Efficient In-Memory Inverted Indexes: Theory and Practice

Joel Mackenzie
The University of Queensland
Brisbane, Australia
joel.mackenzie@uq.edu.au

Sean MacAvaney
University of Glasgow
Glasgow, United Kingdom
sean.macavaney@glasgow.ac.uk

Antonio Mallia
Pinecone
New York, US
antonio@pinecone.io

Michał Siedlaczek
MongoDB
New York, US
michal@siedlaczek.me

Abstract

Inverted indexes are the backbone of most large-scale information retrieval systems. Although conceptually simple, high-performance inverted indexes require a deep understanding of low-level system optimizations, compression techniques, and traversal strategies. With the widespread adoption of in-memory search engines, the rise of learned sparse retrieval (LSR), and the increasing complexity of ranking pipelines, the design space for efficient indexing and retrieval systems has expanded significantly.

This tutorial addresses a critical knowledge gap between textbook-style explanations and advanced techniques required for efficient and optimized retrieval. It aims to equip researchers and practitioners with a comprehensive understanding of how modern in-memory search systems are designed, built, and optimized for high-performance retrieval across large-scale document collections.

CCS Concepts

• **Information systems** → **Information retrieval**; **Search engine architectures and scalability**; **Retrieval efficiency**.

Keywords

Efficiency, Inverted Index, Query Processing, Compression

ACM Reference Format:
Joel Mackenzie, Sean MacAvaney, Antonio Mallia, and Michał Siedlaczek. 2025. Efficient In-Memory Inverted Indexes: Theory and Practice. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), July 13–18, 2025, Padua, Italy.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3726302.3731688

1 Motivation

Efficient indexing and retrieval remain foundational to information retrieval (IR) systems at all scales, from academic testbeds to large-scale commercial search engines. At the heart of these systems lies the inverted index, a decades-old data structure that continues to evolve in response to new retrieval paradigms and performance demands. While many students and early-career researchers are introduced to inverted indexes through textbooks or simplified implementations, the gap between these resources and the requirements of high-performance, in-memory search engines remains substantial. Furthermore, the complexity of these highly optimized solutions presents a significant barrier to entry for those interested in understanding and working in the efficiency space.

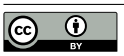
Recent advances in retrieval – particularly the rise of learned sparse retrieval (LSR) models [1, 4, 9, 14, 19, 26, 36], hybrid search pipelines [5, 7, 11, 15], and retrieval-augmented generation (RAG) applications [10] – have renewed interest in efficient sparse retrieval techniques that can be tightly integrated with modern machine learning and cascade ranking workflows. In particular, the use of traditional sparse indexes to serve model-generated representations has led to new challenges in indexing, scoring, and top-k retrieval, motivating the development of specialized pruning strategies such as BM25-guided traversal [8, 20, 28], Block-max Pruning [25], and list decomposition [16] among many others.

Despite these developments, many researchers and practitioners face a steep learning curve when experimenting with efficient retrieval infrastructures, making it difficult to reproduce results, run scalable experiments, or explore new optimization techniques. To address this gap, this tutorial provides both theoretical foundations and practical guidance for building high-performance sparse retrieval systems. Using the open-source *Performant Indexes and Search for Academia* (PISA) engine [22] and its Python bindings via PyTerrier [13], we demonstrate how classical indexing techniques are implemented in practice and how they can be extended to support emerging applications such as LSR and RAG.

The tutorial is particularly relevant to the SIGIR community, where interest in efficient retrieval, reproducible experimentation, and integration with neural models has grown rapidly in recent years. By equipping attendees with both the conceptual background and the practical skills required to build and experiment with state-of-the-art sparse retrieval systems, this tutorial aims to lower the barrier to entry and foster new research directions at the intersection of classical IR and modern machine learning. The tutorial will also improve the visibility of ongoing work and open directions for efficient inverted index-based search systems.

