

Johannes Leveling

Formale Interpretation von Nutzeranfragen
für natürlchsprachliche Interfaces zu Informationsangeboten im Internet

Inhaltsverzeichnis

1	Einführung	1
1.1	Bemerkungen zum Stellenwert der Informationssuche	1
1.2	Entwicklung der computergestützten Informationssuche	3
1.2.1	Traditionelle Katalogrecherche bei der Deutschen Bibliothek	5
1.2.2	Informationsrecherche im Internet: Das Beispiel Google	5
1.3	Motivation für den natürlchsprachlichen Informationszugang	6
1.4	Abgrenzung, Kontext und Einbettung dieser Forschungsarbeit	11
1.5	Thesen	15
2	Natürlchsprachlicher Zugang zu Datenbanken	17
2.1	Natürlchsprachliche Schnittstellen als Gegenstand der Forschung	17
2.2	Argumente für den Einsatz natürlchsprachlicher Interfaces	18
2.3	Gründe für das Ausbleiben des kommerziellen Erfolges	20
2.4	Stand der Technik bei natürlchsprachlichen Systemen	22
2.4.1	Eine Typologie natürlchsprachlicher Systeme	22
2.4.2	Natürlchsprachliche Interfaces	23
2.5	Terminologie und Begriffsdefinitionen	26
3	Sprachen für die Repräsentation von Nutzeranfragen	31
3.1	Aspekte zur Definition von Sprachen	31
3.2	Anfrageverarbeitung für die Informationssuche	32
3.2.1	Informationssuche als iterativer Prozess	32
3.2.2	Syntaktische Elemente von Anfragen und Anfragereformulierung	34
3.3	Natürlchsprachliche Anfragen	37
3.4	Die semantische Darstellung von Nutzeranfragen in MultiNet	38
3.5	Datenbankanfragesprachen und Retrieval-Protokolle	38
3.5.1	Logikorientierte Anfragerepräsentation mit PROLOG	38
3.5.2	Eine datenbankunabhängige Anfragerepräsentation – DIQR	39
3.5.3	Bag-of-words Anfragen	39
3.5.4	Boole'sche Anfragen	41
3.5.5	Die Structured Query Language – SQL	42
3.5.6	Das Wide Area Information Service Protokoll – WAIS	43
3.5.7	Die Anfragesprache von NCSTRL	44
3.5.8	Anfragen im Z39.50-Protokoll	44

3.5.9	Das Simple Object Access Protocol für Web-Services	49
4	Das MultiNet-Paradigma	51
4.1	Mehrschichtige Erweiterte Semantische Netze	51
4.1.1	Relationen, Funktionen und Konzepte	53
4.1.2	Layer-Informationen	55
4.2	Ressourcen für MultiNet	56
4.2.1	Das semantikbasierte Computerlexikon HaGenLex	56
4.2.2	Textcorpora und externe Ressourcen für MultiNet	57
4.3	Werkzeuge für MultiNet	58
4.3.1	Die Werkbank für den Lexikographen LIA	58
4.3.2	Der Wortklassen-gesteuerte Disambiguierende Parser	59
4.3.3	Die Werkbank für den Wissensrepräsentation mit MultiNet	60
4.4	Anwendungen des MultiNet-Paradigmas	60
4.4.1	Das NLI für Informationsanbieter im Internet	61
4.4.2	Logisch-inferentielle Fragebeantwortung	62
5	Das NLI-Z39.50	63
5.1	Architektur des NLI-Z39.50	63
5.1.1	Die Benutzungsschnittstelle – User-Client (Controller)	66
5.1.2	Die Sprachanalyse – WOCADI und WCFA	66
5.1.3	Das Transformationsmodul – TRANSFOR	67
5.1.4	Der Kritiker	68
5.1.5	Der Broker	68
5.1.6	Das Recherchemodul – IR-Client	69
5.1.7	Die Antwortgenerierung – ANSWER	69
5.1.8	Lokale und entfernte Informationsangebote	69
5.2	Erläuterung des Kommunikationsverlaufs	72
5.3	Die Zwischensprache DIQR	77
5.3.1	Zielsetzung für die DIQR	77
5.3.2	Erweiterte BNF für die Syntax der DIQR	77
5.4	Besonderheiten des NLI-Z39.50	79
5.4.1	Besonderheiten der Anfrageaufbereitung	80
5.4.2	Besonderheiten der Mensch-Maschine-Interaktion	84
5.4.3	Besonderheiten der Anfrageverarbeitung	88
5.4.4	Besonderheiten der Ergebnisaufbereitung	104
6	Transformation von Anfragerepräsentationen	107
6.1	Definition der Transformation von Ausdrücken einer Sprache	107
6.2	Stand der Technik bei regelgestützten Transformationen	111
6.3	Der Rete-Ansatz	112
6.4	Transformationsregeln für das NLI-Z39.50	115
6.4.1	Typen von Regeln	115
6.4.2	EBNF für die Syntax der Transformationsregeln	116

