Lecture Notes in Computer Science 7600

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Abdelkader Hameurlain Josef Küng
Roland Wagner Stephen W. Liddle
Klaus-Dieter Schewe Xiaofang Zhou (Eds.)

Transactions on Large-Scale Data- and Knowledge-Centered Systems VI

Special Issue on Database- and Expert-Systems Applications

Springer

ISSN 0302-9743 (LNCS) e-ISSN 1611-3349 (LNCS)
ISSN 1869-1994 (TLDKS)
ISBN 978-3-642-34178-6 e-ISBN 978-3-642-34179-3
DOI 10.1007/978-3-642-34179-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012948782
CR Subject Classification (1998): H.2.4, H.2.8, I.2.3-4, I.2.6, E.2, I.6.5, I.4.8, H.3.1, D.3.2-3, G.3, E.1

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper
Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Following the 22nd International Conference on Database and Expert Systems Applications (DEXA 2011, proceedings published in LNCS vol. 6860 & 6861) the programme chairs invited the authors of 10 selected papers plus the keynote speakers to submit original, extended and revised papers to a special issue of the Transactions on Large-Scale Data- and Knowledge-Centered Systems (TLDKS). Following these invitations nine papers were submitted, each of which was carefully reviewed by two experts in the field. Finally, eight papers were accepted for this special issue.

In their paper **Combination Skyline Queries** Xi Guo, Chuan Xiao, and Yoshiharu Ishikawa investigate a new variation of the skyline problem called the combination skyline problem. In general, the skyline problem is to select for a given collection of data objects the objects that are not dominated by any others. The proposed variation is to find the fixed-size combinations of objects which are skyline among all possible combinations. The technical challenge is that existing skyline approaches cannot handle the huge number of possible combinations. The proposed solution is based on indexing objects with an R-tree in combination with object-selecting patterns that indicate the number of objects to be selected for each MBR. The authors present two major pruning conditions to avoid unnecessary expansions and enumerations, as well as a technique to reduce space consumption on storing the skyline for each rule in the object-selecting pattern. The efficiency of the proposed algorithm is demonstrated by extensive experiments on both real and synthetic datasets.

In their article **Comparing and Evaluating Approaches to Probabilistic Reasoning: Theory, Implementation, and Applications** Gabriele Kern-Isberner, Christoph Beierle, Marc Finthammer, and Matthias Thimm first give a survey on logic-based approaches to probabilistic reasoning and then provide more details about recent developments for relational, respectively first-order, probabilistic methods like Markov logic networks and Bayesian logic programs. In particular, the authors feature the maximum entropy approach as a powerful and elegant method that combines convenience with respect to knowledge representation with excellent inference properties. While comparing the different approaches is a difficult task due to the variety of the available concepts and to the absence of a common interface, both a conceptual and practical point of view is taken into account. On a conceptual layer the authors propose and discuss several criteria by which first-order probabilistic methods can be distinguished, and apply these criteria to a series of approaches. On the practical layer, some systems for probabilistic reasoning are described with details provided for the *KReator* system as a versatile toolbox for various approaches to first-order probabilistic relational learning, modelling, and reasoning.

Furthermore, applications of probabilistic logics in various scenarios are used to illustrate the theoretical concepts.

In his article **The Science and Art of Conceptual Modelling** Bernhard Thalheim argues that conceptual modelling as one of the central activities in complex system construction requires a general theory of modelling as a culture and an art. Conceptual models are schematic descriptions of a system, a theory, or a phenomenon of an origin thus forming a model. A conceptual model is a model enhanced by concepts, and the process of conceptual modelling is ruled by the purpose of both modelling and the models. It is based on a number of modelling acts, correctness conditions, modelling principles and postulates, and on paradigms of the background or substance theories. Purposes determine the added value of a model. Conceptual modelling is performed by modellers who direct the process based on their experience, education, understanding, intention and attitude. The paper discusses aspects of a general theory of modelling considering modelling as apprenticeship and technology, and concluding that it is indeed an art.

