Université Paris Cité

DaiSy: A Library for Scalable Data Series Similarity Search

Francesca Del Gaudio, Manos Chatzakis, Gayathiri Ravendirane, Botao Peng, Themis Palpanas

LIPADE-TR-Nº 12

March 29, 2026

Technical Report

LIPADE
Laboratoire d'Informatique PAris DEscartes

DaiSy: A Library for Scalable Data Series Similarity Search

Francesca Del Gaudio*
Université Paris Cité, LIPADE
F-75006, Paris, France
francescadelgaudio56@gmail.com

Manos Chatzakis*
Université Paris Cité, LIPADE
F-75006, Paris, France
manos.chatzaki@gmail.com

Gayathiri Ravendirane
Université de Bordeaux
Bordeaux, France
gayathiri.ravendirane@etu.u-bordeaux.fr

Botao Peng
Chinese Academy of Sciences
Beijing, China
pengbotao@ict.ac.cn

Themis Palpanas
Université Paris Cité, LIPADE
F-75006, Paris, France
themis@mi.parisdescartes.fr

ABSTRACT

Exact similarity search over large collections of data series is a fundamental operation in modern applications, yet existing solutions are often fragmented, specialized, or tailored to specific execution environments. In this paper, we present DaiSy, a unified library for exact data series similarity search that integrates multiple state-of-the-art algorithms within a single, coherent framework. DaiSy is the first library to support exact similarity search across diverse execution environments, including implementations for disk-based, in-memory, GPU-accelerated, and distributed scalable similarity search. Although designed for data series, DaiSy is also directly applicable to exact similarity search over vector data, enabling its use in a broader range of applications. The library supports interfaces in both C++ and Python, enabling users to easily integrate its functionality into a variety of tasks. DaiSy is open-sourced and available at: `https://github.com/MChatzakis/DaiSy`.

1. INTRODUCTION

Data series constitute one of the most popular data types, appearing across a wide range of application domains, including finance, astrophysics, neuroscience, engineering, seismology, and many others [44, 28]. Such domains typically generate large collections of data series that must be analyzed in order to extract meaningful knowledge [21, 43, 3, 38, 20, 34]. Within these data analytics tasks, similarity search serves as a fundamental operation that enables efficient and effective analysis. Moreover, many similarity search applications are highly sensitive to result accuracy, highlighting the need for similarity search methods that avoid approximation errors, i.e., exact similarity search [13]. In exact query answering, it is crucial to guarantee that the returned data series is indeed the most similar to the query among all data series in the collection.

Scalable Similarity Search. The increasing interest in scalable similarity search across different application domains has led to massive research on search time optimization [12, 10]. To enable scalable similarity search, numerous indexing methods have been proposed [29]. These methods construct index structures over data series collections and employ specialized search algorithms to answer queries. A large portion of prior work focuses on optimizing the core similarity search algorithms themselves, aiming to reduce query latency and improve throughput [11, 19, 24, 26, 15]. At the same time, several approaches extend similarity search to different execution environments based on the available hardware, such as disk-based systems [30], in-memory CPUs [31], GPU-accelerated systems [32], and distributed infrastructures [5, 42]. However, despite this rich body of work, existing methods are largely dispersed, with implementations scattered across different codebases and system designs, which makes practical adoption challenging.

DaiSy. We present DaiSy, the first scalable solution for exact data series similarity search that integrates multiple algorithms into a single, unified library. DaiSy enables efficient similarity search across a wide range of environments and system configurations by supporting scalable algorithms for disk-based, in-memory, GPU-accelerated, and distributed execution. This is achieved by incorporating one representative state-of-the-art algorithm for each execution environment: ParIS+ [30] for disk-based processing, MESSI [31] for in-memory processing, SING [32] for GPU-accelerated processing, and Odyssey [5] for

*These authors contributed equally to this work.

distributed processing. These algorithms were selected due to their demonstrated superior performance over existing alternatives in their respective settings [29, 13]. ParIS+ efficiently supports disk-based similarity search when data do not fit in main memory, MESSI provides highly efficient parallel in-memory search, SING exploits GPU processing to accelerate key search operations, and Odyssey enables exact similarity search over massive datasets using distributed memory and optimized communication. All four above algorithms, Paris+, MESSI, SING, and Odyssey, are guaranteed to always return the exact, correct answers.

