

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.

Functional Dependencies for XML

Axiomatisation and Normal Form
in the presence of
Frequencies and Identifiers

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

MASTER OF SCIENCES
IN
INFORMATION SYSTEMS

at Massey University, Palmerston North, New Zealand

Diem-Thu Trinh
Revised March 2005

Work supervised by
Dr Sven Hartmann

Acknowledgement

I would like to thank all the people who have supported me throughout this year. Although I do not mention anyone specifically, I hope you all know who you are. For some of you, I am thankful for the guidance and help that you have provided me. For others of you, I am also grateful for all the laughter and friendship that we have shared throughout this time - this year would not have been half as enjoyable without this.

I would also like to extend my gratitude to Mrs Clark for the generous Lovell & Berys Clark Scholarship and to the NZVCC for awarding me the William Georgetti Scholarship. The scholarships have provided me with a huge amount of financial support, allowing me to better focus on my study. The scholarships have also been a great source of motivation and encouragement for me this year.

Thu Trinh

December 17, 2004.

Contents

1	Introduction	3
1.1	Outline of the Thesis	4
2	Preliminary Notations	5
2.1	XML Graph Model	5
2.1.1	Rooted Graphs and Rooted Trees	5
2.1.2	XML Graphs	5
2.1.3	Mappings between XML Graphs	6
2.1.4	XML Schema Graphs and XML Data Trees	8
2.1.5	Operators on Subgraphs	11
2.2	Running Example	12
2.3	Functional Dependencies in XML	14
2.3.1	Defining XFDs	14
2.3.2	Implication and Derivation of XFDs	16
3	XFDs in the Presence of Frequencies	18
3.1	Sound Inference Rules	18
3.2	A Sound & Complete Rule System	24
3.3	Additional Inference Rules for XFDs	31
3.4	Armstrong XML Data Trees	33
4	XFDs in the Presence of Frequencies and Identifiers	34
4.1	Revised XML Graph Model	34
4.2	Sound Inference Rules	36

4.3	A Sound & Complete Rule System	41
5	XML with Identifiers Normal Form	44
5.1	Trivial XFDs	44
5.2	Redundancy with respect to XFDs	45
5.3	X^i NF: An XML Normal Form Utilising Identifiers	49
5.4	Elegantly Checking X^i NF	52
5.5	“Redundancy” as a Design Quality	55
6	Related Work	58
6.1	Relational Databases with Null Values	58
6.2	XML and Semistructured Databases	59
7	Conclusion	63
7.1	Future Work	64