

LuminaCite: A Hybrid Knowledge Discovery and Information Retrieval System for Explainable Interdisciplinary Literature Discovery

Müge Akbulut¹ ^a

¹*Department of Information Management, Faculty of Humanities and Social Sciences,
Ankara Yıldırım Beyazıt University, Ankara, Türkiye
mugeakbulut@aybu.edu.tr*

Keywords: Knowledge Discovery, Information Retrieval, Hybrid KD+IR, Explainable Ranking, Relevance–Diversity Trade-off, Topic Modeling, Citation Analysis, Pennant Retrieval, Scholarly Literature Discovery, DergiPark.

Abstract: Scholarly literature discovery requires both focused retrieval and the identification of hidden, cross-disciplinary connections. This paper presents LuminaCite, an explainable hybrid knowledge discovery and information retrieval (KD+IR) platform for exploratory scholarly search in a large Turkish digital library corpus. The system combines Latent Dirichlet Allocation (LDA) as an interpretable topic-based discovery layer with pennant retrieval as a citation-based contextual ranking layer. A user-controlled weighting parameter exposes the relevance-diversity trade-off and displays LDA, pennant, and integrated scores at the result level. LuminaCite was implemented over 633,931 Turkish scholarly articles and 3,768,409 citation links from DergiPark. A scenario-based evaluation with two representative scholarly search tasks compares LDA, pennant retrieval, and the integrated model using NDCG@10, Precision@10, Shannon diversity, and unique topic counts. The results show that pennant retrieval yields the strongest local relevance scores, LDA yields the broadest topical diversity, and the integrated model provides a balanced, transparent compromise between the two objectives.

1 INTRODUCTION

The sustained growth of scholarly publishing has made literature discovery a knowledge discovery problem as much as a retrieval task. Researchers need relevant, current and diverse sources, often across disciplinary boundaries. Relevance is context-dependent and shaped by the user's task, expectations and stage of research (Borlund, 2003; Mizzaro, 1998; Saracevic, 2007). A useful scholarly discovery system should therefore combine complementary evidence and make its ranking logic understandable and adjustable.

This need has become more visible with generative artificial intelligence. Fluent summaries do not remove the need for traceable access to peer-reviewed sources. Researchers still need to locate papers, inspect their relationships and move from direct topical relevance toward broader contextual orientation.

LuminaCite addresses this need by combining LDA, which captures latent thematic structure in titles and abstracts, with citation-based pennant retrieval, which uses co-citation patterns to provide

contextual depth. In this hybrid Knowledge Discovery and Information Retrieval (KD+IR) design, KD uncovers hidden or potentially useful connections in scholarly data, while IR retrieves and ranks scholarly works in response to a user query. A user-controlled weight makes the relevance–diversity trade-off explicit and adjustable (Figure 1). The underlying incremental-refinement framework was evaluated on the iSearch benchmark across 65 queries and shown to be robust to hyperparameter variation, first-stage retriever choice and query reformulation (Akbulut & Tonta, 2026). The present paper focuses on its implementation as an interactive and explainable platform for the DergiPark collection.

The study asks: How does an explainable hybrid ranking model behave across discovery-oriented exploration and in-depth topic analysis when users can adjust the relevance–diversity balance?

^a  <https://orcid.org/0000-0003-0026-6485>

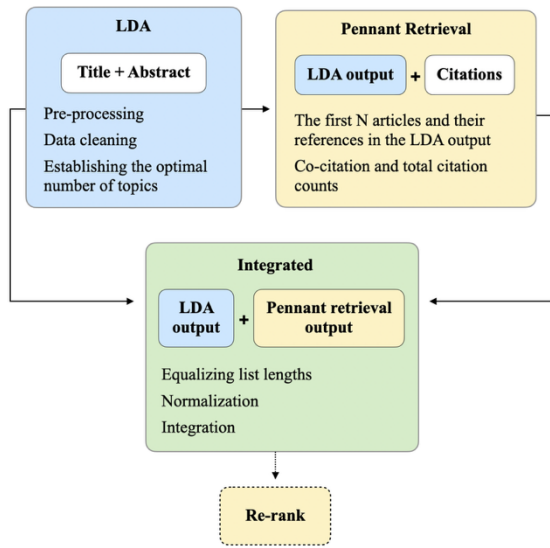


Figure 1: Integrated multi-stage ranking workflow of LuminaCite.

