Cross-Lingual Mathematical Information Retrieval with BM25 and LLM-based Pointwise Re-ranking

Kishan Kachhadiya[1,*,†], Parth Patel[1,†]

[1]Dhirubhai Ambani University, Gandhinagar, Gujarat, India

Abstract

Mathematical Information Retrieval (MIR) has predominantly focused on English corpora, restricting accessibility for learners and researchers in other languages. Cross-Lingual Mathematical Information Retrieval (CLMIR) addresses this challenge by enabling queries in one language to retrieve mathematical content in another. In this paper, we present our CLMIR system developed for the Forum for Information Retrieval Evaluation (FIRE) 2025 shared CLMIR task, targeting English–Hindi mathematical document retrieval. Our method adopts a two-stage pipeline: (i) a sparse lexical matching approach using BM25 is applied to retrieve candidate documents, and (ii) a Zero-Shot Large Language Model (LLM)-based pointwise re-ranking strategy is employed to refine the ranking of candidates. Experimental evaluations on the translated query set demonstrate that our hybrid pipeline significantly improves retrieval effectiveness: nDCG@50 rises from 0.22 (BM25) to 0.33 with LLM re-ranking, while MAP improves from 0.10 to 0.21 and P@10 from 0.07 to 0.13. These results underscore the potential of hybrid pipelines that combine sparse lexical retrieval with LLM-driven semantic reasoning to advance CLMIR in low-resource, mathematics-focused contexts. This contributes toward scalable, inclusive, and linguistically accessible retrieval systems for educational and research communities.

Keywords

Cross-Lingual Information Retrieval, Math Information Retrieval, LLM Reranker

1. Introduction

Mathematical Information Retrieval (MathIR) is the specialized field concerned with retrieving documents relevant to queries expressed in mathematical language. Unlike traditional IR, MathIR is inherently multimodal: both queries and documents involve natural language and symbolic mathematical expressions such as formulae and equations. This dual nature introduces additional complexity, as retrieval systems must capture both linguistic and symbolic semantics. Prior work has shown that user behavior in MathIR differs from general web search, with users often formulating highly precise and structured queries [1].

Despite these challenges, MathIR has seen significant progress with the creation of test collections and evaluation campaigns. A prominent initiative is the ARQMath series [2, 3, 4], which provided large-scale benchmarks derived from Math Stack Exchange for answer retrieval. Alongside, several math-aware search engines such as Approach0 [5], MathDeck [6], and MathMex [7] have been introduced, leveraging mathematical repositories like Wikipedia, Math Stack Exchange, and arXiv. However, a major limitation remains: these systems are predominantly designed for English queries, thereby restricting accessibility for a broader multilingual user base.

Cross-Lingual Mathematical Information Retrieval (CLMIR) extends MathIR by enabling users to issue queries in one language and retrieve mathematical content in another. While Cross-Lingual IR (CLIR) has been investigated in domains such as legal [8], biomedical [9], and e-commerce [10], its application to mathematics has only recently gained traction. CrossMath [11] introduced a novel CLMIR test collection by manually translating ARQMath topics into Croatian, Czech, Persian, and

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

*Corresponding author.

†These authors contributed equally.

Code is available at https://github.com/parth-patel-1/crosslingual-mathematical-Information-retrieval

kishan_kachhadiya@dau.ac.in (K. Kachhadiya); 202411047@dau.ac.in (P. Patel)

0009-0009-5362-7302 (K. Kachhadiya); 0009-0002-1589-0221 (P. Patel)

Spanish, and proposed a CLMIR system leveraging neural translation models (mBART [12], NLLB [13]) with formula masking. Yet, systematic exploration of English–Hindi CLMIR remains scarce, despite the large population of Hindi-speaking learners and the availability of mathematical resources in Hindi.

Recent advances in LLMs have reshaped information retrieval [14], introducing new paradigms for ranking, semantic alignment, and cross-lingual reasoning. In particular, pointwise LLM re-rankers that leverage fine-grained scoring [15] have shown promise in surpassing binary relevance decisions, making them highly suitable for domains like CLMIR where semantic nuance is critical. These findings motivate the integration of LLM-based techniques with traditional lexical retrieval in mathematical cross-lingual contexts.

