Contents

1 Introduction **1**
1.1 Preliminary notions 2
1.2 Problems, hypothesis and goals 5
1.3 Summary of contributions of this thesis 5
1.4 Structure of this thesis. 8

I Preliminaries 9

2 RDF and SPARQL **11**
2.1 The RDF data model 11
2.2 The SPARQL query language 14

II Incomplete data in SPARQL 19

3 Blank nodes as unknown values **21**
3.1 Naive semantics of RDF 24
3.2 RDF graphs as V-tables 25
3.2.1 The V-tables relational algebra 29
3.2.2 The relational algebra of SPARQL 30
3.3 Approximating certain answers in SPARQL 34
3.3.1 Certain answers with null values 34
3.3.2 Under- and over-approximations 35
3.3.3 Approximating relational algebra queries 41
3.3.4 SPARQL rewriting strategies 48
3.4 Evaluation 51
3.4.1 Evaluation Setting 52
3.4.2 TPC–H experiments 52
3.4.3 Wikidata survey 56
3.5 Conclusions 57

4 Unbound values as incomplete data **59**
4.1 A brief review of the landscape of null values 60
4.2 Where do unbound values came from? 61
4.3 Null values in SQL and SPARQL 64
4.3.1 The generalized algebra 64
4.3.2 Mapping SQL and SPARQL operators to the generalized algebra . . 67

4.4 Conclusions . . . 72

III On the semantics of EXISTS 73

5 The notion of substitution under incomplete data 75
5.1 The problem of substitution in SPARQL . . . 76
5.1.1 An overview of the problems of substitution in SPARQL . . . 77
5.1.2 Existing proposals for the semantics of EXISTS . . . 83
5.1.3 How the existing proposals solve the substituion issues . . . 85
5.2 Environment-binding as a substitution proposal . . . 88
5.3 Free, bound, and range restricted variables . . . 91
5.3.1 Substitution in relational calculus . . . 91
5.3.2 Substitution in SQL . . . 91
5.3.3 Substitution in SPARQL . . . 94
5.4 Conclusions . . . 96

6 Expressing SPARQL in Datalog 97
6.1 An overview of nr-Datalog$^{\neg}$ . . . 98
6.2 Translation of the safe SPARQL-0 fragment . . . 100
6.3 Fixing the translation of equality atoms in filter-conditions . . . 104
6.4 Fixing the translation of negation in filter-conditions . . . 106
6.5 Translation of the SPARQL-1 fragment . . . 111
6.5.1 SELECTION queries in nr-Datalog$^{\neg}$ . . . 111
6.5.2 MINUS queries in nr-Datalog$^{\neg}$ . . . 111
6.5.3 VALUES queries to nr-Datalog$^{\neg}$ . . . 112
6.5.4 BIND queries in nr-Datalog$^{\neg}$ . . . 113
6.5.5 The translation of SPARQL-1 fragment to nr-Datalog$^{\neg}$ . . . 113
6.6 Conclusion . . . 114

7 Nested Datalog 117
7.1 Syntax and Semantics of Nested Datalog . . . 118
7.2 Query atoms and external atoms . . . 123
7.3 The expressive power of Nested nr-Datalog$^{\neg}$ . . . 123
7.4 Conclusion . . . 127

8 On the Nested Datalog semantics of the EXISTS clause 129
8.1 Two forms of substitution of SPARQL variables . . . 129
8.1.1 Substitution of parameters . . . 130
8.1.2 Substitution of goal variables . . . 133
8.2 Nested Datalog with nulls . . . 139
8.3 Conclusions . . . 141

IV Conclusions 143

9 Conclusions and future work 145

Bibliography **149**

Appendix A The information lattice of mappings with marked nulls **155**

Appendix B SQL and relational algebra **161**
B.1 SQL and the generalized PRU algebra 162
B.2 SQL and the generalized selection 163
B.3 SQL and the generalized join 163
B.4 SQL and the generalized difference 164