2 Objectives

We organize our tutorial around a set of Intended Learning Outcomes (ILOs) that attendees will be able to achieve by the end of the tutorial. These outcomes are designed to accommodate a broad audience – from those new to information retrieval, to more experienced researchers – by combining foundational theory with



SIGIR '25, Padua, Italy

ACM ISBN 979-8-4007-1592-1/2025/07
https://doi.org/10.1145/3726302.3731688

hands-on practical skills via the PISA engine. Furthermore, we aim to provide valuable content to both *end-users* (those who want to effectively apply inverted indexes in their research) and *tinkerers* (those who want to dig into the implementations and contribute to new research in this domain). The ILOs for the tutorial are:

- **ILO1: Theoretical Understanding of Inverted Indexes** Attendees will develop a solid understanding of how inverted indexes are structured and used in information retrieval systems. They will learn how core components – such as posting lists, lexicons, and skip lists – are implemented and organized in memory, how they are compressed, and how these design choices impact retrieval efficiency and scalability.
- **ILO2: Fast Top-k Retrieval with Dynamic Pruning** Attendees will gain insight into *dynamic pruning-based* retrieval techniques (such as MaxScore [33] and BMW[3, 6, 21]) that accelerate top-k search. They will understand the principles behind these algorithms, including any assumptions, sensitivities to different indexing or parameter choices, and the underlying trade-offs involved with different pruning strategies. Attendees will also gain insights into the current state-of-the-art methods for accelerating inverted index-based algorithms, including threshold estimation [17, 24, 27], advanced skipping or pruning methods [23, 29, 30, 34], and anytime retrieval strategies [18].
- **ILO3: Current Trends and Research Directions** Attendees will understand how traditional inverted index-based retrieval fits into modern IR systems. They will learn how emerging methods – such as learned sparse retrieval – influence traversal strategies, including problems caused by learned sparse data distributions. Then, we will introduce ongoing research to remedy these issues including BM25-guided traversal, list decomposition, and *anytime* retrieval, as well as emerging methods tailored to sparse neural representations such as Block-Max Pruning (BMP) [25] and Seismic [2].
- **ILO4: Building and Running Experiments with PISA** Attendees will gain practical experience using the PISA engine to build indexes, run retrieval experiments, and evaluate performance. They'll learn how to apply different scoring functions and retrieval algorithms to analyze efficiency and effectiveness trade-offs. We plan to offer two levels of access; (1) the PyTerrier-PISA python bindings for users who wish to interact with PISA via the PyTerrier interface; and (2) directly via PISA's command line tool interface.
- **ILO5: Integration of PISA in Modern Applications** Attendees will be able to use PISA as a sparse retrieval backend in applied research contexts, such as incorporating keyword search into retrieval-augmented generation (RAG) pipelines.

3 Relevance

This tutorial is designed for both newcomers to information retrieval and experienced researchers who want to better understand efficient retrieval methods. Although there has been a significant focus on semantic search methods over the last decade, lexical retrieval is still a core component of many search systems at various scales, and the efficient deployment of lexical search systems is a key requirement for the community. Our tutorial complements recent tutorials [31, 35] and surveys [32] by focusing specifically on the efficient design and implementation of such retrieval systems, from traditional models and simple indexes to state-of-the-art techniques. To ensure relevance with modern IR trends and ongoing work, we specifically focus on how traditional inverted index methods can be integrated into modern applications, and dedicate a section of the tutorial to ongoing work and emerging trends.

4 Format and Schedule

The tutorial will be presented as a series of modules, interleaving the theoretical aspects with practical, hands-on activities. We also intend on providing a more advanced "post tutorial" extension task for those who wish to continue to develop their knowledge, but our main goal is to reduce the barrier to entry for attendees. To this end, no "pre-work" is required, and we expect to have both runnable notebooks available to support the python-oriented sections, as well as a Docker image with PISA pre-installed for the command line work. Thus, participants are only expected to have Docker installed and a functioning Python environment. This will facilitate rapid set-up, and these resources will be made available before the workshop with support available from the team prior to, during, and after the tutorial. Note that all timelines listed below are indicative, and are intended to have some level of flexibility to adapt to the audience needs or interests.

Session 1: Indexing and Retrieval [75 min.]

The first session will be focused on exploring the fundamentals of the inverted index. We will begin with a discussion on where the PISA family (including Python bindings) fits into the wider tool landscape, and compare it against other common tools that serve a similar purpose. We will then motivate the need for inverted indexes, followed by a series of brief, visual tutorials on in-memory indexing. These will include details on the representation of inverted indexes, and practical examples of indexing common IR collections. Next, we will move from indexing to retrieval, again using visual tutorials to outline how exhaustive retrieval operates over inverted indexes. Following this, we will move to efficient dynamic pruning algorithms (using the more simple WAND or MaxScore algorithms as exemplars). We will complete this session with a small empirical comparison of top-k retrieval algorithms using PISA on the indexed collection.

