

**Project Name**

Big Data on a Small(er) Campus: Use of Large-Scale Text Analysis by a Comprehensive Primarily Undergraduate Institution

**Principal Investigator** Brian Lowe

**Campus** Oneonta, State University College at

**Year of Project** 2012

**Tier** Tier Two

**Project Team**

- Gregory M. Fulkerson, SUNY Oneonta

**Overview Summary**

Provides access to tools and methods necessary to discern and critically dissect patterns and trends within "Big Data" as it emerges through social media.

**Outcomes Summary**

Has enabled students access to "big data" at the [Center for Computational Research](#) at UB.

**Project Abstract**

The ability to capture and amass huge, complex data sets has advanced rapidly in recent decades, as data generated by social networking, remote sensing and software logging, as well as the storage to handle these inputs, have expanded. Large volumes of data – on the order of petabytes (or  $10^{15}$  bytes) to zettabytes (or  $10^{21}$  bytes) – engender difficulties in the capture, storage, search, sharing, analysis, and visualization of meaningful information. This project will offer instructors across a wide variety of disciplines at primarily undergraduate institutions (PUIs) the tools and methods necessary to discern patterns and trends within "big data" as it emerges through social media.

The capacity to organize and critically dissect claims and information found within social media becomes increasingly important pedagogically for both social scientists and scholars within other fields, as the learning environment for undergraduates is increasingly dominated by social media. For example, The Digital Future Project reports that 60% of respondents have looked “for news online” and 43% have searched for product information online, suggesting that social media is a prime source for information, whether valid or not (Aboujaoude, 2011: 19). Recent scholarship on a variety of contemporary issues, including violent crime (Glassner, 2000), global warming and acid rain (Oreskes & Conway, 2010), and forms of surveillance (Lyon, 2007) suggests that mediated representations have dramatic impacts upon public perception of these concerns. This project will develop the skills and capabilities necessary to conduct similar analyses for claims disseminated within social media. It will benefit undergraduate instructors of courses that include the subject of discourse generated within social media, to promote critical analyses of these claims, and to conduct their own research by collecting and organizing social media data as they inform the construction of social problems and controversies (Best, 2007).

This project will create and subsequently provide instructors and students with readily deployable pedagogical and research strategies for grappling with “big data” as it relates to substantive course materials. Social media and internet sources have deeply penetrated the contemporary learning space; thus, undergraduate instructors must both address this reality and harness it as an educational and research tool. The capacity to coherently organize and analyze claims and information circulated through social media becomes increasingly important in the social sciences and humanities, as the topography of social problems and controversies is influenced by social media.

Phase One of this project will develop the theoretical and methodological skills and access to the technology (DiscoverText and other tools) to create sample universes of data from social media data, and to organize them as a prelude to analysis. They will then be integrated into existing courses, allowing students to participate in the capture, organization, and content analysis of mediated claims (regardless of discipline), which can then be juxtaposed with the current understanding of these claims within academic literature. This will allow undergraduates to develop both timely critical thinking and analytical skills, and an understanding of data management. The project’s specific objectives are to:

- Pilot DiscoverText and associated tools to SUNY Oneonta faculty and students
- Develop a model for integration of DiscoverText and associated tools into existing coursework in sociology and other departments
- Assess students’ ability to use the tools to understand social media data in a broader social science context

Phase Two (if funded by 2013 IITG) will disseminate strategies, findings, and the tools themselves to other SUNY comprehensive colleges to support an environment in which these questions can be investigated by a shared learning community.

## IMPLEMENTATION

DiscoverText is advanced online software for text analysis of large volumes of data; Gnip provides access to raw social media data. Access to these tools has previously been out of reach of the SUNY comprehensive colleges (PUIs), and academia in general, because of both price and an industry focus on business and government clients. This project will not only allow their deployment by SUNY Oneonta, but will develop strategies for the incorporation of these tools into the undergraduate curriculum by the other 12 comprehensive colleges within SUNY. To capitalize on the tools’ flexibility with regard to discipline, the project will be piloted in the following SUNY Oneonta courses:

Fall 2012

SOCL 390 Sr Seminar in Sociology

SOCL 209 Social Research Methods

POLS 200 Approaches to Politics

POLS 277 Immigration & Citizenship  
POLS 284 U.S. Foreign Policy

Spring 2013  
SOCL 242 Rural Sociology  
SOCL 244 Environmental Sociology  
POLS 296 Research Assistantship

In addition to learning to use the tools, students will be given one research question in common: they will conduct a content analysis of captured social media, concerning potential environmental risks posed by hydraulic fracturing and recent international controversies over the hunting of cetaceans (Burnett, 2012), and juxtapose those claims with peer-reviewed literature on the issues. Content analysis is an outstanding tool for learning about social science research, as it provides tangible, understandable data, and is used by both qualitative and quantitative researchers. Smaller-enrollment courses have been selected to facilitate the piloting of DiscoverText prior to use in larger-enrollment courses in Spring 2013. The Co-PIs have received preliminary agreement to pilot the project from additional social science faculty. The Summer 2013 period will be utilized to assess the project's successes and challenges and produce and codify these findings for dissemination to other SUNY campuses and in preparation for Phase Two in Fall 2013.

#### **Reports and Resources**

- [Media and press releases](#)

#### **Faculty Development**

- Organizational Models of Faculty Support

#### **Instructional Design**

- Student Learning Support