

Guest Editor's editorial: ranking in databases

Kaushik Chakrabarti

Published online: 15 June 2012
© Springer Science+Business Media, LLC 2012

We are dealing with increasingly large volumes of data as well as more complex and diverse types of data. Examples of such data include log data like transaction logs and web logs, text, image and multimedia data, scientific data and sensor data. There is an increasing need for effective techniques to search, explore and analyze such datasets. Similarity queries and top- k ranking are important paradigms to search and explore such datasets. The purpose of this special issue on ranking in databases is to cover the new directions of research in this area. This issue contains four research papers, which are briefly discussed as follows.

The paper “Combining CPU and GPU architectures for fast similarity search” proposes a novel, less-explored but nevertheless important direction to speed up similarity search. While most previous works focus on new indexing techniques, this paper studies how to parallelize similarity search using combination of many-core GPU devices and multicore CPU processors. The paper shows how modern architectures can be used to speed up ranking algorithms.

In the paper “On optimality-ratio and coverage in ranking of joined search results”, the authors study a novel ranking problem. Instead of ranking individual items, they consider ranking of combinations of items, e.g., a combination of a hotel and two restaurants. They study the semantics and query processing algorithms in this context. The paper shows the kind of new ranking problems that emerge in these new-age applications.

The paper titled “Distributed top- k query processing by exploiting skyline summaries” studies top- k processing in distributed environments. With increasing volumes of data, the data is typically distributed over multiple servers. Processing top- k

K. Chakrabarti (✉)
Data Management, Exploration and Mining Group, Microsoft Research, One Microsoft Way,
Redmond, WA 98052, USA
e-mail: kaushik@microsoft.com

queries in such settings introduces novel research challenges which are addressed in the paper. The paper proposes a novel approach of summarizing the data in each server and then identifying, based on those summaries, the servers that contain top- k results. The paper shows novel research challenges that arise when the datasets are large and distributed.

While the most common approach to explore datasets in search and retrieval, an alternative approach is dissemination. In the paper “Distributed top- k full-text content dissemination”, the authors study the distributed top- k document dissemination problem. The paper identifies the challenges in the dissemination problem, namely, timeliness, relevance and network cost and proposes novel solutions. The paper shows the potential of this alternate form of data exploration.

These four papers cover several aspects of the ranking problem, starting from the potential of exploiting modern architectures, to novel ranking problems, to challenges in distributed environments, to alternate forms of data exploration like dissemination. Thus, this special issue will appeal to both the experts in the field as well as those who wish a snapshot of the new research directions in the area of ranking and similarity search.