doi:10.1093/comjnl/bxh143

Capsule Reviews

FAIROUZ KAMAREDDINE

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 Kamareddine. Professor Kamareddine is an Associate Editor of *The Computer Journal* and is based in the School of Mathematical and Computer Sciences at Heriot-Watt University, Edinburgh, UK.

A Memory System for Education. J. DJORDJEVIC, B. NIKOLIC AND M. MITROVIC

In the Advanced Computer Architecture and Organisation course taught to third year students at Belgrade university, a memory system is core and aims to provide the students with a theoretical treatment of all fundamental memory system concepts, helps them solve practical problems facing the designer of various parts of a memory system and prepares them for a related course in their fourth year. The authors of this paper have developed a memory system for education and its Web-based simulator. Evaluations of the benefits in using the simulator software for teaching various aspects of memory system were conducted and it was shown that such simulator is beneficial for increasing the level of understanding for students and for raising their average exam marks.

PDetect: A Clustering Approach for Detecting Plagiarism in Source Code Datasets. L. MOUSSIADES AND A. VAKALI

Detecting plagiarism in programming assignments is a time-consuming task since for n programs written, one needs to compare $n * (n - 1)/2$ pairs of programs. This complexity increases further since the students usually modify the plagiarized code in such a way as to avoid detection. Different approaches have been developed to automatically detect plagiarism in programming assignments. However, these approaches fail to detect the actual clusters of plagiarisms. This is unfortunate since this type of plagiarism in programming assignmentstends to occur in clusters and clustering assists the reliability of detection. This paper proposes a clustering-oriented approach for detecting plagiarism in programming assignments. The PDetect software it provides accepts as input a set of plagiarism assignments and outputs the clusters of plagiarism. The plagiarism detection measure is based on the so-called precision and recall. This paper combines both precision and recall and provides performance measures for plagiarism detection as well as an analysis of the effect of clustering on plagiarism detection performance. Furthermore, PDetect is evaluated in comparison with another plagiarism detection tool JPlag and it is concluded that combining both tools may be beneficial.

A Technique for Detecting Plagiarism in Computer Code. C. DALY AND J. HORGAN

In computer science, we have seen various developments of electronic systems and web servers that detect plagiarism in coursework. All of these techniques are successful at detecting similarities in submitted work; however, they fail to identify the original author of a piece of work. This paper addresses this problem and presents a technique for detecting plagiarism which identifies the original author. The idea is to insert a watermark in each submitted program which identifies the person who originally submitted the file. This watermark technique is inserted into an existing electronic system called RoboProf. RoboProf is an automated learning environment for Java programming which has been used to monitor a class of 238 first year students in computer science at Dublin City University. RoboProf has a set of 46 programming exercises which can be accessed on the web and can be completed in the student's own time. Completed exercises are marked automatically. The paper concentrates on the study of the effectiveness of the watermark technique. Of the 283 students, 149 were found to be plagiarists (either copied others' work, or allowed their work to be copied). Of these plagiarists, 48 were only suppliers and did not copy any of the exercises. A more in-depth study of the results is given in the paper together with a comparison with the well-known **MOSS** system that is commonly used to detect plagiarism.

Probabilistic Estimation for Routing Space. F. HE, M. GU, X. SONG, Z. TANG, G. YANG AND L. CHENG

Interconnected congestion estimation is the main theme of study of this paper. The authors argue that since modern integrated circuit designs incorporate millions of cells and connections, an accurate estimation of the wiring space can be better achieved using a probabilistic approach. The probabilistic approach proposed in this paper estimates the wiring space in 2D arrays for top–down hierarchical designs and derives the distribution of net density from which one can predict where the net would be congested in a region. The original routing region is divided by quad-partition or bipartition recursively until any terminate condition is satisfied. Both quadpartition and bipartition are considered in detail where, for each, an important lemma on net density is first given and then the probabilistic density estimation is discussed. Then, a hierarchical probabilistic estimator for wiring space estimation is given. This is achieved by first estimating the net densities in each partition node and then by synthesizing the net densities in all nodes to get the net density distribution and hence to estimate the global density and the

wiring space. This proposed algorithm is implemented in C++ and tested on a number of experimental cases which illustrate its promising performance.

In Place Differential File Compression. D. SHAPIRA AND J. A. STORER

This paper presents algorithms for 'in-place' differential file compression where a target file T is compressed with respect to a source file S. Differential file compression which represents a target file T with respect to a source file S, has many useful applications and is often defined to be the representation of a target file by a sequence of copies and inserts. Both the encoder and decoder have identical copies of the source file S which can be used in any way to compress T. If T is of size n and S is of size m, we say that a differential decompression algorithm is 'in-place' if it uses $m + n + O(1)$ space for encoding and $\mathrm{Max}(m, n) + O(1)$ space for decoding, and that a differential encoder is in place if it also uses $m + n$ space to produce the bits for the compressed stream. The paper considers practical in-place differential file compression algorithms based on traditional sliding window data compression (like gzip). In particular, the paper presents a practical algorithm that employs 'off-the-shelf' compression algorithms, presents the in-place sliding window algorithm which is fast and compares in efficiency with algorithms that are not in place and outlines a move preprocessing algorithm with efficient implementations of key steps.

