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 it to a wider readership. The Capsule Reviews were compiled by Fairouz Kamareddine. Professor Kamareddine is an Associate Editor of *The Computer Journal* and is based in the Department of Mathematical and Computer Sciences at Heriot-Watt University, Edinburgh, UK.

Comparing the Service Component Architecture and Fractal Component Model. PETR HNETYNKA, LIAM MURPHY AND JOHN MURPHY

In component systems, components are assumed to be black boxes with well-defined interfaces, and applications are built by composing these components through the interfaces. Composition can either be flat or hierarchical. Two different systems for development/deployment and efficient composition of components are evaluated and compared. The two systems in question are (i) the service component architecture, which offers a hierarchical model to define service (provided interface), reference (required interface), components and wires, properties, as well as supports any implementation language/platform allowing primitive components in different languages/platforms to be mixed in a single application, and (ii) the fractal hierarchical abstract component model (where concepts are general), which has a number of implementations. To present features of the models, several applications developed in both technologies are presented. Two real-life examples are used to evaluate these systems (the airport internet lounge application and the Common Component Modelling Example research seminar). Comparisons of the features of both systems are given. These features include: specifications, component models, runtime environment and deployment, development of lifecycle of components and extensibility of the models. Evaluation results are reported from different points of view: component model, ease of application development, applicability of the systems, runtime, scalability and performance. A brief comparison with contemporary component-based systems and composition systems is also given.

Closing a System in the Dynamic Input/Output Automata Model. TATJANA KAPUS

The behaviour of a closed system is totally controlled by the system itself, whereas the behaviour of an open system depends on the environment. Input/output (I/O) automata are used for the specification and verification of concurrent and distributed systems and it is said to be closed if it has no input actions in its signature. A closed model is obtained by the composition of the open system model and the model of its environment. This paper is concerned with the dynamic I/O automata (DIOA) where the automata can be dynamically created and destroyed and whose signature is also dynamic. The author shows that it can happen that the composition of an open DIOA and its environment does not necessarily lead to a closed DIOA. Then, the author introduces a restriction operator for DIOA to obtain a closed model. After an introduction to the necessary details concerning the dynamic I/O automata and the travel agent system, the author explains why the DIAO model of the travel agent system is open (although intuitively it is expected to be closed). The paper proposes the solution to closing the DIOA. It does so by changing the definition of the action hiding operator to an action restriction operator—which according to the author— is similar to the restriction operators of CCS or the pi-calculus. The author shows how this change closes the DIOA of the travel agent system.

Recognizing Algorithms Using Language Constructs, Software Metrics and Roles of Variables: An Experiment with Sorting Algorithms. AHMAD TAHERKHANI, ARI KORHONEN AND LAURI MALMI

This paper focuses on algorithm recognition (AR) where the problem is to recognize and classify algorithms to obtain an abstract understanding of the source code. The main goal is to extend the application of program comprehension (PC) in automatic assessment tools, and to develop methods that can automatically recognize algorithms from source code. Since AR is undecidable, the problem and the scope are reduced to a particular group of algorithms and the matching is not required to be exact, but simply to be statistically reasonable. First, the concepts of PC and AR are introduced and compared. Then, the AR method is introduced, and is based on the static analysis of the source code using statistics of language and software complexity metrics. Algorithms are converted into vectors of characteristics, which are identified and extracted from these algorithms. These vectors are then processed by the recognition process. This proposed method has been applied

to five commonly used sorting algorithms and a prototype analyser is developed to automatically compute and convert algorithms into characteristic vectors. Learning data from 70 different versions of the five sorting algorithms were collected and passed to the analyser, and it is shown how the algorithms can be differentiated by their characteristics. An experiment has been carried out on all five sorting algorithms in order to evaluate the performance and accuracy of the method. A total of 217 different algorithms were tested and the results reported and analysed.

On Constructing Minimal Formulae. PAUL DUNNE

The author starts by stating a long-standing problem in computational complexity: 'how difficult is it to determine whether a given formula can be written as an equivalent one whose size is strictly smaller'. Non-trivial lower bounds on the complexity of this problem in its most general form (minimum equivalent expression) have been given before; however there are still no improvements on the obvious upper bound. This paper considers a related problem (where witnessing equivalence is very tightly constrained) and gives its complexity. After a brief overview of computational complexity theory, and of the logical and graph-theoretical notation upon which the techniques of the paper are developed, the author translates his logical formulae into formula graphs and establish necessary properties related to admissible and preferred sets. The author then moves from formula graphs to ideal graphs, which are basic to establishing the upper bound. This leads to a constructive algorithm that is also shown to be correct.

