Graz University of Technology

Faculty of Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science (ISDS)
Group for Data Management

Univ.-Prof. Dr.-Ing.
Matthias Boehm
BMVIT Endowed Chair for
Data Management

Inffeldgasse 13/V
8010 Graz
Austria

+43 (0)316 873 30690

m.boehm@tugraz.at

DVR: 008 1833 UID: ATU 574 77 929

Graz; July 25, 2019

2 PhD Positions at Graz University of Technology, Institute of Interactive Systems and Data Science, Reference 7060/19/ExDRa

The BMVIT-endowed chair for data management at Graz University of Technology, Austria (Institute of Interactive Systems and Data Science) offers two fully funded PhD positions for 3 years and 40 hours per week in the broad area of machine learning (ML) systems and data management. These positions will contribute to the ExDRa project (Exploratory Data Science over Raw Data), a joint research project with Siemens AG, DFKI, and TU Berlin, funded through the BMVIT program "IKT der Zukunft".

Overall Mission: The open positions will further be integrated in the data management for data science group (https://damslab.github.io/) at TU Graz. Our overall mission is to simplify data science by providing high-level, data science-centric abstractions and building systems and tools to efficiently execute these tasks. Over the next years, we will build SystemDS (https://github.com/tugraz-isds/systemds), an open source ML system for the end-to-end data science lifecycle.

ExDRa Project: ML applications are increasingly applied in the enterprise to improve the value chain and gain competitive advantage. In contrast to traditional ML, the objectives are often under-specified, allow for different analysis methods, and can leverage a wide variety of heterogeneous, distributed and partially inaccessible data sources. Therefore, the typical data science process in the enterprise is exploratory, i.e., data scientists investigate hypotheses, integrate the necessary data, and run different analysis pipelines. Since the added value is unknown in advance, very little investments are made into systematic data acquisition, integration, and preprocessing. This lack of infrastructure results in redundant manual effort and computation. Furthermore, the central consolidation might not be technically or economically desirable or even possible (e.g., sensitive personal data). These scenarios share the necessity of federated execution and the need for redundancy elimination. The basic idea of the ExDRa project is an investigation of suitable systems support for this exploratory data science process over heterogeneous and distributed raw data sources, showcased in an application demonstrator. Major research aspects are (1) ad-hoc and federated data integration over raw data, (2) data organization and reuse of intermediates, (3) horizontal optimization over the entire data science lifecycle, and (4) query planning for partially accessible data.

Open Positions: In the context of the ExDRa project, we are looking for two research assistants with strong interest in applied, systems-oriented research for data and ML systems. These positions are expected to contribute to the overall system infrastructure and conduct research in the following areas:

- *Declarative Data Integration and Preparation over Federated Raw Data:* Specification languages and runtime techniques for ad-hoc data integration, cleaning, and preparation, including generation of extractors, support of heterogeneous formats, local and federated data integration, preparation (of numeric representations), and data cleaning, as well as dedicated plan compilation techniques.
- *Adaptive Data Reorganization and Reuse in Federated Environments:* Reuse of redundant intermediates in data preparation, data cleaning, and ML training methods, specialization and reorganization of underlying data structures, federated and multi-tenancy-aware access structures, as well as lossy and lossless compression techniques for efficient communication.

A background in ML algorithms, statistics, and data science is a plus but not necessarily required as our focus is on systems aspects and we are trying to build a team with diverse backgrounds and skills.

Requirements and Benefits: We expect a master or diploma degree in computer science or related fields, a strong interest in systems-oriented research, and good English language skills (fluently spoken and written). This research assistant position is fully funded for 3 years, with a gross salary of at least €2,794.60 (14x yearly). We offer an inspiring research environment, highly practical research questions, and the integration into a diverse team of researchers and developers. Research assistants are expected to pursue their PhD studies at the Institute of Interactive Systems and Data Science at TU Graz.

Application Process: The application deadline is **August 18, 2019**. Please send your application (with a reference to the position ID 7060/19/ExDRa), CV, and additional documents to gabriele.leitner@tugraz.at. For further information regarding research directions or the application process, please contact Matthias Boehm (https://mboehm7.github.io/) via m.boehm@tugraz.at.