2 RELATED WORK AND RESEARCH MOTIVATION

Scholarly search requires both relevance and diversity. Early-stage exploration benefits from breadth and novelty, whereas in-depth topic analysis requires stronger topical and contextual concentration. Topic models such as LDA remain useful because they scale well and provide interpretable document-topic distributions (Blei, 2012; Blei et al., 2003). Their bag-of-words assumptions, however, limit semantic depth and may blur synonymous or polysemous concepts.

Citation-based retrieval provides a complementary signal because scholarly links capture relationships that lexical similarity alone cannot represent. Citation-based pennant retrieval is grounded in Relevance Theory and uses co-citation evidence to balance cognitive effect and processing effort (Sperber & Wilson, 1995; White, 2007). It can reveal contextually important papers around a seed set, but it is affected by citation lag, sparse networks and dependence on the initial seeds.

Recent neural retrievers, including SentenceBERT and SPECTER, can improve semantic representation (Cohan et al., 2020; Reimers & Gurevych, 2019). LuminaCite retains LDA as the first-stage retriever because interpretability, local computation and transparent score decomposition are central design requirements. The architecture is

modular, so alternative first-stage retrievers and fusion methods can be evaluated later.

The motivation is therefore not to treat LDA and pennant retrieval as competing methods. LDA contributes thematic breadth, while pennant retrieval adds citation-based contextual depth. Their explicit integration supports incremental refinement and boundary-spanning discovery, consistent with the larger-scale framework reported by Akbulut and Tonta (2026). This emphasis on indirect, boundary-spanning connections is also consistent with literature-based discovery, where valuable relationships may remain outside direct topical overlap (Swanson, 1986).

3 MATERIALS AND METHODS

3.1 Dataset

The dataset contains 633,931 scholarly articles from DergiPark, represented by titles and abstracts, together with 3,768,409 citation links. DergiPark is a national platform operated by TÜBİTAK ULAKBİM that provides hosting and editorial process management services for scholarly journals published in Türkiye. Its disciplinary breadth and Turkish-language content provide a realistic environment for interdisciplinary discovery in a morphologically rich language.

3.2 Text Preprocessing

Titles and abstracts were lowercased, tokenized and cleaned by removing punctuation, special characters and Turkish stop words using NLTK resources (Bird et al., 2009). Articles with fewer than three words after preprocessing were removed. Conservative morphological normalization reduced inflectional sparsity without aggressively stemming domain terminology. Terms occurring in more than 95% or fewer than 1% of articles were excluded.

3.3 Topic Modeling and Model Selection

The first stage uses LDA trained with online variational Bayes in scikit-learn (Hoffman et al., 2010; Pedregosa et al., 2011). Topic-number selection and convergence were examined before training the operational model.

A central methodological issue in LDA is the selection of the number of topics, denoted by K . To determine an appropriate value, a multi-metric model selection procedure was performed on a 50,000-

document subset. Ten candidate topic numbers between 20 and 200 were tested. The analysis used a combination of perplexity, coherence-based measures and topic-number diagnostics inspired by earlier work (Arun et al., 2010; Cao et al., 2009; Deveaud et al., 2014; Griffiths & Steyvers, 2004). This multi-metric strategy was adopted because no single criterion is sufficient on its own for selecting a stable and interpretable topic structure.

The results indicated that $K=140$ offered the most balanced solution. For this value, the C_v coherence score was 0.3223, while the highest observed C_v value, obtained at $K=200$, was 0.3339. The difference was therefore relatively small. At the same time, $K=140$ reduced training time compared with $K=200$ and preserved better interpretability by avoiding overly fragmented topic structure. For these reasons, $K=140$ was selected as the operational setting for the full corpus (Table 1 and Figure 2).

Table 1: Multi-metric topic number selection results.

