Exploration and Mining of Web Repositories

Gautam Das
Computer Science and Engineering
University of Texas at Arlington
gdas@uta.edu

ABSTRACT

With the proliferation of very large data repositories hidden behind web interfaces, e.g., keyword search, form-like search and hierarchical/graph-based browsing interfaces for Amazon.com, eBay.com, etc., efficient ways of searching, exploring and/or mining such web data are of increasing importance. There are two key challenges facing these tasks: how to properly understand web interfaces, and how to bypass the interface restrictions. In this tutorial, we start with a general overview of web search and data mining, including various exciting applications enabled by the effective search, exploration, and mining of web repositories. Then, we focus on the fundamental developments in the field, including web interface understanding, sampling, and data analytics over web repositories with various types of interfaces. We also discuss the potential changes required for query processing, data mining and machine learning algorithms to be applied to web data. Our goal is two-fold: one is to promote the awareness of existing web data search/exploration/mining techniques among all web researchers who are interested in leveraging web data, and the other is to encourage researchers, especially those who have not previously worked in web search and mining before, to initiate their own research in these exciting areas.

Biography

Gautam Das is a Full Professor in the Computer Science and Engineering Department of the University of Texas at Arlington. Prior to UTA, Dr. Das has held positions at Microsoft Research, Compaq Corporation and the University of Memphis, as well as visiting positions at IBM Research. He graduated with a BTech in computer science from IIT Kanpur, India, and with a PhD in computer science from the University of Wisconsin- Madison. Dr. Das's research interests span social computing, data mining, information retrieval, databases, graph and network algorithms, and computational geometry. His research has resulted in over 150 papers, many of which have appeared in premier conferences and journals. He is the recipient of the IEEE ICDE 2012 Influential Paper Award. His research has been supported by grants from federal and state agencies such as US National Science Foundation, US Office of Naval Research, US Department of Education, Texas Higher Education Coordinating Board, Qatar National Research Fund, as well as industry such as Cadence, Nokia, Apollo, and Microsoft.

The 20th International Conference on Management of Data (COMAD),
17th-19th Dec 2014 at Hyderabad, India.