Report on the 10th International Symposium on Database Programming Languages (DBPL 2005)

Gavin Bierman
Microsoft Research Cambridge, UK

Christoph Koch
Saarland University, Germany

DBPL 2005 was held on August 28–29, 2005, in the charming surroundings of Trondheim, Norway, and was one of the eleven meetings that were co-located with VLDB. DBPL meets every two years and presents the very best work at the intersection of database and programming language research. DBPL 2005 is the tenth symposium in the series.

The program committee selected 17 papers from a total of 63 submissions—an acceptance rate of 27%—and invited Giuseppe Castagna to open the symposium with a lecture. The 18 talks were given over two days, and informal proceedings were distributed at the meeting. In DBPL tradition, the authors of accepted papers were able to improve their papers following discussions and feedback at the symposium, and these papers have been collected and are published by Springer in volume 3774 of the LNCS series.

Invited lecture Giuseppe Castagna opened DBPL 2005 with an invited lecture entitled *Patterns and Types for Querying XML Documents*. During this lecture he surveyed the various approaches for deconstructing XML data from both the database and programming language communities. He identified two main categories: the vertical approach embodied in database query languages such as XPath; and the horizontal approach as found in programming languages such as ℂDuce (a language which he helped design). As the latter is less well-known to the database community, he gave a concise but detailed tutorial on regular expression patterns; a key feature of ℂDuce. His conclusion was that future research should be directed towards developing languages with a unification, or tight integration, of both approaches.

Research papers

The first paper by Claus Brabrand, Anders Møller and Michael Schwartzbach, *Dual Syntax for XML Languages*, describes XSugar which is a system for managing dual syntax for XML languages. Given a specification, the system can both translate between the syntactic alternatives, and check that the transformations are reversible and valid.

The second paper by J. Nathan Foster, Michael Greenwald, Christian Kirkegaard, Benjamin Pierce and Alan Schmitt, *Exploiting Schemas in Data Synchronization*, concerns the authors' synchronization framework, Harmony, which can generate state-based synchronizers for a variety of tree-structured data formats. The authors formalize the synchronization algorithm and show how the synchronization process is driven by the schema of the structures involved.

The third paper by Benny Kimelfeld and Yehoshua Sagiv, *Efficiently Enumerating Results of Keyword Search*, considers the keyword search problem. The authors identify a common class of algorithms and show that they are provably efficient, i.e. they run with polynomial delay.

The next paper by Dario Colazzo and Carlo Sartiani, *Mapping Maintenance in XML P2P Databases*, considers P2P systems where peers manage their own data and where schema mappings exist between peers. The authors consider the problem of corrupted mappings and

provide a technique for maintaining schema mappings based on a semantic notion of correctness.

The paper by Diego Calvanese, Giuseppe de Giacomo, Domenico Lembo, Maurizion Lenzerini and Riccardo Rosati, *Inconsitency Tolerance in P2P Data Integration: an Epistemic Logic Approach*, also considers unstructured P2P systems. The authors extend their previously introduced multimodal epistemic semantics for such systems and study the problem of dealing with inconsistencies in such P2P data integration scenarios.

XML Data Integration with Identification by Antonella Poggi and Serge Abiteboul considers the problem of data integration of XML data where two major issues arise: first, that the global schema can be expressed as a constraint set, and secondly, the complications of node identity. The authors provide a formal framework and consider various problems including globally identifying nodes and answering queries under different mapping assumptions.

The paper by Floris Geerts and Wenfei Fan, *Satisfiability of XPath Queries with Sibling Axes*, considers the impact of sibling axes to the satisfiability problem for XPath fragments. The authors show that in many cases the presence of sibling axes significantly complicates the satisfiability analyses and give bounds for a number of XPath fragments with and without axes.

XML Subtree Queries: Specification and Composition by Michael Benedikt and Irini Fundulaki considers the problem of queries that filter an input document and return a subdocument. Such queries are useful but cannot be naturally specified in either XPath or XQuery. The authors identify a fragment of XPath with an alternative subtree selection semantics and study the query composition problem.

The paper by Jan Hidders, Stefania Marrara, Jan Paredaens and Roel Vercammen, *On the Expressive Power of XQuery Fragments*, provides a broad analysis of the expressive power of various fragments of XQuery. The authors identify 64 different XQuery fragments and

classify them into 17 equivalence classes such that two fragments can express the same fragments iff they are in the same equivalence class.

Peter Thiemann's paper, *A Type Safe DOM API*, considers the W3C recommended, language neutral, API for XML document manipulation, DOM. The assumed type system for DOM is quite simple and a number of constraints are not made explicit at the type level. The author proposes a refinement of the Java type system that makes these constraints explicit and subject to compile-time checking.

The paper by Michael Levin and Benjamin Pierce, *Type-based Optimization for Regular Patterns*, considers pattern matching mechanisms based on regular expressions (such as those described in the invited lecture). The authors consider the problem of compiling pattern matching and propose a method that utilizes the schema of the input value to generate efficient code. They show that generating optimal code is decidable for finite patterns.

The paper by Giorgio Busatto, Markus Lohrey and Sebastian Maneth, *Efficient Memory Representation of XML Documents*, addresses the problem that many query processors use memory to represent XML data whose size far exceeds the size of the XML document itself. The authors present a way of compressing XML trees such that basic tree operations, such as edge traversal, are preserved in the compressed representation.

The paper by Joachim Niehren, Laurent Planque, Jean-Marc Talbot and Sophie Tison, *N-ary Queries by Tree Automata*, generalizes previous work on node-selecting tree automata, and proposes and studies notions of tree automata that select tuples of nodes from trees and their power to express n-ary queries, for both ranked and unranked trees.

The paper by Wim Martens and Joachim Niehren, *Minimizing Tree Automata for Unranked Trees*, considers the problem of efficiently minimizing automata for unranked trees (such as those that form a foundation of XML schema and various query and pattern languages). They show that, surprisingly, unranked tree automata contribute an additional

form of nondeterminism that renders minimial unranked tree automata in general not unique. The paper studies the complexity of minimization and introduces a new model of tree automata that may be of interest in its own right.

Solmaz Kolahi's paper, *Dependency-Preserving Normalization of Relational and XML Data*, considers the familiar problem of normalization for both relational and XML data. First, techniques from information theory are introduced to characterize the amount of redundancy present in 3NF schemas. Then the problem of preserving dependencies while eliminating redundancies is studied for XML and a new normal form is introduced.

The final two papers, *Complexity and Approximation of Fixing Numerical Attributes in Databases Under Integrity Constraints* by Leopoldo Bertossi, Loreto Bravo, Enrico Franconi and Andrei Lopatenko and *Consistent Query Answers on Numerical Databases under Aggregate Constraints* by Sergio Flesca, Filippo Furfaro and Francesco Parisi, both address the consistent query answering problem, i.e., the problem of extracting all consistent answers (w.r.t. a set of integrity constraints) from a database possibly violating these integrity constraints. The first paper provides several interesting complexity results for the general problem while the second introduces and studies the problem under aggregate constraints which are inequalities of queries involving aggregate-sum constructs.

Panel discussion DBPL 2005 closed with a panel discussion that was jointly organized with the co-chairs of The Third International XML Database Symposium (XSym 2005), Ela Hunt and Zachary Ives; and to which the attendees of both symposia were invited. The invited panel discussed *Whither XML, c. 2005?*, and consisted of experts on various aspects of XML: Gavin Bierman (Microsoft Research), Peter Buneman (University of Edinburgh), Dana Florescu (Oracle), H.V. Jagadish (University of Michigan) and Jayavel Shanmugasundaram (Cornell University).