

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

THE MODEL THEORETIC NOTION OF TYPE IN RELATIONAL DATABASES

A thesis presented in partial fulfillment of the
requirements for the degree of

Master
of
Information Sciences

at Massey University,
Wellington,
New Zealand.

Alejandra Lorena Paoletti

2005

Dr. José María Turull-Torres
Thesis' Supervisor

Abstract. It is well known that finite model theory and database theory are two disciplines intimately connected. A topic which has been deeply studied in the context of finite model theory as well as in classic model theory, but which has not received the same attention in the context of databases, is the concept of *type* of a tuple. We believe that it is very important to have a clear understanding of the implications of the concept of type in the field of databases. Therefore, we study here the notion of type in FO (first-order logic) and in FO^k (the restriction of FO to formulas with up to k variables), as well as in their corresponding infinitary extensions $\mathcal{L}_{\infty\omega}$ and $\mathcal{L}_{\infty\omega}^k$, respectively. The FO^k types ($\mathcal{L}_{\infty\omega}^k$ types) are quite relevant in database theory since they characterize the discerning power of the class of reflective relational machines of S. Abiteboul, C. Papadimitriou and V. Vianu with variable complexity k . We examine the three main characterizations of FO equivalence. These characterizations are based in Ehrenfeucht-Fraïssé games, back and forth systems of partial isomorphisms and isolating formulas for types. We study in detail all of them because they give a clear idea of the relevance of the notion of type in the context of databases, and make the fact that the types of the tuples in a database reflect all the information contained in it very clear. We found that the concept of type turned out to be useful to examine from a new perspective a well known problem of the field of databases, consisting on the redundant storage of information. Specifically, we initiate a study towards a method of *normalization* of relational databases based on the detection of *redundant relations*, i.e., relations which can be eliminated from the database without losing information, by using *isolating formulas* for the types of the tuples in the database.

Acknowledgements

This project would have been impossible to finish without the knowledge and encouragement of many people

First, I would like to express a special acknowledgement to my supervisor, Associate Professor, Dr. José María Turull-Torres (Massey University) for sharing his expertise and knowledge and for his valuable guidance.

I would like to thank Massey University, College of Business, for offering me a Graduate Assistant position. Particularly, I would like to thank the Head of Department of Information Systems, Associate Professor, Chris Freyberg, for his support.

I also wish to thank to the reviewers of this thesis, Dr. Alejandro Petrovich (University of Buenos Aires) and Professor, Dr. Klaus-Dieter Schewe (Massey University).

From the beginning of this project I have been in the most encouraged working environment provided by my partner, Flavio, to whom I am deeply grateful for his constant support.

Finally, I would like to thank my family, specially to my parents, Lilia and Lucio, and my sisters, Andrea and Vanesa, for their support from the other side of the world.

Table of Contents

1	Introduction	9
2	Preliminaries	12
2.1	Finite Model Theory and Databases	12
2.2	Computable Queries	17
3	Type of a Tuple	19
3.1	The Notion of <i>FO</i> Type in Databases	21
4	Partial Isomorphisms	23
5	Ehrenfeucht-Fraïssé Games	25
5.1	Inductive construction of the m-isomorphic type	27
6	Pebble Games	34
6.1	Inductive construction of the k-m-isomorphic type	37
6.2	Isolating Formula	39
7	Databases with Redundant Relations	42
7.1	Redundant Relations in Databases	43
7.2	Redundant Relations in Classes of Databases	51
8	Conclusion and Future Work	58