In the paper **Predictive Line Queries for Traffic Prediction** Lasanthi Heendaliya, Dan Lin, and Ali Hurson address the problem of how to enhance real-time tracking systems for moving objects by traffic-jam prediction. In their work traffic prediction is realized by a new type of query, termed as the predictive line query, which estimates the amount of vehicles entering a querying road segment at a specified future timestamp and helps query issuers to adjust their travel plans in a timely manner. The added value is that current methods are targeted to objects moving freely in the Euclidean space without consideration of road-network constraints. Taking the road network topology and object moving patterns into account, the solution proposes a hybrid index structure, the RD-tree, which employs an R*-tree for network indexing and direction-based hash tables for managing vehicles. This is further coupled with a ring-query-based algorithm to answer the predictive line query. Extensive experimental studies demonstrate that the approach significantly outperforms existing work in terms of both accuracy and time efficiency.

The paper **A DHT-Based System for the Management of Loosely Structured, Multidimensional Data** by Athanasia Asiki, Dimitrios Tsoumakos, and Nectarios Koziris presents a DHT-based system called *LinkedPeers* designed for efficient distribution and processing of multidimensional, loosely structured data over a peer-to-peer overlay. Each dimension is annotated with the use of concept hierarchies. The system incorporates large-scale support for partially-structured data and high-performance, distributed query processing including multiple aggregates. To enable the efficient resolution of such queries, *LinkedPeers* utilizes a conceptual chain of DHT rings that stores data in a hierarchy-preserving manner. Moreover, adaptive mechanisms detect dynamic changes in the query workloads and adjust the granularity of the indexing on a per node basis. The pre-computation of possible future queries is also performed during the resolution of an incoming query. Extensive experiments provide evidence that the system achieves high precision in answering queries

while minimizing communication cost and adapting its indexing to the incoming queries.

The paper **A Grammarware for the Incremental Validation of Integrity Constraints on XML Documents under Multiple Updates** by Béatrice Bouchou, Mirian Halfeld Ferrari, and Maria Adriana Vidigal Lima contributes a generic grammarware for validating XML integrity constraints, both from scratch and incrementally during document updates. The authors use an attribute grammar to describe XML documents and constraints, and describe in detail the main parts of novel algorithms dealing with the validation from scratch and the incremental validation under multiple updates, focusing on the case of XML Functional Dependencies (XFDs). The method promises an effective and efficient validation of integrity constraints.

In the paper **Database Support for Enabling Data-Discovery Queries over Semantically-Annotated Observational Data** Huiping Cao, Shawn Bowers, and Mark P. Schildhauer present a formal framework to address the challenges arising from the use of very different structures for observed scientific data with little semantic information about the data itself. The problem is to discover existing data sets based on data semantics (observation and measurement types) and data content (the values of measurements within a data set). The proposed framework consists of a semantic observational model to uniformly represent observation and measurement types, a high-level semantic annotation language to map tabular resources into the model, and a declarative query language that allows researchers to express data-discovery queries over heterogeneous, annotated data sets. In addition, two storage schemes (in-place databases RDB and materialized databases MDB) are proposed to store the source data sets and their annotations, which are complemented by two query schemes (ExeD and ExeH) to evaluate discovery queries and the results of extensive experiments comparing their effectiveness.

The paper **Probabilistically Ranking Web Article Quality Based on Evolution Patterns** by Jingyu Han, Kejia Chen, and Dawei Jiang addresses the problem that user-generated content (UGC), e.g. in Wikipedia, is created, updated, and maintained by various web users, and its data quality is a major concern to all users. Wikipedia pages usually go through a series of revision stages, gradually approaching a relatively steady quality state, and articles of different quality classes exhibit specific evolution patterns. The authors propose assessing the quality of a number of web articles using Learning Evolution Patterns (LEP). First, each article's revision history is mapped into a state sequence using a Hidden Markov Model (HMM). Second, evolution patterns are mined for each quality class, and each quality class is characterized by a set of quality corpora. Finally, an article's quality is determined probabilistically by comparing the article with the quality corpora. Experimental results demonstrate that the LEP approach can capture a web article's quality precisely.

June 2012

Stephen W. Liddle
Klaus-Dieter Schewe
Xiaofang Zhou

Editorial Board

Table of Contents