

Name of Referee: **Pardis Sadatian Moghaddam**

Title: QuEval: **Beyond high-dimensional Indexing a la carte**

INSTRUCTIONS TO REFEREES

For a contribution to be acceptable for publication in a journal/conference, it must comprise novel material not previously published in a technical journal/conference. The novelty will usually lie in original results, methods, observations, concepts, or applications, but may also reside in syntheses of/or new insights into previously reported research. In a regular paper, the title, abstract, introduction, and summary should be sufficiently informative to make the contributions of the paper clear to the broadest possible audience, and to place them in context with the related work.

In addition to these fundamental requirements, acceptance for publication depends on a number of important criteria relating to reader interest, technical content, and presentation. To assist the referee in addressing these criteria, the Review form includes a short answer OVERVIEW (Section III) as well as an open form for DETAILED COMMENTS (Section IV). The principal intent of the Overview is to raise the kind of questions that should be addressed in assessing the paper. In other words, the Overview provides a list of the criteria referred to above and, in this sense, serves as a part of these instructions. In addition, the short answers to these questions provide a uniform synopsis of the review for both the editor and the author(s).

The **essential** part of the evaluation, however, is the information contained in the reviewer's Detailed Comments (Section IV). Please try to provide one page of comments in this section. At minimum, please provide one half-page of comments. It is hoped that these comments will be guided by the responses indicated in the Overview, with emphasis placed on points that substantiate the recommendation to the editor. A recommendation to accept for publication, whether with no changes or with minor revisions, should be reserved for manuscripts that describe novel work and satisfy the readership, content, and presentation criteria indicated in the Overview.

If major revisions are recommended, the referee should point these out as specifically as possible and should differentiate changes regarded as optional from those judged as mandatory. If the revisions required are extensive, it is perhaps best to reject the paper and recommend preparation of a "new", heavily revised manuscript for resubmission to the journal/conference. If a paper is rejected mainly on the basis of reader interest, the reviewer may wish to recommend submission to a more appropriate journal or conference. Papers with little or no salvageable material should be rejected outright and discouraged from later submission.

ABOUT THE JOURNAL/CONFERENCE:

This journal's (or conference's) ambition is to stimulate the exchange of ideas and interaction between these two related fields of interest: Data Engineering and Knowledge Engineering. The journal/conference reaches a world-wide audience of researchers, designers, managers and users. The major aim of the journal/conference is to identify, investigate and analyze the underlying principles in the design and effective use of these systems. The journal/conference achieves this aim by publishing original research results, technical advances and news items concerning data engineering, knowledge engineering, and the interface of these two fields.

The journal/conference covers the following topics:

1. Representation and Manipulation of Data & Knowledge: Conceptual data models. Knowledge representation techniques. Data/knowledge manipulation languages and techniques.
2. Architectures of database, expert, or knowledge-based systems: New architectures for database / knowledge base / expert systems, design and implementation techniques, languages and user interfaces, distributed architectures.
3. Construction of data/knowledge bases: Data / knowledge base design methodologies and tools, data/knowledge acquisition methods, integrity/security/maintenance issues.
4. Applications, case studies, and management issues: Data administration issues, knowledge engineering practice, office and engineering applications.
5. Tools for specifying and developing Data and Knowledge Bases using tools based on Linguistics or Human Machine Interface principles.

I. SUMMARY AND RECOMMENDATION (TO BE WITHHELD FROM AUTHOR)

Summary of Evaluation

- Excellent
- Good
- Fair
- Poor

Recommendation

- Accept without changes
- Accept if certain minor changes are made (see Section IV)
- Author should prepare a major revision (see Section IV) for another round of review
- Rejec

If the paper is rejected, the author(s) should

- Prepare a major revision and resubmit it as a new paper
- Submit it to another journal or conference
- Regard it as not publishable

Contributions: (Please put marks in all columns!)

OVERALL	APPLICATIONS		OVERALL QUALITY	
Excellent		Significant	Excellent	Excellent
	Good	Tutorial		Good
Excellent		Possible		Fair
	Good	Minor		Poor
		None		

II. COMMENTS TO BE WITHHELD FROM THE AUTHOR

Please provide explanations for the marks:

The paper appears to be significant for applications as it presents a framework that can be useful for various application scenarios and provide empirical evidence of its usefulness in real-world data sets.

This paper is not a tutorial but rather a technical research paper. It presents QuEval, a framework for evaluating high-dimensional indexing techniques. However, it can be useful to understand some materials including indexing techniques in detail. So, I consider it a “Good” score.

