

Contents

1	Introduction	11
1.1	Problem Definition	12
1.2	Contribution	14
1.3	Outline of the Thesis	16
2	Foundations	17
2.1	Modeling	17
2.1.1	Ontologies	18
2.1.2	Semantic Web	19
2.1.3	RDF and RDFS	20
2.1.4	SPARQL	24
2.1.5	GLUE Information Model	26
2.2	Peer-2-Peer Networks and DHTs	28
2.2.1	Chord	30
2.2.2	Pastry	32
2.2.3	Common API for DHTs	34
2.3	Subgraph Isomorphism	36
2.4	Bloom Filters	39
2.5	Summary	43
3	Resource Matching	45
3.1	Resource Matching in the Grid	45
3.2	Modeling Resources with RDFS	47
3.2.1	Resource Description	47
3.2.2	Example	48
3.2.3	Integrating Background Knowledge	49
3.2.4	Classifying Properties	51
3.2.5	Static and Dynamic Data	52
3.2.6	Origin of the Information	53
3.2.7	More Semantics	53
3.3	Data Organization	54
3.3.1	Peer-2-Peer	54
3.3.2	Design Choices	54
3.4	Summary	56

4	P2P based RDF Reasoning and Querying	57
4.1	Architectural Overview	57
4.2	The RDF Query Problem	59
4.2.1	Formal Problem Definition	60
4.2.2	Complexity of the RDF Query Problem	61
4.2.3	Further Elements of SPARQL Queries	63
4.3	Knowledge Distribution	63
4.3.1	Life Cycle of Triples	65
4.3.2	Node Departure	65
4.3.3	Node Arrival	66
4.3.4	Accessing the Knowledge	67
4.3.5	Blank Node Labels	67
4.4	RDFS Rules	67
4.5	Query Evaluation	69
4.5.1	Goals of the Query Evaluation Strategy	71
4.6	Exhaustive Query Evaluation	71
4.6.1	Determination of Candidate Sets	71
4.6.2	Look Ahead	78
4.6.3	Bloom Filters	79
4.6.4	Optimization for <code>rdf:type</code>	80
4.6.5	Final Evaluation	80
4.7	Top <i>k</i> Query Processing	81
4.7.1	Introductory Example	83
4.7.2	The Algorithm	85
4.7.3	Caching and Look Ahead	86
4.8	Ordering of the Query Triples	89
4.9	Summary	90
5	Experimental Evaluation	91
5.1	Prototype Architecture	91
5.2	Experimental Setup and Test Data	93
5.2.1	Setup of the Experiments	93
5.2.2	Generation of Test Data	94
5.3	Overall Performance	95
5.4	Exhaustive Evaluation	98
5.4.1	Simulation Results	98
5.4.2	Test Runs	100
5.4.3	Single Query Scenario	102
5.4.4	Performance Factors During Exhaustive Evaluation	103
5.4.5	Influence of Bloom Filter Parameters	104
5.4.6	Load Balancing Issues	106
5.5	Top <i>k</i> Evaluation	109
5.5.1	Performance Compared to Exhaustive	109

5.5.2	Parameter Choice for Look Ahead	111
5.6	Triple Dissemination and Reasoning	112
5.7	Summary	113
6	Related Work	115
6.1	Grid Computing	115
6.2	Semantic Web and Semantic Grid	116
6.2.1	Resource Description Framework	117
6.2.2	Web Ontology Language	118
6.2.3	Semantic Grid	119
6.2.4	Semantic Grid and Services	120
6.3	Relational Databases	121
6.4	Peer-2-Peer	121
6.4.1	P2P Data Integration	122
6.4.2	Semantic Query Routing	124
6.4.3	RDF based P2P Systems	125
6.4.4	Edutella	126
6.4.5	Bibster	127
6.5	Summary	127
7	Conclusion	129
7.1	Applications of P2P based RDF Querying and Reasoning	129
7.2	Limitations of our Approach and Further Work	130
7.3	Summary	131
	List of Figures	133
	Bibliography	135