K	Perplexity	C_v coherence	U_{mass} coherence	Training time (s)
20	621.85	0.2994	-3.77	265
40	693.09	0.3126	-3.67	282
60	756.17	0.2979	-3.80	320
80	798.27	0.3080	-3.77	358
100	817.97	0.3128	-3.75	431
120	864.78	0.3160	-3.74	569
140	877.01	0.3223	-3.67	552
160	927.74	0.3250	-3.71	588
180	955.87	0.3292	-3.67	576
200	930.70	0.3339	-3.62	592

The final model was trained with $max_{iter}=50$ and $learning_{decay}=0.7$. Because convergence is an important practical concern in large-scale topic modeling, an additional convergence analysis was conducted on a 50,000-abstract subset. Separate LDA models were trained with 50, 100, 500 and 1000 iterations. The results suggested that the model had already reached a stable range at 50 iterations. The perplexity change over the final ten iterations was 0.0182% and the difference between the 50-iteration run and the 500–1000 iteration range remained modest (Figure 3). This result supports the claim that the chosen configuration offers a reasonable balance between computational cost and model stability.

Once the model was trained, each user query was projected into the same topic space. Query-to-paper

similarity was then calculated to produce the initial ranking.

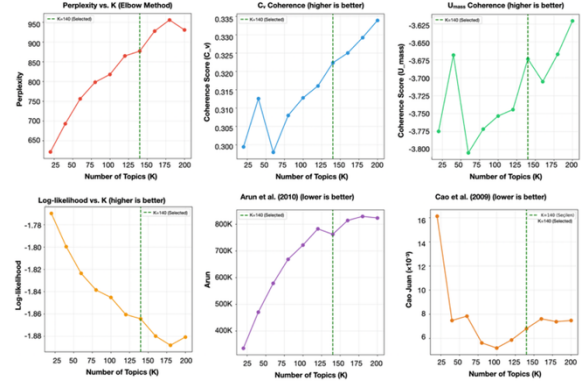


Figure 2: Multi-metric evaluation for topic number selection.

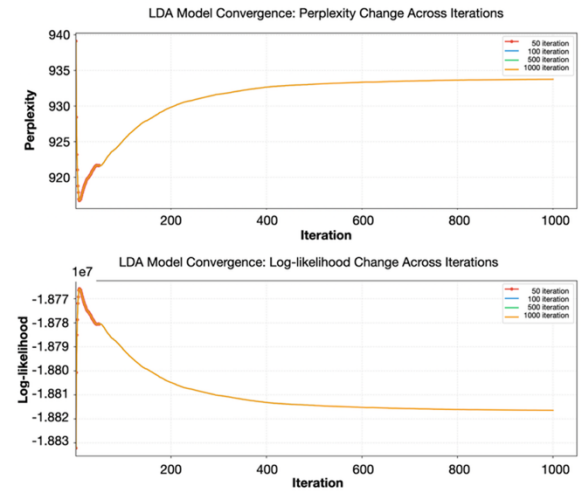


Figure 3: Convergence analysis of the LDA model.

3.4 Pennant-Based Re-Ranking

The second stage incrementally refines the LDA ranking through citation-based pennant retrieval grounded in Relevance Theory and bibliometric context (Sperber & Wilson, 1995; White, 2007). The top five LDA results form the seed set (Figure 1). This operational choice followed preliminary inspection: five seeds provided sufficient citation evidence while limiting topic drift in a collection with uneven citation density. Larger seed sets, such as 10 or 15, may increase coverage but may also introduce weaker thematic anchors.

For each candidate paper, co-citation frequency with the seed set provides local evidence of cognitive effect, while total citation count reflects processing effort and broader visibility. Pennant retrieval therefore adds citation-based contextual depth to the

thematic ranking rather than replacing the initial LDA output.

3.5 Score Integration

The final stage combines min-max normalized LDA and pennant scores in the $[0, 1]$ range:

$$S_{\text{final}} = (1 - \lambda) S_{\text{LDA}} + \lambda S_{\text{Pennant}} \quad (1)$$

Here, λ is user-controlled. At $\lambda = 0$, the ranking is entirely LDA-based; at $\lambda = 1$, it is entirely pennant-based. Intermediate values make the relevance–diversity trade-off explicit and support user-driven personalization. Papers appearing in only one candidate list receive a zero score for the missing component before fusion. Figure 4 illustrates the normalization and fusion workflow.

