

Overview of Cross-Lingual Mathematical Information Retrieval at FIRE 2025

Ayushi Malik^{1,†}, Pankaj Dadure^{1,*} and Sahinur Rahman Laskar^{2,†}

¹UPES Dehradun, Uttarakhand, India

Abstract

This abstract provides a short overview of the first edition of the shared task on Cross-Lingual Mathematical Information Retrieval (CLMIR) organized at the 17th Forum for Information Retrieval Evaluation (FIRE 2025). A more detailed discussion of approaches used by the participating teams is available in the track overview paper. The CLMIR shared task at FIRE 2025 is designed to encourage participants to develop retrieval approaches that can process queries in one language and retrieve relevant results written in another, thereby bridging the accessibility gap between diverse user groups. Although the CLMIR shared task witnessed encouraging interest with 9 teams registering, only 6 teams submitted their results.

Keywords

Information Retrieval, Mathematical Information Retrieval, Cross-lingual Mathematical Information Retrieval, Digital Libraries, Search Engines,

1. Introduction

Natural Language Processing (NLP) is a transformative field that enables computers to understand, interpret, and generate human language [1][2]. With advancements in machine learning and deep learning, NLP applications have expanded from text summarization [3] and question answering [4] to complex tasks like cross-lingual understanding. A key domain where NLP shows practical significance is Information Retrieval (IR) [5], which focuses on searching, ranking, and retrieving relevant information from large repositories. Traditional IR systems relied on keyword or semantic-based retrieval [6][7] of text data, but the growth of digital information has introduced structured, semi-structured, and multimodal data [8][9] [10][11][12]. Among these, mathematical data poses particular challenges due to its symbolic, non-linear structure and representation in languages like LaTeX and MathML. General-purpose IR systems often fail to process mathematical queries effectively, leading to the emergence of Mathematical information retrieval (MIR) [13]. MIR deals with retrieving mathematical expressions, equations, and theorems while addressing issues of structural representation and integration with textual data. Despite advancements, most MIR systems like MathWebSearch [14][15][16], SimSearch [17][18][19], Mathcat [20][21] Tangent [22][23][24][25][26][27][28][29], and Wolfram Alpha [30][31] remain monolingual, predominantly focusing on English. This limits the accessibility of researchers and learners working in multiple languages. To bridge this gap, cross-lingual mathematical information retrieval (CLMIR) comes into the picture. CLMIR extends the scope of MIR by enabling users to query in one language and retrieve mathematical data in another language [32]. In CLMIR, a user may enter the query in one language (i.e., English) and retrieve relevant mathematical data written in another language (i.e., Hindi), thereby bridging the linguistic barriers in accessing mathematical knowledge. However, CLMIR presents a unique set of challenges. These include the scarcity of cross-lingual mathematical datasets, difficulties in aligning mathematical symbols with multilingual textual descriptions, handling structural and semantic variations of mathematical expressions across languages, and ensuring effective ranking of results in cross-lingual settings. Addressing these challenges is essential for developing

Forum for Information Retrieval Evaluation, December 17-20, 2025, Varanasi, India

*Corresponding author.

†These authors contributed equally.

✉ ayushimalik7715@gmail.com (A. Malik); krdadure@gmail.com (P. Dadure); sahinurlaskar.nits@gmail.com (S. R. Laskar)

🆔 0009-0004-5013-9753 (A. Malik); 0000-0002-8003-9671 (P. Dadure); 0000-0002-8413-2718 (S. R. Laskar)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

robust approaches that can support cross-lingual access to mathematical knowledge on a global scale [33].

2. Track Description

The CLMIR shared task at FIRE 2025 is designed to address the challenges of retrieving mathematical data across languages, with a specific focus on the English-Hindi cross-lingual environment. The task encourages the participants to develop retrieval approaches that can process the queries in one language and retrieve relevant results written in another. The primary objectives of CLMIR are to promote research in cross-lingual retrieval of mathematical data and to create a benchmark dataset for English-Hindi MIR, enabling a fair comparison of retrieval models. To encourage the development of novel retrieval approaches that combine symbolic mathematics understanding with cross-lingual natural language processing. This is the first shared task focused on cross-lingual retrieval of mathematical content, making CLMIR an important step toward cross-lingual accessibility in STEM education and research.

