Whitepaper GraphRAG & multi-hop retrieval September 2026 Confidential

From information to evidence-based decisions

The knowledge is already there. Indentia brings it together: with meaning, connections and the sources people need to verify an answer.

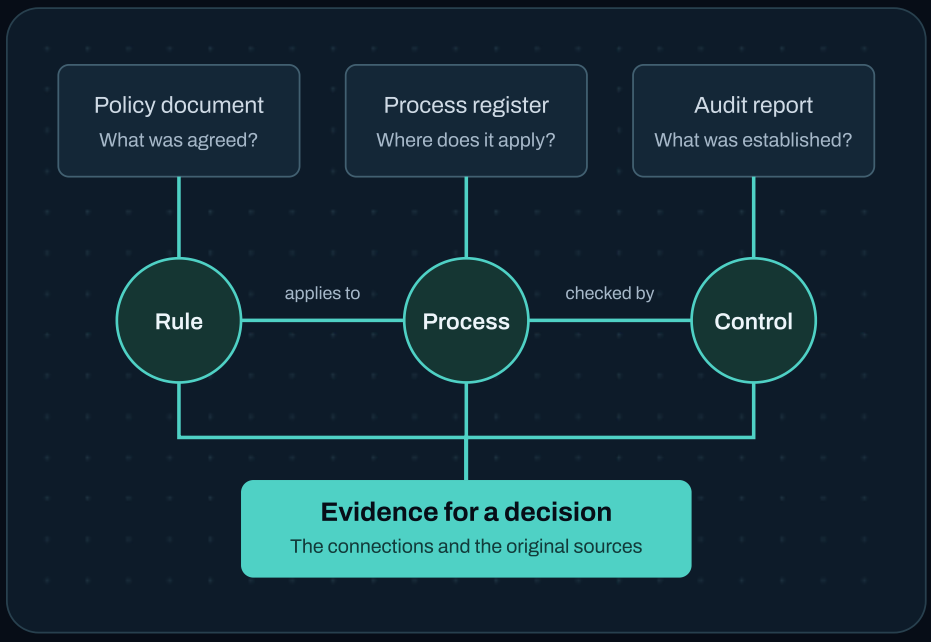


Illustration: several sources form one verifiable body of evidence.

Measured Indentia result · complete tuning run on MultiHop-RAG

98.24%

Evidence recall@10

Evidence coverage at document level

0.9832

MRR@10

Ranking quality on a scale from 0 to 1

A fixed tuning selection of 213 questions from the public MultiHop-RAG dataset, including 180 with reference evidence. Source and selection [1].

The hardest question rarely fits in one document

Organizations store policies, contracts, reports and operational data in different systems. A decision depends on how that information connects.

Finding information is the start

Search helps you find a relevant document. But which version applies? Does another name refer to the same supplier? And which report supports the conclusion about that particular process?

Users often still have to piece the puzzle together. That takes attention and makes the outcome depend on who searches and which source appears first.

Context makes the difference

RAG allows a language model to consult sources before answering. For complex questions, source selection must bring together different facts, periods and perspectives.

A passage can closely resemble the question yet lack the evidence it requires. High similarity becomes useful when the right facts make it into the selection.

> A useful answer connects three things: **what something means, how it relates to other information and what supports the conclusion.**

Illustrative scenario

Which processes are affected by this policy change, and which controls show that we comply?

Policy	Process	Control	Evidence
The requirement and valid version	Its practical application	The check on that application	The finding and its source

The answer requires several steps through the knowledge. This is **multi-hop**: a fact from one source helps place a required fact from another source in context. This example describes an application, rather than a measured customer case.

One knowledge layer for meaning, connections and evidence

Indentia brings information from different sources into a usable knowledge layer. Search, knowledge graphs and AI work together. The question determines which context is needed.

Meaning that persists

Concepts, names and definitions give information a shared frame of reference. Domain knowledge helps identify when sources mean the same thing and when an apparently identical term has a different meaning.

Connections that take questions further

A knowledge graph makes relationships between organizations, documents, processes and events explicit. IndentiaGraph provides that connected knowledge layer. GraphRAG can use those relationships to gather context for an answer.

Sources that support a statement

Hybrid search and evidence selection bring relevant information together. The focus is on concrete support, complementary sources and relevant contradictions. An answer should be traceable to the material on which it rests.

Control over use and change

Source provenance, versions, domain agreements and access policies are part of the platform architecture. They help keep knowledge manageable as documents change and different users work with the same information.

