

Contents

1	Introduction	1
1.1	Motivation & Objectives	1
1.2	Stylometry	2
1.2.1	Authorship Attribution	3
1.2.2	Authorship Verification	4
1.2.3	Author Profiling	4
1.2.4	Authorship Linking and Author Clustering	5
1.3	Achievements	6
1.4	Organization of the Thesis	7
2	Authorship Analysis	9
2.1	Feature Selection	9
2.2	Distance Measures	11
2.3	Classifier Choices	12
2.4	Evaluation Methodology	14
2.5	Performance Measures	15
2.6	From Analyzing to Clustering	19
3	Presentation of the Publications	21
3.1	Distributed Language Representation for Authorship Attribution	23
3.2	A Simple and Efficient Algorithm for Authorship Verification	25
3.3	Distance Measures in Author Profiling	28
3.4	Evaluation of Text Representation Schemes and Distance Measures for Authorship Linking	30
3.5	Author Clustering Using Spatium	32
3.6	Author Clustering with an Adaptive Threshold	35
4	Conclusion	37
4.1	Summary of Contributions	37
4.2	Future Directions	38

A	Papers	41
A.1	Distributed Language Representation for Authorship Attribution . . .	41
A.2	A Simple and Efficient Algorithm for Authorship Verification	59
A.3	Distance Measures in Author Profiling	71
A.4	Evaluation of Text Representation Schemes and Distance Measures for Authorship Linking	89
A.5	Author Clustering Using Spatium	111
A.6	Author Clustering with an Adaptive Threshold	117
B	Publications	131
B.1	Journal Articles	131
B.2	Conference Proceedings	132
B.3	Evaluation Forums	133
C	Voronoi Diagrams of Distance Measures	135
	References	141