CALL FOR PAPERS

ACM Journal of Data and Information Quality

Special Issue on Metadata Discovery for Assessing Data Quality

Guest Editors:
Giuseppe Polese, University of Salerno (gpolese@unisa.it)
Vincenzo Deufemia, University of Salerno
Shaoxu Song, Tsinghua University

Context

Big data sources contain several hidden metadata, such as value patterns and their distributions, functional dependencies, and graph metadata, which can be exploited to assess the quality of data, including non-relational data like those related to social networks or IoT systems. In this context, it might be extremely complex to specify a priori metadata for quality assessment, since the velocity at which big data evolve can quickly make metadata obsolete. To this end, the capability to automatically and incrementally discover metadata from a big dataset represents an important prerequisite for several data quality assessment activities.

The process of automatically discovering metadata from data sources is known as *data profiling*. Current data profiling techniques do not scale well on big data, and are mostly conceived for relational data, whereas in most modern applications there is the need to extract metadata from semi-structured data, column-oriented databases, graph structures, streaming data, and so on. Novel algorithms and methods should comply with the requirements related to the 3(+1) V's (variety, volume, velocity, + veracity) of big data, support incremental metadata discovery, and eventually, exploit distributed computation paradigms to make the metadata extraction process tractable.

This special issue focuses on metadata useful for assessing the quality of data. Thus, the issue is addressed to those members from the data science community proposing novel methods or scalable algorithms capable of extracting metadata from several types of data sources, aiming to assess, monitor, and improve the quality of their data. The benefits of such solutions will not only enhance the monitoring of data quality problems, but it will also provide means to fix them.

The topics of interest of this special issue include:

- Data profiling approaches for data quality
- Profiling geographical, temporal and dynamic data
- Profiling non-relational data
- Profiling streaming data
- Relaxed dependency discovery: conditional, partial, and other relaxed dependencies
- Graph metadata discovery
- Single and multi-column profiling tasks
- Data management architectures for data profiling
- Parallel and distributed dependency discovery algorithms
- Metadata visualization
- Objective data quality metrics based on metadata
- Metadata driven quality assessment
- Design and implementation of metadata-driven tools for data quality monitoring, assessment and improvement
- Scalability and performance of data quality assessment tools

Expected Contributions and Submission Information

We welcome two types of regular contributions:

- Research manuscripts reporting novel methodologies and results (up to 25 pages).
- Experience papers that report on lessons learned from addressing specific issues within the scope of the call. These papers should be of interest to the broad data quality community (10+ pages plus an optional appendix).

JDIQ welcomes manuscripts that extend prior published work, provided they contain at least 30% new material, and that the significant new contributions are clearly identified in the introduction.

Submission guidelines with Latex (preferred) or Word templates are available here: **http://jdiq.acm.org/authors.cfm#subm**. Please submit the paper by selecting as type of submission: "**SI: Metadata Discovery for Assessing Data Quality**."

Important Dates and Timeline:

Initial submission: **October 9, 2019** (extended)
First review: **December 31, 2019**
Revised manuscripts: **February 13, 2020**
Second review: **April 13, 2020**
Camera-ready manuscripts: **June 13, 2020**
Publication: **August 2020**

acm
Association for Computing Machinery