Finally, we note that although these algorithms were designed with data series in mind, they are equally applicable to general high-dimensional vectors (e.g., deep embeddings), where they exhibit state-of-the-art performance, as well [13, 14].

DaiSy exposes a unified interface for all supported algorithms and is available in both C++ and Python to facilitate adoption by users. DaiSy is open sourced and available at: `https://github.com/MChatzakis/DaiSy`.

Contributions. We summarize our main contributions as follows:

- We present DaiSy, the first scalable library for exact data series, as well as vector, similarity search that integrates multiple state-of-the-art algorithms within a unified framework.
- We describe how DaiSy enables efficient data series similarity search across a wide range of configurations and systems, by supporting disk-based, in-memory, GPU-accelerated, and distributed execution environments.
- We design an extensible similarity search library architecture for DaiSy, accompanied by comprehensive benchmarking frameworks, and we provide both C++ and Python interfaces, facilitating easy adoption across diverse domains and applications.
- We release DaiSy as an open-source library and make it publicly available to the community through our GitHub repository.

2. BACKGROUND

Data Series. A *data series* $S = \{p_1, \ldots, p_n\}$ is a sequence of points, where each $p_i = (u_i, t_i)$ represents a value u_i at position t_i. The value n is its *size* (or *dimensionality*).

iSAX Summary. The *iSAX summary* [36] discretizes a data series by dividing the x-axis into equal segments (represented by their mean) and the y-axis into regions based on normal distribution breakpoints. Each region is assigned a symbolic representation with a specific *cardinality*, enabling hierarchical *iSAX-based index trees* [29].

Similarity Search. Given a collection $\mathcal{C}$ and a query S, *similarity search* identifies the top-k series in $\mathcal{C}$ most similar to S according to a distance measure.

Distance Measures. The *L2 (Euclidean) Distance* is $L2S(T, S) = \sqrt{\sum_{i=1}^{n}(t_i - s_i)^2}$, often used as *L2 Squared* for efficiency. The distance between iSAX summaries provides a *lower-bound* to the real distance. *Dynamic Time Warping (DTW)* allows non-linear alignments by computing the minimum cumulative cost over all valid warping paths [22].

2.1. Related Work

Data Series Indexes. Various indices, specific to data series, have been proposed in the literature [12, 10]. DSTree [39] is an index based on the APCA summarization [23]. The DSTree can adaptively perform split operations by increasing the detail of APCA as needed. The iSAX index is based on the SAX summarization, and its extension, iSAX [36]. Several other indices (iSAX-based) have been proposed in the literature [4, 32, 31, 30, 42, 5, 24, 11, 19, 9, 26, 16, 40], summarized in [29] and evaluated in [13].

Similarity Search Libraries. FAISS [8] is a library that implements a wide range of algorithms for approximate vector similarity search. It is conceptually the most closely related system to DaiSy; however, it is designed for vector data and targets application scenarios in which the exactness constraint can be relaxed, e.g. Retrieval Augemented Generation (RAG) [25]. FAISS provides a limited functionality for exact similarity search through a naive bruteforce search implementation. In addition, **UCR Suite** [33] and **TSSEARCH** [17] are libraries for subsequence similarity search in C++ and Python respectively. They both tackle a similar variant of similarity search, but they do not provide various different methods to perform indexing or search. **AEON** [27] and **TSLEARN** [37] are machine learning oriented Python libraries for data series processing, with strong focus on analytics tasks, such as classification, clustering, regression, forecasting, and anomaly detection. Although they provide some functionalities for computing distances between data series, they are not libraries for scalable data series similarity search and they do not expose any interface for indexing or search.

3. SYSTEM DESIGN

DaiSy's architecture follows a layered design centered on a C++ core that implements indexing, distance computation, and search functionality, while exposing the same conceptual model through a Python interface. Foundational primitives provide distance

evaluation, data access, and reusable lower bounds, while an index layer encapsulates the iSAX index as an internal service [4]. On top of these, multiple execution models realize exact similarity search under different assumptions on data placement and available resources. The same abstractions are exposed in both C++ and Python, ensuring a uniform user-facing interface independent of the execution back ends enabled at build time.

