Journal of Intelligent Information Systems, 15, 5–6, 2000

Guest Editor Introduction: Data Warehousing and Knowledge Discovery

MUKESH MOHANIA mohania@cs.wmich.edu
Western Michigan University

A. MIN TJOA tjoa@ifs.tuwien.ac.at
Vienna University of Technology

YAHIKO KAMBAYASHI yahiko@kuis.kyoto-u.ac.jp
Kyoto University

Keywords: data warehouses, knowledge discovery, data mining, multidimensional indexing, view materialization

Introduction

Data Warehousing and Knowledge Discovery is emerging as a key technology for enterprises that wish to improve their data analysis, decision support, and the automatic extraction of knowledge from data. On-Line Analytical Processing (OLAP) tools are well-suited for complex data analysis, such as multidimensional data analysis, and to assist in decision support activities. They access data from a separate repository, called a *data warehouse*, that houses data from many operational, legacy, and possibly heterogeneous data sources. Data mining tools take the process one step further by actively searching the data for patterns and hidden knowledge suitable for decision support.

This special issue contains four papers which were presented in the First International Conference on Data Warehousing and Knowledge Discovery (DaWaK'99), held in Florence, Italy from August 30 to September 1, 1999. These papers have been extended and revised by the authors into full papers.

The first paper titled 'Incremental Design of a Data Warehouse' discusses the problem of designing a data warehouse incrementally, when the users' requirements are either changed or they have new requirements, that is, a set of new queries are posed by the users. For these new requirements (queries), a set of new views needs to be materialized. The problem is to minimize the combined evaluation cost of the new queries, and the maintenance of the new views in a given extra space for materialization. Theodotatos and Sellis have modeled this problem as a state space search problem using an AND/OR directed acyclic graph representation of multiple queries. They present an incremental design approach that selects views to be materialized in the extra space and minimizes the query evaluation and view maintenance costs.

The second paper 'Temporal Coupling Verification in Time Series Databases' proposes a language, called SQL/LPP+, that enables users to cascade multiple patterns using temporal relationships and to obtain the aggregate and meta-aggregate values of the time series data.

Perng and Parker introduce a concept of ‘temporal coupling’ that defines a set of events occurring in a particular time period.

A new approach to implement a multidimensional index structure for a commercial relational database system is discussed in paper entitled ‘Multidimensional Index Structures in Relational Databases’. Such an index structure is necessary for efficient knowledge discovery. Böhm, Berchtold, Kriegel and Michel have mapped the index structure to a relational database design and have simulated the behavior of the index structure using triggers and stored procedures. They have shown the feasibility and efficiency of index structures by implementing an X-tree within Oracle 8.

The last paper ‘Knowledge Discovery from Series of Interval Events’ investigates the problem of mining series of interval events where events are considered to be active for a period of time. Villafane, Hua, Tran and Maulik propose mining techniques to discover temporal containment relationships in such series. Their implementation results show that some interesting relationships can be discovered from events.

We would like to thank Professor Roland Wagner and Ms. Gabriela Wagner for their help and support in organizing the conference. We are very indebted to all the program committee members who reviewed these papers. Finally, we thank Professor Larry Kerschberg and the management team of JIIS (Kluwer Academic Publishers) for giving us this opportunity to complete this special issue.