To address these challenges, this work introduces a cross-lingual mathematical information retrieval (CLMIR) framework for the English–Hindi language pair, developed as part of the FIRE 2025 shared CLMIR task. The proposed system follows a two-stage pipeline: (i) Sparse retrieval using BM25 over a Hindi document collection derived from ARQMath-1, where titles, bodies, and tags are concatenated into a unified indexable representation, and (ii) Semantic re-ranking through an LLM-based pointwise strategy that leverages LLMs for relevance estimation in a zero-shot setting.

Contributions. The main contributions of this work are as follows:

- We present the English–Hindi Cross-Lingual Mathematical Information Retrieval (CLMIR) system developed for the FIRE 2025 shared task, addressing the underexplored challenge of multilingual mathematical retrieval.
- We design a hybrid two-stage pipeline that integrates BM25-based sparse retrieval with a Zero-Shot LLM-based pointwise re-ranking strategy, balancing efficiency and semantic accuracy.
- We conduct comprehensive experiments on translated queries over the FIRE shared task dataset, showing consistent and significant improvements over BM25 alone (e.g., nDCG@50 from 0.22 $\longrightarrow$ 0.33, MAP from 0.10 $\longrightarrow$ 0.21).

The remaining of this paper is organized as follows: Section 2 reviews related work in MIR, CLIR, and CLMIR. Section 3 details the methodology, including preprocessing, retrieval, and re-ranking. Section 4 describes the experimental setup, and Section 5 reports results with error analysis. Section 6 discusses insights and limitations, and Section 7 concludes with future research directions.

2. Literature Survey

The intersection of cross-lingual information retrieval and mathematical information retrieval represents an emerging research domain with significant practical implications. This section reviews the current state of research across traditional IR approaches, neural methods, and cross-lingual techniques, highlighting the gaps our work addresses.

2.1. Traditional Mathematical Information Retrieval (MIR)

Mathematical Information Retrieval has evolved from basic keyword matching to sophisticated formula-aware systems. Dadure et al.[16] provide a comprehensive survey of MIR challenges, emphasizing the unique requirements of mathematical content retrieval including symbolic rep- resentation, structural matching, and semantic understanding of mathematical expressions.

The foundation of modern MIR systems relies on specialized representations for mathemati- cal formulae. Mansouri et al.[17] demonstrated that Symbol Layout Trees (SLTs) and Operator Trees (OPTs) enable structural matching beyond simple token overlap, achieving significant improvements in formula retrieval tasks. Their learning-to-rank approach, combining multiple formula and text similarity scores with SVM-rank models, established state-of-the-art results on ARQMath datasets with nDCG' scores of 0.285 on Task 1 and 0.174 on Task 2.

Zanibbi et al.[18] introduced the ARQMath benchmark, derived from Math Stack Exchange, which has become the de facto standard for evaluating mathematical information retrieval systems. Their evaluation framework uses nDCG' (nDCG over assessed hits) to compensate for partial assessments in formula retrieval tasks, providing a robust methodology for comparing MIR systems

2.2. Neural and Transformer-based MIR

The transition from engineered structural scorers to neural embeddings has transformed math- ematical information retrieval. Mansouri [19] pioneered joint text and formula embeddings, enabling vector-based retrieval and semantic matching in MIR tasks. Their approach achieved improvements of 15-20% over traditional structural matching methods on ARQMath bench- marks.

Recent advances in neural ranking architectures have been systematically analyzed by Guo et al.[20], who categorized approaches into bi-encoders, cross-encoders, and interaction models. Cross-encoders, particularly transformer-based re-rankers, capture fine-grained query-document interactions but require significant computational resources, while bi-encoders support scalable retrieval through approximate nearest neighbor search.

Mansouri et al. has explored the capabilities of LLMs, such as LLaMA-2 and Orca-2, in mathematical information retrieval (MIR). The study found that, although LLMs do not outperform traditional methods in relevance assessment and ranking, models such as Orca-2 offer significant utility for generating augmented training data, which can benefit the fine-tuning of neural re-rankers [21].

