Maqui.ai: Knowledge Graphs and GraphRAG to Manage Legal Documents

Nicolás Barrionuevo[1], Ri Díaz[1], Benjamín Farías[1], Sebastián Ferrada[1,3], Alejandro Grez[1], Juan Reutter[1,2], Hernán Sarmiento[1] and Agustín Urrutia[1]

[1]*Instituto Milenio Fundamentos de los Datos, Chile*
[2]*Pontificia Universidad Católica de Chile, Chile*
[3]*DCC and IDIA, Universidad de Chile, Chile*

Abstract

We present Maqui.ai, a Chilean startup that helps users find and navigate the legal bodies that apply to a given problem by combining a legal knowledge graph with graph-augmented retrieval-augmented generation. We describe the data pipeline behind the system and discuss the main challenges we have faced while deploying it with real users.

1. Introduction

One important problem in the legal industry is *finding the legal bodies that apply* to a certain domain or question. Take, for example, real estate law, and consider a developer considering a new housing project. Architects and contractors know the basic law bodies governing real estate, which, here in Chile, correspond to DFL-458: the so-called General Law of Urban Planning and Construction. But it gets much more complicated. There are the Zoning plans issued by regional divisions. And perhaps some of these documents left some inconsistencies open, which were further resolved by a memo from the Ministry of Housing and Urban Development, or by the Comptroller's office, or maybe even the Supreme Court had something to say. Of course, if this terrain was not developed before, chances are that one would also need to file environmental permits, for which we need different zoning divisions filed by the Environment Ministry, or environmental instructions issued by city councils, etc. The list of necessary documents goes on and on. Looking up all this information can take days or even weeks, and inexperienced lawyers may be prone to make costly mistakes if they fail in this search. At the same time, state employees face the same problem: to issue the correct authorization to a certain project, one also needs to gather and understand all applicable legal bodies, and doing this incorrectly would probably lead to endless litigation with the real estate developer.

Maqui.ai is a Chilean startup that promises a definite fix for this problem. We started looking at real estate law, but the task of identifying and cataloging legal regulations is common across all legal areas; we are now expanding this to the tax and environmental domains. As with most legal tech companies that seek to improve how users manage legal documents, Maqui.ai's solutions are based on the RAG — Retrieval Augmented Generation Framework (see e.g. [1]). However, focusing solely on document retrieval based on a particular content or question may miss important documents, and can also retrieve noisy chunks that are semantically similar to the query yet contextually irrelevant. For example, in Chile, the Supreme Court ruled that certain provisions of the Zoning Plan of a commune in Santiago were invalid. A standard RAG pipeline answering a technical question about zones in that commune would almost never retrieve this ruling alongside the Zoning Plan itself, and the resulting recommendation would therefore not be legally accurate.

The example above immediately suggests that a Knowledge Graph (KG) [2] may be of use. Indeed, with a Knowledge Graph, we can augment documents with edges between each other, and use it to note when a certain ruling or document cites, derogates, or modifies another one. With Maqui's legal Knowledge Graph, users and agents can immediately find all related legal information for a given

ISWC 2026 Companion Volume, October 25–29, 2026, Bari, Italy

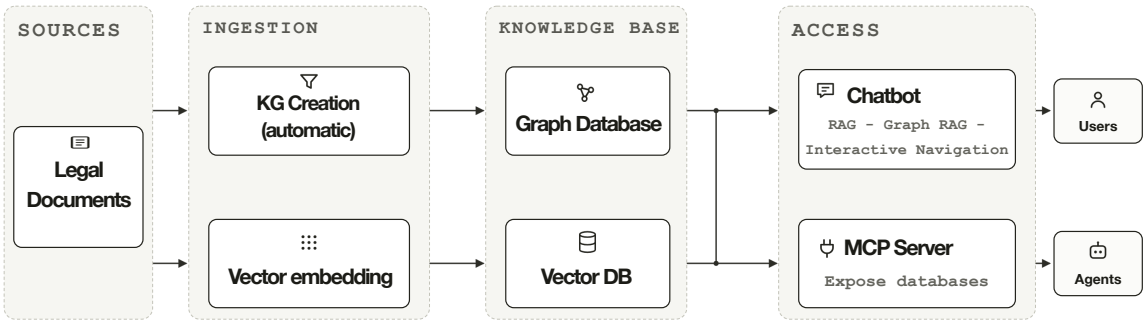

Figure 1: The Pipeline of Maqui. Documents are ingested both as building blocks for a knowledge graph and as vector embeddings. Then, users consume them in either way: as a knowledge graph via interactive navigation, or as a vector search within a RAG/GraphRAG system.

document. So, for example, if one is concerned about a certain memo that restricts the use of a land plot, one can traverse the KG to look up if this memo is still in order, or was superseded by another memo, or if there is a supreme court ruling or comptroller dictum governing this memo, or generally any piece of information that may be valuable. Understandably, this tool alone can save workers in the real estate industry days of work, but it also helps make our RAG-powered chatbots much more robust and less prone to hallucination.

