

INHALTSVERZEICHNIS

Vorwort	v
<u>0. Einleitung</u>	1
0.1 Motivation der Arbeit und der Vorgehensweise	1
0.2 Kurze Zusammenfassung	6
<u>1. Einige Grundaspekte der Informationsverarbeitung</u>	9
1.1 Informationswissenschaft	11
1.2 Abgrenzung des Begriffs:	
Information Retrieval System (IR-System)	15
1.3 Relevanzproblematik	24
1.4 Informationsgenerierungsprozeß	28
1.5 Bezug der Arbeit zur Informationslinguistik	31
1.5.1 Probleme der linguistischen Forschung	31
1.5.2 Quantitative Linguistik	35
1.5.3 Gegenstand der Informationslinguistik	36
1.6 Forschungs- und Labormodelle von Information Retrieval Systemen	37
<u>2. Automatische Klassifikation</u>	44
2.1 Allgemeine Problemstellung	45
2.2 Theoretische Grundlagen	47
2.2.1 Objekte und Merkmale der Klassifikation	47
2.2.2 Objektmengen der automatischen Klassifikation (Cluster)	57
2.2.3 Klassifikationstypen	61
2.3 Methoden der Klassifikation	63
2.3.1 Direkte und rekursive Verfahren	63
2.3.2 Iterative Verfahren	69
2.3.3 Graphentheoretische Verfahren	73
2.3.4 Hierarchische Verfahren	80

2.4	Anwendung der automatischen Klassifikation in IR-Systemen	87
2.4.1	Problemstellung und Motivation	87
2.4.2	Aspekte des Einsatzes von automatischen Klassifikationsverfahren	89
2.4.3	Dokumentenklassifikation (Dokumentenclustering)	90
2.4.4	Termklassifikation	96
2.4.5	Gemischte Verfahren	99
2.5	Notiz zur dynamischen Klassifikation	103
2.6	Evaluierungsproblematik	104
<u>3.</u>	<u>Prinzip des STEINADLER-Ansatzes</u>	109
3.1	Motivation und Grundgedanke des Verfahrens	110
3.2	Notiz zur Voranalyse	112
3.3	Thesaurusaufbau	114
3.4	Klassifikationsprozeß	116
3.4.1	Prioritätsklassenbildung	116
3.4.2	Clusteranalyse innerhalb der einzelnen Prioritätsklassen	118
3.4.3	Verknüpfungen zwischen benachbarten Prioritätsklassen	120
3.5	Updatingprozeß	122
3.6	Übersicht über Dateiorganisation	123
<u>4.</u>	<u>Thesaurusaufbau</u>	125
4.1	Zum Thesaurusbegriff	126
4.2	Informationsaufbereitung	131
4.2.1	Automatische linguistische Analyse (im IR-System CONDOR)	131
4.2.2	Statistische Voranalyse	134
4.3	Thesaurusorganisation	137
4.3.1	Hash-organisierter Stammthesaurus (Thesaurushauptteil)	137
4.3.2	Stammthesauruseintrag	139

<u>5. Ermittlung der Hierarchieebenen der Klassifikation</u>	
<u>(Prioritätsklassenbildung)</u>	141
5.1 Problemstellung	142
5.2 Prinzip der Prioritätsklassenbildung	143
5.2.1 Theoretische Grundlagen und Voraussetzungen (bei Objekten mit binären Merkmalen)	143
5.2.2 Präordnung auf Prioritätsklassen	150
5.2.3 Bestimmung der Prioritätsklassen bei Objekten mit beliebigen Merkmalen	151
5.3 Beschreibung des Algorithmus	153
5.3.1 Einige Hinweise zur Prioritätsklassenbildung	153
5.3.2 Aufbereitung des Klassifikationspools	155
5.3.3 Schematischer Algorithmusablauf	157
<u>6. Clusteranalyse innerhalb der einzelnen Hierarchie-</u>	
<u>ebenen der Klassifikation</u>	159
6.1 Problemstellung	160
6.2 Prinzip des Verfahrens (INPRIOR)	161
6.2.1 Theoretische Grundlagen und Voraussetzungen	161
6.2.2 Operationale Algorithmusbeschreibung	168
6.2.3 Formale Beschreibung des Verfahrensablaufs	176
6.2.4 Thesauruseinträge nach der Clusteranalyse	182
6.3 Alternative Verfahren	184
6.4 Verfahrenseigenschaften	185
<u>7. Verknüpfungen zwischen benachbarten Hierarchieebenen</u>	
<u>der Klassifikation</u>	188
7.1 Problemstellung	189
7.1.1 Theoretische Grundlagen und Voraussetzungen (Ausgangssituation)	190
7.1.2 Grundprinzipien und Basis der Vorgehensweise	198

