Guest Editorial Preface

Special Issue of the French Speaking Conference (EDA 2017) on Business Intelligence and Big Data

Fadila Bentayeb, University of Lyon – Lyon 2, Lyon, France
Hassan Badir, ENSA Tangier, Tangier, Morocco
Omar Boussaid, University of Lyon – Lyon 2, Lyon, France

EDA is a yearly French-speaking conference. The acronym in French stands for Entrepôts de Données et Analyse en ligne whose English translation is Data Warehouses and Online Analytical Processing (OLAP). The conference started in 2005 and has been held in Lyon (2005), Versailles (2006), Poitiers (2007), Toulouse (2008), Montpellier (2009), Djerba (2010), Clermont-Ferrand (2011), Bordeaux (2012), Vichy (2014), Bruxelles (2015), Aix-en-Provence, (2016), Lyon (2017), Tangier (2018), Montpellier (2019) and Lyon (2020) – ADBIS-TPDL-EDA joint conferences.

This volume of IJDWM journal which is a special issue of EDA conference contains five papers: three papers from EDA 2017 conference which held at Lyon (France) during May 23-25, 2017 and two IJDWM standard issue papers. Initially, the program committee of EDA'17 selected five papers and invited their respective authors to submit an extended version of their work for the present volume. After two rounds of reviews, the lecture committee of this special issue retained three papers, considering that these have been substantially expanded. The selected papers present an overview of the research work pursued in the domain of large data management and data analysis as well. In the following, a short description and goals of each paper of this special issue is presented.

The first paper entitled "OCL Constraints Checking on NoSQL Systems Through an MDA-Based Approach" authored by Fatma Abdelhedi, Amal Ait Brahim and Gilles Zurfluh addresses the problem of the automatic storage of Big Data in NoSQL systems. They present an automatic approach where UML conceptual models are translated into NoSQL physical models. A major phase in the translation process is to check the OCL constraints. For this, the authors present a framework to map the OCL constraints defined in the conceptual model to the code that must run at the physical level. Their approach uses two transformations, namely, OCL2JavaModel and JavaModel2JavaCode. To evaluate their approach, the authors implemented a tool that supports the mapping of four fundamental OCL expressions: Iterate-based expressions, OCL predefined operations, If expression and Let expression.

The second paper entitled "Enhancing Data Quality at ETL Stage of Data Warehousing" authored by Neha Gupta and Sakshi Jolly deals with data quality management. Data usually comes into data warehouses from multiple sources having different formats and are specifically categorized into three groups, i.e., structured, semi-structured and unstructured. Various data mining technologies are used to collect, refine and analyze the data which further leads to the problem of data quality management . Data purgation occurs when the data is subject to ETL methodology in order to maintain and improve the data quality. The data may contain unnecessary information and may have inappropriate symbols which can be defined as dummy values, cryptic

values or missing values. The present work has improved the Expectation-Maximization algorithm with dot product to handle cryptic data, DBSCAN method with Gower metrics to ensure dummy values, Wards algorithm with Minkowski distance to improve the results of contradicting data and K-means algorithm along with Euclidean distance metrics to handle missing values in a dataset. These distance metrics have improved the data quality and also helped in providing consistent data to be loaded into a data warehouse.

The third paper by Sami Belkacem, Kamel Boukhalfa, whose title is "Ranking News Feed Updates on Social Media: A Review and Expertise-Aware Approach" deals with users' news feed updates in social media. Due to the large amount of data, users are overwhelmed by updates displayed chronologically in their news feed. The authors propose then to help users quickly to catch up with the relevant updates. They first study approaches proposed in this area according to four main criteria: features that may influence relevance, relevance prediction models, training and evaluation methods, and evaluation platforms. Then, the authors propose an approach that leverages another type of feature which is the expertise of the update's author for the corresponding topics. Experimental results on Twitter highlight that judging expertise, which has not been considered in the academic and the industrial communities, is crucial for maximizing the relevance of updates in news feeds.

The fourth paper by Abdelilah Balamane which title is "Scalable Biclustering Algorithm Considers the Presence or Absence of Properties" presents a biclustering procedure that introduces the negation or absence of object properties within bliclusters. Indeed, it could be beneficial in several application domains such as organized crimes, genetics or digital marketing to identify homogeneous groups of similar objects in terms of both the presence and the absence of attributes. In this paper, the author proposes a scalable and efficient algorithm of biclustering that exploits a binary matrix to produce at least three types of biclusters where the cell's column (i) are filled with 1's, (ii) are filled with 0's, and some columns filled with 1's and/or with 0's (iii). This procedure is scalable and it's executed without having to consider the complementary of the initial binary context. The implementation and validation of the method on data sets illustrates its potential in the discovery of relevant patterns.

The last paper is entitled "Image Retrieval using Intensity Gradients and Texture Chromatic Pattern: Satellite Images Retrieval" and authored by Hoang Long, I.Jeena Jacob, Betty Paulraj, P. Ebby Darney, Tran Tuan, Harold Robinson Yesudhas, Vimal Shanmuganathan and Golden Julie Eanoch. This paper deals with images retrieval domain. Methods to retrieve images involve retrieving images from the database by using features of it. They are colour, shape and texture. These features are used to find the similarity for the query image with that of images in the database. The images are sorted in the order with this similarity. The proposed work uses intra- and inter- texture chrominance and its intensity. Here inter-chromatic texture feature is extracted by LOCTP (Local Oppugnant Colored Texture Pattern). Local Binary Pattern (LBP) gives the intra-texture information. Histogram of Oriented Gradient (HoG) is used to get the shape information from the satellite images. The proposed work gives better results when it is compared with other works.

CONCLUSION

We are grateful to the authors for submitting their research paper to IJDWM journal: those who made the effort to extend their work for this special issue submission and those who submit their paper to the standard issue. We also thank the reviewers for their support in evaluating the submitted papers providing invaluable hints and constructive comments to the authors for ameliorating the quality of the selected papers. We hope that the reader will find this special

issue interesting in opening new, future research. We would like also to express a special thank to Alexis Mille, Assistant Development Editor, for all his help and guidance. And finally, we want to whole heartedly thank David Taniar, Editor-in-Chief of IJDWM, who immediately supported our project of publishing this special issue. This special issue would not have been possible without their support.

Fadila Bentayeb
Hassan Badir
Omar Boussaid
Guest Editors
IJDWM