Evaluating the Impact of Content Representations on Retrieval Performance in Wikibase RAG

Kolja Bailly[1,*], Lozana Rossenova[1] and Ina Blümel[1]

[1]TIB Leibniz Information Centre for Science and Technology - Open Science Lab, Welfengarten 1B, Hannover, 30167, Germany

Abstract

Wikibase has become one of the leading open-source platforms for publishing and managing Linked Open Data (LOD), serving as the foundation for numerous research infrastructures, digital collections, and domain-specific knowledge graphs. As Large Language Models (LLMs) are increasingly combined with Retrieval-Augmented Generation (RAG), understanding how knowledge representation affects RAG performance becomes an important challenge for designing efficient AI-assisted research infrastructures.

Previous work has demonstrated that automated hyperparameter optimization can substantially improve retrieval performance for document-oriented knowledge bases such as MediaWiki by selecting retrieval configurations based on retrieval quality metrics. However, it remains unclear how the underlying representation of knowledge influences retrieval optimization and whether optimization strategies transfer equally well across structured and unstructured knowledge sources.

In this paper, we investigate the impact of knowledge representation for fact retrieval within a RAG pipeline by comparing different textual representations generated from Wikibase entities. Using the same AutoRAG optimization framework and identical benchmark questions, we evaluate retrieval performance and analyze how optimal retrieval configurations differ across these representations. Retrieval quality is measured using Mean Reciprocal Rank (MRR), enabling a controlled comparison focused exclusively on retrieval performance without the influence of language-model answer generation.

The results provide insights into the trade-offs between Linked Open Data and textual knowledge representations for RAG systems, quantify the impact of additional narrative context, and identify how knowledge representation influences optimal retrieval strategies. These findings contribute to the development of efficient, domain-specific RAG systems.

Keywords

RAG, MediaWiki, Wikibase, Semantic Mediawiki, Semantic Wikibase, Retrieval, Hyperparameter Optimization,

CEUR-WS.org/Vol-4312/short38.pdf

1. Introduction

Knowledge graphs have become an important infrastructure for managing and publishing structured research data. Platforms such as Wikibase enable institutions to build domain-specific knowledge graphs based on Linked Open Data principles, while frameworks such as Wikibase4Research facilitate their deployment in scientific environments [1, 2].

Retrieval-Augmented Generation (RAG) enables large language models to exploit such knowledge sources, but retrieval quality depends strongly on the chosen retrieval configuration, including embedding models, chunking strategies, and ranking parameters [3]. Since identifying suitable configurations requires extensive experimentation, automated optimization approaches such as AutoRAG systematically evaluate parameter combinations to maximize retrieval performance while reducing manual effort [4].

While retrieval optimization identifies effective retrieval configurations, an open question remains: **Does the representation of the knowledge itself influence retrieval performance?** Unlike traditional document collections, knowledge graphs expose identical information through multiple representations. A Wikibase entity may be represented as structured JSON, rendered HTML, generated

RAGE-KG 2026: The Third International Workshop on Retrieval-Augmented Generation Enabled by Knowledge Graphs, co-located with ISWC 2026, October 25–26, 2026, Bari, Italy

*Corresponding author.

kolja.bailly@tib.eu (K. Bailly); lozana.rossenova@tib.eu (L. Rossenova); ina.bluemel@tib.eu (I. Blümel)

0009-0001-3533-0275 (K. Bailly); 0000-0002-5190-1867 (L. Rossenova); 0000-0002-3075-7640 (I. Blümel)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-04

text, or human-enriched wiki pages. Although these representations describe the same underlying knowledge, they provide substantially different retrieval signals.

If retrieval quality depends on representation, selecting an appropriate representation becomes a fundamental design decision for knowledge graph-based RAG systems rather than merely a preprocessing step.

In this work, we use AutoRAG [4] to evaluate four representations of identical Wikibase entities under identical retrieval conditions. By optimizing each representation independently, we investigate how representation influences retrieval performance, optimization behavior, and computational cost.