2.1. Use Case/s

The use case example demonstrates how the CLMIR model retrieves relevant mathematical content across languages. It shows a query in English and the corresponding relevant and irrelevant results in Hindi, highlighting the model's ability to match both mathematical expressions and textual meaning accurately.

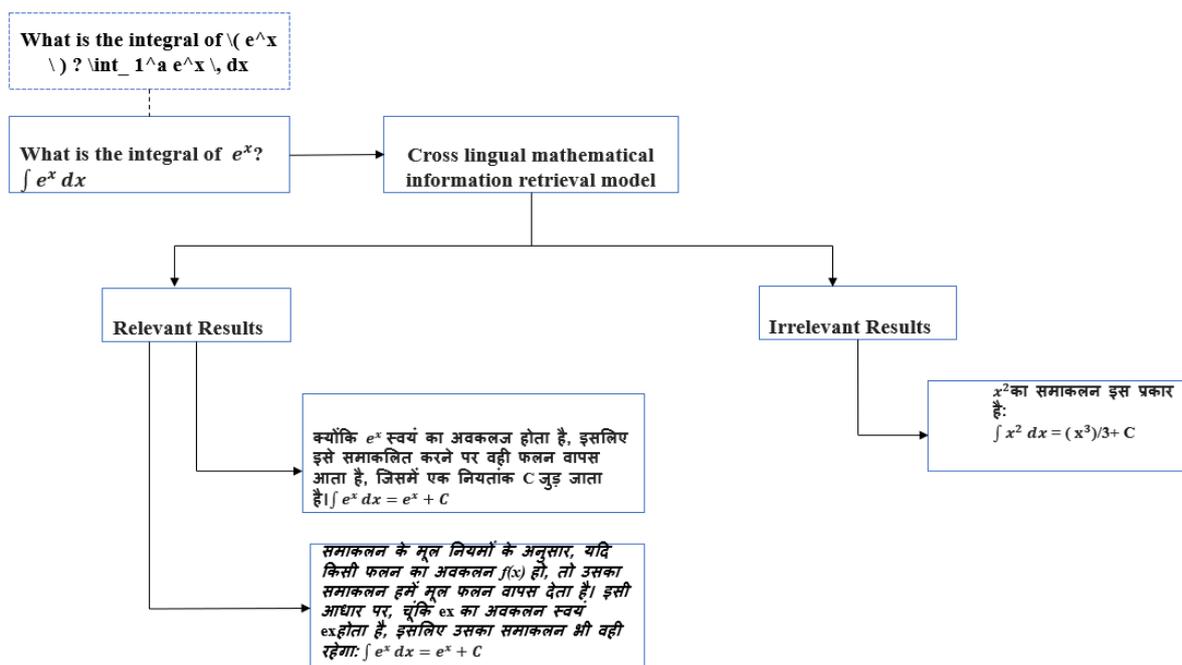


Figure 1: Workflow of the cross-lingual mathematical information retrieval model showing query input, retrieval process, and classification into relevant and irrelevant results.

2.2. Participation

A total of 9 teams, Archisha Dhyani (UPES), CJM (IISER Kolkata), Tends (NIT, Kurukshetra), IReL (IIT BHU Varanasi), j22 (IISER Kolkata), NLP Fusion (Mangalore University), MathQA_AUS (Assam University), DUCS_CLMIR (University of Delhi), Retriever (Dhirubhai Ambani University), have been registered for the task. Out of these 9 teams, only 6 teams, Archisha Dhyani, Tends, IReL, NLP Fusion,

DUCS_CLMIR, and Retriever have been able to submit their results. Among these 6 teams, only 4 Tends, IReL, NLP Fusion, and Retriever, have been successful in submitting the working notes.

