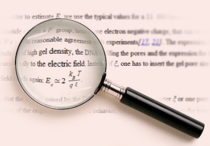
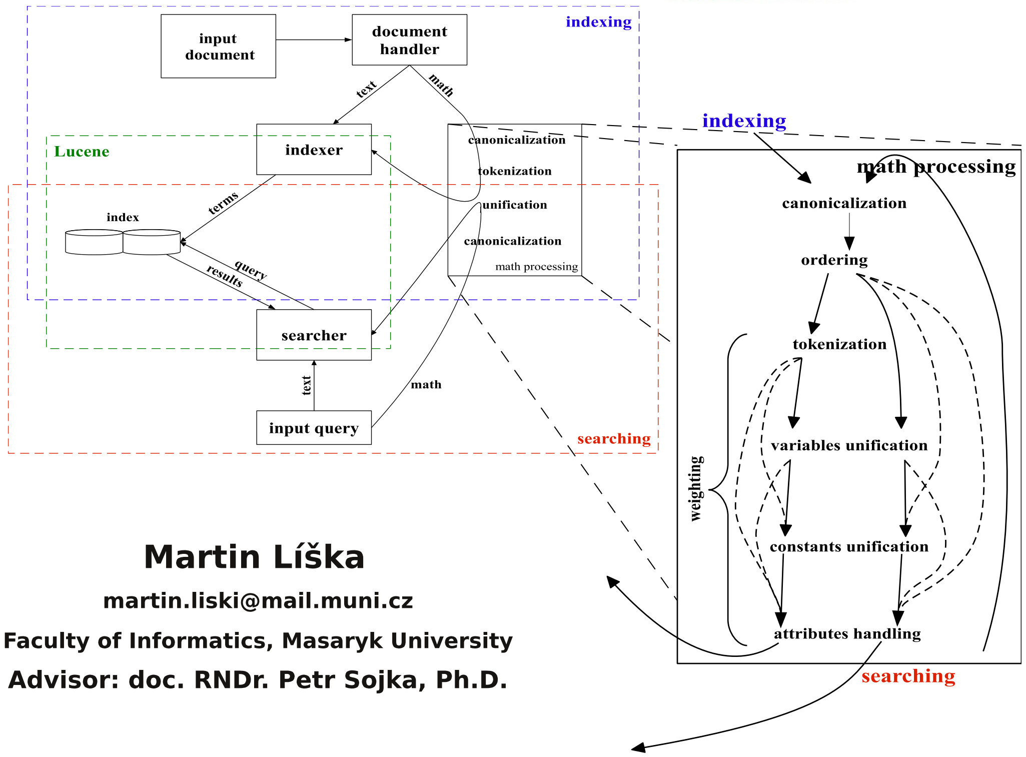
Evaluation of Mathematics Retrieval



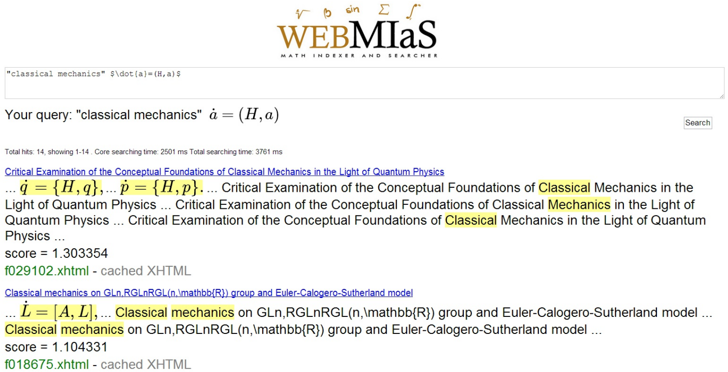
Overview

The thesis deals with the evaluation of mathematics information retrieval (IR). It gives an overview of the history of regular IR evaluation, initiatives that are engaged in this field of research as well as most common methods and measures used for evaluation. The findings are applied to the specifics of mathematics retrieval. This thesis also summarizes the state-of-the-art of MIaS (Math Indexer and Searcher) math search system, which is already being used in a running international digital library EuDML (The European Digital Mathematical Library, https://eudml.org/search). Latest developments aiming towards the second version of the system are described. In addition to participating in the international evaluation conference and workshop, MIaS is tested for effectiveness and efficiency in this work. Measured performance indicators are evaluated and future work is suggested accordingly.