The contributions of this paper are:

- A controlled methodology for evaluating retrieval performance across alternative representations of identical Wikibase knowledge.
- An empirical comparison of four Wikibase-derived representations using automated retrieval optimization.
- Evidence that representation choice may fundamentally influence retrieval performance, optimization behavior, and computational resource requirements.

2. Methodology

2.1. Experiment description

To analyze the influence of knowledge representation on retrieval performance, we design an experiment in which identical Wikibase knowledge is transformed into different retrieval representations. The experiment keeps the underlying knowledge source, benchmark questions, retrieval evaluation procedure, and optimization framework constant while varying only the representation provided to the retrieval pipeline.

The central hypothesis of this study is that different representations of the same knowledge lead to different retrieval behavior and require different retrieval configurations to achieve optimal performance. The evaluation workflow consists of the following steps:

- Create comparable Question-Context pairs from Wikibase entities (ground-truth).
- Generate alternative representations of the same knowledge.
- Apply AutoRAG and evaluate retrieval performance for each representation.

2.2. Knowledge Source and Ground Truth

The experiments are conducted on a Wikibase knowledge graph containing structured information about historical manor houses in the baltic sea region. Wikibase entities provide structured statements consisting of properties, values, qualifiers, and links to related entities. To ensure comparability between entities, we selected 15 manorhouse estate entities as a pseudo ground truth [5], providing the same information (location, list of owners, list of buildings).

The evaluation focuses on entity retrieval rather than answer generation. This approach allows retrieval quality to be measured independently of language model generation, since the objective is to determine whether the correct knowledge unit can be retrieved from the representation.

For each entity, benchmark questions are generated as reverse lookup queries: the question describes one or more characteristics of an entity, while the expected retrieval target is the corresponding Wikibase entity itself. Using this constraints, our method evaluates use cases for fact retrieval of single entities, but not retrieval of complex fact relations over multiple wikibase statements.

We create three questions for each entity, substituting the placeholder (<>) with actual values from the wikibase entity:

- Which estate is located in <LOCATION>?
- Which estate has the following buildings: <HAS_PART>?
- Which estate was owned by <OWNER>?

2.3. Knowledge Representations

To ensure comparability between representations, all evaluated representations are generated from the same set of Wikibase entities. Each representation is indexed independently, resulting in separate retrieval pipelines containing identical knowledge expressed in different forms. The Wikibase entities are converted into the following representations.

Structured JSON representation (WB_JSON) The native Wikibase entity serialization via Mediawiki API method 'wbgetentities?format=json' is used as a structured representation containing labels, descriptions, identifiers, statements, and references in JSON format. The json representation contains Wikibase Q-ids as references to other entities, e.g.: Q1, Q100

Text representation from json (WB_JSON2TEXT) A textual representation is generated by converting entity information into structured text using a template. This representation aims to provide retrieval-friendly textual context while preserving the underlying facts. The text contains the following headings: TITLE, DESCRIPTION, STATEMENTS. With statements written as triples, using the entity labels for references to other entities instead of Q-ids, e.g.'Has_owner: Carl Gustaf Piper' or 'Has_part: gardeners house'.

Text representation from rendered Wikibase UI (WB_HTML2TEXT) The publicly accessible Wikibase item page is used as a web-oriented representation containing formatted statements, labels, and interface elements. Via the python library 'BeautifulSoup', the html is automatically converted into text.

Enriched wiki representation via Semantic Wikibase. (SWB_HTML2TEXT) Semantic Wikibase is able to embed Wikibase content via Semantic Mediawiki queries into MediaWiki article pages. Such manually curated pages are used as an additional representation. These pages contain handwritten custom narrative descriptions, explanations, and contextual information in addition beyond the structured Wikibase facts, see: Example page. Via the python library 'BeautifulSoup', the html is automatically converted into text.

