

1	Einleitung	1
1.1	Thema und Ziel dieser Arbeit	5
2	Grundlagen	7
2.1	Binäre neuronale Assoziativspeicher	7
2.2	Information Retrieval	10
2.2.1	Allgemeine Struktur eines IR-Systems	11
2.2.2	IR-Systeme auf Basis invertierter Dateien	12
	Komprimierung invertierter Listen	13
	Speicherung von Wörterbüchern	15
	Erstellung der invertierten Datei	15
2.2.3	IR-Systeme auf Basis von Signatur- und Bitmap-Dateien	16
2.2.4	Invertierte Dateien im Vergleich zu Signaturdateien	18
2.2.5	Case-folding, Stemming und Stopwörter	20
2.2.6	Paralleles Information Retrieval	22
	MIMD Architekturen	23
	SIMD Architekturen	24
2.3	Zusammenfassung	25
3	Effizienz assoziativer Datenstrukturen	26
3.1	BiNAM	26
3.2	Hashing	35
3.2.1	Wahl der Hashfunktion	36
3.2.2	Hashing mit Verkettung	37
3.2.3	Minimales perfektes Hashing	42
3.2.4	Eignung	48
3.3	B-Bäume	49
3.3.1	Eignung	53
3.4	Vergleich der Speicherplatzeffizienz	54
3.5	Zusammenfassung	60
4	Implementierung und Leistungsbewertung	62
4.1	NeIRiS System-Architektur	62
4.1.1	WWW Spider	63
4.1.2	Indizierung	64
4.1.3	Vor- und Nachbearbeitung	68
4.2	Leistungsbewertung	71
4.2.1	Präzision und Vollständigkeit	71
4.2.2	Speicherplatzeffizienz	75
	Speicherplatzeffizienz von Signatur-Dateien	75
	Speicherplatzeffizienz von MG und NeIRiS (invertierte Listen vs. BiNAM)	79
4.2.3	Zeitbedarf von NeIRiS und MG	86
4.2.4	BiNAM Kenndaten	86
4.2.5	Vergleich mit bisherigen Ansätzen auf BiNAM Basis	90
4.3	Skalierung des Systems	91
4.4	Fehleranalyse	93
4.4.1	Fehler erster Art	93

4.4.2 Fehler zweiter Art	97
4.4.3 Abschätzung der Vollständigkeit	102
4.4.4 Dimensionierung des BiNAM	104
4.5 Diskussion alternativer Codierungsansätze	105
4.5.1 Dokumentpartitionierung	105
4.5.2 Hash-Codierung	107
4.5.3 Ziffernbasierte Hash-Codierung	108
4.6 Zusammenfassung	109
5 Systementwurf eines BiNAM	111
5.1 Prioritätsencoder und Decoder	114
5.2 Neuro-Controller	115
5.3 Virtuelles Speichermanagement	118
5.4 Laufzeitabschätzung	121
5.5 Zusammenfassung	123
6 Zusammenfassung und Ausblick	124
7 Literatur	129