

Inhaltsverzeichnis

1	Einleitung	1
1.1	Leichtgewichtige XML-Datenbanken	1
1.2	Semistrukturierte Daten	2
1.3	Motivation und Zielsetzung	4
1.4	Aufbau dieser Arbeit	5
2	XML	6
2.1	Das World Wide Web Consortium	6
2.2	Historie	8
2.3	Merkmale und Syntax von XML	8
2.3.1	Überblick	8
2.3.2	Syntax	9
2.3.3	Document Type Definitions	16
2.3.4	Wohlgeformtes XML-Dokument	16
2.4	Application Programming Interfaces für XML	19
2.4.1	Simple API for XML	19
2.4.2	Document Object Model	21
2.5	XML-Dokumente und relationale Datenbanken	24
3	Recherche	25
3.1	Anforderungen an Querysprachen für XML	25
3.2	Standard Query Language (SQL)	27
3.3	XML-Query Language (XQL)	31
3.4	XQL versus SQL	33
3.5	Weitere Anfragesprachen	33

3.5.1	XQuery	33
3.5.2	XML-QL	37
3.5.3	Object Query Language (OQL)	39
3.5.4	LoREL	39
3.5.5	XML Path Language (XPath)	41
3.5.6	Quilt	44
3.5.7	YATL	46
3.5.8	XML-GL	47
3.6	Alternative Ansätze	47
3.6.1	Glimpse	47
3.6.2	Tamino	48
4	Systementwurf	50
4.1	Entwicklungsphasen	50
4.2	Anforderungen und Analyse	51
4.2.1	Ist-Analyse	51
4.2.2	Ziele und Anforderungsanalyse	51
4.3	Strategie der Realisierung	53
4.4	Das Schichtenmodell	54
4.5	Das Komponentenmodell	56
4.6	Perl	59
4.6.1	Überblick	59
4.6.2	Reguläre Ausdrücke	59
4.6.3	CPAN	62
4.6.4	POD	63
4.6.5	Vor- und Nachteile von Perl	63
4.6.6	Kriterien für die Auswahl von Perl	64
5	ASIX	65
5.1	Bausteine des Systems	65
5.1.1	Sprite	65
5.1.2	XML-XQL	65
5.1.3	dnotify	66

5.1.4	ASIX-Modul	66
5.2	Die Anfrage	69
5.3	Der Indexer	75
5.4	Das Update	79
5.4.1	Hinzufügen und Ändern	80
5.4.2	Löschen	81
6	Evaluierung der Implementierung	83
6.1	Die Testumgebung	83
6.2	Testreihe 1 - Suchaufwand ohne Index	85
6.2.1	Abhängigkeit der Suchzeiten von Querybedingungen	86
6.2.2	Abhängigkeit der Suchzeiten von der Verschachtelungstiefe	86
6.2.3	Abhängigkeit der Suchzeiten von der Anzahl der XML-Files	87
6.3	Testreihe 2 - Suchaufwand mit Index	88
6.3.1	Abhängigkeit der Suchzeiten von Querybedingungen	88
6.3.2	Abhängigkeit der Suchzeiten von der Verschachtelungstiefe	88
6.3.3	Abhängigkeit der Suchzeiten von der Anzahl der XML-Files	89
6.4	Performancegewinn durch Indexe	90
7	Zusammenfassung	93
7.1	Entwicklungsstand	93
7.2	Ausblick	94
A	Evaluierungsdaten	103
A.1	Benchmark Perl-Modul XML-XQL	103
A.2	Benchmark ASIX	105
A.2.1	Messwerte der Testreihe 1 Der Zeitaufwand ohne Index	105
A.2.2	Messwerte der Testreihe 2 Der Zeitaufwand mit Index	106
A.2.3	Zeitersparnis und Verbesserungsfaktor	108

B Tabellen	111
B.1 XPath	111
B.2 Quilt	114
B.3 Regular Expressions	115
B.4 Das POD-Format	118
C Quellen	119
C.1 Die Konfigurationsdatei	119
C.2 Indextabellen	121
C.3 Die Logdatei	121
C.4 Die XML-Testfiles	121