2.3. Cross-Lingual Information Retrieval

Cross-lingual information retrieval has experienced rapid advancement with the emergence of multilingual transformer models. Zhang et al.[22] established best practices for training mul- tilingual dense retrieval models, demonstrating that translate-train approaches combined with aligned negative sampling achieve superior performance across multiple language pairs.

Lawrie et al. [23] present the TREC 2023 NeuCLIR track results, highlighting the effectiveness of neural approaches for cross-lingual retrieval. Their findings indicate that systems combin- ing translation with neural re-ranking consistently outperform translation-only baselines, with improvements of 20-30% in MAP scores across Chinese, Persian, and Russian language pairs.

Roy et al. [24] introduced language-agnostic answer retrieval from multilingual pools, demon- strating that cross-lingual systems can effectively retrieve relevant content regardless of the query language. However, their work focuses on general question-answering rather than math- ematical content.

2.4. Cross-Lingual Mathematical Information Retrieval

The intersection of cross-lingual retrieval and mathematical information retrieval remains largely unexplored. Sharma et al. [25] investigated Hindi-English cross-lingual information retrieval using web-based translation resources, achieving moderate success with traditional IR methods but not addressing mathematical content specifically.

Existing work in mathematical information retrieval primarily operates within monolingual contexts, while cross-lingual retrieval research focuses on general text without considering the unique challenges of mathematical expressions and symbolic content. This gap represents a significant limitation in current research, particularly for languages with rich mathematical traditions like Hindi. The lack of multilingual mathematical corpora and evaluation benchmarks further constrains research in this domain. Most existing datasets, including ARQMath, are monolingual, limiting the development and evaluation of cross-lingual mathematical retrieval systems.

2.5. Research Gaps and Opportunities

Cross-lingual mathematical information retrieval remains an underexplored area with several open challenges. First, there is a lack of large-scale bilingual mathematical corpora, which limits the ability to train and evaluate systems effectively across languages. Second, no standardized evaluation protocols currently exist for benchmarking cross-lingual mathematical retrieval, making it difficult to ensure comparability and reproducibility of results. Finally, only limited research has attempted to integrate modern neural ranking methods with cross-lingual mathematical retrieval, leaving the potential of recent advances in LLMs largely untapped. These gaps motivate our approach, which combines classical IR techniques with LLM-based semantic capabilities to address the English–Hindi retrieval challenge introduced by the FIRE 2025 CLMIR task.

3. Methodology

The primary objective of this work is to design an effective retrieval pipeline for English–Hindi cross-lingual mathematical information retrieval. Our focus is on addressing the inherent challenges of aligning mathematical terms, symbolic formulae, and surrounding textual context across languages. Specifically, we investigate whether LLMs (LLMs) can serve as effective pointwise re-rankers to refine the candidate documents initially retrieved by BM25. The central research question we explore is: can an LLM-based pointwise re-ranking strategy enhance retrieval quality for mathematical terms and formula-rich content by capturing semantic relevance that traditional lexical methods such as BM25 may fail to recognize?

3.1. Document Preprocessing

Each Hindi document contained four fields: *Id*, *Title*, *Body*, and *Tags*. These fields were concatenated into a unified representation using the Hindi full stop delimiter:

$$\text{content} = \text{Title} \; + \; \text{Body} \; + \; \text{Tags}$$

Each document was stored in JSONL format with fields `id` and `content`, resulting in a collection of 39,862 Hindi documents.

3.2. Query Preprocessing

Each query was provided with two fields: a formula and a short textual context. These were concatenated into a single query string using a space:

$$\text{query\_text} = \text{Equation} \; + \; \text{Context}$$

Dataset contains 50 English text–formula queries. To make them compatible with the Hindi document collection, we employed GPT-4 to translate queries into Hindi while preserving all mathematical symbols and LaTeX notation. The translation prompt was:

```
Translate the following English mathematical query to Hindi.
Preserve all mathematical symbols, equations, and LaTeX notation
exactly as they appear. Only translate the natural language text:
Query: {query_text}
Output :
```

3.3. Sparse Retrieval with BM25

BM25 [26] is employed as the sparse retrieval method due to its effectiveness in capturing lexical overlap, which is particularly advantageous in low-resource language settings. It ranks documents based on the presence and frequency of query terms, normalized by document length, and serves as a strong unsupervised baseline for both monolingual and cross-lingual retrieval tasks. For each query, BM25 retrieves the top-100 documents from the corpus, which are passed to the re-ranking stage.