Blind Watermarking of Colour Images Using Karhunen-Loeve Transform Keying. ALESSANDRO BASSO, DAVIDE CAVAGNINO, VICTOR POMPONIU AND ANNAMARIA VERNONE

Digitak watermarking is an embedding of secret information and has three components: an embedding procedure, an extraction algorithm and a watermark. A watermarking system is non-blind, semi-blind or blind depending on the use it makes of the original image and the watermark in the extraction process. Only blind watermarking schemes can be successfully used in security applications whose purpose is to prove ownership of digital media. Furthermore, if the watermarking scheme is robust, this would provide further protection of ownership rights against illegal distribution. However, there is currently no scheme that is highly secure and able to preserve image quality, since there are tradeoffs between robustness, imperceptibility and security. To guarantee the robustness and still preserve the quality of the original image, a new watermarking scheme for colour image is proposed and is based on the Karhunen–Loeve transform (KLT) whose basis images used in the transform are not related to the images being marked but are from an image that is the secret key used in the watermarking scheme. After an introduction to KLT and related work, the proposed watermarking method is introduced

through a number of steps: (i) the computation of a set of KLT eigenvectors of meaningful secret images, (ii) the embedding of a watermark into a colour image, (iii) the detection of a watermark and (iv) the computation of the pseudo-random sequence of the watermark used in the embedding process. For experimental results, the authors first motivate two parameters (the construction of a feature set and the choice of a strength factor) that help achieve an optimal compromise between the quality of the watermarked image and the robustness against attacks. The robustness of the proposed scheme is then tested for performance against non-geometrical attacks and then, three types of attacks are performed to test the efficacy of the proposal against geometrical attacks. Finally, a discussion of security considerations is given.

A Study of a Positive Fragment of Path Queries: Expressiveness, Normal Form and Minimization. YUQING WU, DIRK VAN GUCHT, MARC GYSSENS AND JAN PAREDAENS

This article focuses on positive XQuery queries in which only the self, parent and child axes are used. These queries can be expressed in an algebraic path query language Path+. In addition to the algebraic formalism, the paper considers a declarative formalism of a class of tree queries. Query expressiveness, resolution expressiveness, minimization, optimization and evaluation are studied in Path+. After a brief discussion of related work, the paper introduces Path+ (the positive path algebra of path queries) with its semantics and the query language T of tree queries again with its semantics. A special path+ expression representing the signature of a pair of nodes of a document is introduced and used for two purposes (representing the unique path connecting the two nodes and also an expression that can be applied to the document under consideration). Then, the paper shows that Path+, two of its fragments and T are all equivalent in expressive power. Furthermore, the authors show using normalization that a Path+ query can be decomposed into sub-queries, which can be expressed in a very small fragment of Path+ and for which efficient evaluation strategies are available. The paper then moves from query expressiveness to resolution expressiveness, which is the language's ability to distinguish a pair of nodes/paths in a document. It is shown that two nodes in a document cannot be resolved by a Path+ expression if and only if the paths from the root of the document to these nodes are of equal length and the corresponding nodes on these paths are perfectly bisimilar. It is also shown that certain path queries cannot be expressed in Path+. The authors then move to minimization where they show that each tree query can be translated into an equivalent unique minimal tree query. In addition, it is shown that the translation of a minimal tree query into a Path+ expression is minimal and that minimal path+ expressions are connected to normal forms. Finally, the authors evaluate their work and discuss the complexity of their proposed algorithms.

Signposting in Documents. HAROLD THIMBLEBY

This paper starts by the observation that many seemingly trivial features of documents can raise serious computational and conceptual difficulties when handled by the computer, and gives signposts as one such document feature. Signposts are directions to other parts of the document (e.g. of where the text is to be continued). This paper studies the support of such signpost mechanisms in documents. As an example of a signpost take 'please turn to page 5'; here five is the destination of the signpost. Various forms of signposts are explained in the paper and a number of examples are given to illustrate a number of features that should be guaranteed (e.g. signposting should be reliable, graceful, without costly overheads on the author, etc.). The paper states that signposting is a natural task for computers to support as part of the document preparation process and discusses issues related to programming signposts (especially cross-reference) raising computational problems with signposting. After a brief comparison between signposts (seen and used by readers) and links (see and understood by computers), the paper moves to signposts in the typesetting language Latex and explains how it works and its shortcomings. Then, the author defines the so-called helpful signposts, explains some problems with these signposts in Latex and proposes a solution. An example of a problem in a helpful signpost is 'see section somewhere earlier in this book' when the section actually appears later in the book. The problems discussed in the paper vary from page numbers, to macro expansion, to moving arguments, to local variables. The final solution is then given with simple and necessary extensions and possible simplifications. Finally, discussions of more variations and further uses of signposts are given.

