

Dipl.-Ing. Abdelkarim Belhoula, Bochum

**Ein regelbasiertes Verfahren
zur maschinellen Graphem-
nach-Phonem-Umsetzung
von Eigennamen
in der Sprachsynthese**

Reihe **10**: Informatik/
Kommunikationstechnik Nr. **432**

Inhaltsverzeichnis

1	Abstract	1
2	Einleitung	2
	2.1 Überblick über vergleichbare Arbeiten	4
3	Datenbasen	7
	3.1 Eignung der Datenbasen für die Sprachsynthese	7
	3.2 Inhalt und Umfang der Datenbasen	9
	3.3 Zusammenfassung	13
4	Ein Grundkonzept für die Umsetzung von Namen	14
	4.1 Unterschiede zwischen Namen und Appellativa	14
	4.2 Regeln für die Generalisierung von Eigennamen	17
	4.3 Gesamtüberblick über das Umsetzungsverfahren	21
5	Die Regelmodule	23
	5.1 Syntax der Regelmodule	23
	5.2 Erstellung der Regelmodule	28
	5.2.1 Manuelle Erstellung der Regelmodule	28
	5.2.2 Regel-Editor	29
	5.2.3 Maschinelle Erstellung der Regelmodule	33
	5.3 Zusammenfassung	42
6	Die Namensvorverarbeitung	43
	6.1 Die Vorverarbeitung von Straßen- und Firmennamen	43
	6.2 Die Klassifizierung der Vornamen nach Herkunft	45
	6.3 Zusammenfassung	51
7	Morphologische Analyse von Eigennamen	52
	7.1 Struktur der Orts- und Nachnamen	53
	7.2 Struktur der Vornamen	55

7.3	Struktur der Firmennamen	57
7.4	Algorithmen für die morphologische Zerlegung	59
7.4.1	Zerlegung der Ortsnamen	63
7.4.2	Zerlegung der Nachnamen	64
7.4.3	Zerlegung der Vornamen	65
7.4.4	Zerlegung von Straßen- und Firmennamen	66
7.5	Ein statistisches Verfahren zur morphologischen Analyse	68
7.5.5	Das Zerlegungsverfahren	69
7.6	Zusammenfassung	71
8	Umsetzung der Rechtschrift in Lautschrift	72
8.1	Statistische Betrachtung der Phoneme in Eigennamen	72
8.2	Erstellung der Transkriptionsregeln	74
8.2.1	Transkription von Orts- und Nachnamen	74
8.2.2	Transkription von Vornamen	77
8.2.3	Transkription von Straßen- und Firmennamen	81
8.3	Zusammenfassung	82
9	Betonung der Eigennamen	83
9.1	Statistische Auswertung der Silben und ihrer Betonung	83
9.2	Betonung der Ortsnamen	85
9.3	Betonung der Nachnamen	88
9.4	Betonung der Vornamen	88
9.5	Zusammenfassung	91
10	Evaluierung des Umsetzungssystems	92
10.1	Evaluierung der Graphem-nach-Phonem-Umsetzung auf symbolischer Ebene	92
10.1.1	Ergebnis der morphologischen Zerlegung	92
10.1.2	Ergebnisse der phonetischen Transkription	93
10.1.3	Ergebnis der lexikalischen Betonung	95
10.2	Einfluß der Transkriptionsfehler auf die Verständlichkeit der Eigennamen	97
10.2.4	Versuchsbeschreibung und Auswertung	98

10.2.5	Ergebnis des Hörversuchs	101
10.2.6	Ergebnis des Tests mit optischer Darbietung der Eigennamen	103
10.2.7	Interpretation der Ergebnisse	106
10.2.8	Zusammenfassung	106
10.3	Evaluierung der synthetisierten Namen	108
10.3.9	Die Generierung des Testmaterials	109
10.3.10	Durchführung und Auswertung des Hörversuchs	117
10.3.11	Versuchsergebnisse und Diskussion	118
10.3.12	Zusammenfassung	121
11	Zusammenfassung	122
Anhang		124
A	Glossar der linguistischen Begriffe	125
B	Phonetisches Alphabet	129
C	Statistische Verteilung der Morphe	131
D	Regelmodule	140
E	Klassifizierung der Vornamen in Herkunftskategorien	155
F	Testmaterial der Hörversuche	158
	Literaturverzeichnis	162