

Contents

List of Figures	vii
List of Tables	xi
Glossary	xv
1 Introduction	1
2 Applications for Toponym Disambiguation	9
2.1 Geographical Information Retrieval	11
2.1.1 Geographical Diversity	18
2.1.2 Graphical Interfaces for GIR	19
2.1.3 Evaluation Measures	21
2.1.4 GeoCLEF Track	23
2.2 Question Answering	26
2.2.1 Evaluation of QA Systems	29
2.2.2 Voice-activated QA	30
2.2.2.1 QAST: Question Answering on Speech Transcripts . . .	31
2.2.3 Geographical QA	32
2.3 Location-Based Services	33
3 Geographical Resources and Corpora	35
3.1 Gazetteers	37
3.1.1 Geonames	38
3.1.2 Wikipedia-World	40
3.2 Ontologies	41
3.2.1 Getty Thesaurus	41
3.2.2 Yahoo! GeoPlanet	43

CONTENTS

3.2.3	WordNet	43
3.3	Geo-WordNet	45
3.4	Geographically Tagged Corpora	51
3.4.1	GeoSemCor	52
3.4.2	CLIR-WSD	53
3.4.3	TR-CoNLL	55
3.4.4	SpatialML	55
4	Toponym Disambiguation	57
4.1	Measuring the Ambiguity of Toponyms	61
4.2	Toponym Disambiguation using Conceptual Density	65
4.2.1	Evaluation	68
4.3	Map-based Toponym Disambiguation	71
4.3.1	Evaluation	72
4.4	Disambiguating Toponyms in News: a Case Study	76
4.4.1	Results	84
5	Toponym Disambiguation in GIR	87
5.1	The GeoWorSE GIR System	88
5.1.1	Geographically Adjusted Ranking	90
5.2	Toponym Disambiguation vs. no Toponym Disambiguation	92
5.2.1	Analysis	96
5.3	Retrieving with Geographically Adjusted Ranking	98
5.4	Retrieving with Artificial Ambiguity	98
5.5	Final Remarks	104
6	Toponym Disambiguation in QA	105
6.1	The SemQUASAR QA System	105
6.1.1	Question Analysis Module	107
6.1.2	The Passage Retrieval Module	108
6.1.3	WordNet-based Indexing	110
6.1.4	Answer Extraction	111
6.2	Experiments	113
6.3	Analysis	116
6.4	Final Remarks	116

7 Geographical Web Search: Geooreka!	119
7.1 The Geooreka! Search Engine	120
7.1.1 Map-based Toponym Selection	122
7.1.2 Selection of Relevant Queries	124
7.1.3 Result Fusion	125
7.2 Experiments	127
7.3 Toponym Disambiguation for Probability Estimation	131
8 Conclusions, Contributions and Future Work	133
8.1 Contributions	133
8.1.1 Geo-WordNet	134
8.1.2 Resources for TD in Real-World Applications	134
8.1.3 Conclusions drawn from the Comparison of TD Methods	135
8.1.4 Conclusions drawn from TD Experiments	135
8.1.5 Geooreka!	136
8.2 Future Work	136
Bibliography	139
A Data Fusion for GIR	145
A.1 The SINAI-GIR System	145
A.2 The TALP GeoIR system	146
A.3 Data Fusion using Fuzzy Borda	147
A.4 Experiments and Results	149
B GeoCLEF Topics	155
B.1 GeoCLEF 2005	155
B.2 GeoCLEF 2006	160
B.3 GeoCLEF 2007	165
B.4 GeoCLEF 2008	170
C Geographic Questions from CLEF-QA	175
D Impact on Current Research	179