

Contents

Abstract	i
Acknowledgments	iii
Contents	iv
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Context	1
1.2 Problem definition	3
1.3 Contributions	6
1.4 Structure of the Thesis	8
1.5 Publications	9
1.6 Artifacts	10
2 Background	13
2.1 Semantic Web and Linked Open Data	13
2.2 Social Media	16
2.3 Representation Learning	17
2.4 Text Representations	18
2.4.1 LSA	20
2.4.2 GloVe and Swivel	20
2.5 Graph Embeddings	22
2.5.1 General Graph Embedding Methods	22
2.5.2 RDF Embeddings	24
2.6 Conclusions	25
3 SocialLink Approach: Linking Knowledge Bases to Social Media Profiles	27
3.1 Introduction	27
3.2 Problem Definition	31
3.3 Approach Overview	33
3.3.1 Data Acquisition	33
3.3.2 Candidate Acquisition	36
3.3.3 Candidate Selection	37
3.4 Graph-based Embeddings	39
3.4.1 Social Graph Embeddings	40
3.4.2 RDF Graph Embeddings	42
3.5 The Embedding-Aware Candidate Selection Model	42

3.6	Evaluation	44
3.6.1	Experimental Setting	45
3.6.2	Overall System Evaluation	48
3.6.3	Candidate Acquisition Evaluation	48
3.6.4	Candidate Selection Evaluation	50
3.6.5	Evaluation by Entity and Profile Type	52
3.6.6	Error Analysis	54
3.7	On the Choice of Word Embeddings	56
3.7.1	Experimental Setting	58
3.7.2	Experimental Results	59
3.8	Related Work	59
3.9	Conclusions and Future Work	61
4	SocialLink Resource	63
4.1	Introduction	63
4.2	SocialLink Pipeline	65
4.2.1	Feature Coverage	66
4.2.2	Scoring and Selection Procedures	67
4.2.3	Populating the Resource	68
4.3	SocialLink Dataset	69
4.3.1	RDF Format	69
4.3.2	Dataset Statistics	71
4.3.3	Availability and Sustainability	72
4.4	Using SocialLink	74
4.4.1	DBpedia to Twitter: User Profiling	74
4.4.2	DBpedia to Twitter: Entity Linking	74
4.4.3	Twitter to DBpedia: Extracting FOAF Profiles and Type Prediction	75
4.4.4	Twitter to Wikidata: Referencing of Crowdsourced Knowledge	76
4.5	Conclusions and Future Work	76
5	Type Prediction Combining Linked Open Data and Social Media	79
5.1	Introduction	79
5.2	Problem Definition	82
5.3	Approach Overview	84
5.4	Ground Truth Acquisition from LOD	85
5.4.1	Methodology	85
5.4.2	DBpedia Type Prediction Tasks	87
5.5	Entity Representation with Social Features	88
5.6	Experiments	91
5.6.1	Experimental Setting	91
5.6.2	Experimental Results	93
5.6.3	Comparison to Wikipedia-based features	95
5.6.4	Dense Social Representations	96
5.7	Related Work	97
5.8	Conclusions	99

6	Concealing Interests of Passive Users in Social Media	101
6.1	Introduction	101
6.2	Related Work	104
6.3	Problem Definition	106
6.4	Interests Inference Pipeline	107
6.5	Concealing Approaches	108
6.6	Evaluation	110
6.6.1	Evaluation against Interest Inference Pipeline	110
6.6.2	Evaluation against Twitter’s Who To Follow	112
6.7	Conclusions and Future Work	113
7	Social Media Toolkit	115
7.1	System Description	115
7.1.1	System API	116
7.1.2	Configuration	118
7.2	MicroNeel: A Tool to Perform Named Entity Detection and Linking on Microposts	121
7.2.1	Background	122
7.2.2	Description of the System	123
7.2.3	Results	128
7.2.4	Discussion	129
7.3	Pokedem: an Automatic Social Media Management Application	129
7.3.1	Background	130
7.3.2	Description of the System	132
7.3.3	Results	134
7.3.4	Discussion	136
7.4	Conclusions	136
8	Conclusions	139
8.1	Summary of Contributions	139
8.2	Future Work	142
8.3	Privacy	143
	Bibliography	146