

Progress To Date

I have made good progress thus far on my independent work for this Spring 2014 semester including gathering all of my data (Hundreds of Twitter Users and 1600 tweets per user) and writing about 9 pages of my final paper including my introduction, background, and related works sections.

Excerpt from introduction:

One mechanism aimed at ebbing the flow of online crime is Computational Stylometry. Stylometric analysis involves identifying and quantifying various linguistic features present in a written text so as to provide a computational profile of that body of writing. The power behind this technique comes from the fact that writing styles are very often unique enough to distinguish one from another. Therefore, computational stylometry makes the task of grouping anonymous texts together by author quite possible in many contexts. Once a group of texts is shown to be linked stylistically, if the author of one such text in the group is known, then the anonymity of all other texts linked to it becomes jeopardized.....

Layout of My Paper:

Introduction

Background

- What is Stylometry
- Feature Set
- Lexical
- Syntactic
- Application-Specific
- Analytic Models

Related Work

Methodology

- Data Retrieval

Plan Going Forward

This weekend I plan on completing the Methodology section of the paper. After talking over it with my advisor to make sure that there are no gaping wholes in my method, I will be able to begin the actual experimentation aspect of my project. I have already done thorough research to make sure that the open source JGAAP package will be able to be used to help with the actual authorship component piece. My plan going forward is to: totally figure out methodology, format data (removing @-replies and URLs), then work on automating the authorship attribution experiments using the JGAAP package.