Computer Science and Information Systems 20(3): v–vii https://doi.org/10.2298/CSIS230300vH

Guest Editorial: Advances in Intelligent Data, Data Engineering, and Information Systems

Mírian Halfeld Ferrari[1], Paolo Ceravolo[2], Sonja Ristić[3], Yaser Jararweh[4], and Dimitrios Katsaros[5]

[1] Université d'Orléans, INSA CVL, LIFO EA, France
[2] Università degli Studi di Milano, Italy
[3] University of Novi Sad, Serbia
[4] Duquesne University, USA
[5] University of Thessaly, Greece

The Special Section on Advances in Intelligent Data, Data Engineering, and Information Systems contains papers selected from the workshops that have been held within the framework of the 25th European Conference on Advances in Databases and Information Systems ADBIS 2021, during August 24–26, 2021, at Tartu, Estonia. ADBIS 2021 conference was aimed at providing a forum where researchers and practitioners in the fields of databases and information systems can interact, exchange ideas and disseminate their accomplishments and visions. Within the scope of the Conference five workshops were held:

- DOING'21: Intelligent Data – from data to knowledge;
- SIMPDA'21: Data-Driven Process Discovery and Analysis;
- MADEISD'21: Modern Approaches in Data Engineering and Information System Design;
- MegaData'21: Advances in Data Systems Management, Engineering, and Analytics; and
- CAoNS'21: Computational Aspects of Network Science.

The authors of the best workshop papers were invited to submit extended versions of their papers in a special section of the journal Computer Science and Information Systems. Extended versions of submitted papers went through a rigorous reviewing procedure, the same as for regularly submitted papers. Finally, we accepted four papers presenting both theoretical and practical contributions. In the following, the accepted papers are briefly outlined.

In the first paper "Multi-perspective Approach for Curating and Exploring the History of Climate Change in Latin America within Digital Newspapers," by the authors Genoveva Vargas-Solar, Jose-Luis Zechinelli-Martini, Javier A. Espinosa-Oviedo, and Luis M. Vilches-Blazquez, an extended description of the Latin American Climate Change Evolution platform called LACLICHEV is proposed. The objective of LACLICHEV is to provide an integrated platform to expose and study meteorological events described in historical newspapers that are possibly related to the history of climate change in Latin America. Exploring the history of climate change through digitalized newspapers published around two centuries ago introduces four challenges: (1) curating content for tracking entries describing meteorological events; (2) processing colloquial language for extracting meteorological events; (3) analyzing newspapers to discover meteorological patterns

possibly associated with climate change; and (4) designing tools for exploring the extracted content. Presented results contribute to data curation and exploration adapted for Spanish textual content within digital newspaper collections. Authors used well-known information retrieval and analytics techniques, within a data exploration environment LACLICHEV that provides tools for curating, exploring, and analyzing historical newspaper articles, their description and location, and the vocabularies used for referring to meteorological events. The platform makes it possible to understand and identify possible patterns and models that can build an empirical and social view of the history of climate change in the Latin American region.

The authors of the second paper entitled "Matching Business Process Behavior with Encoding Techniques via Meta-Learning: An anomaly detection study," Gabriel Marques Tavares and Sylvio Barbon Jr, focus on the detection of anomalous traces in business process event logs that can diminish an event log's quality. They combine the representational power of encoding with a Meta-learning strategy to enhance the detection of anomalous traces in event logs towards fitting the best discriminative capability between common and irregular traces. Their approach creates an event log profile and recommends the most suitable encoding technique to increase anomaly detection performance. They used eight encoding techniques from different families, 80 log descriptors, 168 event logs, and six anomaly types for experiments. The presented results indicate that event log characteristics influence the representational capability of encodings. The authors analyzed the influence of meta-features on the recommended encoding technique. This analysis leveraged the understanding of which features better capture process behavior in the context of anomaly detection.

The authors Sidra Aslam and Michael Mrissa in the third paper "A Framework for Privacy-aware and Secure Decentralized Data Storage" present a decentralized data storage and access framework that ensures data security, privacy, and mutability in the wood supply chain scenario. The proposed framework integrates blockchain technology with Distributed Hash Table (DHT), a role-based access control model, and different types of encryption techniques. Their solution allows authorized actors to write, read, delete, update their data and manage transaction history on a decentralized system. The proposed traceability algorithm enables authorized actors to trace the product data in a decentralized ledger. The main limitations of existing solutions are a single point of failure, data mutability, and public availability of the data. The presented prototype design is flexible to expand and can be easily reused for different application domains such as medicine, and agriculture. The security and privacy analysis of the proposed solution is given, as well as the results of the performance evaluation in terms of time cost and scalability. The experimental results have shown that the proposed solution is scalable, secure, and achieves an acceptable time cost.

The authors Johannes Kastner and Peter M. Fischer in the last paper "Detecting and Analyzing Fine-Grained User Roles in Social Media" have proposed a method on how to determine and label user roles in large-scale social media data sets. This largely automated and scalable detection method combines unsupervised learning (more specifically, hierarchical clustering) to discover the classes of users over a wide range of features and supervised learning – generalizing the knowledge from manually labeled smaller data sets. Presented results of the analysis on a range of large data sets from Twitter show that well-separated roles can consistently be recognized and transferred. The labeling achieves

high accuracy not only within the same data set, but also on new data sets from different event types and/or years apart. The approaches scale well with little need for human intervention and the resource requirements of such analyses are modest, bringing them in the range of commodity hardware.

We sincerely thank the workshop organizers for their support in selecting papers and especially the reviewers for their valuable comments to improve selected papers. We also thank all authors for their contribution to this special section. Special thanks are given to prof. Mirjana Ivanović, the Editor in Chief of ComSIS, for providing us the opportunity to publish this special section, valuable comments in improving the quality of selected papers, and support in the whole process.