

Abstracts-Explorer: Navigating Conferences with Semantic Search and Retriever Augmented Generation



T. Korten, P. Steinbach

Helmholtz-Zentrum Dresden - Rossendorf e. V., Dresden, Germany



SEMANTIC SEARCH

The screenshot shows the Abstracts Explorer interface for NeurIPS 2025. The search bar contains 'uncertainty estimation'. Below it, a card for 'Quantifying Uncertainty in the Presence of Distribution Shifts' by Yuli Slavutsky and David Blei is displayed, along with another card for 'Torch-Uncertainty: Deep Learning Uncertainty Quantification' by Adrien Lafage et al.

Download conference abstracts and search them with LLM-based semantic search.

INTERESTING PAPERS

The screenshot shows the 'Interesting Papers' section of the Abstracts Explorer interface. It lists several papers, including 'Quantifying Uncertainty in the Presence of Distribution Shifts' and 'Torch-Uncertainty: Deep Learning Uncertainty Quantification', each with a small preview and a 'Show more' link.

Interesting papers can be exported as a collection of markdown files for future analysis

RETRIEVER AUGMENTED GENERATION

The screenshot shows the 'AI Research Assistant' feature. A user asks about uncertainty quantification, and the system responds with an overview of relevant papers, mentioning Torch-Uncertainty and other methods.

Example Chat

(Paper 1 to 5 are shown next to the chat):

Q: Give me an overview of the most relevant papers about uncertainty quantification.

A: Several papers focus on improving UQ in deep learning models. Paper 1 introduces Torch-Uncertainty, a PyTorch framework to streamline the training and evaluation of DNNs with UQ techniques. Paper 3 proposes a computationally efficient UQ method using the Empirical Neural Tangent Kernel, approximating the posterior of a Gaussian Process. Paper 4 presents MCNF, a post-hoc method for deep regression that provides both prediction intervals and full predictive distributions without model retraining.

Beyond standard deep learning, Paper 5 addresses UQ specifically for Physics-Informed Neural Networks (PINNs) using Extended Fiducial Inference (EFI), offering a way to build confidence sets without needing prior information.

Finally, Paper 2 tackles UQ in the context of adaptively collected data (e.g., from reinforcement learning), proving asymptotic equivalence between Bayesian UQ and frequentist methods under certain conditions.

CLUSTERING ANALYSIS

The screenshot shows the 'Embedding Clusters' section. It displays a 3D scatter plot of paper embeddings clustered into 60 groups. A legend on the right lists various topics: 3D Scene & Object Understanding (183), Graph Neural Networks & Learning (167), Convex & Gradient Optimization Theory (161), Diffusion Models for Image Generation (157), Multi-Agent LLM Systems (146), LLM Benchmarking & Evaluation (145), LLM Reasoning with Reinforcement Learning (144), Large Language Model Scaling (138), Video Generation & Synthesis (132), and LLM Training & Data (131).

An interactive clustering map gives an overview of the topics that were presented at the conference.

Features

- Download conference data from various sources (NeurIPS, ICLR, ICML, ML4PS)
- Store data in SQL database (SQLite or PostgreSQL) with efficient indexing
- Search papers by keywords, track, and other attributes
- Generate text embeddings for semantic search
- Find similar papers using AI-powered semantic similarity
- Interactive RAG chat to ask questions about papers
- NEW: Cluster and visualize paper embeddings with interactive plots
- Web interface for browsing and searching papers
- NEW: MCP server for LLM-based cluster analysis
- NEW: Multi-database backend support (SQLite and PostgreSQL)
- Environment-based configuration with .env file support