In the following, we give a brief overview of the data pipelines used in Maqui.ai, from knowledge graph construction to graph-augmented RAG [3], chatbots, and visual interfaces that help users navigate our knowledge graph. Maqui.ai is partially funded by the Ministry of Science and is currently preparing for another pre-seed funding round. Our solution is currently used by different actors from the real estate industry, including architects, lawyers, and local government offices. It is publicly available on a free trial on https://Maqui.ai/.

2. Data Pipeline

The Pipeline behind the Maqui.ai system is shown in Figure 1. We continuously scrape legal documents from several official sources, including the Chilean Library of Congress (for laws), ministries (for memos, official letters, rulings, circulars, etc.), the Comptroller's office, the Supreme Court, and many others. Once scraped, documents undergo automatic chunking and embedding and are stored in a vector store. At the same time, chunks are fed together with a target ontology specifying the entity and relation types of interest to WUKONG, a proprietary system that uses an LLM to extract entities and the relations between them, both within and across documents. The resulting graph is exported to MillenniumDB [4][1] and integrated into our legal Knowledge Graph.

The coordination between vector database and knowledge graph allows us to offer both views to users. A simple question to our chatbot or a simple document search prompts the vector database to retrieve similar documents, which are then shown to the user. But then users can navigate the knowledge graph to quickly find associated letters or rulings related to the searched content. Further, we also expose an MCP connection [5] so that users can issue more complex instructions to the chat agent, including automatic navigation through our legal knowledge graph. In the talk, we will present one example (a title examination in the real estate domain) where having access to this complex navigation is key to the precision of our service.

[1] https://github.com/MillenniumDB/MillenniumDB

3. Usage and challenges

During the first semester of 2026, we have been steadily building a community of users. Our chatbot interfaces are offered with a free trial component, granting a number of credits or tokens, and pay-as-you-go top-ups for credits. On the other hand, the MCP plugins to interact with Maqui.ai are unlocked via a monthly subscription.

We've had several demos and live introductions to Maqui.ai, mostly with city councils and real estate companies. It was a major surprise to see how users, in these live demos, used the knowledge graph functionalities to navigate to legal documents that they needed... live during the demo!

At the same time, we learned that most offices working with documents will surely want to add their own set of documents, which has introduced a number of challenges to our text-to-graph pipeline, as now we have a (big) legal knowledge graph that needs to be integrated with a (smaller) proprietary knowledge graph. So far, we have not solved the integration problem, and we currently offer context-based retrieval on the user's own documents, which has been a useful stopgap but is far from the seamless integration we aim for. We summarize below the main lessons and open challenges that have emerged during this period:

- **Precision is the most immediate selling point.** Users have experience with LLMs and are most worried about hallucination.
- **Integrating the client's own documents is our main open problem.** Almost every office wants to bring its private files into the system, but merging this (smaller) proprietary knowledge graph with the (large) legal knowledge graph is far from trivial. For now, we fall back on context-based retrieval over these documents, without incorporating them into the graph itself.
- **One unified graph, or one graph per domain?** As we expand into tax and environmental law, it is tempting to source a single, huge knowledge graph. In practice, however, our real estate graph has been carefully cleaned and tailored to the client's needs, and it is not clear how to preserve this quality at scale. For now, we maintain a separate knowledge graph per domain, and each user accesses only the graph for the domain they work in.
- **MCP as an abstraction layer.** When every capability is exposed as an MCP tool, these architectural details can be hidden from the user, who simply interacts with the final agent. This is, so far, the most promising direction, and it opens a research question we intend to pursue: what is the best way for agents to interact with the semantic information stored in a knowledge graph?

References

[1] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, arXiv preprint arXiv:2312.10997 (2023).

[2] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. de Melo, C. Gutierrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Computing Surveys 54 (2021) 1–37. doi:10.1145/3447772.

[3] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitansky, R. O. Ness, J. Larson, From local to global: A graph RAG approach to query-focused summarization, arXiv preprint arXiv:2404.16130 (2024). URL: https://arxiv.org/abs/2404.16130. arXiv:2404.16130.

[4] D. Vrgoč, C. Rojas, R. Angles, M. Arenas, D. Arroyuelo, C. Buil-Aranda, A. Hogan, G. Navarro, C. Riveros, J. Romero, MillenniumDB: An open-source graph database system, Data Intelligence 5 (2023) 560–610. doi:10.1162/dint_a_00229.

[5] X. Hou, Y. Zhao, S. Wang, H. Wang, Model context protocol (mcp): Landscape, security threats, and future research directions, ACM Transactions on Software Engineering and Methodology (2025).