

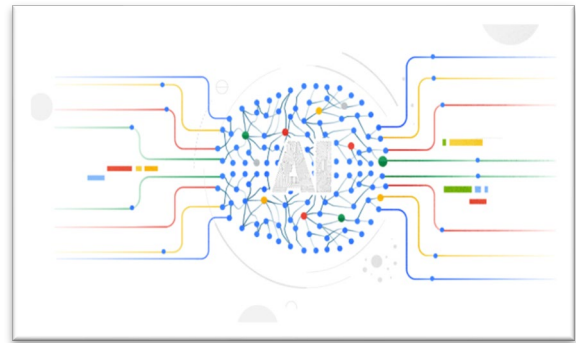
Advanced Topics in Data Management

Summer Term 2024

Lecturers: Prof. B. Mitschang, Prof. H. Schwarz, Dr. P. Reimann, J. Voggesberger
Contact: Julius Voggesberger, julius.voggesberger@ipvs.uni-stuttgart.de
Language: English

Description

The comprehensive analysis of large volumes of data is crucial in modern information systems. Some of the main driving forces are current trends such as Industrie 4.0 and the Internet of Things (IoT). In particular, applying data analysis techniques like machine learning (ML) to data of real-world scenarios imposes various data management challenges in several steps of analytic processes. One challenge in this context is to provision the relevant data in such a way that all user groups (data scientists, domain experts, ...) in an organization are able to find, access and analyze all data and only those data that are important for their specific use cases. Important steps of analytic processes are, for example, the selection of analysis methods, algorithms and their parameters. Often domain knowledge is needed to make proper decisions in these steps. Challenges are how to represent this domain knowledge, how to manage it and how to use it during analysis.



In this seminar, we address current technologies, concepts, algorithms and system infrastructures for above-mentioned challenges for data management in analytic processes:

- Modern data platform architectures: data lake, data lakehouse, data mesh, ...
- Approaches to data modeling in data platforms
- Representation of domain knowledge for analytic processes, e.g., as semantic nets or as knowledge graphs
- Systems to store and provision domain knowledge in analytic processes, e.g., triple stores or graph databases
- Data preprocessing in ML pipelines
- ML lifecycle management and ML Ops
- ...

Remarks

- In this advanced seminar, each student will work on one specific topic. Basic literature will be provided by the advisors. Each student will summarize results in a document of about 20 pages and in a 30 minutes presentation.
- **Seminar topics will be assigned to the participants in a first meeting at the beginning of the summer term. Registered participants will be informed about the date by email.**

Prerequisites

Basic knowledge on database systems and information systems, e.g., from the lectures "Modellierung" or „Data Warehouse, Data Mining and OLAP“, is mandatory.