3. Dataset and Evaluation

The dataset for the CLMIR task is curated from the Math Stack Exchange corpus of ARQMath-1 and contains approximately 39,862 instances. The dataset is formatted to include titles of scientific information and the body, which contains mathematical formulas expressed in LaTeX and supporting textual descriptions. The provided corpus is structured to maintain cross-lingual consistency between two languages, Hindi and English, ensuring that the mathematical content and contextual information are accurately aligned across both languages. The data covers content in Hindi, including mathematical equations, expressions, and related textual information. It is designed to support the development and evaluation of cross-lingual mathematical information retrieval approaches. To assess the performance of participants, 50 formulas and text-based queries will be provided. Participants are expected to submit a results file containing the relevant search results corresponding to each query. The performance of participants' approaches in the CLMIR 2025 task will be evaluated using three key metrics, which are Precision@10 (P@10), Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (nDCG).

4. System Description

The CLMIR shared task attracted diverse participation, with teams proposing a variety of approaches to address the challenges of retrieving mathematical data across languages. This section provides a detailed description of the baseline approach and the methodologies submitted by participating teams.

4.1. Baseline System (Organizer)

The baseline system was developed by initially translating the input query into Hindi using the Google Translate API to ensure consistency in language representation. Following translation, embeddings for both the query and the training data were generated using the sentence-transformers/paraphrase-multilingual-mpnet-base-v2 model. This model was selected due to its proven multilingual capability, robust performance in capturing semantic similarity across languages, and demonstrated effectiveness for Hindi text. Subsequently, cosine similarity was employed as the similarity estimation technique to measure the closeness between the query and training data embeddings. Based on these similarity scores, the top 50 most relevant search results were retrieved.

4.2. Tends

Team tends designed a cross-lingual retrieval pipeline that integrates machine translation, dense retrieval, and re-ranking to address the CLMIR task. The Hindi corpus from ARQMath-1 was first translated into English using the ai4bharat IndicTrans2 model, ensuring both queries and documents resided in the same language space for retrieval. Each document was represented as a composite of its title, body, and tags, which were encoded into contextualized token embeddings using CoBERTv2, a late interaction dense retrieval model optimized for fine-grained query document matching. The embeddings were stored in a Facebook AI Similarity Search (FAISS) index, enabling efficient approximate nearest neighbour search. At retrieval time, FAISS retrieved the top 100 candidate documents, which were further refined through a cross encoder (MiniLM-L6-v2) that jointly encoded each query document pair to generate precise semantic relevance scores. The final ranked list of the top 50 documents was returned as output. Since the corpus was fully translated into English, incoming English queries, including both text and mathematical expressions, were directly processed without any additional translation, making the system both efficient and semantically robust [34].

4.3. IReL

Team IReL suggested a hybrid retrieval system that integrates multilingual embeddings and symbolic formula similarity. Their pipeline first applied preprocessing to separate textual and mathematical components, transliterating Hindi text into Latin script for compatibility with transformer-based encoders. For semantic alignment, they used MiniLM embeddings in Run 1 and MPNet embeddings in Run 2, with candidate retrieval performed through FAISS indexing. Mathematical similarity was modeled using two strategies that are a lightweight token overlap method in Run 1 and a more expressive tree-based structural similarity in Run 2 [35].

4.4. Retriever

Team retriever developed a two-stage English-Hindi CLMIR system combining BM25-based sparse retrieval with LLM-based pointwise re-ranking. English text formula queries were translated into Hindi using GPT-4, and BM25 retrieved the top 100 candidate documents from ARQMath-1 corpora. These candidates were then re-ranked using instruction-tuned Gemma-2 models(4B and 12B) in a zero-shot setting, with binary relevance predictions converted into probabilities for ranking [36].

4.5. NLP_Fusion

Team NLP_Fusion, proposed a semantic retrieval system for the CLMIR 2025 shared task, focusing on English-Hindi cross-lingual mathematical information retrieval. Their approach utilized sentence-transformer/all MiniLM-L6-v2 embeddings to generate multilingual semantic representations of documents and queries, combined with FAISS for efficient vector-based similarity search. The Hindi documents from ARQMath-1 were preprocessed by concatenating titles, bodies, and tags, then split into manageable chunks to preserve contextual coherence. The system retrieved the top 50 most relevant chunks for each query, with additional post-processing to secure scores within the [0,1] range for consistency [37].

