



Guest editorial: special issue on Web and big data 2019

Jie Shao¹ · Man Lung Yiu² · Toyoda Masashi³

Published online: 18 February 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

We are delighted to present this special issue of *World Wide Web* on Web and big data 2019. Recent years have witnessed the explosion need of Web and big data management in various fields. The emergence of big data from different Web applications continues to attract increasing attention from researchers to study new data-related problems. This development poses challenges to traditional data management, data mining, and machine learning paradigms.

We organized this special issue in conjunction with the call for papers and participations of the 3rd Asia Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data (APWeb-WAIM 2019), held in Chengdu, China during August 1–3, 2019. Selected high quality papers (including the best paper and best paper runner-up of APWeb-WAIM 2019) were invited to submit an extended version to this special issue in accordance with customary practice of adding at least 30% new materials for journal publication. The revised papers again went through the review process in accordance with the journal guidelines and are finally presented to the readers in the present form.

The five extended papers in this special issue cover a variety of topics. In the first paper, “Geographical Address Representation Learning for Address Matching”, Shan et al. propose a novel method for matching addresses which has important applications in location-based businesses, by getting rich contexts through search engines. The second paper, “Medical Treatment Migration Behavior Prediction and Recommendation Model Based on Health Insurance Data”, presents an Attention-based Bidirectional Gated Recurrent Unit (AB-GRU) medical treatment migration model to predict which hospital patients will go to in the future. In the third paper, “MDSSES: A Multi-Document Semantic Extraction System using News Data,” Sheng et al. present a system to aid users in quickly discerning salient connections and facts

✉ Jie Shao
shaojie@uestc.edu.cn

Man Lung Yiu
csmlyiu@comp.polyu.edu.hk

Toyoda Masashi
toyoda@tkl.iis.u-tokyo.ac.jp

¹ University of Electronic Science and Technology of China, Chengdu 611731, China

² Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR

³ The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

from a set of related documents, via graph-based visualizations of relationships between concepts even across documents. The fourth paper, “A Framework for Image Dark Data Assessment”, presents an assessment framework combining deep learning, hash technique and graph-based computing, which helps users to detect the potential value of image dark data and avoid unnecessary mining cost. Finally, in “Exploring Nonnegative and Low-Rank Correlation for Noise-Resistant Spectral Clustering”, Wang et al. develop a new spectral clustering method which jointly explores the nonnegative low-rank properties of data correlation and adaptively models the structural sparsity of data noise.

Collectively, these five papers illustrate the diverse range of exciting data-intensive applications, which interact with our online and offline activities more pervasively. We hope that the readers enjoy this special issue. We would like to acknowledge the work done by all authors and their willingness to contribute their papers for this special issue. We thank all the reviewers for their time commitment which is necessary to assure the high quality articles. Finally, we would like to thank Prof. Yanchun Zhang, the Editor-in-Chief of *World Wide Web*, for the guidance and support in the completion of this special issue.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.