Editorial: Special Issue on Quality Assessment and Management in Big Data—Part II

SHADI ALJAWARNEH, Jordan University of Science and Technology (JUST), Jordan
JUAN A. LARA, Madrid Open University, Udim, Madrid, Spain

CCS Concepts: • **Information systems** → **Data management systems**; • **Social and professional topics** → **Quality assurance**;

Additional Key Words and Phrases: Quality assessment, quality management, big data

ACM Reference format:

Shadi Aljawarneh and Juan A. Lara. 2021. Editorial: Special Issue on Quality Assessment and Management in Big Data—Part II. *J. Data and Information Quality* 13, 3, Article 13 (May 2021), 3 pages.
<https://doi.org/10.1145/3449056>

1 INTRODUCTION

It is a pleasure for us to introduce this Special Issue on Quality Assessment and Management in Big Data—Part II, *Journal of Data and Information Quality*, ACM. We have received 27 original submissions from which 11 final papers have been selected for publication (after a rigorous peer review process) in this issue divided into two parts. This editorial corresponds to Part II, in which we included papers related to foundational aspects of quality management in big data scenarios.

In the era of big data [1], organizations are dealing with tremendous amounts of data, which are fast-moving and can originate from various sources, such as social networks [2], unstructured data from various websites [3], or raw feeds from sensors [4]. Big data solutions are used to optimize business processes and reduce decision-making times, so as to improve operational effectiveness. Big data practitioners are experiencing a huge number of data quality problems [5]. These can be time-consuming to solve or even lead to incorrect data analytics. How to manage quality in big data has become challenging, and thus far research has only addressed limited aspects. Given the complex nature of big data, traditional data quality management approaches cannot simply be applied to big data quality management.

2 ARTICLES INCLUDED IN THE SPECIAL ISSUE

In Part II of this special issue, a series of papers addressing issues on foundational aspects of data quality management in big data have been published, including topics such as diffusion and

Authors' addresses: S. Aljawarneh, Madrid Open University, UDIMA, Carretera de La Coruña, Km 38,500 Vía de Servicio 15, 28400, Collado Villalba, Madrid, Spain; email: saaljawarneh@just.edu.jo; J. A. Lara, Jordan University of Science and Technology, JUST, P. O. Box 3030, Irbid 22110, Jordan; email: juanalfonso.lara@udima.es.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).

1936-1955/2021/05-ART13

<https://doi.org/10.1145/3449056>

enhanced exemplar methods, data readiness assurance, the cold-star problem, normalization of noisy text, and repairs explanation in Internet of Things.

After several rounds of review, five papers were selected for publication in Part II of the special issue. Those papers are described next.

In “[A Hybrid Inpainting Model Combining Diffusion and Enhanced Exemplar Methods](#)” by Sreelakshmy and Kovoor, they present an inpainting (image editing technique) model, in which a novel hybrid inpainting algorithm is implemented. This new method adds the benefits of a diffusion-based inpainting method to an enhanced exemplar algorithm. The structure part of the image is dealt with a diffusion-based method, followed by applying an adaptive patch size-based exemplar inpainting. Due to its hybrid nature, the proposed model stands out for the quality of output obtained by applying conventional methods individually. A new term, coefficient of smoothness, is introduced in the model, which is used in the computation of adaptive patch size for the enhanced exemplar method.

In “[Ensuring Data Readiness for Quality Requirements with Help from Procedure Reuse](#)” by Chirkova et al., the authors focus on the problem of determining whether the available data-transforming procedures can be used together to bring about the desired quality characteristics of the data in business or analytics processes. They conduct a formal study in which individual data cleaning, data migration, or other data-transforming tools are abstracted as black-box procedures with only some of the properties exposed, such as their applicability requirements, the parts of the data that the procedure modifies, and the conditions that the data satisfy once the procedure has been applied. As a proof of concept, the authors provide foundational results on sequential applications of procedures abstracted in this way, to achieve prespecified data quality objectives, for the use case of relational data and for procedures described by standard relational constraints.

In “[Ontology-Based E-learning Content Recommender System for Addressing the Pure Cold-Start Problem](#)” by Jeevamol et al., the authors introduce a semantic framework based on ontology to address the pure cold-start problem in content recommenders. The ontology encapsulates the domain knowledge about the learners as well as Learning Objects. The semantic model that they built has been experimented with different combinations of the key learner parameters such as learning style, knowledge level, and background knowledge. The proposed framework utilizes these parameters to build natural learner groups from the learner ontology using SPARQL queries.

In “[An Unsupervised Normalization Algorithm for Noisy Text: A Case Study for Information Retrieval and Stance Detection](#)” by Roy et al., the authors propose an unsupervised algorithm for text normalization that does not need any training data/human intervention. The proposed algorithm is applicable to text over different languages, and can handle both machine-generated and human-generated noise. Experiments over several standard datasets show that text normalization through the proposed algorithm enables better retrieval and stance detection, as compared to that using several baseline text normalization methods.

Finally, in “[Algorithms and Case Study for Explaining Repairs with Uniform Profiles over IoT Data](#)” by Liu et al., the authors present TsClean, a new prototype system for detecting and repairing outliers with explanations over IoT data. The framework defines uniform profiles to explain the outliers detected by various algorithms, including the outliers with variant time intervals, and take approaches to repair outliers. Both batch and streaming processing are supported in a uniform framework. In particular, by varying the block size, it provides a tradeoff between computing the accurate results and approximating with efficient incremental computation. In this article, the authors present several case studies of applying TsClean in industry, e.g., how this framework works in detecting and repairing outliers over excavator water temperature data, and how to get reasonable explanations and repairs for the detected outliers in tracking excavators.

We finally thank the Editor-in-Chief, Tiziana Catarci, and the rest of the editorial team. We also thank all the authors for their interesting contributions, and the reviewers for their valuable help.

REFERENCES

- [1] Juan A. Lara, Aurea A. De Sojo, Shadi Aljawarneh, Robert P. Schumaker, and Bassam Al-Shargabi. 2020. Developing big data projects in open university engineering courses: Lessons learned. *IEEE Access* 8 (2020), 22988–23001.
- [2] Sancheng Peng, Shui Yu, and Peter Mueller, 2018. Social networking big data: Opportunities, solutions, and challenges, *Future Generation Computer Systems* 86 (2018), 1456–1458. <https://doi.org/10.1016/j.future.2018.05.040>
- [3] Sharma Sunny and Manisha Malhotra. 2021. Web usage mining issues in big data: Challenges and opportunities. In *Impacts and Challenges of Cloud Business Intelligence*. edited by Shadi Aljawarneh, and Manisha Malhotra. IGI Global, Hershey, PA, 102–112. <http://doi:10.4018/978-1-7998-5040-3.ch007>
- [4] S. Boubiche, D. E. Boubiche, A. Bilami, and H. Toral-Cruz. 2018. Big data challenges and data aggregation strategies in wireless sensor networks. 2018. *IEEE Access* 6 (2018), 20558–20571.
- [5] M. Abdallah. 2019. Big Data Quality Challenges, In *2019 International Conference on Big Data and Computational Intelligence (ICBDCT'19)*, 1–3.