8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'03)

Mohammed J. Zaki
Computer Science Department
Rensselaer Polytechnic Institute
Troy, NY 12180
zaki@cs.rpi.edu

Charu C. Aggarwal
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
charu@us.ibm.com

Foreword

This is the 8th workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'03), held annually in conjunction with ACM SIGMOD conference. The workshop aims to bring together data-mining researchers and experienced practitioners with the goal of discussing the next generation of data-mining tools. Rather than following a "mini-conference" format focusing on the presentation of polished research results, the DMKD workshop aims to foster an informal atmosphere, where researchers and practitioners can freely interact through short presentations and open discussions on their ongoing work as well as forward-looking research visions/experiences for future data-mining applications.

In addition to research on novel data-mining algorithms and experiences with innovative mining systems and applications, of particular interest in this year's DMKD workshop is the broad theme of "Future Trends in Data Mining." Topics of interest include:

- Privacy & Security issues in Mining: Privacy preserving data mining techniques are invaluable in cases where one may not look at the detailed data, but one is allowed to infer high level information. This also has relevance for the use of mining for national security applications.

- Mining Data Streams: In many emerging applications data arrives and needs to be processed on a continuous basis, i.e., there is need for mining without the benefit of several passes over a static, persistent snapshot.

- Data Mining in Bioinformatics and Biological Database Systems: High-performance data mining tools will play a crucial role in the analysis of the ever-growing databases of bio-sequences/structures.

- Semi/Un-Structured Mining for the World Wide Web: The vast amounts of information with little or no structure on the web raise a host challenging mining problems such as web resource discovery and topic distillation; web structure/linkage mining; intelligent web searching and crawling; personalization of web content.

- Future Trends/Past Reflections: What are the emerging topics/applications for next generation data mining? What lessons have been learned from over a decade of data mining? What areas need more attention?

Out of the 28 paper submissions we were able to select only 8 full papers (an acceptance rate of 28.5%). In addition 5 short papers were accepted. Each paper was reviewed by three members of the program committee. Along with an invited talk, we were able to assemble a very exciting program.

We would like to thank all the authors, invited speakers, and attendees for contributing to the success of the workshop. Special thanks to the program committee for help in reviewing the submissions, and ensuring a high quality program.

Sincerely,
Mohammed Zaki, Charu Aggarwal (editors)

Workshop Co-Chairs

Mohammed J. Zaki, Rensselaer Polytechnic Institute
Charu Aggarwal, IBM T.J. Watson Research Center

Program Committee

Roberto Bayardo, IBM Almaden Research Center, USA
Alok Choudhary, Northwestern University, USA
Gautam Das, Microsoft Research, USA
Venkatesh Ganti, Microsoft Research, USA
Minos N. Garofalakis, Bell Labs, USA
Dimitrios Gunopulos, University of California, Riverside, USA
Jiawei Han, University of Illinois at Urbana-Champaign, USA
Eamonn Keogh, University of California, Riverside, USA
Nick Koudas, AT&T Research, USA
Vipin Kumar, University of Minnesota, USA
Bing Liu, University of Illinois at Chicago, USA
Rosa Meo, University of Torino, Italy
Raymond Ng, University of British Columbia, Canada
Srini Parthasarathy, Ohio State University, USA
Rajeev Rastogi, Bell Labs, USA
Kyuseok Shim, Seoul National University, South Korea
Hannu Toivonen, University of Helsinki, Finland
Philip S. Yu, IBM T.J. Watson Research Center, USA

Workshop Program & Contents

The workshop will foster an informal atmosphere, where participants can freely interact through short presentations and open discussions on their ongoing work as well as discussion on current/future trends in data mining. Questions can be asked freely during any time, and discussion slots have been set aside in the morning and afternoon sessions. These slots are meant to focus on critical evaluation of current approaches and open problems. All speakers and participants are encouraged to think about these issues prior to the workshop and to take active part in the ensuing discussions.

8:00-8:30 Continental Breakfast

8:30-8:35 Opening Remarks

8:35-9:10, Invited Talk.

9:10-9:50 Data Streams I

9:50-10:15 Coffee Break

10:15-11:20 DB Integration

11:20-11:30 Short Discussion Session (Open Floor)

11:30-12:30 FCRC Plenary Talk by James Kurose on Networking

12:30-1:45 Lunch

1:45-2:25 WWW Mining

2:25-3:05 Data Streams II

3:05-3:45 Bioinformatics

3:45-4:00 Coffee Break

4:00-4:50 Privacy & Security

4:50-5:25 Discussion Session (Open Floor)

5:25-5:30 Closing Remarks.