This paper is “Excellent” in terms of its possible contributions, that is because this paper provides a good introduction to QuEval, provides an empirical evaluation of QuEval, helps software engineers and non-database experts to choose the best index structure, and demonstrates the benefits of QuEval in

selecting a suitable index structure in an empirical evaluation using different use cases from various domains.

“Minor” refers to the level of the possible contribution of the paper. I consider it as a “Good” score, The paper has the potential to make some contribution, That is because it provides a new open source name as QuEval. It addresses the challenge of evaluating query performance for high-dimensional data.

III. OVERVIEW

A. Reader Interest

1. Is the paper of current interest to a reasonable segment of JOURNAL/CONFERENCE readership?
 - Yes
 - Perhaps
 - No
2. Relative to the current level of reader interest in the paper, how is this interest likely to change during the next five years?
 - Growing interest
 - Relatively little change
 - Diminishing interest
3. Within its particular field of specialization, is the topic of the paper considered important?
 - Yes, definitely
 - Moderately so
 - Not really

B. Content

1. Is the paper technically sound?
 - Yes
 - Appears to be, but didn't check completely
 - Only partially
 - No
2. How would you describe the technical depth of the paper?
 - Expert level
 - Appropriate for someone working in the field
 - Suitable for the non-specialist
 - Superficial
3. Does the paper make a tangible contribution to the state-of-the-art in its field?
 - Yes, definitely
 - To a limited extent
 - No
4. Is the bibliography adequate?
 - Yes

- Yes, after certain additions and/or deletions are made (see Section IV)
- No

5. To what extent is material in the paper likely to be used by other researchers and practitioners?

- Large
- Average
- Small

C) Presentation

1. Is the abstract an appropriate digest of the work presented?

- Yes
- No

2. Does the introduction clearly state the background and motivation in terms understandable to the nonspecialist?

- Yes
- Probably
- No

3. How would you rate the overall organization of the paper?

- Satisfactory
- Could be improved
- Poor

4. Relative to its technical content, is the length of the paper appropriate?

- Yes
- No, it should be lengthened
- No, it should be shortened

5. Is the language satisfactory?

- Yes
- No

6. How readable is the paper for a computer scientist or engineer who is not a specialist in this particular field?

- Readable with ordinary effort
- Paper is self-contained, but a considerable effort is required
- If the definitions of certain concepts, terms, and symbols were included (noted by "define" in the margins), readability would be improved
- Less than half the paper is readable
- Unreadable

7. Disregarding technical content, how would you rate the quality of this presentation?

- Excellent
- Good
- Fair
- Poor

IV. DETAILED COMMENTS (TO BE RETURNED TO AUTHORS) (Please provide one page of comments in this section.)

Thank you for submitting your paper on the QuEval framework for evaluating high-dimensional indexes. Overall, the paper is well-written and formative, and the framework appears to be useful for selecting appropriate index structures for different use cases. Here are some detailed comments for your consideration:

1. **Abstract:** The abstract is concise and informative, but it could be improved by providing a clearer explanation of the Quval framework and its key features. It would also be helpful to include a brief summary of the main results and conclusions of the study.
2. **Introduction:** The introduction provides a good overview of the challenges related to high-dimensional data storage and access. However, it could benefit from a clearer statement of the paper's main contribution and research questions. Additionally, it would be helpful to include a brief summary of the main results and conclusions of the study.
3. **The QuEval frame work:** Overall, the QuEval framework appears to be a well-designed and extensible sytem for evaluating and comparing indexing structures in a variety of settings. The three methods allow users to add their own indexing structures and distance metrics to the framework. It would be better to give a brief description of the purpose and functionality of each of the three components of the QuEval framework in this section.
4. **QuEval Framework:** The section on the QuEval framework is well-written and provides a good overview of the core components and test case facility. However, it would be helpful to include some examples of the types of index structures and distance metrics that can be evaluated using QuEval.
5. **Purpose of QuEval:** The section on the purpose of QuEval is informative and provides useful examples of different use cases for high-dimensional indexes. However, it could be improved by providing more detail on the specific challenges and requirements of each use case, how QuEval detail the specific challenges and requirements of each use case, and how QuEval addresses these challenges.
6. **Evaluation procedure:** The section on the evaluation procedure is well-written and provides a good overview of the steps involved in evaluating different index structures. However, it could be improved by providing more detail on the specific factors that influenced the performance of each index structure, and how these factors can be used to guide the selection of an appropriate index structure for a given use case.