3.1. Architecture

Design principles and abstractions. DaiSy is structured around a set of design principles that emphasize modularity, extensibility, and clear separation of responsibilities. Its architecture enforces a strict separation of concerns along three orthogonal dimensions. Distance computation is encapsulated in a dedicated component responsible for exact distance evaluation and lower bounds, independent of data storage and execution strategy. Data access is handled through an explicit adapter-based abstraction [1], which decouples indexing and search logic from the physical layout and location of the data and enables support for heterogeneous data sources, including in-memory and on-disk representations. Execution strategy is treated as a separate concern, while search algorithms are implemented as interchangeable components exposing a common interface for index construction and exact k-nearest neighbor query processing. Importantly, exactness is a first-class design constraint in DaiSy, with algorithmic differences expressed solely in terms of execution strategy and resource usage rather than accuracy guarantees.

Core primitives layer. The core primitives layer provides the shared foundation on which all indexing structures and search algorithms in DaiSy are built. It deliberately implements no index or search strategy; instead, it factors out functionality common across algorithms and execution models to ensure consistency and avoid duplication. Distance-related logic is encapsulated in a unified abstraction supporting exact evaluation and lower bounds, and is independent of data placement and execution strategy. Data access is exposed through a stable adapter-based interface that models datasets as streams of data series, keeping higher-level components agnostic to in-memory or on-disk storage. The layer also provides shared reduced representations and lower bounds enabling safe pruning while preserving correctness, forming a stable, metric- and storage-agnostic substrate for higher layers.

Index layer. The index layer encapsulates the iSAX index as an internal service that provides structured access to data series data, without prescribing a search strategy. Index construction is encapsulated by a single configuration object that defines representation parameters, buffer sizes, and storage mode. Importantly, the index layer does not implement search logic. By treating iSAX as a service rather than as an algorithm, DaiSy decouples index management from execution strategy and allows the same index implementation to be reused across multiple execution models.

Similarity Search layer. The similarity search layer in DaiSy is organized by execution model rather than by individual algorithm implementations. Each execution model defines how exact similarity search is carried out under specific assumptions about data placement and available resources, while reusing the same core primitives and index services. Taken together, these models position DaiSy as an execution framework that instantiates a single conceptual search pipeline through multiple back ends. DaiSy supports a variety of similarity search algorithms in this layer, covering use cases in disk-based, in-memory, GPU-accelerated and distributed environments. ParIS+ [30] is an algorithm designed for disk-based exact similarity search, and its tailored for situations where the dataset size exceeds the main memory capacity of the system. MESSI [31] is an exact similarity search algorithm that operates entirely in memory and relies on an iSAX index to guide candidate selection and refinement. It follows a multi-phase search pipeline that combines index-guided exploration with exact refinement while preserving exactness. The SING index [32] leverages GPUs to accelerate the search procedure. It is organized into two main stages: in-memory index construction and multi-phase query answering with GPU offloading. Odyssey [5] is a distributed framework for exact similarity search built on MPI, designed for multi-node environments where datasets can be accommodated in distributed memory. DaiSy also supports two exhaustive search implementations.
They both exhaustively scan the entire dataset for each query (without relying on any index structures). The first one, *Bruteforce*, simply performs all real distance computations over the raw data. The second one, *LbBruteforce*, applies lower-bounding, thus, avoiding real distance computations when the lower bound is larger than the BSF distance.

3.2. Distance Measures

DaiSy supports L2 Squared and DTW for exact similarity search, both exposed through a unified distance computer abstraction. Our library supports both optimized SIMD and scalar implementations,

