

Letter from the Editor-in-chief

Dear members and readers,

It has been customary to collect and publish in the December issue of Explorations reports from organizers of workshops and the KDD cup at the most recent SIGKDD conference event, to give an account of the activities, presentations, results and encounters, and share experiences with our readers. I don't believe we ever managed to get reports from all the workshops and this time is no exception. This year, the SIGKDD conference organized 12 workshops, including a workshop dedicated to the KDD cup. While all organizers were personally invited to send us a report, only ten did. This is already a good outcome and those who could not come to Las Vegas to enjoy the conference or did not have the opportunity to attend some workshops, they have now the option to read the chronicle of most 2008 KDD workshops. The workshop organizers were also invited to send us one selected paper from their meeting. Unfortunately, not counting KDD cup, only two workshops took advantage of this opportunity. The organizers of the workshop on Data Mining in Business Applications (DMBA), who did not provide a report for our issue, did share with us a paper by Michael Giering on sales prediction and item recommendation. WebKDD, which celebrated the tenth year occurrence at SIGKDD this year, organized a best paper award sponsored by Yahoo! Research. The selected article from WebKDD is the winner of this award. Interestingly, this paper by Al Mamunur Rashid, George Karypis, and John Riedl also deals with the topic of recommendation systems.

We started to publish book reviews with last issue of Explorations and I sent a call for more of such informative reviews. In this issue, courtesy of Olfa Nasraoui, we also have another book review of a now prevalent book on web mining written by Bing Liu, who is a past associate editor of our newsletter and a recognized academic researcher in data mining and web content mining. If you would like to share your opinion on a data mining related book, please send me a short review.

We would like to also start publishing in our newsletter abstracts of PhD dissertations in the field of knowledge discovery and data mining. These abstracts could be very informative to our general members but also, for new graduate students, they could be revealing of the trends in data mining research and the challenges of open problems. Coincidentally, this year's SIGKDD conference inaugurated a new award dedicated to PhD dissertations. The four first abstracts published in explorations are abstracts of the winner, Xiaoxin Yin, and runner-ups of this award, Jimeng Sun, David Martens and Pradeep Ravikumar. Congratulations to all of them. I would like to seize this opportunity to call for PhD dissertation abstracts for PhD work that has been defended. Depending on the number of submissions, we will formulate selection criteria to pick the best for publication.

In my June foreword, I reported on the significant impact some papers published in Explorations have in the community based on the number of citations they received. The

citation statistics, collected from Google scholar, that I tried to convey in my letter draw much attention and many commented on this interesting analysis. Among other things, I listed 10 papers cited more than 100 times in June 2008. On this data some pointed out that a number of papers were missing. In particular the paper "*Tools for privacy preserving distributed data mining*" by Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Zhu, published in vol. 4, no. 2, in 2002. In June it was cited 173 times (currently 180 times) according to Google Scholar. This omission is not intentional and I apologize to the authors. The anomaly comes from a not so rare occurrence of mismatch between the paper title in the final copy and the title in the proceedings (or in our case newsletter) table of contents. The paper in question for instance is titled "*Tools for privacy preserving data mining*" in the table of contents of Volume 4 (notice the missing "*distributed*"). It is unfortunately very common that paper authors decide to make small changes to the title just before camera ready copy submission and do not inform the editors or program chairs. I was proceedings chair of SIGKDD 2002 and noticed many cases where the authors changed the title and sometimes the order or even composition of the authors' list without informing the organizers and considering the repercussions and side effects. A recent analysis of paper citations by Google Scholar from a list of conferences extracted from the DBLP database¹ revealed that many papers were never cited. A closer look identified discrepancies between announced and final titles, exposing the fact that this practice is unfortunately not rare. DBLP obtains its lists from table of contents of proceedings while Google Scholar extracts the titles from the on-line published postscripts or PDF files. I hope that at least the readers of this letter become conscious of this issue and avoid changing their paper titles without informing organizers.

Finally, I would like to promote our next issue which will be a special issue on *Open Source Analytics and the Interoperability of Analytic Applications*, guest edited by Robert Grossman. A call for papers can be found at the end of the issue.

Happy reading.

Osmar R. Zaïane

¹ The DBLP Computer Science Bibliography: <http://www.informatik.uni-trier.de/~ley/db/>.