MIaS design overview



WebMIaS interface



```
"classical mechanics" $\dot{a}=(H,a)$
```

Your query: "classical mechanics" $\dot{a} = (H, a)$

Total hits: 14, showing 1-14 . Core searching time: 2501 ms Total searching time: 3761 ms

Critical Examination of the Conceptual Foundations of Classical Mechanics in the Light of Quantum Physics

... $\dot{q} = \{H, q\}$, ... $\dot{p} = \{H, p\}$. ... Critical Examination of the Conceptual Foundations of Classical Mechanics in the Light of Quantum Physics ... Critical Examination of the Conceptual Foundations of Classical Mechanics in the Light of Quantum Physics ... Critical Examination of the Conceptual Foundations of Classical Mechanics in the Light of Quantum Physics ...

score = 1.303354

f029102.xhtml - cached XHTML

Classical mechanics on GLn,RGLnRGL(n,\mathbb{R}) group and Euler-Calogero-Sutherland model

... $\dot{L} = [A, L]$, ... Classical mechanics on GLn,RGLnRGL(n,\mathbb{R}) group and Euler-Calogero-Sutherland model ... Classical mechanics on GLn,RGLnRGL(n,\mathbb{R}) group and Euler-Calogero-Sutherland model ...

score = 1.104331

f018675.xhtml - cached XHTML

MIaS

MIaS is a math-aware full-text based search system. It enables users to search for mathematical formulae and expressions contained within indexed documents encoded in the MathML format. It is a scalable Java-based server application usable as a plug-in for Lucene. It is coupled with a web interface, WebMIaS. The MIaS system evaluated in this thesis is a first ever applied MIR system of a non-trivial scale.

https://mir.fi.muni.cz/mias/

Martin Líška

martin.liski@mail.muni.cz

Faculty of Informatics, Masaryk University

Advisor: doc. RNDr. Petr Sojka, Ph.D.

Evaluation

Mathematics retrieval is a new type of information retrieval. It focuses on searching structured mathematical data to simplify the knowledge management in specialized portals that provide this type of information. The evaluation of MIR (Mathematics Information Retrieval) has not been dealt with until very recently. Raising interest in MIR and its evaluation has so far resulted in two organized events in the fashion of already established evaluation practices in other types of IR.

Precision-recall results

Query	Results retrieved	Relevant docs retrieved	Precision	Recall
Formula 1	0	0	0	0
Formula 2	207	1	0.0048	1
Formula 3	1	1	1	1
Formula 5	0	0	0	0
Formula 6	1	1	1	1
Formula 7	1,045	1	0.00096	1
Full-text 1	1	1	1	1
Full-text 2	1	1	1	1
Full-text 3	1	1	1	1

Efficiency results

	Indexing times [min]		Formulae		Index	Av. query time [ms]	
Docs.	Wall clock	Total CPU	Input	Indexed	size [GB]	Core	Total
10,000	28.8	159.7	7,327,283	155,192,904	3,1	188.2	495.4
20,000	58	325.2	14,736,285	311,258,718			
30,000	85.1	474.2	21,877,907	463,281,808			
40,000	111.5	616.1	29,299,122	618,586,152			
50,000	146.1	821.8	36,801,976	779,487,671	15	182.5	484.1
60,000	177.1	999.4	44,179,606	938,538,811			
70,000	203.1	1,143.6	51,394,938	1,088,869,124			
80,000	231.5	1,306.6	58,633,240	1,241,466,398			
90,000	261.2	1,475.4	66,065,698	1,398,541,881			
100,000	291.8	1,649.0	73,428,180	1,556,839,999	30	199.1	601.9

References

Martin Líška, Petr Sojka and Michal Růžička. Similarity Search for Mathematics: Masaryk University Team at the NTCIR-10 Math Task. Proceedings of the 2013 NTCIR-10 Conference

Sojka, Petr - Líška, Martin. The Art of Mathematics Retrieval. In Matthew R. B. Hardy, Frank Wm. Tompa. Proceedings of the 2011 ACM Symposium on Document Engineering. Mountain View, CA, USA : ACM, 2011. od s. 57--60, 4 pages. ISBN 978-1-4503-0863-2.

Sojka, Petr - Líška, Martin. Indexing and Searching Mathematics in Digital Libraries: Architecture, Design and Scalability Issues. In James H. Davenport, William M. Farmer, Josef Urban, Florian Rabe. Intelligent Computer Mathematics Lecture Notes in Computer Science, 2011, Volume 6824/2011. Berlin / Heidelberg : Springer, 2011. p. 228--243, 15 pages. ISBN 978-3-642-22672-4.