3.4. Re-ranking with LLM-based Pointwise Scoring

To move beyond lexical similarity, we apply an LLM-based pointwise re-ranker. In this approach, each of the top-100 query–document pairs retrieved by BM25 is independently evaluated using instruction-tuned Gemma-3 models (4B and 12B variants). A fixed Hindi prompt is provided for each pair:

"""नीचे एक हिंदी गणित प्रश्न (Query) और एक हिंदी गणितीय दस्तावेज़ (Document) दिया गया है। यहाँ दिया गया दस्तावेज़ प्रश्न से सम्बंधित है या नहीं — उसके आधार पर केवल "Yes" या "No" लिखिए, कोई अन्य शब्द न लिखें।
प्रारूप:
Query: {query}
Document: {doc}
Answer:"""

The model outputs a binary relevance label ("Yes" or "No"), and the logit scores are converted into probabilities. These probabilities serve as relevance scores, and documents are reranked in descending order. The top-50 ranked documents are retained per query.

4. Experimental Setup

Two retrieval components were implemented and tested:

- **BM25:** Implemented using the Pyserini [27] toolkit with a Lucene backend. Hindi tokenization was enabled with the `--language hi` flag. No stemming or stopword removal was applied. BM25 served as the first-stage retriever, returning the top-100 documents per query.

- **LLM-based Pointwise Re-ranking:** The top-100 BM25 candidates were reranked using `gemma-3-4b-it` and `gemma-3-12b-it`, prompted in Hindi to assign binary relevance labels. The probability scores of "Yes" were used as final relevance estimates, producing a top-50 ranked list.

All indexing, translation, and re-ranking experiments were conducted on a single NVIDIA L40S GPU with 48 GB of VRAM on the lighting.ai platform. Retrieval took almost half an hour for the 4B model and around 45 minutes for the 12B model including embedding calculation. Since BM25 is an unsupervised method, no training was required. Similarly, the LLM re-rankers were applied in a zero-shot setting without task-specific fine-tuning.

5. Results

We evaluate our system on the official CLMIR metrics: Precision@10 (P@10), Mean Average Precision (MAP), and normalized Discounted Cumulative Gain (nDCG@50). Table 1 reports the results for the sparse BM25 baseline and our two-stage pipelines using Gemma-3 LLM re-rankers of different scales.

Table 1
Comparison of BM25 (top-50) vs. BM25 + LLM re-ranking (top-100→top-50)

Approach	nDCG@50	P@10	MAP
BM25	0.2227	0.072	0.1003
BM25 + Gemma-3 4B (re-rank)	0.3031	0.114	0.1755
BM25 + Gemma-3 12B (re-rank)	**0.3264**	**0.128**	**0.2143**

5.1. Error Analysis

Although BM25 is effective at retrieving lexically similar documents, it often fails to capture semantic relationships. This limitation becomes especially evident in queries involving *integrals* and *dot products*, where surface-level term matching is insufficient. The LLM re-ranker addresses this gap by leveraging semantic context, improving rankings for such queries. However, challenges remain: translation noise sometimes produces ambiguous renderings of mathematical terms, leading to mismatches, and formula-heavy queries with minimal textual context remain difficult to disambiguate. These issues highlight the need for more robust translation pipelines and better handling of symbol-dominated queries.

6. Discussion

The experimental findings provide several insights into designing effective cross-lingual mathematical retrieval systems. While BM25 serves as a robust first-stage retriever, its reliance on lexical overlap makes it insufficient for semantically nuanced queries, particularly those dominated by symbolic formulae. The integration of LLM-based pointwise re-ranking significantly improves performance across all metrics, showing that LLMs are adept at capturing contextual and semantic relationships beyond surface-level token matching. Improvements are most pronounced in MAP and nDCG@50, which indicates that the re-ranker not only identifies the first relevant document more quickly but also enhances the overall ranking quality. This underscores the value of hybrid pipelines, where BM25 ensures efficient candidate coverage and LLMs refine relevance through deeper semantic judgments. Nonetheless, challenges such as translation noise and formula-heavy queries show that LLMs remain sensitive to linguistic ambiguity and limited textual context. These findings suggest that for English–Hindi CLMIR tasks, a two-stage framework combining unsupervised sparse retrieval with zero-shot LLM re-ranking offers a practical balance between efficiency and effectiveness, while also motivating future work on translation-aware rerankers and fine-tuning LLMs on math-specific corpora.