5. Experimental Results and Analysis

We evaluated all team submissions for the CLMIR 2025 shared task using three metrics, which are P@10, MAP, and nDCG. The results are shown in Table 1. Each team was allowed up to three runs, and the differences in performance reflect the design of their systems. The baseline system used Google Translate to convert queries into Hindi, generated multilingual embeddings with the MPNet model, and ranked documents using cosine similarity. Team Archisha Dhyani showed steady improvements, with Run 3 giving the best results, which suggests that their system became stronger after refinements. DUCS_CLMIR produced stable results across runs, but at a lower level compared to other teams, likely because their method did not add new improvements across submissions. IReL performed much better in Run 2, with a hybrid retrieval system that integrates multilingual embeddings and symbolic formula similarity. NLP fusion gave consistent results across all three runs, with the third run being the strongest. They utilized sentence-transformer/all MiniLM-L6-v2 embeddings to generate multilingual semantic representations of documents and queries, combined with FAISS for efficient vector-based similarity search. Retriever improved a lot in later runs, with Run 3 achieving one of the best MAP and nDCG scores overall. Their combination of BM25 retrieval with large language model re-ranking helped capture both word overlap and deeper meaning. Finally, Tends submitted only one run, but it still achieved strong results. Their system, which used translation, dense retrieval, and re-ranking, showed solid performance even without multiple submissions.

Table 1
Shared Task Results

Team	Run	P@10	MAP	nDCG
Archisha Dhyani	Run 1	0.0280	0.0712	0.1002
	Run 2	0.0440	0.0941	0.1224
	Run 3	0.0500	0.1034	0.1457
DUCS_CLMIR	Run 1	0.0060	0.0145	0.0563
	Run 2	0.0080	0.0159	0.0567
	Run 3	0.0100	0.0153	0.0517
IReL	Run 1	0.0000	0.0064	0.0352
	Run 2	0.1220	0.1650	0.3063
NLP Fusion	Run 1	0.0480	0.0972	0.1521
	Run 2	0.0460	0.0794	0.1966
	Run 3	0.0680	0.1090	0.2523
Retriever	Run 1	0.0000	0.0016	0.0089
	Run 2	0.1140	0.1755	0.3031
	Run 3	0.1280	0.2143	0.3264
Tends	Run 1	0.0800	0.1484	0.2202
Baseline System (Organizer)	Run 1	0.0850	0.1380	0.2450

6. Conclusion

The CLMIR shared task at FIRE 2025 represents the first attempt to systematically evaluate the retrieval of mathematical data across languages, with a focus on the English-Hindi pair. By providing a large-scale cross-lingual dataset, well-defined queries, and a clear evaluation framework, this task establishes a foundation for research in a less-explored area of cross-lingual mathematical information retrieval. In future editions, we aim to expand the dataset to cover additional language pairs and incorporate more diverse query types (including multi-formula and contextual queries). Furthermore, by releasing resources such as baseline models, smaller benchmark datasets, and multilingual extensions, we intend to make the task more accessible and encourage participation from a broader range of researchers. Through these steps, CLMIR seeks to promote innovation in cross-lingual mathematical information retrieval, promote access to STEM resources, and build a long-term research community dedicated to advancing cross-lingual access to mathematical knowledge.

Acknowledgement

This research work is supported under the Visvesvaraya PhD Scheme for Electronics and IT, implemented by the Ministry of Electronics and Information Technology (MeitY), Government of India. The authors also thank UPES Dehradun for providing the necessary support and research infrastructure.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] F. Enríquez, F. L. Cruz, F. J. Ortega, C. G. Vallejo, J. A. Troyano, A comparative study of classifier combination applied to nlp tasks, *Information Fusion* 14 (2013) 255–267. URL: <https://www.>

sciencedirect.com/science/article/pii/S1566253512000425. doi:<https://doi.org/10.1016/j.inffus.2012.05.001>.

