

Text Analytics for Social Research

Stuart W. Shulman

University of Massachusetts Amherst
200 Hicks Way Amherst, MA 01003
stu@polsci.umass.edu

Abstract. This tutorial provides software training in “DiscoverText,” which is text analytic software developed by Professor Shulman. His work advances text mining and natural language processing research. The training links these worlds via straightforward and easy to understand explanations of software features that can be tailored for all experience levels and industries.

Keywords: software, text analytics, archiving, classification, metadata.

1 Introduction

DiscoverText, a new Web-based software application launched by Texifter, LLC, allows users to make sense of email archives, social media content, and other electronic document collections. Utilizing the “Graph API” feature of Facebook, or the public API for Twitter, users of DiscoverText can login using their credentials and begin archiving thousands of posts and comments on selected pages.

In addition to analyzing text from social media and a variety of other sources, the software is designed to improve standard research, government and business processes. Users can securely upload emails or project documents and quickly redact sensitive, confidential, classified, or potentially offensive information before circulating or posting it, thereby saving companies hours of monotonous work. The platform also include de-duplication and near-duplicate clustering features that improve the ability of rule writing agencies to sort through tens or hundreds of thousands of electronic public comments.

With DiscoverText it is also possible to crowd source data analysis in novel ways, leveraging peer relationships and Web-verifiable credentials. Ingesting hundreds of thousands of items from social media, email and electronic document repositories is easier than ever. Advanced social search leveraging metadata, networks, credentials and filters will change the way users interact with text data over time.

This innovative platform brings topic modeling, sentiment detection, and other information retrieval and natural language technologies into an active learning loop where user-created choices customize and improve our text processing algorithms.