CrossMark

Preface to the special issue on big data search and mining

Kai Zheng[1] · Feifei Li[2] · Kyuseok Shim[3]

We are delighted to present this special issue of World Wide Web on Big Data Search and Mining. Recent years have witnessed the explosion of data generated from a wide range of enterprises and applications at an unprecedented way. Big Data management deals with tapping large amount of data that is complex with a wide variety of data types and provides actionable insights at the right time.

The aim of this special issue is to investigate the recent development of new technologies, platforms, and frameworks that can support scalable search and mining over a variety of Big Data. The guest editors selected 5 contributions covers varying topics within this theme, ranging from social media search to semantic analysis, from data stream analysis to spatial query processing. Many problems in this area share common methods including indexing and probabilistic data structures.

The first article, by Xia et al. on "Top-k Temporal Keyword Search Over Social Media Data" investigates a problem of top-k most significant temporal keyword query to enable more complex query analysis over social media data. They design a novel temporal inverted index with two-tiers posting list to index social time series and a segment store to compute the exact social significance of social items.

In "Latent Semantic Diagnosis in Traditional Chinese Medicine", the authors develop a multi-content LDA-based model to find out the pathogenesis based on the latent semantic analysis of symptoms and the corresponding herbs.

✉ Kai Zheng
zhengkai@suda.edu.cn

Feifei Li
lifeifei@cs.utah.edu

Kyuseok Shim
kshim@snu.ac.kr

[1] Soochow University, Suzhou, China

[2] University of Utah, Salt Lake City, UT, USA

[3] Seoul National University, Seoul, South Korea

Fang et al. in "Distributed Stream Join under Workload Variance" present a novel flexible and adaptive scheme partitioning model for stream join operator, which ensures high throughput but with economical resource usages by allocating resources on demand.

In "Novel Structures for Counting Frequent Items in Time Decayed Streams", the authors revisit the problem of identifying frequent item over data streams based on a new streaming model, and propose an innovative heap structure, named Quasi-heap, which maintains the item order using a lazy update mechanism.

In "Discovery of Probabilistic Nearest Neighbors in Traffic-Aware Spatial Networks", the authors study a novel problem of discovering probabilistic nearest neighbors and planning the corresponding travel routes in traffic-aware spatial networks (TANN queries) to avoid potential time delay or traffic congestions.

The special issue was preceded by the 18th Asia-Pacific Web International Conference (APWeb 2016, held at Suzhou (September 2016 in Suzhou, China). All the articles have undergone at least two rounds of rigorous peer-review according to the journal's high standards. We would like to thank all the reviewers involved for their invaluable input.

The guest editors believe the papers appearing in this issue form an accurate representation of current topics in the big data management and hope these articles will stimulate further development in this area. The editors express their appreciation to the authors and reviewers for contributing to this special issue.

We hope you enjoy this special issue and take some inspiration from it for your own future research.

Kai Zheng, Feifei Li and Kyuseok Shim
26/02/2017