- [2] J. C. Hsu, M. Wu, C. Kim, B. Vora, Y. T. Lien, A. Jindal, K. Yoshida, S. Kawakatsu, J. Gore, J. Y. Jin, et al., Applications of advanced natural language processing for clinical pharmacology, *Clinical Pharmacology & Therapeutics* 115 (2024) 786–794.
- [3] S. Panja, Information retrieval systems in healthcare: Understanding medical data through text analysis, in: *Transformative Approaches to Patient Literacy and Healthcare Innovation*, IGI Global, 2024, pp. 180–200.
- [4] R. Zanibbi, B. Mansouri, A. Agarwal, et al., Mathematical information retrieval: Search and question answering, *Foundations and Trends® in Information Retrieval* 19 (2025) 1–190.
- [5] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, G. Pappas, Comparison of pubmed, scopus, web of science, and google scholar: strengths and weaknesses, *The FASEB journal* 22 (2008) 338–342.
- [6] R. Miner, R. Munavalli, An approach to mathematical search through query formulation and data normalization, in: *International Conference on Mathematical Knowledge Management*, Springer, 2007, pp. 342–355.
- [7] P. Sojka, M. Liska, The art of mathematics retrieval, in: *Proceedings of the 11th ACM symposium on Document engineering*, 2011, pp. 57–60.
- [8] Y. Li, J. Ma, Y. Zhang, Image retrieval from remote sensing big data: A survey, *Information Fusion* 67 (2021) 94–115. URL: <https://www.sciencedirect.com/science/article/pii/S1566253520303778>. doi:<https://doi.org/10.1016/j.inffus.2020.10.008>.
- [9] K. T. Ahmed, S. Ummesafi, A. Iqbal, Content based image retrieval using image features information fusion, *Information Fusion* 51 (2019) 76–99. URL: <https://www.sciencedirect.com/science/article/pii/S1566253517307893>. doi:<https://doi.org/10.1016/j.inffus.2018.11.004>.
- [10] J. Wang, J. X. Huang, X. Tu, J. Wang, A. J. Huang, M. T. R. Laskar, A. Bhuiyan, Utilizing bert for information retrieval: Survey, applications, resources, and challenges, *ACM Computing Surveys* 56 (2024) 1–33.
- [11] S. Wu, Y. Li, K. Zhu, G. Zhang, Y. Liang, K. Ma, C. Xiao, H. Zhang, B. Yang, W. Chen, et al., Scimmir: Benchmarking scientific multi-modal information retrieval, *arXiv preprint arXiv:2401.13478* (2024).
- [12] S. Wang, H. Kim, N. Janaka, K. Yue, H.-L. Nguyen, S. Zhao, H. Liu, K.-D. Le, " what's this?": Understanding user interaction behaviour with multimodal input information retrieval system, in: *Adjunct Proceedings of the 26th International Conference on Mobile Human-Computer Interaction*, 2024, pp. 1–7.
- [13] F. Guidi, C. S. Coen, A survey on retrieval of mathematical knowledge, 2015. URL: <https://arxiv.org/abs/1505.06646>. arXiv:1505.06646.
- [14] M. Kohlhase, C.-C. Prodescu, Mathwebsearch at ntcir-10., in: *NTCIR, Citeseer*, 2013.
- [15] R. Hambasan, M. Kohlhase, C.-C. Prodescu, Mathwebsearch at ntcir-11., in: *NTCIR, volume 11*, 2014, pp. 114–119.
- [16] M. Kohlhase, B. A. Matican, C.-C. Prodescu, Mathwebsearch 0.5: Scaling an open formula search engine, in: *Intelligent Computer Mathematics: 11th International Conference, AISC 2012, 19th Symposium, Calculemus 2012, 5th International Workshop, DML 2012, 11th International Conference, MKM 2012, Systems and Projects, Held as Part of CICM 2012, Bremen, Germany, July 8-13, 2012. Proceedings 5*, Springer, 2012, pp. 342–357.
- [17] S. Kamali, *Querying large collections of semistructured data* (2013).
- [18] S. Kamali, F. W. Tompa, Structural similarity search for mathematics retrieval, in: *International Conference on Intelligent Computer Mathematics*, Springer, 2013, pp. 246–262.
- [19] S. Kamali, F. W. Tompa, Retrieving documents with mathematical content, in: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, pp. 353–362.
- [20] G. Topić, G. Y. Kristianto, M.-Q. Nghiem, A. Aizawa, The mcat math retrieval system for ntcir-10 math track, in: *Proceedings of 10th NTCIR Conference, Tokyo, Japan, 2013*, pp. 680–685.
- [21] G. Y. Kristianto, G. Topic, A. Aizawa, Mcat math retrieval system for ntcir-12 mathir task., in: *NTCIR, 2016*.

