## Contents

Acknowledgments i  
Abstract iii  

### Contents vii  

#### 1 Introduction 1  
1.1 Knowledge Representations and Cognitive Science 2  
1.2 Cross-domain and Cross-language Text Representations 5  
1.2.1 Domain Adaptation Text Representations 6  
1.2.2 Cross-language Text Representations 7  
1.3 Motivation and Objectives 9  
1.4 Research Questions 10  
1.5 Contributions of this Thesis 11  
1.6 Structure of this Thesis 12  

#### 2 Cross-domain Polarity Classification using a Knowledge-enhanced Meta-classifier 15  
2.1 Introduction 16  
2.2 Related Work 18  
2.3 Knowledge-enhanced Meta-classifier 19  
2.3.1 Word Sense Disambiguation and Vocabulary Expansion via a Semantic Network 20  
2.3.2 Base Classifiers 25  
2.3.3 Stacked Generalization 28  
2.4 Evaluation 30  
2.4.1 Dataset 30  
2.4.2 Methodology 30  
2.4.3 Evaluation of Base Classifiers 31
## CONTENTS

2.4.4 Single-domain Polarity Classification 36
2.4.5 Cross-domain Polarity Classification 37
2.5 Conclusions 40

3 A Systematic Study of Knowledge Graph Analysis for Cross-language Plagiarism Detection 43
3.1 Introduction 44
3.2 Related Work 47
  3.2.1 Cross-language Plagiarism Detection 47
  3.2.2 Distributed Representations for Conceptual Semantic Relatedness 50
3.3 Knowledge Graphs 51
  3.3.1 BabeNet 51
  3.3.2 Creation of the Knowledge Graphs 53
  3.3.3 Weighting of the Semantic Relations 54
  3.3.4 Characteristics of the Knowledge Graphs 60
3.4 Cross-language Knowledge Graph Analysis (CL-KGA) 63
3.5 Evaluation 66
  3.5.1 Datasets 67
  3.5.2 Methodology 67
  3.5.3 Evaluation of CL-KGA Weighting Schemes for Semantic Relations 70
  3.5.4 Evaluation of the CL-KGA Variants and Characteristics 73
  3.5.5 Comparison with the State-of-the-art 75
3.6 Conclusions 79

4 A Knowledge-based Representation for Cross-language Document Retrieval and Categorization 81
4.1 Introduction 82
4.2 Related Work 83
4.3 A Knowledge-based Document Representation 85
  4.3.1 BabelNet 86
  4.3.2 From Document to Knowledge Graph 87
  4.3.3 Similarity between Knowledge Graphs 90
4.4 A Multilingual Vector Representation 91
  4.4.1 From Document to Multilingual Vector 91
  4.4.2 Similarity between Multilingual Vectors 92
6.2 Future Work .............................................. 159

List of Figures ............................................. 163
List of Tables .............................................. 166
Bibliography ................................................ 169