An indicative timeline is as follows:

- Introduction and motivation [5-10 minutes]
- Inverted indexing and retrieval fundamentals [20 minutes]
- Fast and compact in-memory index representations [20 minutes]
- From exhaustive to fast top-k retrieval [20 minutes]

Practical Aspects. The first session will aim at familiarizing attendees with the notebook/Docker image, and to run some basic indexing and querying. We expect this part to rely on a small and freely available collection such as a subset of an open-source Wikipedia dump. We will start with building a basic index, and running simple

conjunctive and disjunctive queries. Then, we will build additional structures required to support (fast) ranking, and explore exhaustive document-at-a-time top-k retrieval. Finally, we will move to deploying dynamic pruning algorithms like WAND and MaxScore.

Break [15 minutes]

Aligned with the conference coffee break.

Session II: Learned Sparse Retrieval [60 min.]

The second session will focus on the challenges and opportunities of combining traditional inverted index structures with modern learned sparse retrieval (LSR) techniques. We begin with a high-level overview of LSR, covering their motivation, model architectures, and how they differ from statistical rankers like BM25. We then discuss how LSR changes the structure and usage patterns of inverted indexes – including new characteristics of posting lists, term distributions, and score distributions that arise from model-generated representations. Next, we explore how traditional indexing and retrieval algorithms can be adapted to efficiently support LSR, with a particular focus on pruning strategies that help mitigate long and noisy LSR postings.

An indicative timeline is as follows:

- From Statistical to Learned Models [10 minutes]
- LSR: A trouble-maker for Inverted Indexes [20 minutes]
- Accelerating Inverted Indexes in the context of LSR [30 minutes]

Practical Aspects. This session will include a practical walk-through, showing how PISA can be used to index LSR-generated output (e.g., in JSON or CIFF [12]) format and how these indexes can be queried using the same high-performance retrieval infrastructure. These experiments will also demonstrate the slowdowns caused by LSR in practice. The session concludes with a demonstration of how PISA can be integrated into Python-based research pipelines via PyTerrier, enabling its use as a first-stage retriever in modern workflows, including hybrid retrieval and, if time permits, retrieval-augmented generation (RAG) systems.

Session III: Alternatives & New Directions [30 min.]

Our final session will broaden the focus of the tutorial beyond document-ordered inverted indexes with the intent of providing useful pointers to alternative approaches being investigated by the efficiency community. We will also outline some open problems and ongoing work in the efficiency space. Example material includes:

- Impact-Ordered Indexes and Score-at-a-Time Retrieval engines [10 minutes]
- Embedding-Based Retrieval and Hybrid Search [10 minutes]
- The future of the PISA engine [10 minutes]

5 Materials

Attendees will be provided with:

- Slides covering theoretical content and code walk-throughs, including diagrams or animations to improve understanding;
- Access to a Docker image for PISA, and notebooks (Jupyter or Colab) for the PyTerrier interface;
- Sample datasets and prebuilt indexes for experimentation; and
- Links to relevant publications, code repositories, and further readings.

We intend to provide some post-tutorial resources and guides for those attendees interested in learning or engaging further. For example, we plan to provide links to the PISA repository (and the PyTerrier counterpart), detailed documentation, the PISA Slack channel, and a guide on how to contribute further to the project. We will also provide a set of references to recent and ongoing work, and any further material we develop that was not covered during the tutorial.

6 Conclusion

Efficient indexing and retrieval remain central challenges in the design of high-performance information retrieval systems. This tutorial bridges the gap between foundational concepts and cutting-edge developments by offering both theoretical insights and hands-on experience with state-of-the-art tools. It focuses on in-memory inverted indexes and their role in both classical and modern retrieval pipelines — including learned sparse retrieval and retrieval-augmented generation.

This tutorial is designed to support both newcomers to information retrieval and experienced researchers interested in efficiency-focused methods. By the end of the session, attendees will have a strong understanding of core indexing and retrieval techniques, along with practical experience that equips them to apply these approaches in their own research and experimental workflows.

