

Math Indexer and Searcher Web Interface

Towards Fulfillment of Mathematicians' Information Needs

M. Líška, Petr Sojka, M. Růžička

Faculty of Informatics
Masaryk University, Brno, Czech Republic

<http://mir.fi.muni.cz/>

CICM, S&P, July 10th, 2014

Coping with Information Overload by Filtering of *Big Data*

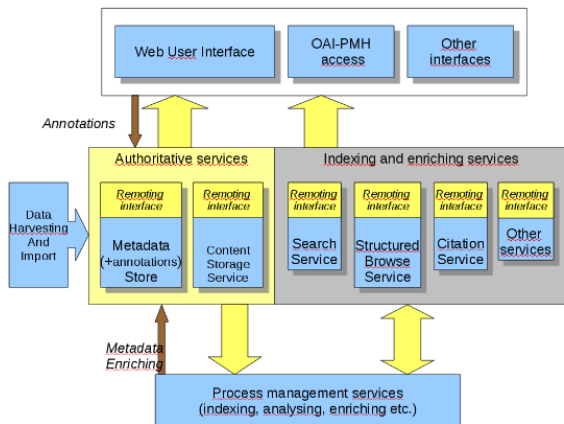


Life is *searching*: group *similar* and narrow focus of search in [your, mathematician's] Big Math Data. Search is 'killer app' of any today's working environments.

Different needs of search: in either formal or informal database of knowledge – in either formal [proof assistant] system of formulae (substitution based MWS for MMT) or for digital library of informal papers (similarity based MlaS for EuDML)

Digital Library Service Architecture and Workflow (EuDML)

Within *European Digital Mathematics Library, EuDML*, project EU CIP-ICT-PSP (2010–2013) we have developed and delivered technology for Math Indexing and Searching MIA-S.



The Need for Scalable Search Solution in EuDML

MlaS reported at CICM 2011: indexing 168,000,000 formulae, having 3,000,000,000 formulae in the index, latency below 1 second. Users like low-latency information systems.

No chance even for linear algorithm for formulae similarity **at runtime**: the method of static index expansion to cover structural (Presentation MathML) or semantic similarity.

<http://eudml.org/search/>

[Help](#) [About](#)



How to write query

```
<math><mi>x</mi> <sup>2</sup> </msup> <mi>y</mi> <sup>2</sup> </msup> </math>
```

Canonicalized MathML query:

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <msup>
    <mi>x</mi>
    <sup>2</sup>
  </msup>
  <math>+</math>
  <msup>
    <mi>y</mi>
    <sup>2</sup>
  </msup>
</math>
```

Search in:

Total hits: 36817, showing 1-30. Searching time: 116 ms

Finite Precision Measurement Nullifies Euclid's Postulates

... and the unit circle $x^2 + y^2 = 1$ are both dense but they do not intersect, in contradiction to Euclid's postulates ...

score = 3.2980976

arxiv.org/abs/quant-ph/0310035 - cached XHTML

COMMENT ON RECENT TUNNELING MEASUREMENTS ON Bi22Sr22CaCu22O88

... gap, (b) s-wave gap, and (c) $s_x^2 + y^2$ gap.

arxiv.org/abs/cond-mat/0304010

Math Search Interface WebMiaS Development

<http://mir.fi.muni.cz/webmias/>

The screenshot displays the WebMiaS search interface. At the top, the logo features mathematical symbols (gamma, pi, sin, sigma, integral) above the text "WEBMiaS" and "MATH INDEXER AND SEARCHER". A search bar contains the query $H^n(X) = Z^n(X)/B^n(X)$. Below the search bar, the text "Your query: $H^n(X) = Z^n(X)/B^n(X)$ " is shown. To the right, there are options for "Debug:" and "Search in:" with a dropdown menu set to "NCTIR-files" and a "Search" button. Below this, the text "Converted and canonicalized query:" is followed by a code block showing XML markup for the query. The search results section indicates "Total hits: 2, showing 1-2" and provides details for two results, including titles like "Line and continuum variability of two intermediate-redshift, high-luminosity quasars" and "Line and continuum variability of two intermediate-redshift, high-luminosity quasars", scores, and file names.

Examples About Help Contact

$H^n(X) = Z^n(X)/B^n(X)$

Your query: $H^n(X) = Z^n(X)/B^n(X)$ Debug: Search in: NCTIR-files Search

Converted and canonicalized query:

```
<math xmlns="" xmlns:xlink="">
  <semantics xmlns="">
    <arrow>
      <msup>
        <mi>H</mi>
        <mi>n</mi>
      </msup>
      <math>=
      </math>
      <msup>
        <mi>Z</mi>
        <mi>n</mi>
      </msup>
      </math>
      <math>/
      </math>
      <msup>
        <mi>B</mi>
        <mi>n</mi>
      </msup>
    </arrow>
  </semantics>
</math>
```

Total hits: 2, showing 1-2. Core searching time: 1691 ms Total searching time: 1811 ms

[Line and continuum variability of two intermediate-redshift, high-luminosity quasars](#)
... $\mu^{(k)}(\lambda) = Q^{(k)}(\lambda)/S^{(k)}(\lambda)$, ...
score = 0.06085243
fo65185.xhtml - cached XHTML

[fo88746.xhtml](#)
... $H^n(\mathcal{R}) = C^n(\mathcal{R})/B^n(\mathcal{R})$...
score = 0.01843161
fo88746.xhtml - cached XHTML

1



Match of the following rules

[remove](#)

[Add clause](#)

Contains the following formula:

k/H_0^2

Math formulae can be entered either in TeX or MathML notation (format will be autodetected). LaTeX math has to be enclosed within \$, AMS packages are supported.

Rendered: k/H_0^2

[Search](#)

Search using:

Search in:

Verbose output:

Total hits: 16, showing 1-16 . Core searching time: 1379 ms Total searching time: 2664 ms

[Giant Vortex Lattice Deformations in Rapidly Rotating Bose-Einstein Condensates](#)

... suggesting the **vortex density** ℓ/R_0^2 to be the dominant factor determining the variation of giant **vortex** core oscillation frequencies. ... (larger ℓ/R_0^2) the core oscillates rapidly, but with increasing core size, the oscillation frequency slows, and approaches the value of the breathing mode, as the size of the giant **vortex** approaches that of the condensate itself. Giant Vortex Lattice Deformations in Rapidly Rotating Bose-Einstein Condensates ... We have performed numerical simulations of giant **vortex** structures in rapidly rotating Bose-Einstein condensates within the Gross-Pitaevskii formalism.

score = 1.2043247

<http://arxiv.org/abs/cond-mat/0307130> - cached XHTML

[Split-merge cycle, fragmented collapse, and vortex disintegration in rotating Bose-Einstein condensates with attractive interactions](#)

... , and $(Nd_0^3)^{1/2}$, respectively, where ... Split-merge cycle, fragmented collapse, and **vortex** disintegration in rotating Bose-Einstein condensates with attractive interactions ... The dynamical instabilities and ensuing dynamics of singly- and doubly-quantized **vortex** states of Bose-Einstein condensates with attractive interactions are investigated using full 3D numerical simulations of the Gross-Pitaevskii equation. ... -fold **density** modulation grows exponentially in time while rotating at frequency ...

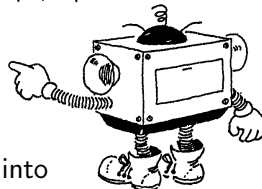
formulae in $\text{T}_{\text{E}}\text{X}$ Mathematicians know and use compact $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ math notation. Auto-detection of MathML is also in place. To convert $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ queries into MlS-supported MathML, we switched the converter from Tralics to $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{XML}$, which is able to convert the user input into mixed Presentation-Content MathML.

on-the-fly formulae rendering Formulae rendering allows quick feedback when writing the query—users know what they want when they see it. Robust live rendering of copy-pasted MathML is provided means of MathJax. Users are also warned when writing an invalid $\text{T}_{\text{E}}\text{X}$ query.

pop-up help Pop-up windows inform users about the interface.

- domain-specific auto-completion Frequent collocations and terms from the DML domain are suggested for text queries.
- facets Adding facets allows natural filtering (by language, author, . . .) of search results to achieve high precision.
- snippets with query coloring Snippets are shown in hit lists. Matched words and formulae are colored for a quicker first look evaluation of the results.
- scoring and debugging Scoring of computed relevance to a query is shown for every hit. In the development interface, one can inspect document score computation.

Conclusions and Future Work



- ▶ embedding MlaS and WebMlaS into Lucene/DSpace/ElasticSearch distributions
- ▶ up and running math-aware interface in EuDML
- ▶ math mining the logs to see user behaviour patterns
- ▶ deploying WebMlaS in further digital libraries, as DML-CZ

Further Readings/ Links

- ▶ WebMlaS: <https://mir.fi.muni.cz/webmias/>
- ▶ Math Information Retrieval: <https://mir.fi.muni.cz/>
- ▶ DML-CZ project: <http://dml.cz/>,
<http://project.dml.cz/>
- ▶ EuDML project: <http://eudml.org/>,
<http://project.eudml.org/>

Yes, we can!

