

## Editors' Foreword

The Seventh International Workshop on Database Programming Languages (DBPL99) took place in Kinloch Rannoch, Perthshire, UK from the 1st to the 3rd of September 1999. This series of workshops focuses on the interaction of theory and practice in the design and development of database programming languages. The workshop has occurred biennially since 1987, and was previously held in:

Roscoff, Finistère, France (1987)  
Salishan, Oregon, USA (1989)  
Nafplion, Argolida, Greece (1991)  
Manhattan, New York, USA (1993)  
Gubbio, Umbria, Italy (1995)  
Estes Park, Colorado, USA (1997)

The workshop, as always, was organised as a mixture of invited speakers, informal paper presentations and discussion. Attendance at the workshop was limited to those who submitted papers and members of the Programme Committee, to ensure a sufficiently small forum for useful discussion. Before finding their way into this volume, papers were refereed by at least three members of the Programme Committee. Sixteen of the 31 submitted papers were accepted for presentation at the workshop. In the tradition of the series, authors were encouraged to improve their papers based on both referees' comments and ensuing discussion at the workshop, and resubmit them for publication in this volume, after which a further stage of refereeing took place. The result, we believe, is a volume of high-quality and well-polished papers.

Two invited presentations were given, by Luca Cardelli (Microsoft Research Labs, Cambridge, UK) and Alon Levy (University of Washington). We are particularly grateful to Luca Cardelli for working his presentation into a full paper for inclusion in the volume, a task well beyond the call of duty!

The sessions of the workshop were arranged under the following headings:

Querying and query optimisation  
Languages for document models  
Persistence, components and workflow  
Typing and querying semi-structured data  
Active and spatial databases  
Unifying semi-structured and traditional data models

It is interesting to note that the subject area of the workshop represents a significant departure from previous workshops. All of the papers are concerned with data-intensive computational systems. However, the number of papers roughly arranged by category of interest are as follows:

Active databases	1
Interoperability	1
Persistence	1
Relational models	2
Semi-structured data	8
Spatial databases	2
Workflow models	1

This is a fairly typical spread of interest for a DPBL workshop, except for the sudden emergence of semistructured data as a major theme. Databases, as defined in any text book, deal with significantly large collections of highly structured data. However, it seems that the DBPL community has implicitly decided that semi-structured data, traditionally viewed as unstructured from a database perspective, is now a major theme within the database research domain.

The workshop sessions contained the following papers:

**Invited talk: semi-structured computation**

In this paper Cardelli shows how his work on mobile ambient systems can be transferred to the domain of semi-structured data. The key observation is that both contexts are based upon imperfect knowledge of labeled graphs, and the paper gives an insight into a radically new model for computation over semi-structured data.

**Querying and query optimisation**

Libkin and Wong discuss conditions under which it is possible to evaluate certain database queries in the context of query languages that do not allow their explicit definition. This may be achieved by the incremental maintenance of the query result over changes to the data, rather than by a defined computation over the current given state.

Aggelis and Cosmodakis show an optimisation method for nested SQL query blocks with aggregation operators, derived from the theory of dependency implication. In some cases this allows the merging of MAX, MIN blocks to allow the same optimisation strategy as tableau equivalence to be used.

Grahne and Waller consider string databases, which they define as a collection of tables, the columns of which contain strings. They address the issue of designing a simple query language for string databases, based on a simple first-order logic extended by a concatenation operator.

**Languages for document models**

Maneth and Neven introduce a document transformation language, with similar expressive power to XSL, using regular expressions. A further language is introduced which replaces simple pattern matching by monadic second-order logic formulae. Various properties of this language are investigated.

Neven contrasts document models defined using extended context-free grammars (in which the right-hand side of expansions may contain regular expressions) with standard context-free grammars. An important difference is the ability to order child nodes. The investigation is into extensions of attribute grammars that may be usefully applied within the extended context.

## Persistence, components and workflow

McIver et al. address the inherent problems of the application of the componentware paradigm in the context of databases. They introduce Souk, a language-independent paradigm for performing data integration, designed to allow the rapid construction of integrated solutions from off-the-shelf components.

Printezis, Atkinson and Jordan investigate the pragmatic issue of the misuse of the *transient* keyword within the Java<sup>1</sup> language. Originally intended to allow explicit closure severance within persistent versions of the language, it is now multiply interpreted by different implementations, allowed because of the loose definition of the language. The paper shows why most current interpretations are inappropriate and describes a more useful one for the context of a persistent Java system.