References

[1] Soyuj Basnet, Jerry Gou, Antonio Mallia, and Torsten Suel. 2024. DeeperImpact: Optimizing Sparse Learned Index Structures. *CoRR* abs/2405.17093 (2024). doi:10.48550/ARXIV.2405.17093 arXiv:2405.17093

[2] Sebastian Bruch, Franco Maria Nardini, Cosimo Rulli, and Rossano Venturini. 2024. Efficient Inverted Indexes for Approximate Retrieval over Learned Sparse Representations. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. ACM, 152–162. doi:10.1145/3626772.3657769

[3] Shuai Ding and Torsten Suel. 2011. Faster top-k document retrieval using block-max indexes. In *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011*. ACM, 993–1002. doi:10.1145/2009916.2010048

[4] Thibault Formal, Benjamin Piwowarski, and Stéphane Clinchant. 2021. SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM, 2288–2292. doi:10.1145/3404835.3463098

[5] Luyu Gao, Zhuyun Dai, and Jamie Callan. 2021. COIL: Revisit Exact Lexical Match in Information Retrieval with Contextualized Inverted List. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*. Association for Computational Linguistics, 3030–3042. doi:10.18653/V1/2021.NAACL-MAIN.241

[6] Adrien Grand, Robert Muir, Jim Ferenczi, and Jimmy Lin. 2020. From MAXSCORE to block-max wand: the story of how Lucene significantly improved query evaluation performance. In *ECIR*. Springer.

[7] Kaili Huang, Thejas Venkatesh, Uma Dingankar, Antonio Mallia, Daniel Campos, Jian Jiao, Christopher Potts, Matei Zaharia, Kwabena Boahen, Omar Khattab, Saarthak Sarup, and Keshav Santhanam. 2025. ColBERT-Serve: Efficient Multi-stage Memory-Mapped Scoring. In *Advances in Information Retrieval - 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6-10, 2025, Proceedings, Part IV (Lecture Notes in Computer Science, Vol. 15575)*. Springer, 21–30. doi:10.1007/978-3-031-88717-8_3

[8] Hrishikesh Kulkarni, Sean MacAvaney, Nazli Goharian, and Ophir Frieder. 2023. Lexically-Accelerated Dense Retrieval. In *Proceedings of the 46th International*

ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023. ACM, 152–162. doi:10.1145/3539618.3591715

[9] Carlos Lassance and Stéphane Clinchant. 2022. An efficiency study for SPLADE models. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2220–2226.

[10] Patrick S. H. Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[11] Jimmy Lin, Xueguang Ma, Sheng-Chieh Lin, Jheng-Hong Yang, Ronak Pradeep, and Rodrigo Nogueira. 2021. Pyserini: A Python Toolkit for Reproducible Information Retrieval Research with Sparse and Dense Representations. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM, 2356–2362. doi:10.1145/3404835.3463238

[12] Jimmy Lin, Joel Mackenzie, Chris Kamphuis, Craig Macdonald, Antonio Mallia, Michał Siedlaczek, Andrew Trotman, and Arjen de Vries. 2020. Supporting interoperability between open-source search engines with the common index file format. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2149–2152.

[13] Sean MacAvaney and Craig Macdonald. 2022. A Python Interface to PISA!. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*. ACM, 3339–3344. doi:10.1145/3477495.3531656

[14] Sean MacAvaney, Franco Maria Nardini, Raffaele Perego, Nicola Tonellotto, Nazli Goharian, and Ophir Frieder. 2020. Expansion via Prediction of Importance with Contextualization. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*. ACM, 1573–1576. doi:10.1145/3397271.3401262

[15] Craig Macdonald, Nicola Tonellotto, Sean MacAvaney, and Iadh Ounis. 2021. PyTerrier: Declarative Experimentation in Python from BM25 to Dense Retrieval. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*. ACM, 4526–4533. doi:10.1145/3459637.3482013

[16] Joel Mackenzie, Antonio Mallia, Alistair Moffat, and Matthias Petri. 2022. Accelerating Learned Sparse Indexes Via Term Impact Decomposition. In *Findings of the ACL: EMNLP 2022*. 2830–2842.

[17] Joel Mackenzie and Alistair Moffat. 2020. Examining the additivity of top-k query processing innovations. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1085–1094.

[18] Joel Mackenzie, Matthias Petri, and Alistair Moffat. 2021. Anytime ranking on document-ordered indexes. *ACM Transactions on Information Systems (TOIS)* 40, 1 (2021), 1–32.