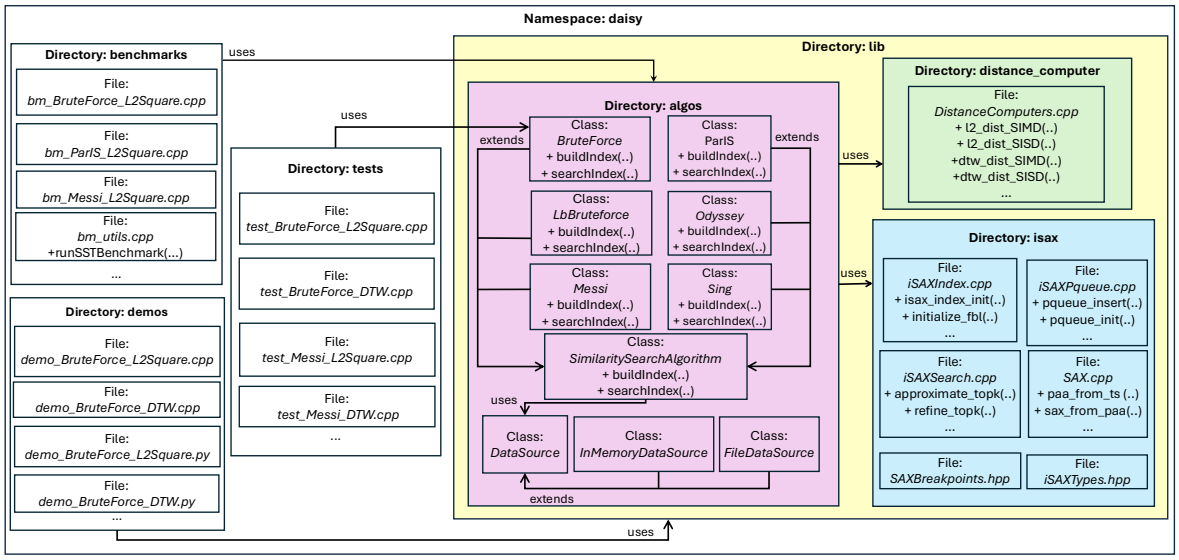


Figure 1: Component Diagram of DaiSy.

to ensure compatibility with different hardware configurations. Distance computations support early termination via an optional upper bound, allowing evaluation to stop once the partial distance exceeds the current best-so-far value [33]. In addition, DaiSy provides exact lower-bound functions for both measures.

3.3. Z-normalization

DaiSy supports both z-normalized (i.e., each data series, or vector, has mean 0 and standard deviation 1) and non z-normalized data. Though, the algorithms are more suited to Z-normalized data (we use the default iSAX breakpoints [36], which have been optimized for z-normalized data).

4. USING DAISY

4.1. API and Tooling

DaiSy exposes a simple API to enable easy use in both C++ and Python.

- `buildIndex`. Initializes and builds the index structure for the selected algorithm.
- `searchIndex`. Performs the top-k index search for the given queries, returning the most similar data series from the index.

Furthermore, reproducibility and systematic evaluation are supported through a set of shared utilities, demos, benchmarks, and tests. Demos are provided in both C++ and Python and exercise the same pipelines exposed to end users. Benchmarks enable controlled performance evaluation across execution models and configurations, while tests validate correctness and integration. All tooling components consume the same APIs as the core library, ensuring consistency between usage, evaluation, and experimentation.

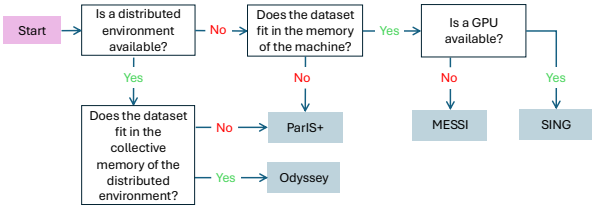


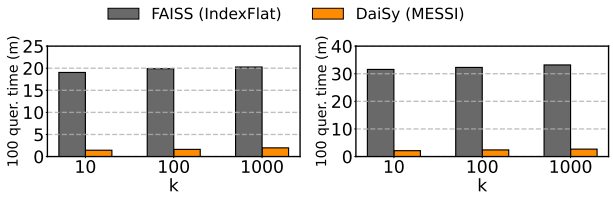
Figure 2: Algorithm selection decision tree of DaiSy.

4.2. Hyperparameter Setting

As expected, the methods implemented in DaiSy involve a number of algorithm-specific hyperparameters. The library allows experienced users to explicitly configure all relevant hyperparameters for each supported algorithm. At the same time, DaiSy does not require the user to manually specify these parameters, as it initializes them to the values recommended in the original publications of the corresponding algorithms.

4.3. Algorithm Selection

As we demonstrated, DaiSy provides efficient similarity search algorithms tailored for different execution environments. Figure 2 summarizes the algorithm selection logic based on dataset size and available resources. When the dataset fits in main memory, DaiSy first considers whether a distributed environment is available and, if so, selects Odyssey to exploit distributed memory. In the absence of a distributed setting, the choice is driven by the available hardware accelerators: SING is used when a



(a) Deep100M (b) Seismic100M

Figure 3: Query answering time for 100 queries when varying k (48 hyperthreads).