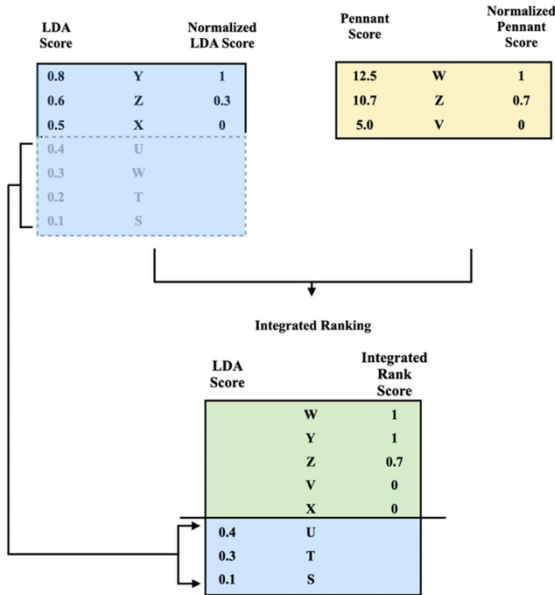


Figure 4: Score normalization and integration workflow.

Figure 5 presents the three modular layers of LuminaCite. Flask manages requests and query processing. scikit-learn, NLTK and NumPy support topic modeling, citation-based re-ranking, normalization and integration (Bird et al., 2009; Harris et al., 2020; Pedregosa et al., 2011). SQLite and JSON store structured records, article metadata and intermediate objects. Trained models and processed structures are cached for reuse.

Figure 6 shows the interface, which displays LDA, pennant and integrated scores together with topic filters and pennant-based visualizations. Users can inspect why papers appear together and observe

how rankings change as λ is adjusted. Turkish queries are supported with or without diacritics.

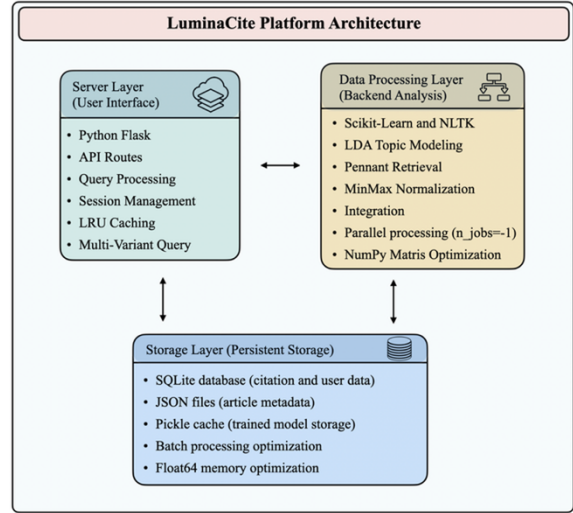


Figure 5: LuminaCite platform architecture.

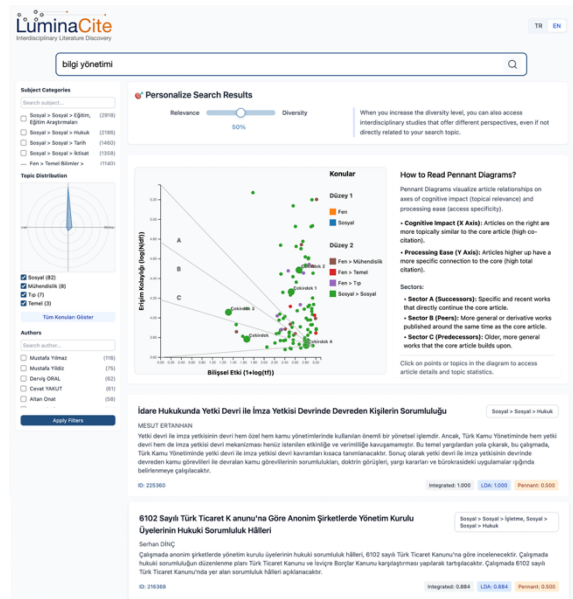


Figure 6: Search interface and ranking output screen.

