Reminiscences on Influential Papers

Kenneth A. Ross, editor

Unfortunately, this will be my last influential papers column. I've been editor for about five years now (how time flies!) and have enjoyed it immensely. I've always found it rewarding to step back and look at why we do the research we do, and this column makes a big contribution to the process of self-examination. Further, I feel that there's a strong need for ways to publicly and explicitly highlight "quality" in papers. Criticism is easy, and is the more common experience given the amount of reviewing (and being reviewed) we typically engage in. I look forward to seeing this column in future issues.

Ken Ross.

AnHai Doan, University of Illinois, `anhai@cs.uiuc.edu`.

[Richard Hamming: You and Your Research. Seminar Talk at Bell Communication Research, 1986; transcribed by J. F. Kaiser.]

In this paper Richard Hamming discusses what it takes to do great research. He considers multiple topics, ranging from well-known ones such as problem selection, courage, hard work, and communication skills, to less familiar ones such as the need to tolerate ambiguity. The topics were discussed in an engaging manner, and vividly illustrated with personal anecdotes. The paper as a whole was a lot of fun to read. It can be found on the Internet (and a 3-page summary is available by googling for "striving for greatness Hamming").

I first read this paper during my Ph.D. years, and have periodically reread it ever since. The paper has influenced me in three ways. First, it confirmed some of my vague ideas about the research process, and suggested new "tricks-of-the-trades". Second, it has and continues to inspire me to do my best in research. Hamming stresses the need to continually ask ourselves: "What am I doing? And what are the important problems in my field?". Reading this part periodically does help me to keep an eye on the big picture, and to put day-to-day concerns, such as the strong pressure to publish, in perspective.

Finally, the "imperfections" of this paper do provoke me to think deeper about the research process. For example, Hamming claims that first-class research is worth the effort because it is as good as wine, the opposite sex, and song put together. I found this claim ...um... not terribly convincing, but it did make me think about what makes many of us "suffer" long hours in this business. My own theory (developed after some wine) is that research is an *art*, and researchers are artists. Artists do not have jobs, but rather *callings*: the call to create lasting, beautiful messages, and to communicate them to (influence) others. Such messages are packaged in various forms: novels for writers, paintings for painters, and papers (and students) for us. Hence, as artists, we endure as we strive to create lasting work.

Sihem Amer-Yahia, AT&T Laboratories, `sihem@research.att.com`.

[Janet L. Wiener, Jeffrey F. Naughton: OODB Bulk Loading Revisited: The Partitioned-List Approach. VLDB 1995: 30–41,]

I read this paper when I was at INRIA, right at the beginning of my Ph.D. I did not know then what it meant to do database research. All I knew was the relational model, SQL and the existence of database systems. This paper had a great impact on my research in different ways.

First, although the focus its was on loading data into object-oriented databases, the paper connected what I then knew of databases altogether. I understood that data queried using SQL was not sitting in a database by miracle as it needed to be loaded from the outside world. I also understood that something beyond SQL was needed to do it.

Second, the paper connected the database industry world with the world of database research, which at that time, gave me a sense for research and, meant that designing loading algorithms was a tough research question.

Third, the algorithms proposed in the paper were not only implemented and tested inside Shore, a real system, but several times in the paper, the authors compared their design and performance with other loading solutions in other real systems.

Finally, although I had not fully realized it then, the paper and its preceding version, published in VLDB 1994, are what I consider thorough and really informative database performance papers. It looks at a real, apparently simple problem that many of us thought is solved and that many of us still encounter, maybe without realizing it. The algorithms are very elegant and make use of all the features of a database system such as sorting and indexing. The experimental evaluation is full of exciting details and looks at every aspect of loading: CPU time, buffer pool space, I/Os. Finally, the authors did a not-so-typical thing which is to learn from the results of their experiments and propose in the same paper, new ideas and actually give enough details to implement them as opposed to throwing in some future work directions. I really appreciate that now as it became the basis for my Ph.D work.

This paper convinced me to look at algorithmic issues in database systems. Unfortunately, I seem to not have managed to learn how to do such a detailed database performance evaluation yet!
