

# Capsule Reviews

---

The Capsule Reviews are intended to provide a short succinct review of each paper in the issue, in order to bring the content to a wider readership. This issue's Capsule Reviews were compiled by Fairouz Kamarreddine. Professor Kamarreddine is an Associate Editor of *The Computer Journal* and is based in the Department of Computing and Electrical Engineering at Heriot-Watt University, Edinburgh, UK.

---

## **Persistent views—a mechanism for managing ageing data.** J. SKYT AND C. S. JENSEN

Data storage and management technologies bring about some of the most fascinating challenges. Often, data to be stored increases at an explosive rate. Some data becomes obsolete after some time and will need to be discarded. Other data is of absolute importance and needs to be preserved. It is even required by law that some data be removed after some time and that some other data be retained. In addition, new data needs to be incorporated and this may lead to the need of updating old data. The fast explosion of data nowadays makes all the above points about data storage and management very challenging.

The paper of Skyt and Jensen introduces the so-called persistent views or p-views, which are a flexible mechanism that helps in managing increasing amounts of data and in aiding to reduce the volume of data present. P-views are similar to conventional views. However, physical deletions of data from p-views have no effects on these p-views. This allows one to delete or weed-out some old and inaccurate data, while retaining select and aggregate information. This retention is accomplished via the so-called shadow relations.

The paper is an interesting step towards this difficult task of data management that continues to challenge us in most applications in the computation of information.

## **The efficiency of histogram-like techniques for database query optimization.** B. J. OOMMEN AND L. G. RUEDA

The paper of Oommen and Rueda deals with the challenging problem of information retrieval in database systems. It provides a rigorous prototype model that helps analyse different methods of query optimization, which amounts to choosing the most economical query evaluation plan. It is well-known in information retrieval that the same query can be answered by invoking different intermediate operations and this leads one to look for an optimal query evaluation plan QEP (i.e. one in which the number of operations performed in the intermediate joins is minimal). However, the problem of choosing the optimal QEP is combinatorially explosive, and most of the current solutions are based on histogram-based estimation methods. The paper presents a prototype that is used to benchmark these histogram-based estimation methods. The results of the paper justify the fact that current database systems use histogram methods to determine QEPs. Furthermore, the paper demonstrates that more accurate estimation

techniques lead to greater probability of choosing more efficient QEP. The paper is a very useful theoretical and experimental step in the analysis of query optimization techniques.

## **How weak categorizers based upon different principles strengthen performance.** V. S. UREN AND T. R. ADDIS

In modern computational systems, the principle of compositionality or of combining results is vital. Often, larger results are the combination of smaller ones. In some systems (e.g. machine learning), combining results has much promise. However, in other systems (e.g. text categorizers), combining results has led to little progress in general. This paper argues that combining text categorizers can be successful provided essential differences are considered. It explores a theoretical probabilistic framework approach to predicting the performance of combined systems and uses its predictions to benchmark the real results of a number of combined systems. The paper carries out two experiments that illustrate the failure of combining text categorizers. From the first experiment, the paper hypothesises that a workable difference must be created between the various algorithms in question. The second experiment capitalizes on this hypothesis by creating two new categorizers each of which break the rules used to develop typical categorizers in some way.

## **User tracking and mobility management algorithm for wireless networks.** I. KOZATCHOK AND S. PIERRE

The past decade has seen an explosive amount of research on mobile computing. This is hardly surprising because mobile communications have become a basic daily necessity in many parts of the world. However, this explosive use of mobile communications networks brings with it various problems related to network overload and the need for efficient and low-cost management of these mobile networks. The paper of Kozatchok and Pierre proposes an algorithm for minimizing the cost associated with the management of a user's mobility in these networks. It analyses the principles of tracking mobile users, which consists of updating the location area of a mobile unit, and paging to locate the base station serving that mobile unit, all within a reasonable delay time. The result of this analysis leads to the identification of various cost components associated with mobility management and to the design of an algorithm (based on the search for an optimal

grouping of cells) for minimizing these costs. The algorithm proposed in this paper guarantees the global minimum of the cost function and takes into account the average probabilities of received calls, movements, updates and paging costs. The analysis of the simulation results of the proposed management model and algorithm is particularly interesting. This paper is promising for further extensions and a good read for anyone interested in the problem of mobility management.

**POEMS: A parallel object-oriented environment for multi-computer systems.** W. JIE, W. CAI AND S. J. TURNER

This paper is one of a series of papers by the authors in their efforts to design a parallel object-oriented system called POEMS, the major objectives of which are high system performance and flexibility in programming. The starting point of POEMS is that many of the concurrent and parallel systems support either single program multiple data (SPMD) or multiple program multiple data (MPMD) types of programming, but not both. POEMS addresses this shortcoming and supports the parallel object replication model (POR) and the parallel object collection model (POC). Earlier papers by these authors describe POEMS in detail. This current paper, however, concentrates on the programming aspect of POEMS where both the POR and POC models are described and the programming interface that enables their use is given. Simulation results related to performance and examples demonstrating the flexibility of the interface are given. The design and implementation of a prototype runtime system are discussed. Of course the system is still under development and as such there are many questions that need to be answered. However, this paper, as well as the earlier ones, provides interesting and useful design issues and ideas in the field of parallel computation.

The paper concludes with a discussion of the current state of the system and of future work.

**Randomized receiver initiated load-balancing algorithms for tree-shaped computations.** P. SANDERS

Many computational algorithms (such as game tree search and the satisfiability problem) can be seen as tree traversal. Moreover, the paradigms of theorem proving and programming languages benefit greatly from efficient tree traversal algorithms. However, as the applications of tree traversal algorithms are very time consuming, there is a huge interest in parallelizing these algorithms. But, parallelizing involves splitting, and splitting can be arbitrarily uneven and it is not possible to estimate the size of a subproblem. This paper considers generic load-balancing algorithms that efficiently parallelize a large class of applications based on traversing implicitly defined trees with irregular shape. The paper departs from the so-called random polling dynamic load-balancing algorithm which avoids the above-mentioned problem due to splitting. Random polling belongs to a family of receiver initiated load-balancing algorithms which have the advantage of splitting subproblems only on demand by idle processing elements. The paper surveys earlier results on this and other kinds of algorithms and concentrates on analysing random polling where it establishes a bound on the expected parallel execution time of asynchronous random polling dynamic load balancing. Then the paper introduces an algorithm with mixed local and global communication, which is more efficient on interconnection networks because it replaces global communication of random polling with local communication and occasional random permutations. Not only does this paper give a theoretical explanation of the experimental studies demonstrating the efficiency of random polling, but it also provides extra experiments that support further the theoretical results.