RAG-Anything: An All-in-One Retrieval-Augmented Generation Framework

Vishal Bairagi
vishalba499@gmail.com

January 22, 2026

Abstract

Retrieval-Augmented Generation (RAG) has proven essential to extend the knowledge capabilities of large language models beyond their static training limits. However, current RAG methods are restricted by being text-only, ignoring rich multimodal information environments. In this white paper, we present **RAG-Anything**, a unified framework designed to enable comprehensive retrieval from multimodal knowledge sources. The system introduces dual-graph construction and cross-modal hybrid retrieval, achieving superior performance on benchmarks and establishing a new benchmark for multimodal knowledge access. :contentReference[oaicite:1]index=1

1 Introduction

Modern knowledge repositories contain heterogeneous multimodal data including text, images, tables, and equations. Traditional RAG frameworks largely neglect these sources, focusing on plain textual indexing and retrieval. This misalignment with real-world data leads to significant information loss and limited comprehension in complex domains. RAG-Anything addresses this gap by reconceptualizing multimodal content as interconnected knowledge entities instead of isolated data types. :contentReference[oaicite:2]index=2

2 Problem Statement

Standard RAG systems operate under the assumption that all knowledge is text-based. This assumption results in:

- Loss of visual semantics in document understanding,
- Inadequate interpretation of structured content such as tables,
- Limited extraction from mathematical or non-text elements.

These limitations make traditional frameworks unsuitable for domains where non-textual features are crucial, such as scientific literature, financial reports, and technical documentation. :contentReference[oaicite:3]index=3

3 Core Contributions

RAG-Anything incorporates:

1. **Unified Multimodal Knowledge Representation**: Decomposing documents into atomic units regardless of modality.

2. **Dual-Graph Construction**: Building separate knowledge graphs for cross-modal context and text semantics, and fusing them for unified retrieval.

3. **Cross-Modal Hybrid Retrieval**: Combining structural navigation and semantic matching to find relevant evidence across modalities.

These components enable RAG-Anything to reason over heterogeneous data and significantly outperform previous multimodal baselines. :contentReference[oaicite:4]index=4

4 Technical Framework

4.1 Multimodal Decomposition

Every document is decomposed into atomic units containing modality-specific content, e.g., text paragraphs, image metadata, structured table elements, and mathematical expressions. Each unit retains contextual integrity to preserve original semantic richness. :contentReference[oaicite:5]index=5

4.2 Dual Knowledge Graph

Two complementary graphs are constructed:

- **Cross-Modal Graph**: Represents entities derived from all modalities.
- **Text-Based Graph**: Captures conventional text relationships.

These are aligned through entity matching to generate a single, holistic representation. :contentReference[oaicite:6]index=6

4.3 Hybrid Retrieval

For an input query, the system performs:

1. Modality-aware query encoding,
2. Graph-based structural reasoning,
3. Dense semantic matching,

ensuring retrieval captures both explicit structure and semantic relevance. :contentReference[oaicite:7]index=7

5 Experimental Results

RAG-Anything was benchmarked on multimodal datasets such as DocBench and MMLongBench. It showed **substantial performance gains** compared to state-of-the-art multimodal RAG baselines. Performance improvements were especially significant in long-context settings where heterogeneous content spans multiple modalities. :contentReference[oaicite:8]index=8

6 Conclusion

RAG-Anything advances the field of retrieval-augmented models by integrating multimodal information effectively within a single framework. By eliminating the architectural fragmentation of previous systems, it achieves superior performance in realistic data environments where visual, structured, and textual content coexist. :contentReference[oaicite:9]index=9

7 Future Work

Future extensions may explore:

- Dynamic learning of entity alignments,
- Efficient indexing for very large corpora,
- Application of the framework in domain-specific environments such as healthcare or law.

References

References

[1] Zirui Guo et al., “RAG-Anything: All-in-One RAG Framework,” arXiv, Oct. 2025. :contentReference[oaicite:10]index=10