

Formula Markup Algorithm for Search in Scientific Documents

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 IEEE. This paper discusses the approach to a semantic search of mathematical expressions which allows seeking formulae by textual definitions of their variables. The proposed method allows establishing relations between terms, variables, and formulae. This method was tested in two different search services with different preprocessing approaches. The first service uses full-text indexing and responds to a query, the second uses ontology concepts and establishes relations in advance. In both cases, the proposed method produced high-relevant results. The method is applicable for search in documents containing mathematical expressions regardless the preliminary markup approach.

<http://dx.doi.org/10.1109/EWDTs.2018.8524804>

Keywords

formulae markup, mathematical knowledge, ontology, semantic search

References

- [1] Birialtsev, E. V., Gusenkov, A. M., Zhibrik, O. N., "Some approaches to markup of scientific texts containing mathematical expressions", *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2014, vol. 156, no 4, pp. 133-148. (2014)
- [2] Birialtsev, E. V., Gusenkov, A. M., Zhibrik, O. N., "Search of mathematical expressions in scientific texts. Experimental evaluation of the relevance", *Proceedings of the Kazan School on computational and cognitive linguistics TEL-2014*, pp. 34-37. Tatarstan Academy of Sciences, Fen, Kazan (2014)
- [3] Gusenkov, A. M., "Intelligent search of complex objects in large data sets", *Russian Digital Libraries Journal*, vol. 19, no 1, pp. 3-39 (2016)
- [4] Elizarov, A. M., Lipachev, E. K., "Lobachevskii-DML: Object Approach to the Creation of a Digital Mathematical Library", *Proceedings of the XIX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2017)*, Moscow, Russia, October 10-13 (2017)
- [5] Elizarov, A., Kirillovich, A., Lipachev E., Nevzorova O., "Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management", *Communications in Computer and Information Science*, Springer, vol. 706, pp. 33-46 (2017)
- [6] Elizarov, A. M., Zhizhchenko A. B., Zhil'tsov, N. G., Kirillovich, A. V., Lipachev, E. K., "Mathematical knowledge ontologies and recommender system for collections of documents in Physics and Mathematics", *Computer Science*, vol. 93, issue 2, pp. 231-233. Springer, Berlin, Heidelberg (2016).
- [7] Nevzorova, O., Zhiltsov, N., Zaikin, D., Zhibrik, O., Kirillovich, A., Nevzorov, V., Birialtsev, E., "Bringing Math to LOD: a semantic publishing platform prototype for scientific collections in mathematics", *12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I. Lecture Notes in Computer Science*, vol. 8218, pp. 379-394. Springer, Berlin, Heidelberg (2013)

- [8] Nevzorova, O., Zhiltsov, N., Kirillovich, A., Lipachev, E., "OntoMathPRO Ontology: A Linked Data Hub for Mathematics", Knowledge Engineering and the Semantic Web Communications in Computer and Information Science Volume 468, 2014, pp. 105-119, <http://link.springer.com/chapter/10.1007/978-3-319-11716-4-9>, arXiv:1407.4833 (2014)
- [9] Birialtsev, E. V., Elizarov, A. M., Zhiltsov, N. G., Lipachev, E. K., Nevzorova, O. A., Solov'ev, V. D., "Methods for Analyzing Semantic Data of Electronic Collections in Mathematics", Automatic Documentation and Mathematical Linguistics, 2014, vol. 48, no. 2, pp. 81-85 (2014)
- [10] Google Scholar, <https://scholar.google.com>, last accessed 2018/06/01.
- [11] Microsoft Academic Search, <http://academic.research.microsoft.com>, last accessed 2018/06/01.
- [12] The Springer LaTeX Search, <http://latexsearch.com>, last accessed 2018/06/01.
- [13] (Uni)quation. Math expression search engine, <http://uniquation.com>, last accessed 2018/01/07.
- [14] Misutka, J., Galambos, L., "Extending Full Text Search Engine for Mathematical Content", Proceedings of DML, pp. 55-67 (2008)
- [15] Sojka P., Líška M., "Indexing and Searching Mathematics in Digital Libraries", Davenport J. H., Farmer W. M., Urban J., Rabe F. (eds), Intelligent Computer Mathematics. CICM 2011. Lecture Notes in Computer Science, vol 6824. Springer, Berlin, Heidelberg
- [16] The Wolfram Functions Site, <http://functions.wolfram.com>, last accessed 2018/06/01.
- [17] Apache Lucene, <http://lucene.apache.org>, last accessed 2018/06/01
- [18] Developing a 21st century global library for mathematics research, The National Academies Press, 131 p. (2014)
- [19] World Digital Mathematics Library (WDML), <https://www.mathunion.org/ceic/library/world-digitalmathematic-library-wdml>, last accessed 2018/06/01.
- [20] Olver, P. J., "The World Digital Mathematics Library: report of a panel discussion", Proceedings of the International Congress of Mathematicians, August 13-21, 2014, Seoul, Korea, vol. 1, pp. 773-785. Kyung Moon SA (2014)
- [21] GATE: general architecture for text engineering, <https://gate.ac.uk>, last accessed 2018/06/01.
- [22] Elizarov, A., Kirilovich, A., Lipachev, E., Nevzorova, O., "Mathematical Knowledge Management: Ontological Models and Digital Technology", CEUR Workshop Proceedings, 2016, vol. 1752, pp. 44-50, <http://ceurws.org/Vol-1752/paper08.pdf> (2016)
- [23] Finding Concepts in Mathematical Formulas alpha, <http://lobachevskii-dml.ru:8890/mathsearch>, last accessed 2018/06/01.
- [24] Birialtsev E., Gusenkov A., Zhibrik O., Gusenkova P., Palacheva Y. (2018) "Search in Collections of Mathematical Articles", Rocha Á., Adeli H., Reis L. P., Costanzo S. (eds), Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018. Advances in Intelligent Systems and Computing, vol 745. Springer, Cham
- [25] SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query>, last accessed 2018/06/01.