GPU is present, while MESSI is selected for CPU-based in-memory execution. When the dataset does not fit in main memory, DaiSy resorts to ParIS+ for disk-based execution.

We would like to stress that, apart from data series, DaiSy is equally applicable to vector data, as well, where it exhibits state-of-the-art performance. For example, DaiSy is the solution of choice for applications that need exact answers in vector search, or for analysis tasks that involve vector approximate search and need to compute the groundtruth answers. As evidence, we present the results of the following experiment (rigorous experimental evaluations can be found in previous studies [13, 14, 5]).

We conduct an in-memory evaluation against FAISS-IndexFlat, the most widely used library for such tasks, focusing on exact similarity search for two datasets: Deep100M [2] (100 million 96-dimensional image embeddings) and Seismic100M [18] (100 million 256-dimensional seismic data series). IndexFlat is the FAISS solution used to compute exact answers: it stores the full vectors and performs an exhaustive search. Both libraries are compiled with identical settings using `g++ 11.4.0`, configured for in-memory execution, and execution times are measured using the GoogleBenchmark [1] framework. Figure 3 reports the total query execution time for 100 queries on each dataset, when varying k between 10-1000, using 48 hyperthreads. The results show that DaiSy-MESSI consistently outperforms FAISS-IndexFlat, calculating the exact answers for the entire workload 12× faster for the Deep100M vector collection. The same observation is true for the Seismic100M data series collection, where DaiSy-MESSI is 14× faster. Speed-up increases with the number of threads. Similar results hold for other vector datasets, as well.

4.4. Examples

We provide two examples of using DaiSy in C++ and Python to demonstrate the convenience our library provides for tasks involving similarity search. Without loss of generality, in our examples we use the MESSI algorithm and the L2 Squared distance. In C++, a typical use case of DaiSy is depicted below.

```cpp
#include <vector>
#include "daisy/daisy.hpp"
int main() {
    daisy::idx_t d = 256; // dimensionality
    daisy::idx_t n_db = 10000; // dataset size
    daisy::idx_t n_q = 100; // queries size

    // Load or generate input data series.
    float* db = /* load or generate database series */;
    float* q = /* load or generate query series */;

    // Build index using squared Euclidean distance.
    daisy::Messi messi(daisy::DistanceType::L2_SQUARED);
    messi.buildIndex(db, n_db, d);

    // Execute exact k-NN similarity search.
    const daisy::idx_t k = 5;
    std::vector<daisy::idx_t> I(static_cast<size_t>(n_q)*k);
    std::vector<float> D(static_cast<size_t>(n_q)*k);
    messi.searchIndex(q, n_q, k, I.data(), D.data());

    return 0; }
```

Using DaiSy with Python follows the same API, as shown below.

```python
from daisy import DistanceType, Messi

d = 256 # dimensionality
n_db = 100000 # dataset size
n_q = 1000 # queries size

# Load or generate input data.
db = ... # load or generate database series
q = ... # load or generate query series

# Build the index using squared Euclidean distance.
index = Messi(DistanceType.L2_SQUARED)
index.buildIndex(db)

# Execute exact k-NN similarity search.
k = 5
I, D = index.searchIndex(q, k)
```

5. FUTURE WORK

DaiSy serves as the foundation for several planned future efforts aimed at further improving and extending the library. In particular, we plan to incorporate additional algorithms for exact similarity search [41, 35], including methods drawn from different data series summarization families [9]. We also intend to introduce automatic hyperparameter tuning capabilities for the supported algorithms, leveraging modern and effective optimization techniques such as Bayesian optimization, which has been widely adopted for hyperparameter auto-tuning [7]. Additionally, we aim to extend the library to support subsequence similarity search [26, 33], streaming data series [24], and early termination [6, 11, 19].

6. CONCLUSIONS

We present DaiSy, a novel library for scalable exact similarity search on large collections of data series,

[1] `https://github.com/google/benchmark`

as well as general high-dimensional vectors (e.g., deep embeddings). It supports efficient search algorithms across a wide range of hardware configurations, including disk-based, in-memory, GPU-accelerated, and distributed environments. We open source DaiSy and provide carefully designed C++ and Python interfaces that enable straightforward interaction with the library.