Table 1
Overview of evaluated knowledge representations

Representation	Source	Structured	Human enriched
WB_JSON	Wikibase API	✓	×
WB_HTML2TEXT	Wikibase frontend	partial	×
WB_JSON2TEXT	Wikibase API + custom script	partial	×
SWB_HTML2TEXT	Semantic Wikibase page	partial	✓

2.4. Retrieval Optimization

For each representation, retrieval configurations are optimized using AutoRAG and equal starting configurations. Each representations content is indexed in an empty vector database to enable efficient semantic retrieval of relevant content. Most parameters optimized in our approach affect the information extraction and indexing process, determining how textual information is segmented, enriched, and represented before retrieval. For vectorization, we used the embedding model Qwen3-embedding:0.6B.

Optimization Parameters The following parameters were optimized within the RAG Pipeline:

- Chunk Size: Defines the size of text segments that can later be retrieved individually.
- Chunk Overlap: Defines the degree of overlap between neighboring text segments.
- Top K: Defines the number of retrieved chunks considered as relevant search results.

Optimization Strategy We employed a greedy optimization strategy as suggested in Orbach et al. [6]. This iterative approach optimizes one parameter at a time. While this strategy does not guarantee globally optimal solutions, it enables sufficiently good retrieval configurations with significantly lower computational costs.

Evaluation Metric As the primary evaluation metric, we used the Mean Reciprocal Rank (MRR), which measures the ranking position of relevant retrieval results within the returned result lists. A major advantage of this methodology is that the process is performed entirely without answer generation by LLMs. We evaluate only the retrieval quality, not the quality of generated answers. This substantially reduces computational costs as well as the influence of LLM temperature and bias during answer generation. Each representation is optimized independently to avoid bias caused by using a configuration tuned for another representation. This enables the following analyses:

- Comparison of maximum achievable retrieval performance across representations.
- Comparison of optimization resource requirements across representations
- The analysis of whether different representations require different retrieval configurations.

Experimental Environment The optimization workflow was implemented within the Wikibase4Research ecosystem using LlamaIndex, Ollama, and a Docker-based Python API. The system extends the existing RAG module of Wikibase4Research, which supports MediaWiki, Wikibase, and hybrid semantic knowledge infrastructures like Semantic Wikibase. [2]

3. Results

3.1. Retrieval Performance Across Representations

Table 2
Retrieval performance across knowledge representations.

Representation	Min MRR	Median MRR	Best MRR
WB_JSON2TEXT	**0.844**	**0.917**	**0.942**
WB_HTML2TEXT	0.800	0.889	0.938
SWB_HTML2TEXT	0.694	0.833	0.886
WB_JSON	0.600	0.678	0.778

The results show substantial differences between representations derived from the same Wikibase knowledge. The automatically generated textual representation achieved the highest retrieval performance with a maximum MRR of 0.942, while the structured JSON representation achieved the lowest maximum MRR of 0.778. The generated textual representation consistently outperformed the remaining representations. In contrast, directly indexing Wikibase JSON resulted in substantially lower retrieval quality. This indicates that retrieval-oriented textual representations provide stronger signals for semantic retrieval than directly indexing structured entity data. This could be because of the usage of Wikibase Q-ids instead of entity labels in the provided json data as well as the syntactical overhead of json representation compared to the template based text representation.

3.2. Impact of Retrieval Optimization

Automated optimization improved retrieval performance for all representations. The largest relative improvement was observed for the structured JSON representation, indicating that representations with weaker retrieval signals benefit more from parameter optimization. However, optimization did not eliminate the performance differences between representations.

