Editorial

This JIDM Special Issue brings twelve extended versions of papers published as full papers on the 2021 SBBD Dataset Showcase Workshop (SBBD-DSW). The purpose of SBBD-DSW is to provide a forum to share and discuss ways of building and organizing useful data sets which can be used for data analysis and exploration in Computer Science and other Sciences. Sixteen full papers were accepted for presentation at the SBBD-DSW 2021 and invited to submit an extended version to this JIDM special issue, incorporating at least 30% of new content with respect to the original paper. From the sixteen invited papers, twelve submitted an extended version, which have been reviewed by at least three reviewers, from a committee composed of senior researchers from the database community.

The works presented in this Special Issue can be grouped into five topics: (i) Documents, containing a Brazilian corpus of essays repository, and a repository of research questionnaires extracted from HTML documents; (ii) Society, containing data sets with fact-checking instances collected from six different Brazilian fact-checking agencies, a criminal data set for the generation of Police Patrol routes, and a data set for analyzing elections in Brazil; (iii) Culture, containing an organized public domain Portuguese-language literature data set, an open data set with temporal musical success information, and a data set of Brazilian popular songs of success and non-success; (iv) Covid, bringing cured and enriched data sets from the national vaccination campaign against COVID-19, and a repository of data from multiple sources on the COVID-19 pandemic in Brazil; and (v) Other applications, presenting data sets generated from the compilation of visual features extracted from many public image data sets, and the data set of stock quotes for Machine Learning on all companies from B3.

We know that collecting and processing data for scientific experiments reports to a database community can be exhausting, laborious, and very challenging. On the other hand, if the data set is well organized, with a well-designed execution methodology, the implementation and comparison of algorithms can be highly facilitated. The main purpose of our work is to provide data sets of quality and that follow the best methods of collection, cleaning and organization. All the data sets presented in here are publicly available for use in any suitable experiment. So, take advantage of the data organized by the authors of this special edition and have fun!

We would like to thank all the authors and reviewers that contributed in putting together this special issue.

Enjoy your reading, and data sets.

Carina F. Dorneles
Eric Araújo
DSW 2021 - Guest Editors