The Data Engineering Conference: Repeat from March, 2003

The Technical Committee on Data Engineering, in addition to publishing the Bulletin, also sponsors the Data Engineering Conference, referred to as ICDE ("International Conference on Data Engineering"). The most recent conference (for 2003) was held in Bangalore, India. The 2004 conference will be held in Boston, MA. A "Call for Papers" for this conference appears on the back inside cover of this issue.

The Data Engineering Conference is one of three large and prestigous annual database conferences, the others being SIGMOD and VLDB. It is the IEEE Computer Society's flagship conference in the database area. The program for the conference is excellent, the result of a very competitive paper selection process. Because of the quality of the conference, many of the leading researchers in our field regularly attend the conference. Further, any paper published in the conference is included in the SIGMOD Anthology's CD or DVD collection of database papers. So ICDE papers have a very wide readership.

I hope that many Bulletin readers will submit papers to the Data Engineering Conference, not only in 2004, but in subsequent years as well. Perhaps I shall have the pleasure of meeting you at the conference, as I very frequently attend.

The Current Issue

One of the more curious facts about citations in the database literature is that there are more citations to the original R-tree paper than there are to the original B-tree paper (see the DBLP web cite). There may be a number of reasons for this, but one of them has to be that it is much harder to assess the strength of an access method when dealing with multi-dimensional data than it is when dealing with a single dimension. With B-trees, dealing with a single dimension, it is relatively easy to characterize both storage utilization and query costs. All regions have at most two neighbors, and all queries partially overlap only with the two neighbors of a contiguous set of regions. And all B-trees can guarantee minimum storage utilization and are well characterized with respect to average utilization.

With multi-dimensional data, things are much more difficult. It is harder to divide the data into regions and each region has more boundaries, hence bordering on more neighbors. Hence there are many more ways to partition the search space. As well, there are many more ways to query the search space. And, of course, there is a strong connection between the querying and the search space partitioning. Finally, there is enormous variation in how data may be distributed over the multiple dimensions. All this makes it much harder to assess how good or appropriate a multi-dimensional access method is.

The current issue does not introduce new access methods or new query processing techniques. Rather it seeks to tackle the task of providing an infrastructure in which good research can be done for multi-dimensional query processing. While it is perhaps too much to hope for that this infrastructure will bring closure to this area (this is a difficult topic), nonetheless, such infrastructure will make it possible, both for those creating new techniques and for potential users, to judge the appropriateness of any given method for a given task, and to start the process of seriously evaluating the strengths and weaknesses of proposed multi-dimensional approaches. This is essential to progress in our field! I want to thank Christian Jensen, the issue editor for the current issue, for proposing and following through on this important area. This issue is very important reading for the multi-dimensional database research community.