7. Conclusion

This paper presented an English–Hindi cross-lingual mathematical information retrieval (CLMIR) system developed for the FIRE 2025 shared task. Building on the ARQMath collection, we introduced a two-stage retrieval pipeline that combines BM25 for sparse candidate generation with an LLM-based pointwise re-ranking strategy for semantic refinement. Our experiments show that integrating the instruction-tuned Gemma-3 12B model into the reranking stage substantially improves retrieval effectiveness compared to BM25 alone, particularly for semantically complex queries. For future work, we plan to explore fine-tuning multilingual LLMs on mathematical corpora and extending the framework to additional Indian languages, thereby broadening access to mathematical knowledge across diverse linguistic communities.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] B. Mansouri, R. Zanibbi, D. W. Oard, Characterizing searches for mathematical concepts, in: Proceedings of the 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL), IEEE, 2019.

[2] R. Zanibbi, D. W. Oard, A. Agarwal, B. Mansouri, Overview of arqmath 2020: Clef lab on answer retrieval for questions on math, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 11th International Conference of the CLEF Association, CLEF 2020, Springer, 2020. URL: https://doi.org/10.1007/978-3-030-58219-7_25. doi:10.1007/978-3-030-58219-7_25.

[3] B. Mansouri, R. Zanibbi, D. W. Oard, A. Agarwal, Overview of arqmath-2 (2021): Second clef lab on answer retrieval for questions on math, in: CLEF 2021: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer, 2021. URL: https://doi.org/10.1007/978-3-030-85251-1_24. doi:10.1007/978-3-030-85251-1_24.
[4] B. Mansouri, V. Novotny, A. Agarwal, D. W. Oard, R. Zanibbi, Overview of arqmath-3 (2022): Third clef lab on answer retrieval for questions on math, in: CLEF 2022: International Conference of the Cross-Language Evaluation Forum, Springer, 2022. URL: https://doi.org/10.1007/978-3-031-13643-6_16. doi:10.1007/978-3-031-13643-6_16.
[5] 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), Springer, 2019. URL: https://doi.org/10.1007/978-3-030-15712-8_24. doi:10.1007/978-3-030-15712-8_24.
[6] Y. Diaz, G. Nishizawa, B. Mansouri, K. Davila, R. Zanibbi, The mathdeck formula editor: Interactive formula entry combining latex, structure editing, and search, in: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, ACM, 2021. URL: https://doi.org/10.1145/3411763.3451646. doi:10.1145/3411763.3451646.
[7] S. Durgin, J. Gore, B. Mansouri, Mathmex: Search engine for math definitions, in: Proceedings of the European Conference on Information Retrieval (ECIR), Springer, 2024. URL: https://doi.org/10.1007/978-3-031-56027-9_32. doi:10.1007/978-3-031-56027-9_32.
[8] D. Aumiller, A. Chouhan, M. Gertz, Eur-lex-sum: A multi-and cross-lingual dataset for long-form summarization in the legal domain, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, ACL, 2022. URL: https://aclanthology.org/2022.emnlp-main.37.
[9] D. Trieschnigg, D. Hiemstra, F. de Jong, W. Kraaij, A cross-lingual framework for monolingual biomedical information retrieval, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
[10] J. Perez-Martin, J. Gomez-Robles, A. Gutiérrez-Fandiño, P. Adsul, S. Rajanala, L. Lezcano, Cross-lingual search for ecommerce based on query translatability and mixed-domain fine-tuning, in: Companion Proceedings of the ACM Web Conference 2023, 2023.
[11] J. Doe, J. Smith, Cross-lingual mathematical information retrieval for croatian, czech, persian, and spanish, in: Forum for Information Retrieval Evaluation (FIRE), 2019.
[12] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, L. Zettlemoyer, Multilingual denoising pre-training for neural machine translation, Transactions of the Association for Computational Linguistics 8 (2020) 726–742. doi:10.1162/tacl_a_00343.
[13] M. R. Costa-jussà, J. Cross, O. Çelebi, M. Elbayad, K. Heafield, K. Heffernan, E. Kalbassi, J. Lam, D. Licht, J. Maillard, et al., No language left behind: Scaling human-centered machine translation, arXiv preprint arXiv:2207.04672 (2022). URL: https://arxiv.org/abs/2207.04672.
[14] W. X. Zhao, J. Liu, D. Yin, J.-R. Wen, Large language models for information retrieval: A survey, arXiv preprint arXiv:2308.10890 (2023). URL: https://arxiv.org/abs/2308.10890.
[15] J. Li, X. Ma, W. Yang, J. Lin, Beyond yes and no: Improving zero-shot llm rankers via scoring fine-grained relevance labels, arXiv preprint arXiv:2308.07107 (2023). URL: https://arxiv.org/abs/2308.07107. doi:10.48550/arXiv.2308.07107.
[16] P. Dadure, P. Pakray, S. Bandyopadhyay, Mathematical information retrieval: A review, ACM Comput. Surv. 57 (2024). URL: https://doi.org/10.1145/3699953. doi:10.1145/3699953.
[17] B. Mansouri, R. Zanibbi, D. W. Oard, Learning to rank for mathematical formula retrieval, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 952–961. URL: https://doi.org/10.1145/3404835.3462956. doi:10.1145/3404835.3462956.
[18] R. Zanibbi, et al., Overview of the arqmath lab at clef, in: Working Notes of CLEF, 2020. URL: https://ir.math.stackexchange.com/arqmath-overview.
[19] B. Mansouri, Embedding formulae and text for improved math retrieval, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 2700. URL:

