Available online at www.sciencedirect.com
ScienceDirect
Data & Knowledge Engineering 63 (2007) 1–3
DATA & KNOWLEDGE ENGINEERING
www.elsevier.com/locate/datak

Editorial

Data warehouse and knowledge discovery (DAWAK'05)

Since 1999, the aim of the International Conference on Data Warehousing and Knowledge Discovery (DaWaK) is to bring together researchers, developers and practitioners to discuss research issues and experience in developing and deploying data warehousing and knowledge discovery systems, applications, and solutions. The 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK'05), held in Copenhagen, Denmark, on August, 22nd–26th, continued these series of successful conferences dedicated to these topics. Compared to the past conferences, where more data mining papers were presented in DAWAK, in this edition, the conference tried to provide the right, logical balance between data warehousing and knowledge discovery. Moreover, this can be regarded as a natural evolution of these two research topics as more and more works covers both areas.

This year, DAWAK'05 received 196 abstracts, and finally received 162 full papers from 38 countries, and the Program Committee selected 51 papers, making an acceptance rate of 31.4% of submitted full papers. The authors of the best papers were invited to extend their papers and re-submit them for this special issue. These extended papers had two more rounds of reviews where reviewers made strong revisions paying special attention on the new material. In summary, in this special issue, the first two papers are focused on aspects directly related to data warehouses and OLAP technologies, then, two papers cover data mining techniques with other technologies such as data warehouses and ontologies, and finally, the last paper presents a data mining algorithm. In the following, we summarize these selected papers:

The first paper, "*Progressive Ranking of Range Aggregates*", by Hua-Gang Li, Hailing Yu, Divyakant Agrawal and Amr El Abbadi, argues that although ranking-aware queries have recently been gaining much attention in many applications such as multimedia databases, search engines or data streams; they have received less attention on the field of On-Line Analytical Processing (OLAP) applications. For this reason, in this paper the authors introduce aggregation ranking queries for OLAP data cubes. The importance of these queries is that they explicitly support the ranking of aggregate information (highly common in OLAP queries) over user-specified ranges. Thus, the authors propose a progression of three different algorithms (*Complete scan, Bi-directional transversal* and *Dominant-Set Oriented*) to handle the aggregation of ranking queries. Their *Dominant-Set Oriented algorithm* is efficient and realistic, since it exploits pre-computed cumulative information. Finally, the authors empirically evaluate their algorithm on an on-line advertising tracking data warehouse application where their experimental results show a significantly improved query cost.

The second paper, "*An Approach towards an Event-fed Solution for Slowly Changing Dimensions in Data Warehouses with a Detailed Case Study*", authored by Tho Manh Nguyen, A Min Tjoa, Jaromir Nemec and Martin Windisch, relies on the fact that the incoming information in data warehouses can be generally classified into (i) state-oriented data and (ii) event-oriented data or transactional data, which contains information about the change performed by processes on the instances of information objects. The authors argue that on the way towards achieving the goal of a full fledged active data warehouse it becomes increasingly important to provide data with minimal latency to solve the hard and well-known problem of the Slowly Changing Dimensions. In this paper, the authors focus on dimensional data which is provided by general data warehouse applications and propose an Event-fed comprehensive Slowly Changing Dimension (SCD)

0169-023X/$ - see front matter
doi:10.1016/j.datak.2006.10.003

approach to overcome the limitation of existing SCD approaches and of snapshot based solutions. In their approach, the information transfer is performed via messages containing the change of information on the dimension instances. The proposed approach is able to validate the event-messages, to reconstruct the complete history of the dimension, and to provide a well applicable "comprehensive slowly changing dimension" (cSCD) interface for queries on the historical and current state of the dimensions. The paper ends with a description of a prototype implementation for this kind of an "active integration" in a data warehouse and a case study at the T-Mobile company.

The third paper, entitled "*A UML 2.0 Profile for Designing Association Rule Mining Models for Data Warehouses*", by José Jacobo Zubcoff and Juan Trujillo evidences that with the use of data mining techniques, the data stored in data warehouses (DW) can be analyzed for the purpose of uncovering and predicting hidden patterns within the data. In this paper, the authors present a novel approach to integrate data mining models into multidimensional models in order to accomplish the conceptual design of data warehouses with association rules (AR). To achieve this aim, the authors propose a UML profile that allows designers to specify the association rule mining models on the multidimensional modelling of DWs at the conceptual level in a clear and expressive way. The main advantages of their novel proposal is that the association rules are specified on the main multidimensional terms (i.e. facts, measures, dimensions, classification hierarchies, etc.) used by users to analyze data warehouses. Therefore, the approach of defined ARs are closer to the data warehouse goals and user requirements than the traditional method of specifying the ARs on the final database implementation structures such as tables, rows or columns. In this way, the authors claim that considering the ARs definition in the early stages of a data warehouse project reduces the total developing time and cost. Finally, in order to show the benefits of their approach, the authors apply their approach to a case study and implement the specified ARs on a commercial database management server.

