

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Entstehung und Inhalt der Arbeit . . . . .	1
1.2	Theoretische Grundlagen und Terminologie . . . . .	4
1.2.1	Wissensrepräsentation . . . . .	4
1.2.2	Frames als formale Strukturen zur Darstellung von Wissen . . . . .	6
1.2.3	Parallelen zur Semantik . . . . .	10
1.2.4	Die "Weltordnung" der abzubildenden Dinge: Ontologien . . . . .	13
1.3	Aufbau der Arbeit . . . . .	17
<b>2</b>	<b>Das Corpus-Sammelprogramm</b>	<b>20</b>
2.1	Das <i>World Wide Web</i> als Corpus . . . . .	21
2.2	Funktionsweise von Suchmaschinen . . . . .	24
2.3	Das Sammeln von Corpora mit dem Programm <i>LingKonnet</i> . . . . .	25
2.3.1	<i>LingKonnet</i> – ein zweiphasiger Web-Wrapper . . . . .	25
2.3.2	Algorithmus des Seitenwrappers . . . . .	31
2.4	Evaluation des Corpus-Sammelprogramms . . . . .	37
2.4.1	Wrapper-Entwicklung als iterativer Verbesserungsprozeß . . . . .	37
2.4.2	Tauglichkeit einzelner HTML-Tags zur Kontextbestimmung . . . . .	38
2.4.3	Weitere Parameter . . . . .	44
2.4.4	Vorschläge zur Weiterentwicklung . . . . .	45
2.5	Zusammenfassung und vergleichbare Ansätze . . . . .	47
2.5.1	<i>Information Retrieval</i> und <i>Information Extraction</i> . . . . .	49
2.5.2	Das Web als Corpusquelle für ein <i>American National Corpus</i> . . . . .	49
2.5.3	Begriffsdefinitionen aus dem Web . . . . .	52
2.5.4	Natürlichsprachliche Benutzungsoberflächen für Suchmaschinen . . . . .	53
2.5.5	Konzepte und "typische" Wörter von Webseiten . . . . .	54
2.6	Technische Entwicklungen im <i>World Wide Web</i> . . . . .	55
<b>3</b>	<b>Eine Struktur und eine Hierarchie für Konzeptframes</b>	<b>57</b>
3.1	Charakteristika unterschiedlicher Framestrukturen . . . . .	58
3.1.1	Ein Beispiel aus der Linguistik . . . . .	58
3.1.2	Ein Beispiel aus der Wissensrepräsentation . . . . .	61
3.1.3	Erwünschte Eigenschaften der zu erstellenden Frame-Struktur . . . . .	63
3.2	Verwendete Framestruktur . . . . .	63
3.2.1	Subslots: Relationen mit Verben als Kernbestandteil . . . . .	66
3.2.2	Filler: Konzeptuelle Inhalte für den Frame . . . . .	74
3.2.3	Slots: Eine weitere Strukturebene innerhalb von Frames . . . . .	77
3.3	Propositionen in Frames: Zusammenfassung und Vergleich . . . . .	79