4 SYSTEM DEMONSTRATION

4.1 Scenarios

The system demonstration uses two deliberately contrasting scenarios. They are illustrative cases selected to contrast different information needs, not a statistically representative benchmark query set.

Scenario 1 uses the query 'explainable artificial intelligence' and represents discovery-oriented exploration, where topical breadth and potentially unexpected connections are valuable. Scenario 2 uses 'privacy-preserving federated learning methods in health data' and represents in-depth topic analysis, where stronger contextual concentration is prioritized while adjacent discovery remains useful.

Ranking behavior was summarized using NDCG@10 and P@10 as local relevance-oriented diagnostics and Shannon diversity and the number of unique topics as topical diversity indicators. The relevance reference was derived from degree centrality within each query-specific pennant subgraph. Because this proxy is based on the same citation structure used by pennant retrieval, the comparison is internally consistent but not independent. The values should therefore be interpreted as local diagnostics rather than external validation.

$$H' = - \sum_{i=1}^s p_i \ln(p_i) \quad (2)$$

5 RESULTS

5.1 Scenario-Level Performance

Table 2 summarizes the two illustrative scenarios. In both, pennant retrieval yields the strongest relevance-oriented values, whereas LDA provides the greatest topical diversity. The integrated ranking occupies an intermediate position. This is consistent with its intended role: it does not dominate both components on every measure, but preserves much of the citation-based concentration while retaining more topical breadth than pennant retrieval alone.

Table 2: Comparative performance of LDA, pennant retrieval and the integrated model.

Scenario	Method	NDCG @10	P@10	Shannon H'	Topics
Discovery-oriented exploration	LDA	0.68	0.55	2.8	9
Discovery-oriented exploration	Pennant	0.82	0.75	1.6	4
Discovery-oriented exploration	Integrated	0.79	0.70	2.4	7
In-depth topic analysis	LDA	0.72	0.60	3.0	8
In-depth topic analysis	Pennant	0.88	0.80	1.9	5
In-depth topic analysis	Integrated	0.85	0.78	2.6	7

For discovery-oriented exploration, the integrated ranking improves the relevance-oriented values relative to LDA while retaining substantially greater diversity than pennant retrieval. The same pattern is observed for in-depth topic analysis. These observations demonstrate the intended relevance–diversity trade-off, but they should not be generalized beyond the two cases without broader independent evaluation.

6 DISCUSSION

LuminaCite operationalizes a long-standing tension in scholarly search. Highly concentrated rankings may omit useful conceptual alternatives, whereas highly diverse rankings may dilute focus. Its contribution is an explainable, user-controlled mechanism that lets researchers move between thematic breadth and citation-based contextual depth rather than accepting a fixed, opaque ranking.

The two components are complementary. LDA supports orientation across themes; pennant retrieval concentrates the result set around a citation-defined context. A weighted linear combination makes the relevance–diversity trade-off visible and adjustable. Alternatives such as Reciprocal Rank Fusion, learning-to-rank, or neural fusion may produce different ranking behavior, but would require independent training data and evaluation. They are appropriate extensions rather than claims tested in the present system paper.

The system also demonstrates the feasibility of hybrid scholarly discovery over a large Turkish

collection. This matters because much retrieval research remains centered on English-language benchmarks. LuminaCite does not attempt to compete with large language models on text generation. Its role is to provide document-grounded, inspectable discovery. The broader retrieval framework and its robustness were evaluated independently on iSearch (Akbulut & Tonta, 2026).

6.1 Relationship to Prior Validation

The DergiPark scenarios reported here are deliberately limited to demonstrating the behavior of the implemented platform. The retrieval architecture builds on the author’s doctoral research and an earlier method proposal (Akbulut, 2022; Akbulut & Tonta, 2022) and it should be interpreted together with the independent evaluation of the underlying ranking framework reported by Akbulut and Tonta (2026). That study used the iSearch test collection, which contains approximately 435,000 physics papers, 3.7 million citation links and 65 realistic search topics with expert relevance assessments. The framework was evaluated through relevance, diversity and boundary-spanning indicators, including degree centrality, arXiv subject categories, maximal marginal relevance and normalized discounted cumulative gain.