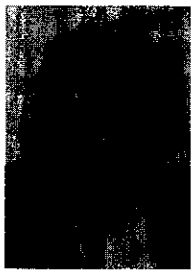
The fourth paper is "*Integration of Association Rules and Ontologies for Semantic Query Expansion*", by Min Song, Il-Yeol Song, Xiaohua Hu and Robert B. Allen. In this paper, the authors propose a novel semantic Query Expansion (QE) technique, called SemanQE, that combines association rules with ontologies and Natural Language Processing techniques. SemanQE is a hybrid QE technique that applies semantic association rules to the information retrieval problem. The authors argue that their approach automatically discovers the characteristics of documents that are useful for extraction of a target entity. Then, by using these seed instances, their system retrieves a sample of documents from the database. Finally, they apply machine learning and information retrieval techniques to queries that will tend to match additional useful documents. This technique is different from others in that (i) it utilizes the explicit semantics as well as other linguistic properties of unstructured text corpus, (ii) it makes use of contextual properties of important terms discovered by association rules, and (iii) ontology entries are added to the query by disambiguating word senses. Finally, the authors present a series of experiments accomplished with TREC ad hoc queries and they achieved an improvement from 13.41% to 32.39% for P@20 and from 8.39% to 14.22% for the F-measure, by comparing their results with other experiments conducted by other similar techniques.

The fifth paper entitled "*ARMADA – An Algorithm for Discovering Richer Relative Temporal Association Rules from Interval-based Data*" is authored by Edi Winarko and John F. Roddick. This paper relies on the fact that temporal association rule mining promises the ability to discover time-dependent correlations or patterns between events in large volumes of data. The authors argue that so far most temporal data mining research has focused on events existing at a point in time rather than over a temporal interval. The authors claim that in comparison to static rules, mining by accommodating temporal intervals rules with respect to time points provides semantically richer rules. In this paper, the authors present a new algorithm, ARMADA, to discover frequent temporal patterns and to generate richer interval-based temporal association rules. They illustrate the proposed ARMADA algorithm by using an example and the method to generate richer temporal association rules from the frequent temporal patterns. Furthermore, they also introduce a maximum gap time constraint that can be used to get rid of insignificant patterns and rules, so that the number of generated patterns and rules can be reduced. Finally, the authors utilize synthetic datasets to assess the performance of the algorithm.

Finally, we would like to thank all the authors who revised and extended their papers for this special issue and the reviewers for their hard work in the two phase revising process of the extended conference papers and for providing their critical and useful comments which helped the authors in improving their papers.

Absolutely, all of them have contributed striving towards a special issue of a high quality. We hope the readers will enjoy reading this issue and find the content beneficial for their research work.

Juan Trujillo is an associated professor at the Computer Science School at the University of Alicante, Spain. He received a Ph.D. in Computer Science from the University of Alicante (Spain) in 2001. His research interests include database modeling, data warehouses, conceptual design of data warehouses, multidimensional databases, data warehouse security and quality, mining data warehouses, OLAP, as well as object-oriented analysis and design with UML. He has published many papers in high quality international conferences such as ER, UML, ADBIS, CAiSE, WAIM or DAWAK. He has also published papers in highly cited international journals such as IEEE Computer, Decision Support Systems (DSS), Data and Knowledge Engineering (DKE) or Information Systems (IS). He has served as a Program Committee member of several workshops and conferences such as ER, DOLAP, DAWAK, DSS, JISBD and SCI and has also spent some time as a reviewer of several journals such as JDM, KAIS, ISOFT and JODS. He has been Program Chair of DOLAP'05 and BP-UML'05, and Program Co-chair of DAWAK'05, DAWAK'06 and BP-UML'06. His e-mail is http://jtrujillo@dlsi.ua.es.



A. Min Tjoa is since 1994 director of the Institute of Software Technology and Interactive Systems at the Vienna University of Technology. He is currently also the head of the Austrian Competence Center for Security Research. He received his Ph.D. in Engineering from the University of Linz in 1979. He was a Visiting Professor at the Universities of Zurich, Kyushu and Wroclaw (Poland) and at the Technical Universities of Prague and Lausanne (Switzerland). He was the president of the Austrian Computer Society from 1999 to 2003. He is a member of the IFIP Technical Committee for Information Systems. His current research focus areas are Data Warehousing, Grid Computing, Semantic Web, Security, and Personal Information Management Systems. He has published more than 150 peer reviewed articles in journals and conferences. He is the author and editor of 15 books.

Juan Trujillo
Department of Language and Information Systems
University of Alicante
Apto. correos 99. E-03080
03690 Alicante, Spain
E-mail address: jtrujillo@dlsi.ua.es

A. Min Tjoa
Vienna University of Technology
Institute of Software Technology
Favoritenstr. 9 - 11 /188
A-1040 Wien, Austria
E-mail address: amin@ifs.tuwien.ac.at

Available online 9 November 2006