

## Letter from the Editor-in-chief

Dear members and readers,

You probably noticed that this issue is rather small in size compared to some previous ones. We nevertheless have six exciting contributions. In the past, issues of Explorations either covered a specific topic, as in a special issue, or compiled collections of data mining related workshop reports and reports associated to our famous KDD cup. This current issue does neither. In the transition from my predecessor to me as editor-in-chief, we did not secure guest editors for a special issue in June. However, we are discussing some possible topics for the coming issues. Stay tuned. These topics will be announced soon. I expressly welcome proposals for special issues on any relevant data mining topics that would be of interest to the KDD community. The special issues are very appreciated by our members and attract significant attention, at least by academics, as attested to by the citations we can find. I shall report on this later. Here is a list of past special issues:

v2.1 (2000)	Internet Data Mining
v2.2 (2000)	Scalable Data Mining
v3.2 (2002)	Online, Interactive, and Anytime Data Mining
v4.1 (2002)	Constraints in Data Mining
v4.2 (2002)	Privacy and Security
v5.1 (2003)	Multi Relational Data Mining
v5.2 (2003)	Microarray Data Mining
v6.1 (2004)	Learning from Imbalanced Datasets
v6.2 (2004)	Web Content Mining
v7.1 (2005)	Natural Language Processing and Text Mining
v7.2 (2005)	Link Mining
v8.1 (2006)	Successful Real World Data Mining Applications
v9.1 (2007)	Data Mining for Health Informatics
v9.2 (2007)	Visual Analytics

**Table 1: List of SIGKDD Explorations Special Issues**

I have been associate editor of SIGKDD Explorations since 2004 and I have noticed a trend. We are receiving more and more contributed articles outside the scope of a special issue. These are generally welcome as long as they cover problems of interest to the data mining community, and certainly enhance our periodical. Such articles are important especially for issues such as this one that do not cover a special topic. They undergo a review process and not all are accepted. For instance, less than 50% were accepted this time. I believe this trend is due to the success of SIGKDD Explorations. It is perceived by many as an archival journal. But is it?

It is important to note that articles submitted to special issues are thoroughly reviewed by a committee set by the guest editors while regular papers are reviewed by colleagues assigned by the associated editors or myself. While up to three reviewers may assess a submission, reviews are admittedly not as rigorous and systematic as for a regular journal.

Perhaps it is because of the quality of papers in the special issues, perhaps due to other SIG newsletters, such as SIGMOD records, being indexed with an impact factor in the Thomson ISI index, or perhaps it is the wide distribution Explorations gets. The reason for this perception, that Explorations is a full-fledged journal, is not clear, but the fact is

that Explorations has a measurable impact. Querying Google scholar<sup>1</sup> for each paper from previous issues, we can notice that some have been cited substantially. The most popular articles are survey papers, but other specific and technical papers are also appreciated and referenced. In the last 18 issues, ten articles were cited in the literature more than 100 times (See Table 2). The paper *Web usage mining: discovery and applications of usage patterns from web data* by Srivastava, Cooley, Deshpande and Tan was even cited 842 times, a very respectable success. 21 other papers were cited between 50 and 100 times, an enviable achievement for our newsletter. This attests to the quality of our publications.

Article	Issue	Citations
Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data	V1, N2, 2000	842
Web Mining Research: A Survey	V2, N1, 2000	608
Algorithms for Association Rule Mining - A General Survey and Comparison	V2, N1, 2000	298
Information diffusion through blogspace	V6, N2, 2004	166
Mining Frequent Patterns with Counting Inference	V2, N2, 2000	150
Support Vector Machines: Hype or Hallelujah?	V2, N2, 2000	149
A Survey of Data Mining and Knowledge Discovery Software Tools	V1, N1, 1999	128
Subspace clustering for high dimensional data: a review	V6, N1, 2004	126
State of the art of graph-based data mining	V5, N1, 2003	119
Mining with rarity: a unifying framework	V6, N1, 2004	105

**Table 2: Papers cited more than 100 times according to Google Scholar – June 2008 data**

Given the obvious impact factor of our newsletter and the precedence of other SIG newsletters in the ISI science citation index, I am in the process of discussing with ACM press to intervene for the inclusion of SIGKDD Explorations in the Thomson ISI index. However, it is a long process and controversial outcome.

Conceivably, we should not consider Explorations as a journal, or question the role of the newsletter. For now, the role is undoubtedly not for archival and simply a means to inform members and provide relevant, revealing coverage of pertinent topics to the data mining community. This could be articles on novel methods, surveys on the state-of-the-art, conference reports or interviews as we had in the past. I would like to also start having reviews for books, tools or datasets as these are also very informative and useful to our members and readers. We start this new theme in this very issue with one book review by Bruce D. McCullough.

To close, I would like to thank my associate editors, Charu Aggarwal and Bart Goethals, to whom I am very grateful, and all the contributors of this issue. The editorial team is looking forward to receiving your special issue proposals, your reports, book reviews, articles and other relevant contributions.

Happy reading.

Osmar R. Zaiane

<sup>1</sup> <http://scholar.google.com/> (queried in June 2008).

# Appendix

Article	Issue	Citations
Scalability for Clustering Algorithms Revisited	V2, N1, 2001	94
A survey of kernels for structured data	V5, N1, 2003	86
Mining Frequent Patterns by Pattern-Growth: Methodology and Implications	V2, N2, 2000	85
Cryptographic Techniques for Privacy-Preserving Data Mining	V4, N2, 2002	83
Editorial: special issue on learning from imbalanced data sets	V6, N1, 2004	83
Concept-Based Knowledge Discovery in Texts Extracted from the Web	V2, N1, 2000	82
Probabilistic logic learning	V5, N1, 2003	80
A Perspective on Inductive Databases	V4, N2, 2002	78
A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research	V1, N1, 1999	78
Why so many clustering algorithms: a position paper	V4, N1, 2002	71
A study of the behavior of several methods for balancing machine learning training data	V6, N1, 2004	69
The Inference Problem: A Survey	V4, N2, 2002	69
Mining Data Streams under Block Evolution	V3, N2, 2002	65
Multi-relational data mining: an introduction	V5, N1, 2003	65
Theoretical Frameworks for Data Mining	V1, N2, 2000	61
Link mining: a survey	V7, N2, 2005	60
A Survey on Wavelet Applications in Data Mining	V4, N2, 2002	59
Learning by Googling	V6, N2, 2004	58
Understanding the Crucial Differences Between Classification and Discovery of Association Rules	V2, N1, 2000	56
Link mining: a new data mining challenge	V5, N1, 2003	53
Constrained frequent pattern mining: a pattern-growth view	V4, N1, 2002	52

**Table 3: Papers cited between 50 and 100 times according to Google Scholar – June 2008 data**

