 4. Second image unpaired
- Comparison passed if only groups 1 (+2) are not empty

$$P = 3B_0 \left(\frac{1 - \eta}{\eta^2} \right) e^{\frac{3}{2}((B_0 - 1)(1 - \eta))}$$



Overview

1. Introduction to Wikipedias new math rendering using mathoid
 2. Improvements over the old math rendering using texvc
 3. Comparing different math rendering engines using mathpipe
 1. Automation Framework
 2. Comparison of math images
- **1st Result**
4. Future Work



1st Result

$$R_{ix}(t) = M_i A_{ix}(t)$$

$$R_{ix}(t) = M_i A_{ix}(t) \cdot$$

$$R_{ix}(t) = M_i A_{ix}(t) \frac{\quad}{\quad}$$

```
<mfrac xml:id="p1.1.m1.1.14a" xref="<apply xml:id="p1.1.m1.1.14.cmml" xref="
  <mi xml:id="p1.1.m1.1.14.2" />    <divide xml:id="p1.1.m1.1.14.1.cmml"/>
  <mi xml:id="p1.1.m1.1.14.3" />    <csymbol cd="latexml">absent</csymbol>
</mfrac>                            <csymbol cd="latexml">absent</csymbol>
```



Overview

1. Introduction to Wikipedia's new math rendering using mathoid
2. Improvements over the old math rendering using texvc
3. Comparing different math rendering engines using mathpipe
 1. Automation Framework
 2. Comparison of math images
- 4. Future Work**



Future work

- Improve reliability
- Investigate all results
- Complete web service

<http://png.formulasearchengine.com>

- Integrate to the standard Wikimedia development workflow



Contact

Moritz Schubotz (now at Universität Konstanz)

moritz@schubotz.de

+49 7531 88 4438

Mobile: +49 1578 047 1397

www.isg.uni-konstanz.de

www.formulasearchengine.com