Image Retrieval System Based on Adaptive Colour Histogram and Texture Features. CHUEN-HORNG LIN AND WEI-CHIH LIN

This paper focuses on building an efficient and accurate image retrieval system. One possible technique in image retrieval systems is the use of colour histograms. However, this technique ignores the texture of the image. On the other hand, texture-based image retrieval systems that incorporate feature descriptors and feature selection techniques have been proposed. In this paper, colours and textures are employed as features in similarity retrieval to develop an effective image retrieval system. Three image features are used in this proposal to characterize different properties of the image. One feature is a colour distribution feature (called adaptive colour histogram), the second is also a colour feature (called adaptif motifs co-occurrence matrix) and the third is a texture feature (called the gradient histogram for adaptive motifs). Each of these features is explained in detail. Since these features are complementary, they are integrated in the paper to establish a colour-texture and colour-histogram-based image retrieval system (AMCGH). This combined system is introduced and is validated by four image databases. The performance of AMCGH on each image set is described.

A Self-enriching Methodology for Clustering Narrow Domain Short Texts. DAVID PINTO, PAOLO ROSSO AND HECTOR JIMENEZ-SALAZAR

Document clustering consists of the assignment of documents to unknown categories. The paper states that there is a lack of clustering methods that perform well on different data collections and that clustering is especially difficult for Internet documents where raw data are used. The paper proposes to study clustering narrow domain short texts. It starts by outlining the difficulties of dealing with narrow domain short texts that include the domain broadness degree and the fact that the combination of both a narrow domain and a short text leads to a higher level of complexity when trying to obtain accurate clustering. The related work is discussed and the basis for the analysis and evaluation of textual data (especially narrow domain short texts) is given. The paper concentrates on three corpus features (domain broadness, shortness and stylometry) and presents evaluation measures for each of these features. These assessment measures enable the analysis of a given corpus to determine whether it is composed of narrow domain short texts. If correct corpus features are discovered, ad hoc techniques are used to improve the corpus clustering task. Before the clustering phase, the representation of the short length documents is improved using a term enrichment procedure (this is term-expansion). Then, term-selection techniques (TSTs) are used to improve the terms used in the clustering processing (by filtering discriminate ones and decreasing the number of terms). The proposed methodology (self-term expansion and TSTs for clustering narrow domain short text corpora) is tested.

On Codd Families of Keys over Incomplete Relations. SVEN HARTMANN, UWE LECK AND SEBASTIAN LINK

Most commercial database systems are founded on Codd's relational model of data. Since relations allow the storage of inconsistent data, assertions (data dependencies) are given to restrict relations to meaningful ones. Most research on keys assume that relations are total. However, the paper argues that most relations are partial and moreover that keys behave and interact differently on partial relations than on total ones. For this reason, the authors introduce a class of keys over partial database relations, which is a proper generalization of the class of keys over total relations and is motivated by Codd's principle of entity integrity. The authors show that some well-known axiomatization principles are not sound for partial relations and give a set that forms a finite ground axiomatization for the implication of keys over relations. This axiomatization is used to construct an algorithm that decides the implication problem. Then the authors concentrate on a special class of relations (the Armstrong relations), give a necessary and sufficient condition

for a relation to be Armstrong and show how to compute an Armstrong relation for an arbitrary set of keys. The time complexity of finding Armstrong relations is studied and then the authors move to discuss the question of the maximal number of keys that need to be specified explicitly.

Fast Calculation of Electronic Potential on the GPU or the ASIC MD-GRAPE-3. TETSU NARUMI, KENJI YASUOKA, MAKOTO TAIJI, FRANCESCO ZERBETTO AND SIEGFRIED HOFINGER

This paper argues that it would be an advantage to have molecular surfaces be given in a standard format like the Gaussian cube format but that for this to be the case, the molecular surface would need to be translated into some volumetric data. For this reason, the authors present an algorithm for the translation of molecular surfaces into volumetric data sets, analyse the computation of electrostatic potentials (ESPa) and show that calculations can be efficiently carried out on the graphics processing unit or the specialized computer chip MD-GRAPE-3. First, ESPs and surfaces are formally described and it is shown that volumetric data are represented in Gaussian cube format and how a molecular surface can be mapped onto a cubic grid. Then, it is shown how ESP computation can be hardware-accelerated and an example is given to analyse the results.

