



ΠΑΡΟΥΣΙΑΣΗ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

ΗΜΕΡΟΜΗΝΙΑ: Τρίτη, 30 Σεπτεμβρίου 2025

ΩΡΑ: 12:00 – 14:00

ΑΙΘΟΥΣΑ: Αίθουσα Σεμιναρίων ΤΜΗΥΠ

ΟΜΙΛΗΤΗΣ: Αθανάσιος Γεωργιάδης

Θ έ μ α

« Scalable Management of Complex Spatial Data Types »

Επταμελής Εξεταστική Επιτροπή:

1. **Νικόλαος Μαμουλής**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
2. **Ευαγγελία Πιτουρά**, Καθηγήτρια Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
3. **Βασίλειος Β. Δημακόπουλος**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
4. **Παναγιώτης Βασιλειάδης**, Καθηγητής Α' Βαθμίδας, Τμήμα Μηχανικών Η/Υ & Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
5. **Ελένη Τζιρίτα Ζαχαράτου**, Assistant Professor, Hasso Plattner Institute for Digital Engineering
6. **Γεωργία Κούτρικα**, Κύρια Ερευνήτρια, Ινστιτούτο Πληροφοριακών Συστημάτων, Ερευνητικό Κέντρο Καινοτομίας στις Τεχνολογίες της Πληροφορίας των Επικοινωνιών & της Γνώσης «ΑΘΗΝΑ»
7. **Μανόλης Κουμπάρκης**, Καθηγητής Α' Βαθμίδας, Τμήμα Πληροφορικής & Τηλεπικοινωνιών, Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών



Περίληψη:

Scalable spatial data management is crucial in both scientific and commercial domains, particularly in Geographic Information Systems (GIS), which handle massive volumes of geographic data. As spatial data continues to grow rapidly, the demand for efficient spatial data analytics tools has become increasingly pressing. A core functionality of such tools is the computation of spatial and topological joins over large collections of objects. These operations aim to identify intersecting object pairs (i.e., objects that share at least one common point), a fundamental task with applications in geospatial interlinking, spatial databases, and beyond. However, intersection testing is computationally intensive, especially for polygonal objects, which often contain a large number of vertices and require costly geometric processing.

This dissertation investigates approximation techniques for handling high-complexity polygons, with the aim of making processing faster and more efficient. The central objective is to minimize reliance on original geometries, using them for computations only as a last resort. Our proposed solutions introduce efficient polygon approximation methods with a low memory footprint, along with filtering techniques that enable spatial joins to be evaluated without directly accessing the original geometries. The work addresses both scalability and accuracy challenges while striving to deliver solutions that are directly applicable to modern in-memory spatial databases.

Scalable spatial data management has two key aspects. First, query processing algorithms must be highly parallelizable and independent, enabling them to fully leverage distributed and parallel spatial databases for both vertical and horizontal scalability. Second, they must maintain efficiency as geometric complexity increases, since complex shapes often become a major bottleneck in spatial query processing. In the second part of this dissertation, we design and implement a prototype distributed spatial data management framework that operates independently of underlying engines, focusing specifically on the performance and scalability of spatial query evaluation in tightly coupled clusters. The prototype integrates state-of-the-art indexing, approximation, and filtering techniques while carefully minimizing both communication overhead and memory usage.

With the rapid advancement of Large Language Models (LLMs) and their expanding use across diverse domains, questions arise regarding their ability to handle complex tasks, particularly spatial reasoning over text. While LLMs excel at inferring and extracting information from large text collections, spatial knowledge is often



domain-specific and not inherently intuitive for them. The first limitation is commonly mitigated through Retrieval-Augmented Generation (RAG), where external databases provide context at inference time, enhancing factual accuracy in responses. However, without fine-tuning or re-training, which can be costly and counterproductive to the goal of broad generalization, LLMs have consistently demonstrated weak performance on spatial reasoning tasks.

In the final part of this dissertation, we investigate how topological relations can help LLMs generate correct responses to spatial reasoning questions expressed in text. We first employ our efficient spatial topology algorithms to scalably compute key inter-dataset spatial relations and represent them as RDF (text) triplets. These are then leveraged through RAG-based mechanisms and indexing techniques to enable fast and accurate spatial context retrieval at inference time, with the dual objective of (i) supplying the LLM with domain-specific spatial knowledge and (ii) supporting it in producing factually correct responses.

In summary, this dissertation presents a comprehensive study of scalable spatial data management. It introduces in-memory solutions that are both efficient and accurate, addressing key challenges across a broad spectrum of use cases. The proposed approaches are directly applicable to modern spatial databases and well-suited for data-intensive geospatial applications.