



Guest editorial: special issue on Web data querying, mining, and privacy preserving

Yunjun Gao¹ · Lu Chen²

Published online: 14 May 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

With the rapidly growing volume of webs, various types of web data (e.g., location, text information, graph data, etc.) are produced, and thus, corresponding areas of web technologies, methodologies, and applications are investigated in order to analyze and process those kinds of data. Specifically, efficient and effective querying and mining algorithms are needed to satisfy the requirements of analyzing the web data in real-life applications. Moreover, the web data could leak out during the processing, and hence, privacy preserving is another important challenge for analyzing and processing web data.

This special issue was preceded by the 18th International Conference on Web Information Systems Engineering (WISE 2017), held in Moscow region, Russia, during October 7–11, 2017. We send invitations to the authors of five selected high quality papers accepted in WISE 2017. All the 5 submissions are accepted and comprise this special issue. As guest editors of the special issue, we carefully examined the feedback and discussions provided by the authors and reviewers throughout the revision cycle. The issue presents a collection of articles that illustrate the diversity and richness of current research in web data processing, which is organized as follows. The first paper studies the topic of querying web data, while the next two papers investigate the topic of mining on web data. Last but not least, the last two papers focus on the privacy preserving on the web data processing.

The first paper “Spatio-Temporal Top- k Term Search over Sliding Window” by Lisi Chen, Shuo Shang, Bin Yao, and Kai Zheng discovers top- k most frequent nearby terms over a sliding window, as massive volumes of geo-tagged streaming text messages are becoming available on social media. The authors develop a novel and efficient mechanism to solve the problem, including a quad-tree based indexing structure, indexing update technique, and a best-first based searching algorithm.

✉ Yunjun Gao
gaoyj@zju.edu.cn

Lu Chen
luchen@cs.aau.dk

¹ College of Computer Science, Zhejiang University, 38 Zheda Road, Zhejiang 310027 Hangzhou, China

² Department of Computer Science, Aalborg University, Selma Lagerlöfs Vej 300, 9220 Aalborg, Denmark

The second paper “Mining Maximal Sub-Prevalent Co-location Patterns” by Lizhen Wang, Xuguang Bao, Lihua Zhou, and Hongmei Chen explores the problem of spatial prevalent co-location pattern mining to discover interesting and potentially useful patterns from spatial data. The authors introduce a new concept called sub-prevalent co-location patterns, and propose two efficient algorithms.

The third paper “Large-scale Holistic Approach to Web Block Classification: Assembling the Jigsaws of a Web Page Puzzle” by Andrey Kravchenko investigates the problem of web block classification, whereas web blocks (e.g., navigation menus, advertisements, headers, footers, and sidebars) are ubiquitous across the web. The authors propose to take a holistic view of the page where all block classifiers in the classification system interact with each other, so that accuracies of individual classifiers can be improved through this interaction.

The fourth paper “Towards Secure and Truthful Task Assignment in Spatial Crowdsourcing” by Dongjun Zhai, Yue Sun, An Liu, Zhixu Li, Guanfeng Liu, Lei Zhao, and Kai Zheng aims to protect the location privacy for the task assignment in spatial crowdsourcing. The authors present a privacy-preserving reverse auction based assignment model that consists of two parts. The first part generalizes private location to travel cost, and protects it by an anonymity based data aggregation protocol. The second part utilizes a reverse auction task assignment algorithm to encourage workers to offer authentic data.

The last paper “Adapting HTML5 Web Applications to User Privacy Preferences” by Georgia M. Kapitsaki and Theodoros Charalambous tries to address the issue that users can give their consent on the use of this sensitive information, but should have the right to express their privacy preferences. The authors specify a privacy preferences language for users tailored to HTML5 web applications employing the eXtensible Access Control Markup Language, and introduce a mechanism that adapts the web application considering these user preferences.

There exists a fertile ground for innovative research at the aspect of querying, mining, and privacy preserving on web data. The five papers of this special issue demonstrate how to use querying and mining to analyze the web data efficiently in order to provide more types of web services, and how to perform privacy preserving to improve the quality of web service. More work is needed to address the challenges of processing and analyzing the web data such as those discussed in the special issue.

The guest editors express their appreciation to the authors for their high quality work and their contribution to the state-of-the-art of this emerging domain. We would like to thank all reviewers for their helpful feedback and comments. Special thanks go to Editor-in-Chief Yanchun Zhang for his unwavering support. We are also grateful to the Springer staff who worked tirelessly in helping to bring this project to fruition.

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