Authorship Attribution

Patrick Juola

Department of Mathematics and Computer Science
Duquesne University
Pittsburgh, PA 15282
USA
juola@mathcs.duq.edu

now
the essence of knowledge

Boston – Delft
## Contents

1 Introduction 1  
1.1 Why “Authorship Attribution”? 1  
1.2 Structure of the Review 2  

2 Background and History 5  
2.1 Problem Statement 5  
2.2 Theoretical Background 7  
2.3 Historical Overview 8  
2.4 Methodological Issues in Authorship Attribution 13  
2.5 What Would an Ideal Authorship Attribution System Look Like? 17  

3 Linguistic and Mathematical Background 21  
3.1 Mathematical Linguistics 21  
3.2 Information Theory 25  
3.3 Matrices and Vector Spaces 27  
3.4 A Theoretical Framework 29  

4 Linguistic Features 31  
4.1 Vocabulary as a Feature 31  
4.2 Vocabulary Properties as Features 33  
4.3 Syntactic Properties as Features 34  
4.4 Miscellaneous Features 35