Letter from the Special Issue Editor

Over the past years there has been a growing recognition of the increasing role of people in the data life cycle. People massively contribute data and share opinions, people also extensively analyze data and consume the derived information. In this issue, we have a slate of very interesting articles discussing the different roles of human in data management. In particular, we discuss three complementary aspects. First, we consider the role of people as crowdsourcing workers, assisting in data generation and in data-centric computation and analysis. The focus here is on devising methods for cost-effective, meaningful usage of human capabilities and knowledge. Second, we consider the role of people as data analysts. The focus here is on interactive data analysis and mining and the development of tools that facilitate such effective interaction. Finally, we consider the role of people as data consumers and acknowledge the importance of considering ethics in many aspects of data creation, access, and usage. The focus here is on finding new ways for maximizing the benefits of massive data while nevertheless safeguarding the privacy and integrity of citizens and societies.

We start with three papers that investigate different aspects of crowdsourcing. In "Toward Worker-Centric Crowdsourcing", Amer-Yahia and Roy argue that accounting for human factors, in particular workers' characteristics, in task assignment benefits both workers and requesters, and discuss new opportunities raised by worker-centric crowdsourcing. In the second paper, "Spatial Crowdsourcing: Challenges and Opportunities", Chen and Shahabi consider spatial aspects of crowdsourcing, in particular those related to the usage of mobile devices and users, and discuss the challenges and opportunities related to this important setting. In "Optimizing Open-Ended Crowdsourcing: The Next Frontier in Crowdsourced Data Management", Parameswaran, Sarma and Venkataraman examine open-ended crowdsourcing and survey existing work on formally reasoning about and optimizing this important, but relatively understudied class of crowdsourcing.

The next three papers focus on interactive data analysis. In "Interactive Data Exploration via Machine Learning Models", Papaemmanouil, Diao, Dimitriadou and Peng examine how learning-based exploration techniques can automatically steer the user towards interesting data areas, based on relevance feedback on database samples. In "Runtime Support for Human-in-the-Loop Feature Engineering Systems", Anderson, Antenucci and Cafarella take a closer look at feature selection for machine learning and discuss two projects that accelerate feature engineering by applying domain insights to engineer high-impact features. Finally, in "Towards a Benchmark for Interactive Data Exploration", Eichmann, Zgraggen, Zhao, Binnig and Kraska discuss the metrics that should be used for evaluating interactive data exploration systems and present ideas towards a new benchmark that simulates typical user behavior and allows such systems to be compared in a reproducible way.

Next, in "The Values Challenge for Big Data", Jagadish explores the characteristics of human involvement in Big Data management and proposes a research agenda to address associated challenges. Finally, we conclude by "HILDA 2016 Workshop: A Report", by Nandi, Fekete and Binnig that overview the Human-in-the-Loop Data Analytics (HILDA) workshop, that was held in association with ACM SIGMOD 2016.

I hope that you enjoy the issue as much as I enjoyed putting it together!

Tova Milo
Tel Aviv University