3.4	Standardinformationen für Frames und Framehierarchie . . . . .	84
3.4.1	Gewinnung von Standardinformationen . . . . .	86
3.4.2	Erstellung der <i>Top-Level</i> -Framehierarchie . . . . .	89
3.5	Vergleichbare Ansätze . . . . .	95
3.5.1	Die lexikalische Datenbank EuroWordNet . . . . .	95
3.5.2	Cyc – ein Projekt zur Erfassung des gesamten Weltwissens . . . . .	99
3.5.3	Formale <i>Top-Level</i> -Ontologien . . . . .	103
3.6	Zusammenfassung . . . . .	105
<b>4</b>	<b>Annotation von Web-Corpora mit Frames</b>	<b>107</b>
4.1	Erstellung, Speicherung und Ausgabe von Annotationen . . . . .	107
4.1.1	Graphische Benutzungsoberfläche . . . . .	108
4.1.2	Status einzelner Corpusannotationen und der annotierten Frames . . . . .	114
4.1.3	Die LingKonnet-Datenbank . . . . .	116
4.1.4	Ausgabedescripte . . . . .	120
4.2	Erstellte Konzeptframes . . . . .	122
4.2.1	Untersuchte Konzepte . . . . .	122
4.2.2	Einordnung der annotierten Konzepte in die Framehierarchie . . . . .	132
4.3	Evaluation verwendeter <i>Top-Level</i> -Frames und ihrer Relationen . . . . .	137
4.3.1	Relevanz der Standardsubslots . . . . .	138
4.3.2	<i>Top-Level</i> -Subslots und neue Subslots in den Annotationen . . . . .	143
4.3.3	Neue Subslots für die <i>Top-Level</i> -Framehierarchie . . . . .	144
4.3.4	Evaluationsergebnis . . . . .	159
4.4	Zusammenfassung . . . . .	162
<b>5</b>	<b>Sprachliche Realisierung konzeptueller Relationen</b>	<b>164</b>
5.1	Methoden der Ermittlung und Beschreibung von Mustern . . . . .	165
5.1.1	Muster für eine corpusbasierte terminologische Wissensbasis . . . . .	165
5.1.2	Muster zum Auffinden von <i>knowledge rich contexts</i> . . . . .	167
5.1.3	Muster und Regeln im System SEEK-JAVA . . . . .	168
5.1.4	Vergleich der Methoden . . . . .	170
5.2	Eigene Methode . . . . .	171
5.3	Einzelanalysen . . . . .	176
5.3.1	Muster für das Subslot + <i>inclure</i> . . . . .	177
5.3.2	Muster für das Subslot + <i>ressembler à</i> . . . . .	185
5.3.3	Muster für das Subslot + <i>avoir (Les caractéristiques extérieures)</i> . . . . .	190
5.3.4	Muster für das Subslot + <i>se trouver, à</i> . . . . .	197
5.3.5	Muster für das Subslot + <i>avoir (L'importance pour l'homme)</i> . . . . .	203
5.4	Auswertung . . . . .	211
<b>6</b>	<b>Intrakonzeptuelle Strukturen und FrameNet-Frames</b>	<b>214</b>
6.1	Die intrakonzeptuelle Struktur der OPFER-Rolle . . . . .	214
6.2	Die intrakonzeptuelle Struktur der WAREN-Rolle . . . . .	218
6.2.1	LA MAISON EN TANT QUE MARCHANDISE . . . . .	218
6.2.2	LE LIVRE EN TANT QUE MARCHANDISE . . . . .	221
6.2.3	Die WAREN-Rolle: Zusammenfassung . . . . .	222
6.3	Das FrameNet-Projekt . . . . .	225
6.4	Methodenvergleich mit FrameNet . . . . .	227

6.5	Intrakonzeptuelle Strukturen und FrameNet: Einzelvergleiche . . . . .	229
6.5.1	WAREN-Rolle und <i>Commercial Transaction</i> -Frame . . . . .	230
6.5.2	OPFER-Rolle, <i>Cause_harm</i> -Frame und <i>Risk</i> -Frame . . . . .	233
6.6	Zusammenfassung . . . . .	239
<b>7</b>	<b>Ergebnisse und Ausblick</b> . . . . .	<b>241</b>
7.1	Ergebnisse . . . . .	241
7.2	Anwendungsgebiete . . . . .	243
7.3	RDF-Konzeptframes für das <i>Semantic Web</i> . . . . .	245
7.4	Schlußbemerkung . . . . .	248
	<b>Literaturverzeichnis</b> . . . . .	<b>249</b>
<b>A</b>	<b>Material und Evaluation</b> . . . . .	<b>262</b>
A.1	Framedefinitionen . . . . .	262
A.2	Grundausrüstung des Frame-Annotationsprogramms . . . . .	278
A.3	Quantitative Evaluation . . . . .	287
A.4	Vier verbesserte <i>Top-Level</i> -Frames . . . . .	299
<b>B</b>	<b>Annotierte Frames</b> . . . . .	<b>302</b>
B.1	Ein Subframe von PERSONNE DANS UN RÔLE: ENFANT . . . . .	302
B.2	Ein Subframe von PERSONNE AVEC UNE PROFESSION: JOURNALISTE . . . . .	311
B.3	Ein Subframe von ÉVÉNEMENT: DANGER . . . . .	321
B.4	Ein Subframe von ARTEFACT: MAISON . . . . .	322
B.5	Subframes von ORGANISME: INSECTE, ausgewählte Subframes . . . . .	331
B.6	Drei Frames für NATURE . . . . .	335
	<b>Index</b> . . . . .	<b>338</b>