
CONTENTS

1	Introduction	1
1.1	Data Model	1
1.2	XML query	3
1.2.1	From XPath and XQuery query to twig pattern query	4
1.2.2	Twig pattern matching	6
1.3	Document labeling and inverted list	7
1.4	Our research scope and contributions	11
1.5	Organization	13
2	Literature Review	14
2.1	Query processing over XML tree	14
2.1.1	The relational approach	14
2.1.2	The native approach	19
2.1.3	Comparison between the relational approach and the native approach	24
2.1.4	Hybrid management of relational data and XML data	25
2.2	Query processing over XML graph	26
2.3	Summary of related work	28

3	A semantic approach for twig pattern query processing	31
3.1	Introduction and motivation	32
3.2	VERT algorithm	35
3.2.1	Object-related semantics in XML data	35
3.2.2	An overview of VERT	38
3.2.3	Document parsing in VERT	39
3.2.4	Query processing in VERT	42
3.2.5	Analysis of VERT	45
3.3	Semantic optimizations	47
3.3.1	Optimization 1: object/property table	47
3.3.2	Optimization 2: object table	49
3.3.3	Optimization 3: relationship table	52
3.4	Query across multiple twig patterns	55
3.4.1	Query plan selection	57
3.5	Experiments	59
3.5.1	Settings	59
3.5.2	Comparison with Schema-based Relational Approach	60
3.5.3	Comparison with TwigStack	62
3.6	Summary	65
4	Enhancing twig pattern semantics for complex output information	67
4.1	Introduction	68
4.2	Query node characteristics	71
4.2.1	Purpose of query nodes	71
4.2.2	Optionality of query nodes	72
4.2.3	Occurrence of output information	72
4.3	TP+Output: an extension of twig pattern	73
4.3.1	Predicate node	74
4.3.2	Optional-predicate node	74

4.3.3	Output node	75
4.3.4	Optional-output node	76
4.3.5	Predicated-output node	76
4.3.6	Optional-predicated-output node	77
4.3.7	Discussion	78
4.4	VERT ^O to process TP+Output queries	78
4.4.1	Analysis	83
4.5	Experiments	84
4.5.1	Experimental settings	84
4.5.2	Compare TP+Output with TP and GTP	85
4.5.3	Scalability of VERT ^O	87
4.5.4	Comparison with XQuery processors	87
4.6	Summary	88
5	Performing grouping and aggregation in XML queries	90
5.1	Introduction	91
5.2	Related work on XML grouping	93
5.3	Query expression	95
5.4	VERT ^G algorithm	96
5.4.1	Data structures and output format	97
5.4.2	Query processing	99
5.4.3	Early pruning	104
5.4.4	Extension flexibility	104
5.4.5	Discussion on semantic optimization	106
5.4.6	Combining VERT ^O and VERT ^G	107
5.5	Experiments	108
5.5.1	Experimental settings	109
5.5.2	Comparison between VERT ^G without and with optimizations . . .	109
5.5.3	Comparison with other approaches	111

5.6 Summary	114
6 Conclusion	116
6.1 Conclusion	116
6.2 Future work	118
Bibliography	120