[19] Antonio Mallia, Omar Khattab, Torsten Suel, and Nicola Tonellotto. 2021. Learning Passage Impacts for Inverted Indexes. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*. ACM, 1723–1727. doi:10.1145/3404835.3463030

[20] Antonio Mallia, Joel Mackenzie, Torsten Suel, and Nicola Tonellotto. 2022. Faster Learned Sparse Retrieval with Guided Traversal. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*. ACM, 1901–1905. doi:10.1145/3477495.3531774

[21] Antonio Mallia, Giuseppe Ottaviano, Elia Porciani, Nicola Tonellotto, and Rossano Venturini. 2017. Faster BlockMax WAND with Variable-sized Blocks. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*. ACM, 625–634. doi:10.1145/3077136.3080780

[22] Antonio Mallia, Michal Siedlaczek, Joel M. Mackenzie, and Torsten Suel. 2019. PISA: Performant Indexes and Search for Academia. In *Proceedings of the Open-Source IR Replicability Challenge co-located with 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, OSIRRC@SIGIR 2019, Paris, France, July 25, 2019 (CEUR Workshop Proceedings, Vol. 2409)*. CEUR-WS.org, 50–56. https://ceur-ws.org/Vol-2409/docker08.pdf

[23] Antonio Mallia, Michał Siedlaczek, and Torsten Suel. 2021. Fast disjunctive candidate generation using live block filtering. In *Proceedings of the 14th ACM international conference on web search and data mining*. 671–679.

[24] Antonio Mallia, Michal Siedlaczek, Mengyang Sun, and Torsten Suel. 2020. A Comparison of Top-k Threshold Estimation Techniques for Disjunctive Query Processing. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*. ACM, 2141–2144. doi:10.1145/3340531.3412080

[25] Antonio Mallia, Torsten Suel, and Nicola Tonellotto. 2024. Faster Learned Sparse Retrieval with Block-Max Pruning. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*. ACM, 2411–2415. doi:10.1145/3626772.3657906

[26] Thong Nguyen, Sean MacAvaney, and Andrew Yates. 2023. A Unified Framework for Learned Sparse Retrieval. In *Advances in Information Retrieval - 45th European Conference on Information Retrieval, ECIR 2023, Dublin, Ireland, April 2-6, 2023, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 13982)*. Springer, 101–116. doi:10.1007/978-3-031-28241-6_7

[27] Yifan Qiao, Parker Carlson, Shanxiu He, Yingrui Yang, and Tao Yang. 2024. Threshold-driven Pruning with Segmented Maximum Term Weights for Approximate Cluster-based Sparse Retrieval. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 19742–19757.

[28] Yifan Qiao, Yingrui Yang, Haixin Lin, and Tao Yang. 2023. Optimizing guided traversal for fast learned sparse retrieval. In *Proceedings of the ACM Web Conference 2023*. 3375–3385.

[29] Michał Siedlaczek, Antonio Mallia, and Torsten Suel. 2022. Using conjunctions for faster disjunctive top-k queries. In *Proceedings of the fifteenth ACM international conference on web search and data mining*. 917–927.

[30] Gabriel Tolosa and Antonio Mallia. 2023. Many are Better than One: Algorithm Selection for Faster Top-K Retrieval. *Information Processing & Management* 60, 4 (2023), 103359.

[31] Nicola Tonellotto and Craig Macdonald. 2018. Efficient Query Processing Infrastructures: A half-day tutorial at SIGIR 2018. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 1403–1406. doi:10.1145/3209978.3210191

[32] Nicola Tonellotto, Craig Macdonald, and Iadh Ounis. 2018. Efficient Query Processing for Scalable Web Search. *Foundations and Trends in Information Retrieval* 12, 4-5 (2018), 319–500.

[33] Howard R. Turtle and James Flood. 1995. Query Evaluation: Strategies and Optimizations. *Inf. Process. Manag.* 31, 6 (1995), 831–850. doi:10.1016/0306-4573(95)00020-H

[34] Erman Yafay and Ismail Sengor Altingovde. 2023. Faster Dynamic Pruning via Reordering of Documents in Inverted Indexes. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2001–2005.

[35] Andrew Yates, Carlos Lassance, Sean MacAvaney, Thong Nguyen, and Yibin Lei. 2024. Neural Lexical Search with Learned Sparse Retrieval. In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region* (Tokyo, Japan) *(SIGIR-AP 2024)*. Association for Computing Machinery, New York, NY, USA, 303–306. doi:10.1145/3673791.3698441

[36] Puxuan Yu, Antonio Mallia, and Matthias Petri. 2024. Improved Learned Sparse Retrieval with Corpus-Specific Vocabularies. In *Advances in Information Retrieval - 46th European Conference on Information Retrieval, ECIR 2024, Glasgow, UK, March 24-28, 2024, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 14610)*. Springer, 181–194.