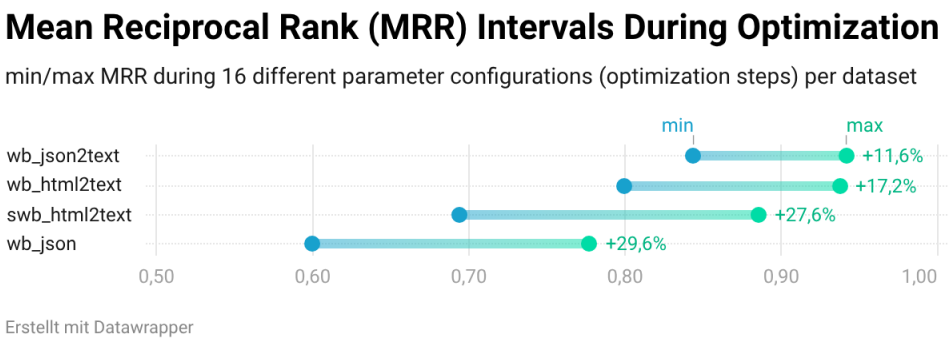


Figure 1: Range of MRR values across all evaluated retrieval configurations for each knowledge representation.

We further observed that the retrieval advantage of stronger representations was not limited to their individually optimized configurations. In several cases, a stronger representation achieved higher retrieval performance using the same retrieval configuration that was optimized for a weaker representation, outperforming the weaker representation even at its optimal configuration. This indicates that retrieval optimization can improve a given representation, but the choice of representation establishes a fundamental constraint on achievable retrieval quality.

These results indicate that representation quality has a stronger influence on retrieval performance than retrieval configuration alone. While AutoRAG can identify effective configurations for each representation, optimization cannot compensate for missing semantic context or limited retrievability introduced by the underlying representation.

Retrieval Configuration Sensitivity Textual representations showed more stable retrieval performance across configurations, while structured representations exhibited larger performance variations. This suggests that representations with explicit textual context may require less retrieval-specific tuning.

3.3. Optimization Cost

The optimization process also differed substantially in computational cost. The JSON representation required considerably more optimization time than textual representations, likely due to differences in document structure and retrieval space. This demonstrates that representation choice affects not only retrieval quality but also the effort required to identify suitable retrieval configurations.

Table 3
Optimization time (Evaluation of 16 parameter settings for 15 pages, using 3 questions (retrievals) per page).

Representation	Optimization time (min)
WB_JSON2TEXT	5
WB_HTML2TEXT	5
SWB_HTML2TEXT	23
WB_JSON	80

Limitations The benchmark intentionally focuses on entity-level fact retrieval. Consequently, the conclusions cannot yet be generalized to transitive reasoning or context-dependent question answering. The used Dataset is small and only featuring a single knowledge domain. Extended experiments are necessary to further generalize the results.

Related Work [7] Evaluated the effect of content transformations after retrieval, leading to the conclusion, that representation structure and length is not essential as input for LLM answer generation,

as long as answer retention rate in the transformation result remains at same level. [8], showing advantages in performance using structured data instead of text representations in LLM answer generation phase of RAG pipelines.

4. Conclusion

Our experiments demonstrate that retrieval performance is strongly influenced by how Wikibase knowledge is represented. While automated optimization consistently improves retrieval quality, it cannot compensate for limitations introduced by the representation itself. Consequently, selecting an appropriate retrieval representation should be considered a primary design decision in knowledge graph-based RAG systems, alongside retrieval configuration. RAG Optimization Frameworks like Wikibase4Research + AutoRAG should integrate representation selection into the optimization process.

After retrieval optimization, the manually enriched SWB_HTML2TEXT representation achieved retrieval performance quite comparable to the best performing representations WB_JSON2TEXT and WB_HTML2TEXT but required substantially more resources for retrieval optimization (and content creation). However, the enriched representation contains manually curated information that is not available in the underlying Wikibase data. Therefore, their value may become more apparent for broader question types beyond fact retrieval, such as questions requiring contextual, narrative, or domain-specific knowledge. Human-enriched representations could provide additional retrieval value, especially for knowledge-intensive tasks where structured facts alone are insufficient.