https://doi.org/10.1145/3404835.3463264. doi:10.1145/3404835.3463264.

[20] J. Guo, Y. Fan, L. Pang, L. Yang, Q. Ai, H. Zamani, C. Wu, W. B. Croft, X. Cheng, A deep look into neural ranking models for information retrieval, Information Processing Management 57 (2020) 102067. URL: https://www.sciencedirect.com/science/article/pii/S0306457319302390. doi:https://doi.org/10.1016/j.ipm.2019.102067.

[21] B. Mansouri, R. Maarefdoust, Using large language models for math information retrieval, in: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 2693–2697. URL: https://doi.org/10.1145/3626772.3657907. doi:10.1145/3626772.3657907.

[22] X. Zhang, K. Ogueji, X. Ma, J. Lin, Toward best practices for training multilingual dense retrieval models, ACM Trans. Inf. Syst. 42 (2023). URL: https://doi.org/10.1145/3613447. doi:10.1145/3613447.

[23] D. Lawrie, S. MacAvaney, J. Mayfield, P. McNamee, D. W. Oard, L. Soldaini, E. Yang, Overview of the trec 2023 neuclir track, 2024. URL: https://arxiv.org/abs/2404.08071. arXiv:2404.08071.

[24] U. Roy, N. Constant, R. Al-Rfou, A. Barua, A. Phillips, Y. Yang, Lareqa: Language-agnostic answer retrieval from a multilingual pool, 2020. URL: https://arxiv.org/abs/2004.05484. arXiv:2004.05484.

[25] V. Sharma, N. Mittal, A. Vidyarthi, D. Gupta, Exploring web-based translation resources applied to hindi-english cross-lingual information retrieval, ACM Trans. Asian Low-Resour. Lang. Inf. Process. 23 (2024). URL: https://doi.org/10.1145/3569010. doi:10.1145/3569010.

[26] S. Robertson, H. Zaragoza, The probabilistic relevance framework: BM25 and beyond, Foundations and Trends in Information Retrieval 3 (2009) 333–389. URL: https://doi.org/10.1561/1500000019. doi:10.1561/1500000019.

[27] J. Lin, X. Ma, S.-C. Lin, J.-H. Yang, R. Pradeep, R. Nogueira, Pyserini: A python toolkit for reproducible information retrieval research with sparse and dense representations, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), ACM, 2021, pp. 2356–2362. URL: https://doi.org/10.1145/3404835.3463238. doi:10.1145/3404835.3463238.