Dong et al. show a method for translating distributed workflow schemata into a family of communicating flowcharts, which are essentially atomic and execute in parallel. Semantics-preserving transformations over these sets of flowcharts can be used to optimise the overall workflow according to the physical infrastructure available for its execution.

## Typing and querying semi-structured data

Bergholz and Freytag discuss the querying of semi-structured data. They propose that queries may be divided into two parts, the first part deriving a match between the data and a partial schema, the second part manipulating that part of the data that matches the schema. The first part of the query can be re-used for a number of different queries requiring the same structure.

Buneman and Pierce investigate a new use of the unlabelled union type for typing semi-structured data. This overcomes the problems of the normal strategy of combining typed data sources in a semi-structured collection, which is to throw away all the existing type information. The union treatment shown allows type information, albeit in a weakened form, to be maintained without losing the inherent flexibility of the semi-structured format.

Buneman, Fan and Weinstein concentrate on a restricted semi-structured data model, where outgoing edges are constrained to have unique labels. In this model, which is representative of a large body of semi-structured collections, many path constraint problems, undecidable in the general model, are decidable. The limits of these results are studied for some different classes of path constraint language.

---

<sup>1</sup> Java is a trademark of Sun Microsystems.

## Active and spatial databases

Geerts and Kuijpers are interested in 2-dimensional spatial databases defined by polynomial inequalities, and in particular in the issue of topological connectivity. This is known not to be first-order expressible in general. They show a spatial Datalog program which tests topological connectivity for arbitrary closed and bounded spatial databases, and is guaranteed to terminate.

Kuper and Su show extensions to linear constraint languages which can express Euclidean distance. The operators under study work directly on the data, unlike previous work which depends upon the data representation.

Bailey and Pouloussis consider the termination of rules, which is a critical requirement for active databases. This paper shows an abstract interpretation framework which allows the modeling of specific approximations for termination analysis methods. The framework allows the comparison and verification of different methods for termination analysis.

## Unifying semi-structured and traditional data models

Granhe and Lakshmanan start from the observation that the state-of-the-art in semi-structured querying is based on navigational techniques, which are inherently detached from standard database theory. First, the semantics of querying is not entirely defined through the normal input/output typing of queries. Second, the notion of genericity is largely unaddressed within the domain, and indeed the emerging trend is for query expressions to be dependent on a particular instance of a database.

Lahiri et al. investigate an integration of structured and semi-structured databases. They describe Ozone, a system within which structured data may contain references to semi-structured, and vice versa. The main contribution is towards the unification of representing and querying such hybrid data collections.

## Acknowledgements

DBPL99 was co-chaired by Richard Connor and Alberto Mendelzon. The Programme Committee members were:

Luca Cardelli	(Microsoft Cambridge)
Richard Connor	(University of Strathclyde)
Alan Dearle	(University of St Andrews)
Stephane Grumbach	(INRIA, Versailles)
Laks Lakshmanan	(Concordia University)
Leonid Libkin	(Bell Labs)
Gianni Mecca	(Università di Basilicata)
Alberto Mendelzon	(University of Toronto)
Fausto Rabitti	(CNUCE-CNR, Pisa)
Peter Schwarz	(IBM Almaden)
Dan Suciu	(AT&T)
David Toman	(University of Waterloo)

We would also like to thank: David Lievens and Steve Neely for practical help during the workshop and with the preparation of the preprint proceedings; Elizabeth MacFarlane for organising the finances, and the staff of the Kinloch Rannoch Hotel for providing a traditional warm Scottish welcome. Advice from Tony Printezis on single malt whisky was well received by the workshop participants, as was the whisky itself.

The next DPBL workshop will be co-chaired by Giorgio Ghelli and Gösta Grahne, and will take place in Italy in September 2001.

August 2000

Richard Connor  
Alberto Mendelzon

Research Issues in Structured and Semistructured  
Database Programming  
7th International Workshop on Database Programming  
Languages, DBPL'99 Kinloch Rannoch, UK, September  
1-3, 1999 Revised Papers  
Connor, R.; Mendelzon, A. (Eds.)  
2000, XII, 328 p., Softcover  
ISBN: 978-3-540-41481-0