An Architectural Framework for the Design, Analysis and Implementation of Interactive Systems. ALEXANDRE R.J. FRANCOIS

Immersipresence combines immersion and interactivity. This paper presents the Software Architecture for immersipresence (SAI) framework for the design, analysis and implementation of interactive systems. The principles and primitives are explained and include (i) a concurrent asynchronous processing model where architectural designs explicitly specify costly synchronization structures only when needed, (ii) a data model that distinguishes between volatile and persistent data allowing for the explicit and consistent modelling of feedback loops and (iii) explicit elements of the model include data flow and synchronization. The SAI architectural style is introduced in terms of components, connectors and constraints. Data are held in pulses (volatile data in active pulses and persistent data in passive pulses). An active pulse triggers a series of operations which lead to the processing of the pulse. The processing model is given and followed by the run-time data binding which searches (filters) a pulse for relevant data. The architectural design specification and the desirable style properties are given and then a Modular Flow Scheduling Middleware (MFSM) is introduced as an open-source architectural middleware for the SAI style. The visual tracking with CAMSHIFT application is used to illustrate the SAI framework. The use of SAI patterns, including a generalization of the tracking pattern in the earlier application, is illustrated in the design case study of Stevi, a computer vision system that performs real-time people detection and tracking from stereo colour video. Other applications are also reported. Finally, related works are discussed.

Some Results on Minimum Discrete Bending Energy Path in Simple Polygon. YULAI XIE

This paper studies computing minimum discrete bending energy (DBE) in simple polygon in two-dimensional plane possibly with holes. The author reviews work on computing bending and discrete bending energy and states that in a polygonal region with holes but that does not edge-cross itself, it is hard to find the shortest path between arbitrary points. After an analysis of the lower bound of the MinBEP problem, the author gives the minimum DBE problem (MDBEP) as the decision problem: given a simple polygon P, is there a polyline S within P such that the bending energy E_b on S satisfies $E_b \leq k$ for a constant k? The author proves that MDBPE in a simple polygon possibly with holes is NP-hard.

A Partitioned Cellular Automaton Approach for Efficient Implementation of Asynchronous Circuits. JIA LEE, SUSUMU ADACHI AND FERDINAND PEPER

Asynchronous cellular automata (ACAs) are cellular automata (CAs) that randomly time their state transitions. Stochastic updating schemes have been used to emulate the randomly timed behaviour of ACAs. In this approach, each cell assumes a certain probability of being updated at a time step. However, asynchronous timing of ACAs cause unexpected behaviour. To solve this problem, computation on ACA is implemented by forcing all cells into approximate synchronization, which has the side-effect of increasing the number of cell states and the number of transition rules. The authors propose a new computationally universal two-dimensional ACA in which the number of transition rules is reduced. This is done by partitioning each cell into four parts. After setting out the preliminaries including the delay-insensitive (DI) circuit and a set of primitives that is universal for DI-circuits, the authors introduce their asynchronous partitioned cellular automaton, which is a two-dimensional grid of cells in which each cell is connected to a neighbour in the von Neumann neighbourhood of the cell, and is partitioned into four sub-cells along its neighbours. The authors explain how to realize DI-circuits in this ACA. Due to the universality of the set of primitives, any DI-circuit can be embedded into the ACA model provided the cellular space is large. Finally, the arbitration function of asynchronous systems is discussed.

Joint Utilization of Appearance, Geometry and Chance for Scene Logo Retrieval. MEDENI SOYSAL AND A. AYDIN ALATAN

Logo detection has mostly been done in constrained environments. Addressing logo detection from the natural scene uses local interest point features where matches are filtered by a

geometric consistency check-up. This approach treats matching appearance and geometry as two different steps and suffers from a number of disadvantages. To overcome these disadvantages, the paper proposes an approach that combines the evaluation of appearance and geometry. This avoids matching each test descriptor to each template descriptor and renders one-to-many matching possible. The advantages include an algorithm that is robust to both significant appearance changes and to random false matches. The authors solve the robust appearance representation problem using clustered local descriptors, and filters the side effect of decreased discriminate power using geometrical constraints. Candidate groups of interest points are identified based on the unlikeliness of being matched by chance. To evaluate the proposed algorithm, the Coca-Cola logo is selected as a typical example and the algorithm is tested in terms of both positive and negative samples. Performance values are measured in terms of true positive and false positive rates. An extension simulation denoted 'artificial template extension' is also given. The performance of the algorithm is compared to another algorithm that is prominent in the literature.