

## REVIEWER'S REPORT

Manuscript No.: IJAR-58896

**Title: Query-Type Sensitivity and Selective Benefits of Hybrid Lexical–Semantic Retrieval: An Experimental Analysis with BM25 and Multilingual SBERT .**

### Recommendation:

Accept as it is .....

Accept after minor revision.....

Accept after major revision .....yes.....

Do not accept (*Reasons below*) .....

| Rating         | Excel. | Good | Fair | Poor |
|----------------|--------|------|------|------|
| Originality    |        |      | Y    |      |
| Techn. Quality |        |      | Y    |      |
| Clarity        |        |      | Y    |      |
| Significance   |        |      | y    |      |

**Reviewer's ID: JPR-Dr.Shaweta Sachdeva**

### Detailed Reviewer's Report

1. The manuscript addresses a relevant problem in information retrieval, particularly the heterogeneous performance of hybrid lexical–semantic retrieval across different query types. However, the novelty and specific contribution of the study should be articulated more clearly in comparison with existing hybrid retrieval and query-adaptive retrieval studies.
2. The experimental benchmark contains only 50 queries and 19 documents, resulting in a relatively small evaluation setting. The authors should provide a stronger justification for the dataset size and discuss how the limited benchmark may affect the reliability and generalizability of the findings.
3. The query-type distribution is highly unbalanced, with only 2 comparative and 1 definitional query, while other categories contain substantially more queries. The authors should explain the rationale behind this taxonomy and consider whether conclusions based on such small subgroups are meaningful.
4. The manuscript relies on a fixed  $\alpha = 0.5$  as the primary hybrid configuration. Although the authors appropriately state that this is not necessarily a universal optimum, a stronger experimental justification or comparison with an optimized/adaptive fusion strategy would improve the study.
5. The use of query-wise min–max normalization should be justified more extensively. The authors should discuss the sensitivity of the results to alternative normalization techniques, since normalization can substantially influence lexical–semantic score fusion.
6. Only one multilingual SBERT model, paraphrase-multilingual-mpnet-base-v2, is evaluated. Including additional multilingual or state-of-the-art dense retrieval models would strengthen the robustness of the conclusions.
7. The study reports descriptive performance but does not perform statistical significance testing. The authors should consider appropriate statistical tests or confidence intervals to determine whether the observed differences between BM25, SBERT, RRF, and the hybrid approach are statistically meaningful.

## REVIEWER'S REPORT

8. The reported overall improvement of the hybrid model over BM25 is relatively modest, while the large gain for PARAPHRASE queries appears to contribute substantially to the overall result. The authors should provide a more detailed analysis of how individual queries influence the aggregate nDCG@5.
9. The manuscript should provide more details about the construction of the relevance ground truth, including how the 73 PRIMARY positive judgments were generated, who performed the judgments, whether multiple annotators were involved, and whether inter-annotator agreement was measured.
10. The reproducibility of the experiments could be improved. Details such as BM25 parameter settings, SBERT implementation/version, preprocessing procedures, hardware/software environment, and random seeds should be reported.
11. The rationale for selecting RRF with  $k = 60$  should be explained. Since RRF performance can vary with the value of  $k$ , evaluating additional values would provide a more comprehensive comparison.
12. The manuscript uses nDCG@5 as the primary metric and Precision@5 and Recall@5 as secondary metrics. The authors should report the secondary metrics in the results section rather than merely stating that they were retained.
13. Figure 1, Figure 2, and Figure 3 should be made more informative by providing clear axis labels, legends, readable font sizes, and numerical values where appropriate. The figures should also be sufficiently high resolution for publication.
14. The discussion appropriately avoids making causal claims regarding why hybrid retrieval performs better for paraphrase queries. Nevertheless, the authors could strengthen the discussion by providing qualitative examples of representative queries where semantic retrieval contributes additional relevant documents.
15. The fusion-weight sensitivity analysis is useful, but the interpretation should be made more rigorous. The authors should clarify whether the optimal  $\alpha$  values were determined independently for each query type and whether selecting the best value from the same evaluation set introduces potential overfitting.
16. The manuscript would benefit from a per-query result table containing BM25, SBERT, RRF, and hybrid nDCG@5 values. This would make the claim of selective benefit more transparent and independently verifiable.
17. The conclusion is appropriately bounded by the limitations of the benchmark. However, the practical implications of the findings could be developed further, particularly regarding how query-type information might be incorporated into future adaptive retrieval or dynamic fusion systems.
18. The reference section should be carefully checked for citation-reference consistency, formatting, DOI accuracy, and completeness. In particular, references cited in the methodology should clearly correspond to the appropriate retrieval models and techniques.
19. There are several presentation and formatting issues, including inconsistent spacing, punctuation, capitalization, and section labeling. For example, "Resultats" should be corrected to "Results," and the manuscript should be carefully proofread for typographical and grammatical errors.
20. The keywords section currently appears as "Key words: -" and should be corrected to provide appropriate keywords consistent with the research topic. The manuscript later lists relevant terms, which should be formatted consistently.

Overall reviewer recommendation

Major Revision

# International Journal of Advanced Research

**Publisher's Name: Jana Publication and Research LLP**

*www.journalijar.com*

---

## **REVIEWER'S REPORT**

The manuscript presents an interesting and potentially useful analysis of query-type sensitivity in hybrid lexical-semantic retrieval. The emphasis on query-level behavior rather than only aggregate retrieval effectiveness is a positive aspect. However, the small benchmark, highly imbalanced query categories, limited model comparisons, absence of inferential statistical analysis, and insufficient details concerning relevance judgments and reproducibility need to be addressed before the work can be considered for publication.