

## Letter from the Special Issue Editor

The field of data analytics has flourished and continues to grow with unabated momentum. Insights derived from rich and diverse data sets drive business decisions and modern software. The race to derive more insights, faster, and more effectively has intensified, and research has followed suit. At the same time, data and analytics tools have grown complex, and analysis results are often poorly understood. In this issue, we bring together an exciting collection of recent and ongoing work, investigating methods for accelerating the derivation of insights in data analysis and technologies for understanding the insights derived from data.

We start with “The Case for a Visual Discovery Assistant: A Holistic Solution for Accelerating Visual Data Exploration”, by Doris Jung-Lin Lee and Aditya Parameswaran. The paper considers the use of visualizations in exploratory data analysis, and presents a vision for automating the navigation across large collections of visualizations. It provides a review of existing and ongoing work, and outlines open research questions and challenges as we are moving towards the direction of expecting less input from the human users and the system bears more of the onus in visual discovery.

In the second paper, “Towards Quantifying Uncertainty in Data Analysis & Exploration”, Yeounoh Chung, Sacha Servan-Schreiber, Emanuel Zraggen, and Tim Kraska consider the many facets of uncertainty and its impact on insights derived from analytics and exploration pipelines. The paper presents a toolset for quantifying uncertainty with the objective of making data analysis reliable and safe, guarding against incorrect analyses and false discoveries.

The third paper focuses on methods for understanding the validity of data insights. In “Query Perturbation Analysis: An Adventure of Database Researchers in Fact-Checking”, Jun Yang, Pankaj K. Agarwal, Sudeepa Roy, Brett Walenz, You Wu, Cong Yu, and Chengkai Li point out that even though data can be used to back a claim, that claim may still be misleading and essentially incorrect. Discovering the query that produced the claim and perturbing it appropriately can shed light on the claim’s validity. The paper highlights results in query perturbation and computational fact checking that largely predate the rise of the so-called “fake news”, as well as practical deployments of the relevant systems, including during the 2016 US federal elections.

The next paper, “Explanation Tables”, by Kareem El Gebaly, Guoyao Feng, Lukasz Golab, Flip Korn, and Divesh Srivastava, considers deriving data summaries; such summaries can aid human understanding of a dataset’s properties. The paper focuses on the problem setting of a relation with categorical dimension attributes and one binary outcome attribute, where the objective is to derive an interpretable explanation of the factors affecting the binary attribute.

In “Ontology-Based Natural Language Query Interfaces for Data Exploration”, Chuan Lei, Fatma Özcan, Abdul Quamar, Ashish Mittal, Jaydeep Sen, Diptikalyan Saha, and Karthik Sankaranarayanan argue for the use of natural language as the query interface to data exploration. Exploratory tasks are varied and often require different querying capabilities of the underlying data stores. Natural language can become a unifying interface, bypassing the need for users to learn and use other complex languages, thus increasing the accessibility of analytics. The paper describes an end-to-end NLQ system, overviews ongoing work and challenges, and highlights practical use cases.

We conclude with a paper that reexamines a fundamental component of data processing pipelines: data structures. “The Periodic Table of Data Structures”, by Idreos et al., points out that data structure decisions impact the types of analysis that one can perform, the efficiency of that analysis, and, ultimately, the insights that can be achieved. The paper describes a vision that calls for breaking data structures into first principles, reexamining the design space, speculatively estimating a data structure’s impact on a workload, and automating data structure design to cater to specific analysis tasks.

Thank you to all the authors for their insightful contributions, and thank you to Dave Lomet for valuable guidance on putting together the issue. I hope you enjoy this collection.

Alexandra Meliou  
University of Massachusetts Amherst