

Editor's Notes

Welcome to the December 2015 issue of the ACM SIGMOD Record!

This issue opens with a Database Principles article by Fagin et al., which presents a relational framework for Information Extraction (IE), namely, discovering structured information in textual content. In particular, the article presents a framework, called *document spanners*, for examining the expressiveness of rule languages for IE. The spanner representation systems include regex formulas, spanner algebra, basic extraction programs, and automata. The article gives important results on the expressiveness of these representation systems. It further offers a declarative language for specifying policies for conflict resolution among different rules. The article closes by discussing other formalisms related to spanners as well as some open research questions.

The Research and Vision Articles Column features a vision article, by Kumar et al., on “Model Selection Management Systems: The Next Frontier of Advanced Analytics”. This article is motivated by the observation that advanced analytics often requires running machine learning (ML) algorithms, which is an iterative process involving feature engineering, algorithm selection, and parameter tuning, collectively referred to as the *model selection* problem. Model selection, while being a highly time-consuming yet crucial task for advanced analytics, has been largely overlooked in the database community. This article envisions a new class of analytics systems called model selection management systems (MSMS), and discusses how time-tested ideas from database research offer new avenues to improving model selection.

The Surveys Column features a survey by Pournajaf et al. on “Participant Privacy in Mobile Crowd Sensing Task Management”. The article focuses on participant privacy concerns and solutions in the context of task management, in contrast to privacy issues related to data collection as studied in previous work. It presents a detailed classification of task management, identifies the categories of privacy threats to participants, and provides a detailed discussion of privacy mechanisms for each type of threat. The article finally discusses ongoing research and additional challenges regarding participant privacy in Mobile Crowd Sensing task management.

The Systems and Prototypes column features a data cleaning system, “Cleanix: a Parallel Big Data Cleaning System,” by Wang et al. As data cleaning is becoming a crucial task in big data analytics, Cleanix supports data cleaning at a large scale, with key features including: *scalability* on a shared-nothing cluster; *unification* of various automated data repairing tasks in a single parallel dataflow; and *usability* where users are offered with a simple and friendly graphical user interface for selecting data cleaning rules and visualization utilities for better understanding errors and fixing them.

The Distinguished Profiles column features Rick Snodgrass, Professor of Computer Science at the University of Arizona and an ACM Fellow. Rick has served as Editor-in-Chief of ACM Transactions on Database Systems, Chair of ACM SIGMOD, the ACM Pubs Board and the ACM History Committee. He has received the SIGMOD Outstanding Contributions Award and ACM Outstanding Contribution Award. Rick has been best known for his work on temporal databases. In this interview, he shared with us his thoughts on standards, branding, and his new research on “ergalics”.

This issue includes three event reports. The first article reports on the Second International Workshop on Exploratory Search in Databases and the Web (ExploreDB 2015), co-located with SIGMOD 2015. The workshop included two keynote talks and six peer-reviewed research papers, which investigated a wide range of topics including explore-by-example, reformulation of database queries,

ranked search, and query personalization. The second article reports on the PhD Workshop in Information and Knowledge Management (PIKM) co-located with ACM CIKM 2014. The PIKM workshop included a regular paper track and a short paper track, both with oral and poster presentations, to increase interaction between the presenters and the audience. It also included a special track with invited talks by experienced researchers as well as a keynote speech, providing additional guidance and advice to early PhD students.

The third article in the events column reports an interesting recent study, by Benevenuto et al., on whether ACM SIG conferences have indeed promoted collaborations in a variety of research communities. More specifically, this study investigates two questions: (1) How structured are the ACM SIG conference communities? and (2) Who are the individuals responsible for connecting each ACM SIG conference community? By examining 24 ACM SIG communities and datasets from DBLP and SHINE, the article reports findings including: (1) ACM SIGMOD ranks the first among 24 communities in terms of the coverage of the largest connected component of the coauthor graph, indicating that our community is well connected in terms of collaboration; (2) a set of researchers have contributed significantly to connecting the coauthor graphs, which are well aligned with those individuals who have won research awards in their respective communities.

On behalf of the SIGMOD Record Editorial board, I hope that you all enjoy reading the December 2015 issue of the SIGMOD Record!

Your submissions to the Record are welcome via the submission site:

<http://sigmod.hosting.acm.org/record>

Prior to submission, please read the Editorial Policy on the SIGMOD Record's website:

<http://www.sigmod.org/publications/sigmod-record/sigmod-record-editorial-policy>

Yanlei Diao

December 2015

Past SIGMOD Record Editors:

Ioana Manolescu (2009-2013)
Ling Liu (2000-2004)
Arie Segev (1989-1995)
Thomas J. Cook (1981-1983)
Daniel O'Connell (1971-1973)

Alexandros Labrinidis (2007-2009)
Michael Franklin (1996-2000)
Margaret H. Dunham (1986-1988)
Douglas S. Kerr (1976-1978)
Harrison R. Morse (1969)

Mario Nascimento (2005-2007)
Jennifer Widom (1995-1996)
Jon D. Clark (1984-1985)
Randall Rustin (1974-1975)