ACM Journals

CALL FOR PAPERS

Journal of Data and Information Quality

Special Issue on Human in the Loop Data Curation

Guest Editors:

- **Gianluca Demartini,** The University of Queensland (Australia), g.demartini@uq.edu.au
- **Shazia Sadiq**, The University of Queensland (Australia), shazia@itee.uq.edu.au
- **Jie Yang**, TU Delft (Netherlands), J.Yang-3@tudelft.nl

Although data quality is a long-standing and enduring problem, data quality problems have recently received a resurgence of attention due to the fast proliferation of data analytics, machine learning, and decision-support applications built upon the wide-scale availability and accessibility of (big) data. Particularly, the success of machine learning heavily relies on the quality and quantity of training data. Data curation which may include ingestion, annotation, cleaning, integration, etc., is a critical step to provide adequate assurances on the quality of analytics and machine learning results. Such data preparation activities are recognized as time and resource intensive for data scientists as data often comes with a number of challenges that need to be tackled before it can be used in practice. Data re-purposing and the resulting distance between design and use intentions of the data, is a fundamental issue behind many of these challenges. These challenges include a variety of data issues such as noise and outliers, incompleteness, representativeness or biases, heterogeneity of format or semantics, etc. Mishandling these challenges can lead to negative and sometimes damaging effects, especially in critical domains like healthcare, transport, and finance.

An observable distinct feature of data quality in these contexts is the increasingly important role played by humans, being often the source of data generation and the active players in data curation. This special issue looks at the interdisciplinary overlap between manual, automated, and hybrid human-machine methods of data curation. The need for new research effort on involving humans in the loop of the data curation process is exacerbated by the importance of developing methods that can scale to large amounts of data while also maintaining a human touch. This means designing processes that can deliver high level of transparency in the data curation process (e.g., explaining why certain values have been dropped), deal with ethical data challenges like the decision to use or discard certain attributes (e.g., applicants' gender) in decision making processes, and, overall, increase the quality and trust in the outcome.

Topics

The topics of interest are inspired from the themes above and include, but are not limited to:

- Improving the quality of crowdsourcing outcomes
- Supporting crowd workers in task completion
- Interaction techniques for manual, collaborative, and hybrid
- human-machine data curation
- Data worker engagement drawing techniques from citizen science and
- collective intelligence
- Crowdsourcing and human computation studies into the transparency,
- reliability, and biases in manual data curation
- Human intervention in data cascades
- Benchmarks in machine learning, AI, and related areas
- Privacy and security issues of data quality, e.g., data poisoning attacks

Expected Contributions

We welcome five types of research contributions:

- **Survey papers**: the core focus of the work should be to generate new insights about human in the loop data curation, rather than the specific method applied (up to 25 pages).
- **Methodology papers**: should have a core focus on the test of the effectiveness of a proposed method (up to 25 pages).
- **Reproduction papers**: should provide new insights by presenting a holistic view of a topic, and should try to reproduce results documented in prior work (up to 25 pages).
- **Resource papers**: should present a new resource, such as a dataset or tool, or an interesting compilation of multiple datasets (up to 15 pages).
- **Use case papers**: could present new insights about a specific use case, such as an event or a community (up to 15 pages).

Important Dates

- Submission deadline: 28 February 2023
- First-round review decisions: 1 June 2023
- Deadline for revision submissions: 1 August 2023
- Notification of final decisions: 1 October 2023
- Camera-ready Manuscript: 15 November 2023
- Tentative publication: December 2023

Submission Information

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 at https://dl.acm.org/journal/jdiq/author-guidelines#subm

Please submit the paper by selecting, as the type of submission: "**SI: HITL-DC**".

For questions and further information, please contact **Gianluca Demartini, g.demartini@uq.edu.au.**

Anonymity requirements*: papers will be subject to single blind reviewing by three reviewers. Please submit anonymous manuscripts that do not contain information that identifies the authors or their organization.*