

A master's Thesis University of Applied Science and Art, Lucerne Data Science Department

# NL2GeoSQL - Agentic information retrieval from Geospatial-Temporal information systems



## Problem Formulation

Interacting with relational databases necessitates the capability to efficiently execute queries, which frequently requires specialized knowledge of SQL. The process of text-to-SQL parsing for Large Language Models (LLMs) seeks to bridge this gap, making database access more user-friendly for non-experts by converting natural language queries into SQL commands. This goal has prompted the creation of innovative methods, primarily centered on prompt engineering and fine-tuning. This survey offers a thorough overview of LLMs in Text-to-SQL tasks, encompassing the generation of custom datasets, advancements in prompt engineering, fine-tuning techniques, and potential future research directions.

## Problem Action

The following steps are planned. First, a spatial database will be created locally using IBM-provided data. Next, candidate models will be identified (IBM Granite 8B, Meta LLaMA 3.1). To optimize the models' performance, a spatial dataset will first be generated, followed by QLoRA fine-tuning to enhance their capabilities. Additionally, in-context learning techniques will be employed to further improve their performance. A LangChain-based pipeline will then be developed to enable seamless interaction with the spatial database, providing a robust framework for evaluation. Finally, the system will be tested on spatial queries to assess its accuracy, efficiency, and overall effectiveness.

## Result

The evaluation demonstrated that the highest average accuracy was achieved with Granite hosted in RITS, guided by an optimized meta-prompt. This highlights that prompt engineering can be a highly promising

approach for guiding the model in Text-to-SQL tasks, especially when the prompt is well-structured and incorporates few-shot examples. Furthermore, embedding a chain-of-thought process within the prompt can effectively guide the model in navigating from the input question to the correct path in the database.

**Advisor:**  
Prof. Dr. Andrew Paice

**Expert**  
Prof. Dr. Peter Kolb

**Cooperation Partner**  
IBM Research

