Lecture Notes in Computer Science 595

Edited by G. Goos and J. Hartmanis

M. Levene

The Nested Universal Relation Database Model

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Series Editors

Gerhard Goos
Universität Karlsruhe
Postfach 69 80
Vincenz-Priessnitz-Straße 1
W-7500 Karlsruhe, FRG

Juris Hartmanis
Department of Computer Science
Cornell University
5149 Upson Hall
Ithaca, NY 14853, USA

Author

Mark Levene
Department of Computer Science, University College London
Gower Street, London WC1E 6BT, UK

CR Subject Classification (1991): H.2.1, H.2.3

ISBN 3-540-55493-9 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-55493-9 Springer-Verlag New York Berlin Heidelberg

Printed in Germany

Typesetting: Camera ready by author/editor
Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
45/3140-543210 - Printed on acid-free paper

Preface

During the 1980's the *flat relational model* (relational model), which was initiated by Codd in 1970, gained immense popularity and acceptance in the market place. One of the main reasons for this success is that the relational model provides *physical data independence*, i.e. changing the physical organization of the database does not require alteration of the database at the conceptual level. However, the relational model does not provide *logical data independence*, since users must navigate amongst the flat relations in the database when posing queries to the database. Logical data independence would imply that changing the database at the conceptual level does not have an effect on the user's view of the database.

The *universal relation model* (UR model) endeavours to achieve logical data independence in the relational model by allowing the user to view the flat database as if it were composed of a single flat relation. To this end, the user is provided with a UR interface - with all the semantics embedded into the attributes - encapsulating the user's view of the flat database at the external level, on top of the conceptual level. The UR model was firmly established in mainstream relational database theory during the mid 1980's with the introduction of the *weak instance* approach. In the weak instance approach to the UR model, the *representative instance* becomes the underlying data structure of the UR model, which is suitable for storing all the data in the flat database in a single flat relation. Although the application areas of the UR model are slightly restricted by several underlying assumptions it could serve well as the foundation of a natural language interface to a database.

In recent years there has been a growing demand to use databases in non-business applications, such as: office automation, computer aided design (CAD), computer aided software engineering (CASE), image processing, text retrieval, expert systems and geographical and statistical analyses. The flat structure of relations imposed by the first normal form assumption on relational databases poses a severe restriction on the modelling capabilities of the relational model for such non-business applications.

In order to facilitate the modelling of the above non-business applications the *nested relational model* was developed during the 1980's as an extension of the relational model. The nested relational model achieves this wider applicability by allowing hierarchically structured objects, also referred to as *complex objects*, to be modelled directly, whilst maintaining the sound theoretical basis of the relational model.

One of the problems with the nested relational model is that it may prove to be too complex for non-technical users to interact with. This *usability problem* arises because of the fact that queries posed to a nested database involve navigation both amongst and within the structure of nested relations in the nested database. Thus, as in the flat relational model, the nested relational model does not provide logical data independence. Moreover, posing queries to the nested database is much more difficult in the nested relational model than in the flat relational model due to the hierarchical structure of nested relations.

In this monograph we propose to alleviate the usability problem by providing logical data independence to the nested relational model. To this end we extend the UR model to nested relations by defining a new database model, called the *nested Universal Relation model* (nested UR model). Logical data independence is achieved by allowing users to view the nested database as if it were composed of a single nested relation. Moreover, the nested UR model allows users to interact with the nested database without having to know its structure, which may be complex.

In order to formalize the nested UR model we present a comprehensive formalization of the nested relational model, which incorporates null values into the model. We provide semantics to the nested relational model in terms of *null extended data dependencies*. These dependencies are obtained by extending from flat relations to nested relations both functional data dependencies and the classical notion of lossless decomposition. Furthermore, we define the *extended chase* procedure in order to test the satisfaction of the said null extended data dependencies and to infer more information from a given nested relation. The theory of the nested UR model is established by extending the weak instance approach to the classical UR model to the *nested weak instance* approach to the nested UR model. The nested weak instance approach leads naturally to the definition of the underlying data structure for the nested UR model, namely, the *nested representative instance* (NRI) over the *nested universal relation scheme* (NURS).

A major result of the monograph is that the NRI over the NURS is a suitable model for storing the data in a nested database in a single nested relation. Thus, the classical UR model becomes a special case of the nested UR model. An important implication of this result is that a UR interface can be implemented by using the nested UR model, thus gaining the full advantages of nested relations over flat relations. In particular, redundancy is minimized, query processing becomes more efficient, semantics are often explicitly represented within the nested relations, and we gain more expressive power as both flat and hierarchical data may be presented to the user.

We believe that usability of complex object databases is one of the challenges of the 1990's in the area of database management, as the research into database models for complex objects is gaining maturity. Therefore, both database researchers and practitioners can benefit from the approach of the nested UR model, which is to formally show how one can reduce the complexity of the user interface at the external level of a database in order to gain usability.

This monograph is a slightly revised version of my thesis, which was submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the University of London in November 1989 and was obtained in June 1990. The thesis was written at the Computer Science Department of Birkbeck College, which is part of the University of London.

I would like to thank my supervisor Professor George Loizou for the many hours he devoted to discussing and carefully reading the thesis. It has been a pleasure for me to work with George who has always had the patience to find and painstakingly correct my mistakes and give me the necessary guidance. I am also grateful to my parents and the rest of the family for giving me the moral and financial support I needed during my studies. In particular, I would like to thank my brother Dan for designing the illustration on the cover of the monograph. I also wish to thank the Wingate Foundation for the financial support that was provided during my last year of study. Finally, I would like to dedicate this monograph to Sara.

London
February 1992

Mark Levene

Table of Contents