
QDB 2016

***International Workshop on
Quality in Databases***

<http://www.dbis.rwth-aachen.de/QDB2016/>

***in conjunction with VLDB 2016
(<http://vldb2016.persistent.com/index.php>)***

New Delhi, India

Monday, September 5, 2016

Call for Papers

Data quality problems arise frequently when data is integrated from disparate sources. In the context of Big Data applications, data quality is becoming more important because of the unprecedented volume, large variety, and high velocity. The challenges caused by volume and velocity of Big Data have been addressed by many research projects and commercial solutions and can be partially solved by modern, scalable data management systems. However, variety remains to be a daunting challenge for Big Data Integration and requires also special methods for data quality management. Variety (or heterogeneity) exists at several levels: at the instance level, the same entity might be described with different attributes; at the schema level, the data is structured with various schemas; but also at the level of the modeling language, different data models can be used (e.g., relational, XML, or a document-oriented JSON representation). This might lead to data quality issues such as consistency, understandability, or completeness.

The heterogeneity of data sources in the Big Data Era requires new integration approaches which can handle the large volume and speed of the generated data as well as the variety and quality of the data. Thus, heterogeneity and data quality are seen as challenges for many Big Data applications. While in some applications, a limited data quality for individual data items does not cause serious problems when a huge amount of data is aggregated, data quality problems in data sources are often revealed by the integration of these sources with other information. Data quality has been coined as 'fitness for use'; thus, if data is used in another context than originally planned, data quality might become an issue. Similar observations have been also made for data warehouses which lead to a separate research area about data warehouse quality.

The workshop QDB 2016 aims at discussing recent advances and challenges on data quality management in database systems, and focuses especially on problems related to Big Data Integration and Big Data Quality.

Research Topics

Topics covered by the workshop include, but are not restricted to, the following

Big Data Quality

- Data quality in Big Data integration
- Data quality models
- Data quality in data streams
- Data quality management for Big Data systems
- Data cleaning, deduplication, record linkage
- Big Data Provenance, Auditing

Big Data Integration

- Big Data systems for data integration
- Real-time (On-the-fly) data integration
- Graph-based algorithms for Big Data integration
- Integration and analytics over large-scale data stores
- Data integration for data lakes
- Efficiency and optimization opportunities in Big Data Integration
- Data Stream Integration

Management of Heterogeneous Data

- Query processing, indexing and storage for heterogeneous data
- Information retrieval over semi-structured or unstructured data
- Efficient index structures for keyword queries
- Query processing of keyword queries
- Data visualization for heterogeneous data
- Management of heterogeneous graph structures
- Knowledge discovery, clustering, data mining for heterogeneous Data

Schema and Metadata Management

- Innovative algorithms and systems for "Schema-on-Read"
- Schema inference in semi-structured data
- Pay-as-you-go schema definition
- Schema & graph summarization techniques
- Metadata models for Big Data
- Schema matching for Big Data

Important Dates

- **Submission: June 3, 2016 (EXTENDED)**
- **Notification: July 6, 2016**
- **Camera-Ready Version: July 15, 2016**
- **Workshop Date: September 5, 2016**

Paper Submission

QDB welcomes full paper submission of original and previously unpublished research. All submissions will be peer-reviewed, and once accepted will be included in the workshop proceedings.

Submission Guidelines:

- Full-length papers are accepted through the online submission system of the workshop. Full papers can be up to **8 pages** in length including all figures, tables and references. It should be submitted as a PDF according to the VLDB format. Templates can be found at http://vldb2016.persistent.com/formatting_guidelines.php
- We also encourage submission of short papers (up to 4 pages) reporting work in progress.
- Submissions in PDF are to be uploaded to the workshop's EasyChair submission site : <https://easychair.org/conferences/?conf=qdb16>

Workshop Proceedings

The proceedings of the workshop will be published online as a volume of the CEUR Workshop Proceedings (<http://www.ceur-ws.org>, ISSN 1613-0073), a well-known website for publishing workshop proceedings. It is indexed by the major publication portals, such as Citeseer, DBLP and Google Scholar.

Furthermore, the authors of the best papers of the workshop will be invited to submit an extended version of their work to a special issue of the ACM Journal of Data and Information Quality (<http://jdiq.acm.org/>).

Workshop Organizers

Laure Berti, Qatar Computing Research Institute, Qatar
Verikat N. Gudivada, East Carolina University, Greenville, USA
Rihan Hai, Information Systems, RWTH Aachen University, Germany
Christoph Quix, Fraunhofer FIT & RWTH Aachen University, Germany
Hongzhi Wang, Harbin Institute of Technology, China

Website

<http://www.dbis.rwth-aachen.de/QDB2016/>

Contact

qdb2016@dbis.rwth-aachen.de