The prior study also examined sensitivity to key design choices. Its robustness analyses considered hyperparameter variation, first-stage retriever choice and query reformulation. These results provide evidence for the retrieval framework itself, whereas the present paper contributes a different layer of evidence: implementation on a large national scholarly collection, transparent score presentation, interactive weighting and user-facing visual explanation. The two studies therefore address related but non-identical questions and should not be treated as duplicate evaluations.

This distinction also clarifies the role of the two DergiPark scenarios. Their purpose is to show how the same ranking architecture behaves in contrasting use contexts and how the interface communicates that behavior. They do not replace benchmark-based evaluation, nor do they establish statistical generalizability for DergiPark. Instead, they connect the previously validated method to a deployable platform and expose practical issues that are less visible in an offline benchmark, including multilingual preprocessing, uneven citation density, caching, score inspection and direct user control of the relevance–diversity balance.

6.2 Design Implications

The implementation suggests three design implications for scholarly discovery systems. First, explainability can be provided without claiming that every component is intrinsically interpretable in a complete cognitive sense. In LuminaCite, explanation is operational: users can inspect component scores, view citation-based relationships and observe how the final ranking changes when the weighting parameter changes. This makes the ranking process traceable at the result level and allows users to connect interface actions with changes in retrieval behavior.

Second, diversity should not be treated as a fixed system objective. Different stages of research require different balances. A researcher entering an unfamiliar area may value a wider topical spread, while a researcher conducting in-depth topic analysis may prefer a tighter citation-defined neighborhood. The slider does not infer these needs automatically; instead, it supports user-driven personalization by allowing researchers to adjust the relevance–diversity balance themselves. The trade-off is explicit and reversible. Users can move between settings, compare rankings and decide which configuration better fits the current task.

Third, modularity is important for future development. Alternative first-stage retrievers such as SciBERT, ColBERT and SPECTER and fusion methods such as Reciprocal Rank Fusion, learning-to-rank and neural fusion, are compatible with the architecture, but they represent new experimental conditions rather than minor implementation substitutions. Replacing LDA would change the first-stage representation, while replacing linear weighting would change the fusion objective and the meaning of the slider. Each alternative therefore requires independent tuning, comparison and relevance judgments. The current version retains a transparent baseline so that the contribution of citation-based refinement remains visible and reproducible.

The same reasoning applies to seed-set size. Five seeds were used because the operational DergiPark network has uneven citation density and because preliminary inspection indicated that a small thematic core reduced drift. Increasing the set to 10 or 15 could improve citation coverage for some topics, but it could also introduce weaker initial matches and broaden the co-citation neighborhood beyond the intended query context. A defensible comparison should therefore test several seed sizes across many queries and report both relevance and diversity effects.

7 LIMITATIONS AND FUTURE WORK

The main limitation is evaluation scope. The two scenarios demonstrate platform behavior but do not provide statistical evidence across heterogeneous queries. In addition, degree centrality in the query-specific pennant subgraph is method-related and cannot serve as fully independent ground truth. The reported values should therefore be read as local diagnostics, not as definitive effectiveness estimates.

Future evaluation should use a larger randomly sampled query set, independent relevance judgments and preferably blinded assessment by domain experts. It should also compare seed-set sizes, test contextual retrievers such as Sentence-BERT, SciBERT, ColBERT, or SPECTER and examine fusion methods including RRF and learning-to-rank. User studies are needed to evaluate whether the slider and explanations improve task performance, trust and discovery. Citation lag, temporal dynamics and cold-start behavior for recent articles also remain open issues.

8 CONCLUSION

This paper presented LuminaCite, an explainable hybrid KD+IR platform applied to a large Turkish scholarly collection. LDA provides thematic breadth, citation-based pennant retrieval provides contextual depth and a user-controlled weight enables researchers to personalize relevance rankings by adjusting the relevance–diversity balance.

The two illustrative scenarios show the expected complementary pattern: LDA yields greater topical diversity, pennant retrieval yields stronger relevance-oriented values and the integrated ranking provides a transparent and controllable balance. These cases demonstrate platform behavior rather than establish general effectiveness. Larger independent evaluations are required.