7. **Overall suggestions and conclusions:** The overall suggestions and conclusions section is well-written and provides useful recommendations. It would be better to mention that they want to “evaluate the impact of new trends in databases, such as GPU co-processing or cloud computing,” but it would be helpful to provide more specific details on how they plan to evaluate these trends and what they hope to learn from the evaluation.

Identify your view of the major contributions (or potential thereof) of the paper [summary].

QuEval is a framework for evaluating the performance of high-dimensional indexing techniques. It consists of three major components: a data generator that creates datasets with specified properties, a query-point generator that creates query points for evaluation, and an HDI tester that evaluates test cases using a user-specified configuration that includes selecting a dataset, query parameters, distance metric, storage device, number of test runs, and index structure. QuEval offers advanced features such as visualization of data space partitioning, C++ implementation compatibility, and an extension for producing datasets via R.

The purpose of QuEval is to provide a framework for evaluating high-dimensional indexing techniques for various application scenarios. This paper discusses the use cases for high-dimensional indexing in various domains.

This paper conducted an empirical evaluation of QuEval using datasets from motivating examples. They selected different index structures such as R-Tree, k-d tree, and VA File, as well as Pyramid Technique and Prototype-based approach. The evaluation setup includes parameters of index structures and stochastic data distributions. The evaluation metrics used to compare the index structures include Tresp, Acc, Nhdd, Mheap, and Tbuild.

Motivating examples and empirical evaluations in this paper demonstrate the usefulness of the framework in various domains.

Overall, this paper has the potential to aid researchers and practitioners in selecting and optimizing suitable index structures for their data-intensive systems.

Specify the major reason(s) for acceptance/resubmit/reject.

Here are some reasons why I accept this paper:

1. **Novelty:** The paper presents a new open-source framework called QuEval, which addresses the challenge of evaluating query performance for high-dimensional data. It offers a flexible and extendable environment to address the challenge of choosing the appropriate index structure and gives specific use case requirements and resolved research questions.
2. **Relevance:** The paper addresses a significant problem faced by software engineers, data scientists, and researchers in efficiently accessing high-dimensional data, particularly for similarity or range queries.
3. **Usefulness:** The paper demonstrates the usefulness of the QuEval framework in optimizing index performance through empirical evaluations. It provides recommendations for selecting the best index structure for specific use cases, which can help software engineers create new solutions or

optimize existing ones. It also provides a detailed explanation of parameters and factors related to indexing high-dimensional spaces, which can be useful for non-database experts.

4. Extensibility: QuEval offers extensibility options for integrating additional index structures or distance metrics, making it a valuable tool for database researchers.
5. Collaboration: The authors invite the community to design and implement new index structures, tailor existing approaches, or provide new/improved implementations of existing index structures that can be benchmarked using QuEval.

Overall, the paper presents a valuable tool for evaluating index performance for high-dimensional data and provides recommendations for selecting the best index structures for specific use cases. The paper is well-written, easy to follow, and can be useful for a broad range of audiences.

Itemize specific revisions needed/suggested.

Suggested revisions:

In the Abstract section, clarify the meaning of “indexes” in the abstract (e.g., “data indexes” or “search indexes”).

Considering adding a sentence to summarize the empirical evaluation results.

In the introduction section, consider these steps:

1. The introduction could benefit from a brief overview of the existing literature on high-dimensional index structures and their limitations, which provide context for the proposed framework.
2. The section on motivating examples could benefit from a clearer connection to the proposed framework and its objectives.
3. The evaluation procedure could be more clearly defined and explained, with specific details on the experimental setup and methodology used in the case study.
4. Including a brief summary of the main results and conclusions of the study would enhance the clarity of the paper. It can provide readers with a quick understanding of the key findings and implications of the research.

The result section completely shows the complexity and scope of the study and presents the key findings of the study in a clear and concise manner, using tables, figures, and graphs to illustrate the data.

However, excessive examples make the paper appear cluttered and difficult to read. I suggest presenting your results in a more concise manner.

In the Conclusion and Future Work try to clarify the contribution of QuEval: The first sentence of the conclusion states “Due to versatile characteristics of indexing high-dimensional data, choosing a suitable index structure for a specific use case is very difficult”. However, it is not clear how QuEval addresses this challenge. Consider adding a sentence or two to explain how QuEval helps researchers and practitioners to choose suitable index structures.