Editorial

This issue of the fourteenth volume of the Journal of Information and Data Management (JIDM) comprises extended versions of full papers published in the proceedings of the Brazilian Symposium on Databases (SBBD) in 2022. SBBD is the official event on databases of the Brazilian Computer Society (SBC) and the largest venue in Latin America. SBBD covers issues in data management, database and information systems research, data science, Big Data, and other related topics, in the light of considering data as a technological cornerstone of emerging applications.

This special issue consists of 7 articles. All the articles have at least 30% of new content material to the original paper (SBBD 2022) and were extensively revised by active members of the Brazilian Database community. Covered topics include data mining, information and data integration, data quality, and data science. This special issue starts with the article entitled *"Adaptive Fast XGBoost for Multiclass Classification"*, by Fabiano Baldo, Julia Grando, Yuji Yamada Correa, and Deivity Amorim Policarpo. Their article presents an XGBoost-based classifier called AFXGB-MC to quickly classify non-stationary data streams with multiple classes. They compare the classifier with six state-of-the-art algorithms for multiclass classification found in the literature.

The second article, entitled *"Improved Generalization of Cyclist Detection on Security Cameras with the OpenImages Cyclists dataset"*, by Ednilza Nardi, Bruno Padilha, Leonardo Tadashi Kamaura and João Eduardo Ferreira provide a new OpenImages Cyclists dataset, built through the pre-selection of images from the OpenImages set. They also present a new algorithm for the semiautomatic generation of cyclist annotation aided by people and bicycle detectors.

The third article, entitled *"Using Active Learning for Segmentation and Semantic Classification of Legal Acts Extracted from Official Diaries"*, addresses an effort to deal with two essential aspects of official government diaries: text segmentation and semantic classification. To this end, it presents an active learning strategy that allows repeated interactions between the classifier and the annotator, selecting a reduced number of the most valuable instances for training. The article is authored by Kattiana Constantino, Thiago H. P. Silva, João Vítor B. Silva, Victor Augusto L. Cruz, Otávio M. M. Zucheratto, Marcos Carvalho, Welton Santos, Celso França, Claudio M. V. de Andrade, Alberto H. F. Laender, and Marcos André Gonçalves.

The fourth article, entitled *"Hurricane: a Dataflow-oriented Data Service for Smart City Applications"*. Hurricane executes multiple dataflows to gather, pre-process, integrate, and share data. The article is authored by Maicon Banni, Maria Luiza Falci, Isabel Rosseti, and Daniel de Oliveira.

The next article of this special issue discusses data quality in a data warehouse with governmental information of the Brazilian state of Minas Gerais. To this end, it presents a brief comparison of eight open-source data quality tools and then chooses the Great Expectations tool for analyzing such data in two real applications: public bids and public expenditure. The article, entitled *"Assessing Data Quality Inconsistencies in Brazilian Governmental Data"*, is authored by Gabriel P. Oliveira, Bárbara M. A. Mendes, Clara A. Bacha, Lucas L. Costa, Larissa D. Gomide, Mariana O. Silva, Michele A. Brandão, Anisio Lacerda and Gisele L. Pappa.

The sixth article, entitled *"Contextual Reinforcement, Entity Delimitation and Generative Data Augmentation for Entity Recognition and Relation Extraction in Official Documents"*, is authored by Fabiano Muniz Belém, Claudio Valiense, Celso França, Marcos Carvalho, Marcelo Ganem, Gabriel Teixeira, Gabriel Jallais, Alberto H. F. Laender and Marcos A. Gonçalves. The article proposes new techniques for contextual reinforcement and entity delimitation based on pre- and post-processing techniques to provide a richer semantic context, improving SpERT, a state-of-the-art Span-based Entity and Relation Transformer.

The final article, entitled *"Class Schema Discovery from Semi-Structured Data"*, presents CoFFee, an approach to class schema discovery. The approach provides a summarized set with core attributes

by applying a strategy that combines attributes co-occurrence and frequency. The article is authored by Everaldo Costa Neto, Johny Moreira, Luciano Barbosa, and Ana Carolina Salgado.

The JIDM Editorial board is very thankful to all authors and reviewers for their valuable contributions towards the creation of this rich special issue. We wish you all pleasant and insightful readings.

Damires Souza
Special Issue Editor

Daniel de Oliveira
Editor-in-Chief

Ticiana Linhares
Associate Editor