

Contents

1	Introduction	1
2	Background Models and Formalisms	3
2.1	Tree Grammars and Tree Automata	3
2.2	Attribute Grammars	8
2.3	XML Data Model	12
2.4	Document Type Definition	17
2.5	XPath Query Model	19
2.6	LDAP Model	24
2.6.1	LDIF Model	30
2.7	LDAP Query Model	31
3	Single-Server Caching Architecture	35
3.1	Overall Architecture	36
3.2	Internal Proxy Architecture	37
3.2.1	Proxy Caching Server	38
3.2.2	Query Engine	40
3.2.3	Distributed Cache Engine	41
3.2.4	XML Parser	44
3.2.5	XMLDAP Cache	45
3.2.6	Traditional Cache	47
4	Data and Query Representation Models	49
4.1	Data Representation	49
4.1.1	Query Representation	56
4.1.2	Meta-data Representation	57
4.2	Query Model	58
4.2.1	LDAPQL Model	59
4.2.2	Evaluation of XPath Queries using HLCACHES	60
4.3	Advantages of LDAPQL	73
4.3.1	Cache Answerability	74
4.3.2	Partial Query Evaluation	78
4.3.3	Subquery Preprocessing	80

4.3.4	Parallel Subquery Evaluation	87
5	Complexity Analysis	93
5.1	Formal Preliminaries	93
5.1.1	Monadic Second Order Logic for Graphs	93
5.1.2	Query Automata	95
5.2	XPath Complexity	96
5.2.1	XPath Time Complexity	98
5.2.2	XPath Size Complexity	99
5.2.3	XPath Decidability	102
5.3	LDAPQL complexity	104
6	Distributed Caching Architecture	107
6.1	PLEXOR Overview	107
6.2	PLEXOR Internal Architecture	110
6.2.1	The Network Management Daemon: <code>plexord</code>	111
6.2.2	The Application Management Daemon: <code>appd</code>	114
6.2.3	<code>proflerd</code>	117
6.3	Integration with HLCACHES	119
7	Experimental Evaluation	121
7.1	Single-server Evaluation	121
7.1.1	Experimental Setup	122
7.1.2	Document Storage Overhead	127
7.1.3	Average Storage and Retrieval Time	128
7.1.4	Query Execution Performance Improvement	129
7.1.5	Experimental Conclusions	130
7.2	Distributed Framework Evaluation	131
7.2.1	Communication Module Characteristics	131
7.2.2	Protocol Overhead	133
7.2.3	Experimental Conclusions	135
8	Assessment of Contributions	137
8.1	Generic XML and XPath Model Contributions	137
8.1.1	Related Work	138
8.2	HLCACHES	139
8.2.1	Related Work	140
8.3	PLEXOR	141
8.3.1	Related Work	142
9	Conclusions and Future Work	143
9.1	Theoretical Work	143
9.2	Experimental Work	144

CONTENTS

iii

A Query Automata

147

Bibliography

155

Index

171