7.2	Beschreibung des Verfahrensansatzes	203
7.2.1	Auflösung zu kleiner Dokumentengruppen	204
7.2.2	Behandlung der nichtzuzuordnenden Knotenteile	211
7.2.3	Schematischer Algorithmusablauf	212
7.3	Thesauruseinträge nach der Vernetzung	216
7.4	Verfahrensmodifikation und alternative Verfahren	218
7.5	Interpretation und Eigenschaften des Verfahrens	219
8.	<u>Updatingmöglichkeiten der Klassifikation</u> (dynamische Klassifikation)	222
8.1	Problemstellung	223
8.2	Updatingprinzip beim STEINADLER-Verfahren	225
8.3	Funktionen der Updatingkomponente	228
8.3.1	Ausgangssituation und Funktionsübersicht	228
8.3.2	Klassieren	233
8.3.3	Korrektur der Termcluster im Klassifikationsnetz	238
8.3.4	Sonderbehandlung einiger Einzelterme	240
8.3.5	Löschen von Informationsobjekten	243
8.3.6	Sonstige Modifikationsmöglichkeiten	246
8.4	Veränderung der Thesauruseinträge nach der Updatingkomponente	247
8.5	Aufwandschätzung (der Klassifikation mit STEINADLER)	249
9.	<u>Interpretation der Ergebnisse der Klassifikation</u> bei der Grobrecherche	251
9.1	Einführung in die Rechercheproblematik	252
9.1.1	Recherchearten und Suchtechniken	253
9.1.2	Traditionelle Suchlogiken	256
9.1.3	Prinzip der Clustersuche	258
9.2	Spezielle Problemstellung (STEINADLER)	262

9.3	Grobrecherchestrategien	264
9.3.1	Einleitung und Definitionen	265
9.3.2	Suchauftragprojektion in das Klassifikationsnetz	269
9.3.3	Retrievalfunktionen	274
9.3.4	Rankingalgorithmen	275
9.4	Rahmenfunktionen der Grobrecherche	282
<u>10. Präzisierung der Rechercheergebnisse</u>		
<u>(als spezielles Relevanzfeedback)</u>		283
10.1	Grundlagen zu Relevanzfeedback-Strategien (interaktive Retrievalstrategien)	284
10.2	Problemstellung der Präzisierung als eines speziellen Relevanzfeedback	288
10.3	Präzisierung	290
10.3.1	Grundgedanke und Einteilung der Vorgehensweise	291
10.3.2	Potentielle Erweiterungsmöglichkeiten	294
10.4	Rahmenbedingungen der Präzisierungsfunktion	295
<u>11. Evaluierungsproblematik</u>		298
11.1	Bewertungskriterien und -maße	299
11.1.1	Bewertungsproblematik: kritische Betrachtung	299
11.1.2	Übersicht über die wichtigsten Effektivitätsmaße	303
11.2	Aspekte der Bewertungsproblematik	306
11.3	Evaluierungsmethoden	310
11.3.1	Zum Vorgang der Evaluierung von Retrievalergebnissen	311
11.3.2	Zur Bewertung der Retrievalleistung	314
11.3.3	Bewertung der Klassifikationsanwendungen	319
11.3.4	Zur CONDOR-Evaluierung (bzgl. STEINADLER)	326
11.4	Evaluierungsproblematik: Schlußbemerkung	327

<u>12. Wechselwirkungen der automatischen Klassifikation mit anderen Erschließungsverfahren und ihr Anwendungsgebiet</u>	329
12.1 Abgrenzung der Anwendungsziele der automatischen Klassifikation in Information Retrieval Systemen	330
12.2 Grenzen der Einflüsse sonstiger Erschließungsinstrumente	333
12.2.1 Thesaurusproblematik	333
12.2.2 Automatische linguistische Analyse	334
12.2.3 Formale Dokumentanalyse	335
12.2.4 Verhältnis zwischen automatischer Klassi- fikation und automatischer Indexierung	336
12.3 Automatische Klassifikation und Datenmodell	337
12.4 Intellektuelle und sonstige Einflüsse und Wechselwirkungen	339
12.5 Zur Anwendung der automatischen Klassifikation in Information Retrieval Systemen	341
Anmerkungen	343
Literatur	368
<u>Anhänge</u>	391
A. Statistische Daten	393
B. Beispiele der Knotenbildung	397
C. Recherchebeispiel	406
Verzeichnis der mathematischen Bezeichnungen, der sonstigen Notation und Abkürzungen	412