

Contents

1	Introduction	1
1.1	Lexical Semantics and Text Understanding	3
1.2	Semantic Domains: Computational Models for Lexical Semantics	4
1.3	Structure of the Book	5
1.3.1	Semantic Domains	6
1.3.2	Domain Models	7
1.3.3	Semantic Domains in Text Categorization	8
1.3.4	Semantic Domains in Word Sense Disambiguation	9
1.3.5	Multilingual Domain Models	11
1.3.6	Kernel Methods for Natural Language Processing	12
2	Semantic Domains	13
2.1	The Theory of Semantic Fields	14
2.2	Semantic Fields and the <i>meaning-is-use</i> View	18
2.3	Semantic Domains	20
2.4	The Domain Set	22
2.5	WORDNET DOMAINS	23
2.6	Lexical Coherence: A Bridge from the Lexicon to the Texts ...	25
2.7	Computational Models for Semantic Domains	29
3	Domain Models	33
3.1	Domain Models: Definition	33
3.2	The Vector Space Model	34
3.3	The Domain Space	36
3.4	WORDNET-Based Domain Models	38
3.5	Corpus-Based Acquisition of Domain Models	40
3.6	Latent Semantic Analysis for Term Clustering	41
3.7	The Domain Kernel	44
3.7.1	Domain Features in Supervised Learning	44
3.7.2	The Domain Kernel	46

4	Semantic Domains in Text Categorization	49
4.1	Domain Kernels for Text Categorization	49
4.1.1	Semi-supervised Learning in Text Categorization	50
4.1.2	Evaluation	51
4.1.3	Discussion	55
4.2	Intensional Learning	56
4.2.1	Intensional Learning for Text Categorization	56
4.2.2	Domain Models and the Gaussian Mixture Algorithm for Intensional Learning	58
4.2.3	Evaluation	62
4.2.4	Discussion	67
4.3	Summary	68
5	Semantic Domains in Word Sense Disambiguation	69
5.1	The Word Sense Disambiguation Task	70
5.2	The Knowledge Acquisition Bottleneck in Supervised WSD	73
5.3	Semantic Domains in the WSD Literature	74
5.4	Domain-Driven Disambiguation	76
5.4.1	Methodology	76
5.4.2	Evaluation	77
5.5	Domain Kernels for WSD	79
5.5.1	The Domain Kernel	80
5.5.2	Syntagmatic Kernels	81
5.5.3	WSD Kernels	82
5.5.4	Evaluation	82
5.6	Discussion	86
6	Multilingual Domain Models	89
6.1	Multilingual Domain Models: Definition	90
6.2	Comparable Corpora	91
6.3	Cross-language Text Categorization	92
6.4	The Multilingual Vector Space Model	93
6.5	The Multilingual Domain Kernel	95
6.6	Automatic Acquisition of Multilingual Domain Models	96
6.7	Evaluation	98
6.7.1	Implementation Details	98
6.7.2	Monolingual Text Categorization Results	99
6.7.3	Cross-language Text Categorization Results	99
6.8	Summary	100
7	Conclusion and Perspectives for Future Research	101
7.1	Summary	101
7.2	Future Work	103
7.2.1	Consolidation of the Present Work	103
7.2.2	Domain-Driven Technologies	104

7.3 Conclusion	105
A Appendix: Kernel Methods for NLP	107
A.1 Supervised Learning	107
A.2 Feature-Based vs. Instance-Based Learning	110
A.3 Linear Classifiers	111
A.4 Kernel Methods	115
A.5 Kernel Functions	118
A.6 Kernels for Text Processing	118
References	125