7. REFERENCES

[1] M. G. Al-Obeidallah, D. G. Al-Fraihat, A. M. Khasawneh, A. M. Saleh, and H. Addous. Empirical investigation of the impact of the adapter design pattern on software maintainability. In *International Conference on Information Technology, ICIT 2021, Amman, Jordan, July 14-15, 2021*, pages 206–211. IEEE, 2021.

[2] A. Babenko and V. Lempitsky. Efficient indexing of billion-scale datasets of deep descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2055–2063, 2016.

[3] P. Boniol and T. Palpanas. Series2Graph: Graph-based Subsequence Anomaly Detection for Time Series. *PVLDB*, 2020.

[4] A. Camerra, T. Palpanas, J. Shieh, and E. Keogh. isax 2.0: Indexing and mining one billion time series. In *2010 IEEE international conference on data mining*, pages 58–67. IEEE, 2010.

[5] M. Chatzakis, P. Fatourou, E. Kosmas, T. Palpanas, and B. Peng. Odyssey: A journey in the land of distributed data series similarity search. *Proc. VLDB Endow.*, 16(5), 2023.

[6] M. Chatzakis, Y. Papakonstantinou, and T. Palpanas. Darth: Declarative recall through early termination for approximate nearest neighbor search. *Proceedings of the ACM on Management of Data*, 3(4), 2025.

[7] S. Daulton, M. Balandat, and E. Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. *Advances in Neural Information Processing Systems*, 33, 2020.

[8] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou. The faiss library. *IEEE Transactions on Big Data*, 2025.

[9] K. Echihabi, P. Fatourou, K. Zoumpatianos, T. Palpanas, and H. Benbrahim. Hercules against data series similarity search. *arXiv preprint arXiv:2212.13297*, 2022.

[10] K. Echihabi and T. Palpanas. Scalable analytics on large sequence collections. In *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2022.

[11] K. Echihabi, T. Tsandilas, A. Gogolou, A. Bezerianos, and T. Palpanas. Pros: data series progressive k-nn similarity search and classification with probabilistic quality guarantees. *The VLDB Journal*, 32(4).

[12] K. Echihabi, K. Zoumpatianos, and T. Palpanas. New trends in high-d vector similarity search: ai-driven, progressive, and distributed. *PVLDB*, 14(12), 2021.

[13] K. Echihabi, K. Zoumpatianos, T. Palpanas, and H. Benbrahim. The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art. *PVLDB*, 2018.

[14] K. Echihabi, K. Zoumpatianos, T. Palpanas, and H. Benbrahim. Return of the lernaean hydra: Experimental evaluation of data series approximate similarity search. *Proc. VLDB Endow.*, 13(3), 2019.

[15] P. Fatourou, E. Kosmas, T. Palpanas, and G. Paterakis. Fresh: a lock-free data series index. In *2023 42nd International Symposium on Reliable Distributed Systems (SRDS)*, pages 209–220. IEEE, 2023.

[16] P. Fatourou, E. Kosmas, T. Palpanas, and G. Paterakis. Fresh: A lock-free data series index. In *42nd International Symposium on Reliable Distributed Systems, SRDS 2023, Marrakesh, Morocco, September 25-29, 2023*, pages 209–220. IEEE, 2023.

[17] D. Folgado, M. Barandas, M. Antunes, M. L. Nunes, H. Liu, Y. Hartmann, T. Schultz, and H. Gamboa. Tssearch: Time series subsequence search library. *SoftwareX*, 18:101049, 2022.

[18] I. R. I. for Seismology with Artificial Intelligence. Seismic Data Access. `http://ds.iris.edu/data/access/`, 2018.

[19] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos. Progressive similarity search on time series data. In *BigVis 2019-2nd International Workshop on Big Data Visual Exploration and Analytics*, 2019.

[20] P. Huijse, P. A. Estevez, P. Protopapas, J. C. Principe, and P. Zegers. Computational intelligence challenges and applications on large-scale astronomical time series databases. *CIM*, 2014.

[21] K. Kashino, G. Smith, and H. Murase. Time-series active search for quick retrieval of audio and video. In *ICASSP*, 1999.

[22] E. Keogh and C. A. Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3):358–386, 2005.

