

Editor's Notes

Welcome to the December 2012 issue of the ACM SIGMOD Record!

The issue opens with the Database Principles column, where Arenas, Gutierrez, Miranker, Pérez and Sequeda summarize a set of results around the SPARQL standard established by the W3C for querying Semantic Web (RDF) data and point to a set of open interesting problems within the area of Semantic Web data management. Regarding SPARQL, the authors provide an algebraic semantics, and summarize known evaluation complexity results. Among the challenges, raised by the Linked Data applications, the authors discuss publication (of existing databases into linked RDF), data discovery, trust, provenance, and more.

In the survey column, Grandi proposes an annotated bibliography of temporal and evolution aspects of the Semantic Web. As models and languages for Semantic Web data mature, representing time and changes are natural extensions. This short-format survey provides a very comprehensive set of works, an online version of which is being maintained online by the author.

In the Systems and prototypes column includes an overview of the Deco system for answering declarative queries over relational databases and crowd sourced data, by Park, Pang, Parameswaran, Garcia-Molina, Polyzotis and Widom. The paper explains how to bring small and elegant extensions to the SQL language, optimization, and execution framework, in order to optimally exploit the crowd-sourced data.

Daniel Abadi is the recipient of the SIGMOD 2009 Doctoral Dissertation award and the guest of this month's the Distinguished Database Profiles column. Daniel discusses the benefits and drawbacks of column stores, interviewing for a job in academia versus company research labs, and how he regrets not having gone on an internship while he was a grad student! Given that Daniel is now a professor at Yale, it is probably too late for him now. However, it is certainly a good time for ongoing PhD students to absorb the advice.

The Research Centers column features two columns this month. First, Beskales and eight co-authors from the Qatar Computing Research Institute (QCRI), in Doha, Qatar, present their Data Analytics group, whose research spans over data quality (data cleaning, repair, and more), data profiling, some concrete data analytics applications explored on a project related the World Bank, and social media analytics. Second, Lu, Pedersen, Saltenis, Thomsen, Thomsen and Torp present the Center for Data-Intensive Systems (Daisy, in short, <http://daisy.aau.dk>) at Aalborg University in Denmark. The group is currently 28-strong and works on topics such as spatio-temporal data management, mobile services, data warehousing and BI. To the best of my knowledge, these two groups had not been the focus of the Record's Research Centers column recently, and I am very glad to see how our new editor has kickstarted his new position in style, with these two very interesting reports!

The Open Forum column opens with an article by Aggarwal, devoted to two important clustering techniques: projected clustering, and probabilistic latent semantic indexing (PLSI). These techniques were independently introduced in the SIGMOD 1999, respectively, SIGIR 1999 conferences, and have since been pursued independently in numerous follow-up works. The paper shows that the two techniques are fundamentally equivalent, assuming a probabilistic interpretation of the projected clustering problem. This opens the way for cross-domain adaptation of the techniques developed separately in previous research, and hopefully will provide better solutions to both problems.

Next in the Open Forum column is a very refreshing and informative contribution of Hristidis, from UC Riverside. Based on several years' worth of experience setting up collaborations with MD (Doctor of

Medicine) researchers, the author comments on the different cultural perspectives, expectations, and scientific practice of the medical and respectively the CS-oriented side of Medical Informatics. The article closes with a set of suggestions to ease the pain and increase the gains of such collaborations in the future. Three event reports appear in this issue. Dong and Dragut outline the discussion and sessions of the 10th International Workshop on Quality in Databases (QDB 2012), held next to the 2012 VLDB Conference in Istanbul, Turkey. The workshop sessions focused on the performance of entity resolution, data cleaning and truth discovery, and experience with real-life cleaning problems and systems.

Second, Hidders, Sroka and Missier present the works of the 1st workshop on Scalable Workflow Enactment Engines and Technology (SWEET 2012), held in conjunction with ACM 2012 in Scottsdale, Arizona. Two main application areas (and accordingly lines of work) emerged in the workshop: on one hand, systems for data-intensive computational science, and on the other hand, workflow and data analytics infrastructure for social media analysis. The workshop had a mix of reviewed and invited papers, and keynote speakers from Yahoo and Twitter.

The third report by Meng and Want focuses on the First Extremely Large Databases (XLDB) conference at Asia, held at Beijing, China in June 2012. The conference featured an impressive array of invited talks, lightning presentations and poster, focusing on technologies for “BigData”, on the merits and promises of the “NoSQL” movement, and other topics related to very large-scale data management.

The issue closes with a call for participation to the Heidelberg Forum of Mathematics and Computer Science, a novel yearly forum where young researchers will get an opportunity to interact with laureates of the Turing Award, and of the the Abel prize and the Fields medal in mathematics. Check it out!

2012 ACM fellows from the data management community

Last but certainly not least, we are extremely proud to have nine distinguished members of our community recently appointed ACM 2012 fellows! Congratulations to Gustavo Alonso, Rick Catell, Ahmed Elmagarmid, Wenfei Fan, Masaru Kitsuregawa, Leonid Libkin, Tova Milo, Rajeev Rastogi and Patrick Valduriez for the distinction they bring to our scientific area.

Your contributions to the Record are welcome via the RECESS submission site (<http://db.cs.pitt.edu/recess>). Prior to submitting, be sure to peruse the Editorial Policy on the SIGMOD Record’s Web site (<http://www.sigmod.org/publications/sigmod-record/sigmod-record-editorial-policy>).

Ioana Manolescu

December 2012

Past SIGMOD Record Editors:

Harrison R. Morse (1969)
Daniel O'Connell (1971 – 1973)
Randall Rustin (1974-1975)
Douglass Kerr (1976-1978)
Thomas J. Cook (1981 – 1983)
Jon D. Clark (1984 – 1985)
Margaret H. Dunham (1986 – 1988)
Arie Segev (1989 – 1995)
Jennifer Widom (1995 – 1996)
Michael Franklin (1996 – 2000)
Ling Liu (2000 – 2004)
Mario Nascimento (2005 – 2007)
Alexandros Labrinidis (2007 – 2009)