**Editorial for the Special Section Related to C-Big**

*Prashant Krishnamurthy and Vladimir I. Zadorozhny*

. Many applications and industries apply analytics, computing and data mining techniques with data coming from disparate data sources. The challenges with utilizing data from disparate sources are many - the data may be in differing formats, may have conflicts and gaps, may come with varying levels of quality and trust, may need varying levels of security and privacy, and in general may need very different approaches for management and handling. Selecting the appropriate data, cleaning and merging them efficiently and in a timely manner, using the available resources in the best possible manner, are additional challenges. At the same time, there are obvious benefits to using data from a variety of sources in a collaborative manner. In many cases the outcomes of the analytics will be very different depending on the availability or lack thereof of certain data.

This special issue of Information Systems is devoted to a small selection of the problems associated with collaborative big data. We have included three papers here, each addressing a problem in the collaborative big data world. The special issue was a result of the first workshop on Collaborative Big Data (C-Big) that was held in Pittsburgh in 2012. While the papers are not all from this workshop, the workshop did provide an impetus for this special issue. We had an open call for papers and selected a subset of the submissions for this special issue.

The first paper in the special issue by Rexit, Tsui, Espino, Chrysanthis, Wesaratchakit, and Ye considers one challenge in fusing data from different sources, quickly and in an automated manner. As an example of this problem, the authors examine selecting the data that are health indicators for a given population. The authors use aggregate sales of *over the counter* medicines with aggregate counts of *visits* to hospitals' emergency rooms to identify the medicines, whose sales may accurately predict disease outbreaks. The problem is how to accomplish this in a speedy and efficient manner. The paper evaluates several approaches and presents a preferable "greedy search" algorithm for selecting the subset of over the counter medicines that are good predictors of disease outbreaks.

The second paper in the special issue by Fabian, Ermakova, and Junghanns presents an architecture for sharing data across organizations, while maintaining the required security and privacy levels. The threat model considers cloud service providers who may be curious, but it is a first step at carefully addressing the serious security and privacy concerns that arise in collaborative big data analytics. The authors do present a comprehensive list of security and privacy requirements in the paper. The application area that is considered in this paper is again healthcare, but the use of business process notation makes it possible to extract the architectural aspects that are suitable for other application scenarios.

Finally, the paper by Ulltveit-Moe and Oleschuk describes an approach to provide privacy or anonymization using XML messages, with the ability to reveal identities where needed by using secret-key encryption. XML messages provide a standard approach for data sharing across organizations, making them suitable for collaborative data analytics. The work in this paper provides an early step for maintaining privacy (and reducing it where appropriate) based on specific policies.

All of the papers in the special issue have a sub-focus on the performance aspects of the approaches, i.e., how efficiently can the techniques be implemented and whether they are actually feasible (of course under the assumptions stated in the paper).

The editors of this special issue would like to thank the authors for their submissions and the reviewers for their valuable efforts in making this special issue a reality.