Together with the benchmark-scale validation reported by Akbulut and Tonta (2026), the DergiPark implementation shows how an interpretable hybrid method can be translated into a working discovery environment. Such systems remain valuable where traceability, user control, multilingual scholarly infrastructure and interdisciplinary exploration matter.

ACKNOWLEDGEMENTS

This study is one of the outputs of the project titled Incremental Refinement of Relevance Rankings in Machine Learning-Based Information Retrieval, supported by the TÜBİTAK 2218 National Postdoctoral Research Fellowship Program and hosted by MANTIS Software. The work is based in part on the author's doctoral dissertation completed at Hacettepe University. The author gratefully acknowledges Prof. Dr. Yaşar Tonta for his supervision, guidance and support throughout the doctoral research on which this study builds.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process. During the preparation of this work, the author used ChatGPT (OpenAI) to improve the language and readability of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

Availability of Data and Code. The platform source code is available at <https://github.com/mugeakbulut/LuminaCite>. The DergiPark data were provided by TÜBİTAK ULAKBİM for local research use and cannot be redistributed publicly.

REFERENCES

- Akbulut, M. (2022). *Bilgi erişimde ilgi sıralamalarının artırımı olarak geliştirilmesi* [Unpublished doctoral dissertation]. Hacettepe University. https://bby.hacettepe.edu.tr/yayinlar/Muge_Akbulut_PhD_Tez.pdf
- Akbulut, M., & Tonta, Y. (2022). İlgi sıralamalarının artırımı olarak geliştirilmesi: Pennant erişimle desteklenen yeni bir yöntem önerisi. *Türk Kütüphaneciliği*, 36(2), 169–203. <https://doi.org/10.24146/tk.1062751>
- Akbulut, M., & Tonta, Y. (2026). Incremental refinement of relevance rankings: Balancing relevance depth and scope. *Journal of the Association for Information Science and Technology*, 1–24. <https://doi.org/10.1002/asi.70101>
- Arun, R., Suresh, V., Madhavan, C. E. V., & Murthy, M. N. (2010). On finding the natural number of topics with latent Dirichlet allocation: Some observations. In M. J. Zaki, J. X. Yu, B. Ravindran, & V. Pudi (Eds.),

- Advances in Knowledge Discovery and Data Mining* (LNCS, Vol. 6118, pp. 391–402). Springer. https://doi.org/10.1007/978-3-642-13657-3_43
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://dl.acm.org/doi/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. <https://www.jmlr.org/papers/v3/blei03a.html>
- Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913–925. <https://doi.org/10.1002/asi.10286>
- Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, 72(7–9), 1775–1781. <https://doi.org/10.1016/j.neucom.2008.06.011>
- Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020). SPECTER: Document-level representation learning using citation-informed transformers. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the Association for Computational Linguistics* (pp. 2270–2282). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.207>
- Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique*, 17(1), 61–84. https://shs.cairn.info/article/DN_171_0061
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1), 5228–5235. <https://doi.org/10.1073/pnas.0307752101>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hoffman, M., Bach, F. R., & Blei, D. M. (2010). Online learning for latent Dirichlet allocation. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 23* (pp. 856–864). Curran Associates. https://proceedings.neurips.cc/paper_files/paper/2010/file/71f6278d140af599e06ad9bflba03cb0-Paper.pdf
- Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting with Computers*, 10(3), 305–320. [https://doi.org/10.1016/S0953-5438\(98\)00012-5](https://doi.org/10.1016/S0953-5438(98)00012-5)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1410>
- Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 2126–2144. <https://doi.org/10.1002/asi.20681>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Blackwell.
- Swanson, D. R. (1986). Undiscovered public knowledge. *Library Quarterly*, 56(2), 103–118. <https://doi.org/10.1086/601720>
- White, H. D. (2007). Combining bibliometrics, information retrieval and relevance theory, part 1: First examples of a synthesis. *Journal of the American Society for Information Science and Technology*, 58(4), 536–559. <https://doi.org/10.1002/asi.20543>