Handling Synchronization Problem in Petri-Net-Based System Design by Property-Preserving Transition-Reduction. L. JIAO, T.-Y. CHEUNG AND W. LU

The synchronization of events is an important yet difficult and complex task in the modelling and design of distributed and concurrent systems. It is important during synchronizations to make sure that the desirable properties of the various subsystems are not destroyed. This paper formulates a property-preserving transition–reduction transformation to handle the synchronization problem in Petri-net system design. This is much needed since there is a lack of general verification techniques for synchronizations in Petri-net system design. The approach of the paper is in three steps: (i) Modelling the system in general Petri-nets; (ii) Formulating the synchronization problem as transition reduction transformations; and (iii) verifying the system to check whether it possesses certain properties such as conservativeness, consistency, repetitiveness, liveness, boundedness, reversibility etc. The paper shows that many properties can be preserved automatically and that some conditions are needed for ensuring preservation of other properties. The main results are illustrated by an example.

Clustered RAID Arrays and Their Access Costs. A. THOMASIAN

RAID stands for Redundant Arrays of Inexpensive/ Independent Disks. RAID was developed to cope with the high cost of mainframe disks of the time by replacing them with inexpensive disks to be included with personal computers. There are different RAID levels varying from 0 to 6. This paper concentrates on the last two levels (RAID5 and RAID6). RAID5 and RAID6 disk arrays utilize the capacity equivalent of one respectively two disks to tolerate single respectively double disk failures. The load increase in RAID5 and RAID6 disk arrays operating in the co-called degraded mode results in elevated disk utilizations and higher disk response time. The clustered RAID paradigm solves this load increase and provides a continuum of redundancy levels to minimize the impact of disk failures on performance in degraded mode. This paper concentrates on the performance analysis of clustered RAID5 and RAID6 disk arrays in both normal and degraded modes.

A Methodology to Define External Schemas in ODMG Databases. M. TORRES AND J. SAMOS

Object oriented databases (OODB) were developed to overcome the limitations of relational databases (RDB). However, this meant that some features that were handled well in RDBs, have become a problem in OODBs. In particular, OODBs lack (i) a standard for object definition and management, and (ii) mechanisms for defining external schemas. The ODMG standard solves the lack of standard for object definition and management. Much work has been carried out to solve the lack of mechanisms for defining external schemas in OODBs. However, these various proposals do not offer a global solution. This paper addresses this problem and proposes a global solution to the definition of external schemas in ODMG databases, providing a language to specify external schemas as well as functionalities for schema closure (guaranteeing that no class of a schema has reference to other classes that are not included in it) and schema generation (which deals with inheritance relationships between the classes of a schema). Adding to the ODMG standard the functionality of defining external schemas allows ODMG databases to fulfill the ANSI/SPARC architecture just like RDBS do.

Static Program Slicing Algorithms are Minimal for Free Liberal Program Schemas. S. DANICIC, C. FOX, M. HARMAN, R. HIERONS, J. HOWROYD AND M. R. LAURENCE

Program slicing is an important area of computing with a number of applications and variations. Roughly, program slicing amounts to deleting various statements from a program, leaving the so-called 'slice' which must preserve the effect of the original program on a certain set of variables of interest. The size of the slice is crucial and the smaller the slice the better. However, statement minimal slices are not computable. This paper reformulates the dataflow minimal slicing question in terms of schematology and proves that slicing algorithms produce minimal slices for free liberal program schemas where all expressions have been replaced by certain symbolic expressions. The paper starts with a description of the syntax and semantics of program schemas and their classes together with some basic results. Then, the paper shows how the syntax and semantics of slicing can be defined using schemas, and develops a schema theory

which is used to show that Weiser's algorithm produces dataflow minimal slices for schemas which are liberal and free.

A Jumping Gene Algorithm for Multiobjective Resource Management in Wideband CDMA Systems. T. M. CHAN, K. F. MAN, K. S. TANG AND S. KWONG

Since second-generation mobile communication systems only support voice and low-speed data, third generation mobile communication systems are being developed to support many types of multimedia services like high-quality voice, video and interaction. The so-called code division multiple access (CDMA) is a promising multiple access technique capable of supporting high capacity. Therefore, the so-called direct sequence wideband CDMA (DS-WCDMA) has been proposed in the development of 3G systems. However, the total capacity that the WCDMA network can support depends on the requested media type of each user and on various other factors especially since users make several choices on a media type. Furthermore, the quality of service (QoS) of a user depends on various factors as well and there is a tradeoff between the capacity and the QoS. Resource management attempts to allocate resources optimally while still satisfying the QoS of each user. This paper studies centralized resource management in DS-WCDMA systems using an approach based on finding the appropriate transmission power and transmission rate for each connecting user while still satisfying their each user's QoS. Since total transmission power and total transmission rate are conflicting, the paper proposes a new MultiObjective Evolutionary Algorithm (MOEA) called the Jumping-Gene Genetic Algorithm (JGGA) which emulates a jumping-gene phenomenon discovered by the Nobel laureate Barbara McClintok from her work on corn plants. The performance of a number of MOEAs are compared and used to assess the proposed JGGA.