- [22] K. D. V. Wangari, R. Zanibbi, A. Agarwal, Discovering real-world use cases for a multimodal math search interface, in: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014, pp. 947–950.
- [23] D. Stalnakar, R. Zanibbi, Math expression retrieval using an inverted index over symbol pairs, in: Document recognition and retrieval XXII, volume 9402, SPIE, 2015, pp. 34–45.
- [24] R. Zanibbi, K. Davila, A. Kane, F. Tompa, The tangent search engine: Improved similarity metrics and scalability for math formula search, arXiv preprint arXiv:1507.06235 (2015).
- [25] R. Zanibbi, K. Davila, A. Kane, F. W. Tompa, Multi-stage math formula search: Using appearance-based similarity metrics at scale, in: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, 2016, pp. 145–154.
- [26] K. Davila, R. Zanibbi, Layout and semantics: Combining representations for mathematical formula search, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 1165–1168.
- [27] D. Fraser, A. Kane, F. W. Tompa, Choosing math features for bm25 ranking with tangent-l, in: Proceedings of the ACM Symposium on Document Engineering 2018, 2018, pp. 1–10.
- [28] W. Zhong, R. Zanibbi, Structural similarity search for formulas using leaf-root paths in operator subtrees, in: Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14–18, 2019, Proceedings, Part I 41, Springer, 2019, pp. 116–129.
- [29] B. Mansouri, S. Rohatgi, D. W. Oard, J. Wu, C. L. Giles, R. Zanibbi, Tangent-cft: An embedding model for mathematical formulas, in: Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval, 2019, pp. 11–18.
- [30] V. E. Dimiceli, A. S. Lang, L. Locke, Teaching calculus with wolfram|alpha, International Journal of Mathematical Education in Science and Technology 41 (2010) 1061–1071.
- [31] W. N. S. Wan Mohd Rosly, S. S. Syed Abdullah, F. N. Ahmad Shukri, The uses of wolfram alpha in mathematics, Teaching and Learning in Higher Education (TLHE) 1 (2020) 96–103.
- [32] J. Gore, J. Polletta, B. Mansouri, Crossmath: Towards cross-lingual math information retrieval, in: Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval, 2024, pp. 101–105.
- [33] A. Malik, P. Dadure, S. R. Laskar, A review of mathematical information retrieval: Bridging symbolic representation and intelligent retrieval, Archives of Computational Methods in Engineering (2025) 1–35.
- [34] A. Sur, V. Shukla, A. Rai, L. Garg, Enhancing multilingual mathematical document retrieval: A hindi to english translation and colbert based approach, in: Proceedings of the Forum for Information Retrieval Evaluation (FIRE), ACM, 2025.
- [35] K. Tewari, S. Chanda, R. Tripathi, S. Pal, Miracle: Multilingual information retrieval with cross-lingual embeddings for mathematical expressions, in: Proceedings of the Forum for Information Retrieval Evaluation (FIRE 2025), To appear, 2025. Accepted for publication.
- [36] K. Kachhadiya, P. Patel, Cross-lingual mathematical information retrieval with bm25 and llm-based pointwise re-ranking, in: Proceedings of the Forum for Information Retrieval Evaluation (FIRE 2025), To appear, 2025. Accepted for publication.
- [37] K. S. Coelho, A. Hegde, M. Z. Taljeh, A. Ahmad, Math beyond language barriers: Retrieving mathematical content using sentence transformers, in: Proceedings of the Forum for Information Retrieval Evaluation (FIRE 2025), To appear, 2025. Accepted for publication.