From source to usable context

Collect
Make documents and data accessible

Connect
Capture meaning and relationships

Substantiate
Bring complementary evidence together

Use
Verify and apply answers

This knowledge layer can be reused for search, research and agents. An organization can build on the same concepts and sources when the next question arises.

The benchmark in this paper measures Indentia's **hybrid retrieval and evidence selection route**. The figures do not separately measure graph traversal or test the full governance and security chain.

More of the required evidence. Stronger sources first.

On MultiHop-RAG, the optimized Indentia route improves both coverage and ranking of reference sources. Both are needed to give a reader useful context.

Evidence recall@10

What proportion of the required source documents appears in the first ten results? We calculate this per question and take the mean.

MRR@10

How high does the first relevant document rank? Rank 1 scores 1; rank 2 scores ½. No match in the first ten scores 0.

The same tuning questions, the same maximum final context

Indentia configuration	Evidence recall@10	MRR@10	All evidence@10
Hybrid retrieval with source diversification Internal reference configuration	88.15%	0.8607	73.33%
Optimized evidence selection **Complete tuning run**	**98.24%**	**0.9832**	**95.00%**

+10.09 percentage points evidence recall **+0.1225** MRR on the 0 to 1 scale

213 questions, including 180 with reference evidence. The optimized run has zero failed questions. “All evidence” is the share of questions for which all required documents are in the selection. In that run, at least one required source was found for 100% of the 180 questions.

The optimized route examines more candidates and uses additional model-assisted selection. This compares two complete configurations with a maximum of ten documents and 200,000 characters of source content for the reader; search and compute budgets differ.

Broader validation: preliminary results

Of the remaining 2,343 questions, 569 have been processed, including 498 with reference evidence. The same selected configuration achieves:

96.12% evidence recall@10 | 0.9774 MRR@10

These figures measure finding and ranking source documents. They are not a percentage of correct answers, a measure of hallucinations or an independent leaderboard.

Value where information needs to come together

The results show that carefully designed retrieval can find much of the required evidence. In an organization, that becomes meaningful through real questions, its own sources and clear responsibilities.

Research and analysis	Bring together news, reports and case files on the same subject. Make developments over time and differences between sources visible, so an analyst can assess the evidence.
Policy, risk and compliance	Connect a requirement to processes, measures and findings. This makes it clear which evidence supports a conclusion and where further checks are needed.
Knowledge for employees and agents	Give people and AI applications access to the same concepts and sources. Reusable context supports follow-up questions, while control over access and actions remains part of the setup.

Start with questions that take effort today

1. **Choose a defined domain.** Gather representative questions, the required sources and the people who can assess an answer's substance.
2. **Agree on the measures in advance.** Measure evidence coverage, answer quality, turnaround time and cost. Define when the system lacks sufficient grounds to answer.
3. **Validate on separate examples.** Test the chosen configuration on questions that were not used for tuning, retaining failures and difficult cases.
4. **Support day-to-day use.** Account for source changes, permissions, definitions and human review before expanding the application.

> A strong benchmark is a starting point. The next step is to demonstrate what Indentia delivers for **your questions and your knowledge.**

The figures in context

Dataset and source

The public MultiHop-RAG dataset [1] contains 609 news articles and 2,556 English-language questions. Tasks cover inference, comparison, temporal relationships and questions without reference evidence.

Tuning and validation

Indentia made a fixed, reproducible selection of 213 of the 2,556 questions for tuning: 180 with reference evidence and 33 without. This is an Indentia split within the public benchmark. The remaining 2,343 questions form the broader evaluation set; the reported preliminary results cover 569 questions, including 498 with reference evidence.

What was measured?

Document-level retrieval, with up to ten documents and 200,000 characters of source content for the reader. The route combines hybrid search with model-assisted evidence selection. Answer generation is outside these scores.

How are questions counted?

Evidence recall and MRR are averaged over questions with reference evidence. Questions without reference evidence remain reported but are excluded from these two means. Failed questions with reference evidence score zero.

---

[1] Dataset and benchmark. Tang, Y. & Yang, Y. (2024). *MultiHop-RAG: Benchmarking Retrieval-Augmented Generation for Multi-Hop Queries.* COLM 2024. arXiv:2401.15391.

Dataset: github.com/yixuantt/MultiHop-RAG · licence: ODC-BY 1.0. The scores shown here were measured by Indentia, September 2026.

www.indentia.ai/reports/indentia-graphrag.html

Bring your hardest knowledge question.

Explore with Indentia how to turn scattered information into connected knowledge and answers your organization can verify.

www.indentia.ai/contact