Editorial: Smart Data Management and Applications

Edgar Bisset Alvarez[1]

Editorial

The scientific and technological development achieved by humanity in recent years has led the human being to face a high increase in the amount of data and information available in their surroundings. The large volume of data generated by a diversity of sources, such as mobile applications, sensors, vehicles, buildings, and human beings, also brought about a demand for the correct storage, processing, publication, and verification of the same, to allow each individual to have access to quality information to make the right decisions at certain times in their life.

The content produced and disseminated on the web, and especially on online social networks, has posed great challenges for people. At the same time, these social networking environments have allowed for an increase in the sharing, production, use, and consumption of information, opening new perspectives for the evaluation of science, through the so-called social media metrics. Because of this reality, the scientific community has been questioning, are the current data processing techniques sufficient to interpret the content of social media? That is why, researchers from different fields of knowledge such as data science, natural language processing, data engineering, big data, research assessment, network science, sociology of science, and communication are working to face technological, economic, political, and social challenges imposed by the social media metrics of science.

This special issue includes eight high-quality articles presented at the First International Conference on Data and Information in Online Environments (DIONE 2020).

The first article presents a study of how recurrent neural networks (RNN) are used to make forecasts in the stock market, as well as finding an explanation for the behavior of financial assets and dealing with large volumes of sequences of values. The study points out that most of the studies are intended to predict the result of a few stocks. The survey brought together similar market stocks based on their correlation and using the K-averages algorithm, groups of similar stocks were grouped. Subsequently, an RNN was applied to predict possible stock prices. The results showed that the inventory grouping did not influence the network's efficiency, asthe trend was predicted correctly on average 48% of the time. These results can be of great help to investors, allowing them to buy shares with the same behavior. Besides, the work pointed out that future research includes analyzing the impact of different measures of similarity and how this can influence the process of grouping stocks.

The second article proposes an analysis of the privacy principles of personal data prescribed in the General Data Protection Regulation (GDPR) and the treatment of data in blockchain technology in its primary version, guided by the phases of the Data Life Cycle (DLC). The research aimed to identify the influence that the phases of the DLC have on the adequacy of the principles of the GDPR and in the treatment of data carried out in blockchain technology to lead the adaptation of the technology to this conformity. During the data collection, some data treatment options were presented, which made it possible to conclude that the results achieved during the research can be used as a guide in the systematization of the processes of legal adaptation in the process of adopting Blockchain technology.

The third article covers another form of application of blockchain technology, presenting a study on how permissioned blockchain networks can safely accommodate collaborative business models to provide thriving business alliances. The paper describes examples of cooperative business models and business relationship guidance and how they create value when combined with allowed blockchain networks - a more business-oriented blockchain variety. The research draws on case studies to highlight how simple it is to put an authorized blockchain into practice so that companies achieve closer ties with business partners. The paper highlights how companies are beginning to take advantage of the intrinsic characteristics that have made blockchain so notorious in their operations, such as security, integrity, and

✉ Edgar Bisset Alvarez
edgar.bisset@ufsc.br

[1] Federal University of Santa Catarina, R. Eng. Agronômico Andrei Cristian Ferreira, s/n - Trindade, Florianópolis, SC 88040-900, Brazil

transparency, allowing companies to cooperate with business partners in a protected and secure environment.

The fourth article presents a study on mobile data science and smart apps, and how AI-based modeling can be used to design and develop smart mobile apps for human life enhancement, based on the use of artificial intelligence (AI) techniques in learning machine, deep learning methods, and natural language processing. The survey conducted tests with multiple AI-based smart applications across multiple application domains. The study presents a brief discussion of how successful intelligent mobile systems should have the relevant AI-based modeling, depending on the characteristics of the data, as well as the need for sophisticated algorithms to be trained through collected data and application-related knowledge destination before the system can help users with suggestions and decision making. Finally, the study highlights several research questions and relevant paths for analysis in the field of mobile data science and smart applications.

The fifth article proposes the use of a generic source data collection method based on the W3C Provenance Data Model (PROV-DM) that aims to create a way to capture, store and analyze data from the anemia index by applying a workflow scientific research, together with the management of knowledge provenance. The research is part of a larger study carried out by the same authors in a Brazilian hemotherapy center. The study was developed using an exploratory, practical and deductive method carried out with real data from 197,551 blood donor candidates, extracted from reports from 2000 to 2018. The research was able to identify that women had the highest levels of inappropriate blood donations product of the high rates of anemia in blood. The results achieved during the research, allow us to confirm that the generic method of data collection used can be applied in several areas of knowledge.

The sixh article presents a study on the use of technologies associated with intelligent data processing in real-time can be of great value in the health field, including artificial intelligence, data analysis technologies, and flow processing technologies. The research presents a systolic blood pressure prediction system that can predict SBP in real-time and, therefore, can prevent health problems that can result from sudden hypertension. A system was proposed using models of Recurrent Neural Network (RNN), Long Term Memory (LSTM), Bidirectional Short-Term Memory (BI-LSTM), Models of Recurring Units with Gated (GRU), and Medical Information Center for Intensive Care (MIMCII) SBP time series data set. At the end of this stage of the research, it was realized that the use of the proposed system achieved a better performance using three hidden layers and can be used to predict the near future of the SBP in real-time, which allowed identifying the need to develop a system health-based cloud services to provide a systolic blood pressure prediction service to users/patients.

The seventh article seeks to extract and classify the entities detected in a text written in natural language using natural language processing (NLP) techniques. The research proposes a methodology to improve the Portuguese-based Named Entity Recognition (NER), which uses journalistic corpora as a basis for the formation of textual corpora. Based on the hypothesis that journalistic writing has the best adherence to the contemporaneity of any language, since it preserves characteristics such as objectivity, simplicity, impartiality, and is a reference for transmitting information without ambiguities. The paper provides a model for extracting entities and evaluate the results obtained with the use of Recurrent Neural Network architectures. The results achieved show that the proposed methodology applied to the Portuguese language not only exceeds the average accuracy found in the literature by increasing it from 83.38% to 85.64%, but it can also decrease the computational costs related to NER processing tasks.

The eighth article describes the terminological composition of information management (IM) and knowledge management (KM) to show how much these areas overlap in the library and information sciences (LIS) space. To fulfill the proposed objectives, the researchers used two approaches: 'overlay visualization' and 'topic detection', by which conducted a term co-occurrence analysis using documents extracted from the Web of Science indexed during the period 1980–2015. According to the authors, the results might allow to (re) define the curricular contents of the LIS programs, as well as the research and practical activities carried out by academics and managers.

Edgar Bisset Alvarez is graduate and M. Sc in Library and Information Science from the University of Havana, Cuba, and the Ph.D. degree in Information Science from the São Paulo State University, Brazil. He is currently a professor with the Federal University of Santa Catarina, Brazil, and serving as Editor of Encontros Bibli Journal. Associated with the Cuban Library Association (ASCUBI), also with the Working Group GT-8 Information and Technology of the National Association for Research and Graduate Studies in Information Science (ANCIB), Brazil. He participates as Co-Leader of the Research Group in Philosophy of Information and is a researcher of the research group ITI-RG (Intelligence, Technology, and Information - Research Group) at the Federal University of Santa Catarina.