[23] E. J. Keogh, K. Chakrabarti, S. Mehrotra, and M. J. Pazzani. Locally adaptive dimensionality reduction for indexing large time series databases. In S. Mehrotra and T. K. Sellis, editors, *SIGMOD*, 2001.

[24] H. Kondylakis, N. Dayan, K. Zoumpatianos, and T. Palpanas. Coconut palm: Static and streaming data series exploration now in your palm. In *Proceedings of the 2019 International Conference on Management of Data*, 2019.

[25] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[26] M. Linardi and T. Palpanas. Scalable, variable-length similarity search in data series: The ulisse approach. *Proceedings of the VLDB Endowment*, 11(13):2236–2248, 2018.

[27] M. Middlehurst, A. Ismail-Fawaz, A. Guillaume, C. Holder, D. Guijo-Rubio, G. Bulatova, L. Tsaprounis, L. Mentel, M. Walter, P. Schäfer, and A. Bagnall. aeon: a python toolkit for learning from time series. *Journal of Machine Learning Research*, 25(289):1–10, 2024.

[28] T. Palpanas. The parallel and distributed future of data series mining. In *HPCS*, 2017.

[29] T. Palpanas. Evolution of a data series index: The isax family of data series indexes: isax, isax2. 0, isax2+, ads, ads+, ads-full, paris, paris+, messi, dpisax, ulisse, coconut-trie/tree, coconut-lsm. In *Information Search, Integration, and Personalization: 13th International Workshop, ISIP 2019, Heraklion, Greece, May 9–10, 2019, Revised Selected Papers 13*. Springer, 2020.

[30] B. Peng, P. Fatourou, and T. Palpanas. Paris: The next destination for fast data series indexing and query answering. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018.

[31] B. Peng, P. Fatourou, and T. Palpanas. Messi: In-memory data series indexing. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020.

[32] B. Peng, P. Fatourou, and T. Palpanas. Sing: Sequence indexing using gpus. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. IEEE, 2021.

[33] T. Rakthanmanon, B. Campana, A. Mueen, G. Batista, B. Westover, Q. Zhu, J. Zakaria, and E. Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270, 2012.

[34] U. Raza, A. Camerra, A. L. Murphy, T. Palpanas, and G. P. Picco. Practical data prediction for real-world wireless sensor networks. *TKDE*, 2015.

[35] P. Schäfer, J. Brand, U. Leser, B. Peng, and T. Palpanas. Fast and exact similarity search in less than a blink of an eye. In *41st IEEE International Conference on Data Engineering, ICDE 2025, Hong Kong, May 19-23, 2025*, pages 2464–2477. IEEE, 2025.

[36] J. Shieh and E. J. Keogh. *i*sax: indexing and mining terabyte sized time series. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008*, pages 623–631. ACM, 2008.

[37] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, et al. Tslearn, a machine learning toolkit for time series data. *Journal of machine learning research*, 21(118):1–6, 2020.

[38] Q. Wang, S. Whitmarsh, V. Navarro, and T. Palpanas. iEDeaL: A Deep Learning Framework for Detecting Highly Imbalanced Interictal Epileptiform Discharges. *PVLDB*, 16(2), 2023.

[39] Y. Wang, P. Wang, J. Pei, W. Wang, and S. Huang. A data-adaptive and dynamic segmentation index for whole matching on time series. *PVLDB*, 6(10), 2013.

[40] Z. Wang, Q. Wang, P. Wang, T. Palpanas, and W. Wang. Dumpyos: A data-adaptive multi-ary index for scalable data series similarity search. *VLDB J.*, 33(6):1887–1911, 2024.

[41] Z. Wang, Q. Wang, P. Wang, T. Palpanas, and W. Wang. Dumpyos: A data-adaptive multi-ary index for scalable data series similarity search. *VLDB J.*, 33(6):1887–1911, 2024.

[42] D. E. Yagoubi, R. Akbarinia, F. Masseglia, and T. Palpanas. Dpisax: Massively distributed partitioned isax. In *IEEE International Conference on Data Mining, ICDM*. IEEE Computer Society, 2017.

[43] L. Ye and E. Keogh. Time series shapelets: a new primitive for data mining. In *SIGKDD*. ACM, 2009.

[44] K. Zoumpatianos and T. Palpanas. Data series management: Fulfilling the need for big sequence analytics. In *ICDE*, 2018.