Leibniz Centre for
Agricultural Landscape Research
(ZALF)

Research Area 4 "Simulation and Data Science"

A methodological framework to align Large Language Models for forest systems information retrieval

Stephan Playfair; Dr. Ferreol Berendt; Prof. Dr. Tobias Cremer; Prof. Dr. Gunnar Lischeid

Introduction

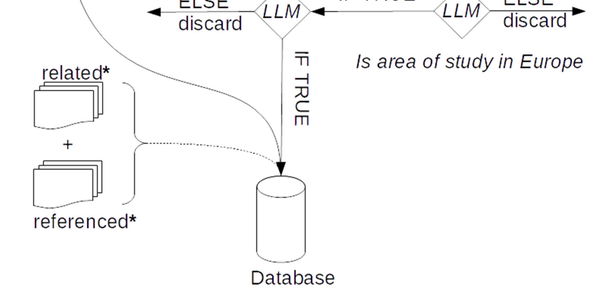
In ecological disciplines, relying on accurate, concise, and relevant information is crucial for solving complex problems and making data-driven decisions that affect real-world systems, policies, or public perception.

Large Language Models (LLMs) can provide support. However, this may:

- Introduce hallucinations (Huang et al., 2025).
- Have no access to real-world data (Zuzu'arregui et al., 2025).

A custom framework that ensures the curation of a qualitatively robust knowledge base minimizes these shortcomings.

We propose a framework to allow AI systems to source information from a reliable knowledge base prior to inference.

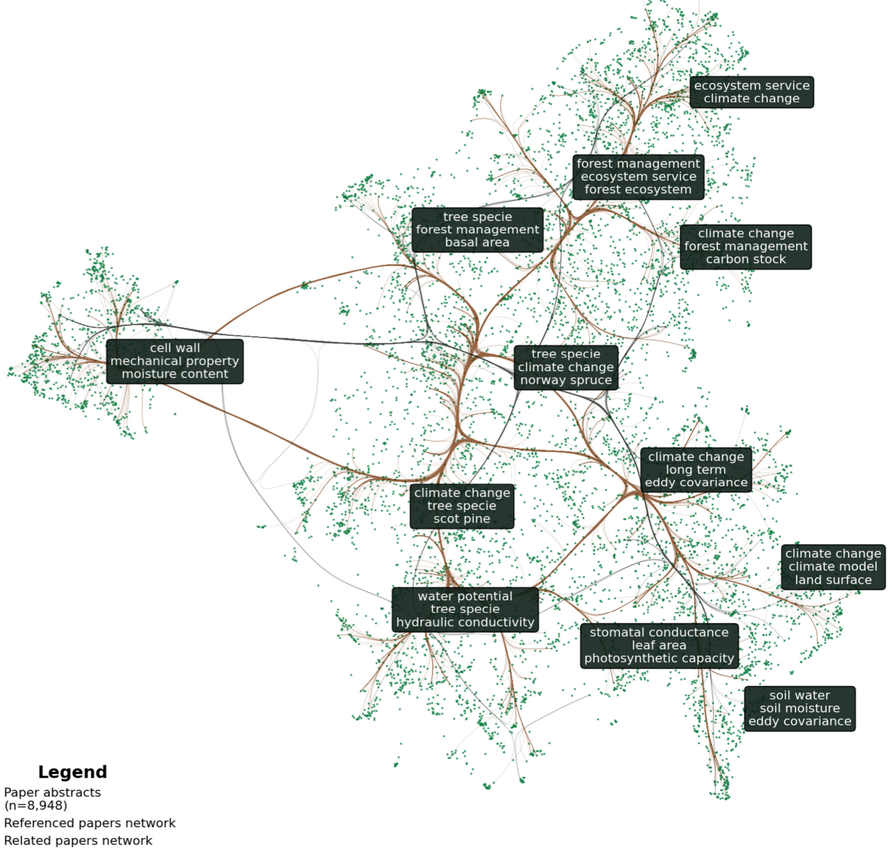


Framework's workflow: from ontology to knowledge graph ingestion

Methods

Populating the knowledge base through systematic curation:

- Sourcing papers from a reliable ontology using human selection criteria
- 2-stage LLM-based automatic filtration processes before
- Inserting research papers into a hierarchical, navigable, directed knowledge graph.
- Visualize the knowledge base with Natural Language Processing tools and methods.



Spatial relationship of 8,948 research papers

Results

Closely examining the visualization of the knowledge base, we can infer the following:

- Each dot in Figure 2 represents a research paper. Its position is based on its semantic meaning.
- We gain insight into which information may be over- or underrepresented in the knowledge graph based on the density of the scatterplot.
- We can understand the relationship between the various clusters based on the connections between the clusters.

Discussion

This framework's flexibility allows for:

- Amend settings to restrict or relax rules in the workflow to either increase or decrease the size of the knowledge base
- Possible adoption in other domains or fields.
- Allows any LLM model to be used based on design preference.

Next steps

- Users can consult an interactive platform with queries to get a more coherent response.
- Fine-tune the model and optimize the architecture based on user interaction.

Visit
chat.scifor.de
To try it out

References

Huang, et al., (2025). https://doi.org/10.1145/3703155

Zuzu'arregui, et al., (2025). https://doi.org/10.48550/arxiv.2506.10106.

Leibniz Centre for Agricultural Landscape Research (ZALF) | Eberswalder Straße 84 | 15374 Müncheberg | Germany
Contact: Stephan Playfair Stephan,playfair@zalf,de
Date: [17.03.2026]

With funding from the:
Federal Ministry of Research, Technology and Space
ADAPT WALD HOLZ
30 Years Leibniz Association