Ethics and Environmental Footprint The experiments use publicly available knowledge base content and do not process personal or sensitive data. Automated question generation and optimization may inherit biases or inaccuracies from the underlying knowledge representation, requiring careful evaluation of generated datasets. Optimization runs required an estimated 0.1–0.14 kWh of energy (2h computing time on laptop), corresponding to roughly 300-400 g CO_2 emissions depending on the electricity mix.

Future Work Future work will extend the evaluation beyond fact retrieval towards context-aware question answering, where additional narrative and domain-specific knowledge can contribute to answer quality. To achieve this, we plan to compare enriched text-based retrieval approaches, such as SWB_HTML2TEXT, with Wikibase-based GraphRAG approaches to investigate whether explicitly modeled knowledge graph relationships can complement or outperform enriched textual representations for complex knowledge queries.

Acknowledgments

Wikibase4Research developments and RAG extension are partly funded by NFDI4Culture, funded by the Deutsche Forschungsgemeinschaft (DFG) under grant no. 441958017, and the EU HORIZON project ECHOLOT funded by the European Union under Grant Agreement n.101233096.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5.5 in order to: Drafting content, Text Translation, Improve writing style, Peer review simulation. After using these tool(s)/service(s), the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. Rossenova, P. Duchesne, I. Blümel, Wikidata and wikibase as complementary research data management services for cultural heritage data. proceedings of the 3rd wikidata workshop 2022 co-located with the 21st international semantic web conference (iswc2022), ISWC2022 (2022). URL: https://doi.org/10.25968/opus-2573.

[2] L. Günther, Wikibase4research, MediaWiki Users and Developer Conference Fall 2025 (2025). URL: https://www.semantic-mediawiki.org/wiki/MediaWiki_Users_and_Developers_Conference_Fall_2025/Wikibase4Research.

[3] J. Gao, Q. H. Pham, S. Varghese, S. Saurav, V. Hoskere, Ragalyst: Automated human-aligned agentic evaluation for domain-specific rag, arxiv.org:2511.04502 (2025). URL: https://arxiv.org/abs/2511.04502.

[4] K. Bailly, L. Rossenova, I. Blümel, Efficient rag for mediawiki and wikibase: Hyperparameter tuning for semantic retrieval under resource constraints, 2026. To appear.

[5] Z. Yuheng, W. Yiran, T. Yuzhu, Z. Min, H. Yanhua, Seara: An automated approach for obtaining optimal retrievers, arxiv.org:2507.06554v2 (2025). URL: https://arxiv.org/abs/2507.06554.

[6] M. Orbach, O. Eytan, B. Sznajder, A. Gera, O. Boni, Y. Kantor, G. Bloch, O. Levy, H. Abraham, N. Barzilay, E. Shnarch, M. E. Factor, S. Ofek-Koifman, P. Ta-Shma, A. Toledo, An analysis of hyperparameter optimization methods for retrieval augmented generation, arXiv arXiv:2505.03452v3 (2025). URL: https://arxiv.org/abs/2505.03452.

[7] J. J. Ross, B. Koopman, A. van der Vegt, G. Zuccon, On the impact of retrieved content representations in rag pipelines, 2026. URL: https://arxiv.org/abs/2605.30790. arXiv:2605.30790.

[8] J. Tan, Z. Dou, W. Wang, M. Wang, W. Chen, J.-R. Wen, Htmlrag: Html is better than plain text for modeling retrieved knowledge in rag systems, in: Proceedings of the ACM on Web Conference 2025, WWW '25, ACM, 2025, p. 1733–1746. URL: http://dx.doi.org/10.1145/3696410.3714546. doi:10.1145/3696410.3714546.

5. Online Resources

The software implementations described in this paper are available as open-source code repositories under the following links:

- Wikibase4Research git repository: https://gitlab.com/nfdi4culture/wikibase4research/wikibase4research
- WikibaseRAG / AutoRAG git repository: https://gitlab.com/nfdi4culture/wikibase4research/wikibase-RAG
- Manorhouses of the baltic sea region Wiki: https://wb.manorhouses.tibwiki.io