6.4.3	Die Transformationsregeln	123
6.4.4	Wissensbasen und Frageliterale	128
6.5	Transformation für das NLI-Z39.50	128
6.5.1	Die Übersetzung der Transformationsregeln in ein Rete- Netzwerk	130
6.5.2	Kontrollstrategien zur Regelverarbeitung	135
6.5.3	Transformationsoperationen	136
6.5.4	Einbeziehung von Hintergrundwissen in die Transformation	141
6.6	Aufbau der TRANSFOR-Komponente	143
6.6.1	Der Regelauswertungszyklus in TRANSFOR	143
6.6.2	Die Erkennungsphase ('Recognize')	144
6.6.3	Die Auswahlphase ('Select')	146
6.6.4	Die Ausführungsphase ('Act')	149
6.7	Die Transformation einer Beispielanfrage	150
6.7.1	Transformation einer natürlichsprachlichen Anfrage in einen MultiNet-Ausdruck	150
6.7.2	Transformation eines MultiNet-Ausdrucks in einen DIQR-Ausdruck	151
6.7.3	Transformation eines DIQR-Ausdrucks in einen Ausdruck des Retrievalprotokolls	160
6.8	Werkzeuge und Ressourcen für die Transformation	163
6.8.1	Der Regeleditor	163
6.8.2	Die Akquisition und Wartung der Regelmenge	165
6.9	Zusammenfassung der Leistungen der Transformation	167
7	Die Mensch-Maschine-Interaktion im NLI-Z39.50	169
7.1	Anforderungen an eine Mensch-Maschine-Schnittstelle	169
7.1.1	Probleme der Nutzer mit herkömmlichen Systemen	171
7.1.2	Probleme der Nutzer mit NLI's	173
7.1.3	Das Vokabularproblem	175
7.1.4	Das Nutzerverhalten im NLI-Z39.50	176
7.2	Linguistische Phänomene in Bibliotheksanfragen	177
7.3	Die sprachliche Abdeckung der Transformation	179
7.4	Die Abdeckung durch die Datenbanken	180
7.5	Antwortverhalten des NLI-Z39.50	181
7.5.1	Antwortaufbereitung	181
7.5.2	Antwortzeiten des NLI	181
7.6	Ansätze für ein mehrsprachiges Retrieval	182
7.7	Technische Realisierung	183
7.7.1	Technische Grundlagen für die Implementierung des NLI	183
7.7.2	Programmiersprachen und Entwicklungsumgebung	184

8	Evaluationen des NLI-Z39.50	187
8.1	Ziele einer Evaluation	187
8.2	Kernbegriffe für Evaluationen der Informationssuche	190
8.3	Evaluationsinitiativen für die Informationssuche	192
8.4	Erste Phase der Evaluation	193
8.4.1	Die technische Evaluation	194
8.4.2	Die quantitative Evaluation	200
8.4.3	Die qualitative Evaluation	202
8.5	Die externe Evaluation durch die Universität Leipzig	207
8.6	Das bibliographische Anfragecorpus – MBQ	208
8.7	Experimente zur Informationssuche	210
8.7.1	GIRT 2003	210
8.7.2	GIRT 2004	211
8.7.3	GIRT 2005	212
8.7.4	GeoCLEF 2005	212
8.8	Eine Evaluation der Rückzugsstrategie (MAKL)	213
8.8.1	Grundlagen für die Wortformklassifikation	214
8.8.2	Die Trainingsphase von MAKL	215
8.8.3	Die Klassifikationsphase von MAKL	216
8.8.4	Die Kombination von SYMATRA und MAKL im NLI-Z39.50	217
8.9	Schlussfolgerungen aus den Evaluationen des NLI-Z39.50	220
9	Fazit	221
9.1	Zusammenfassung der erreichten Leistungen	221
9.2	Gültigkeit der Thesen	224
9.3	Ausblick und weitere Forschung	228
A	Anhang	245
A.1	Auszug aus dem Hintergrundwissen	245
A.2	Auszug aus der Regelmenge	247
A.3	Elementare Tests in den Transformationsregeln	257
A.4	Aktionen der Transformationsregeln	259