



Machine-Learning Text and Social Data Analytics



**Contract Holder**  
Contract GS35F002CA

## Federal Government Capabilities Statement

Texifter is a small business that hosts cloud computing solutions for gathering, searching, filtering, clustering, human coding, and machine classifying unstructured text and associated metadata. The flagship product DiscoverText provides text analytics capabilities tailored to the needs of federal rule writers, FOIA staff, social media listeners, survey researchers, or any enterprise feedback manager with a large amount of digitized text. Texifter provides self-serve search and retrieve access to every undeleted Tweet in history.

With DiscoverText, analysts work alone or collaboratively in a web-based text analytics network. Key capabilities include: data ingest and export via Application Programming Interface (API), advanced keyword search; metadata filters, duplicate detection; near duplicate clustering; topic modeling; interactive tag clouds; multi-coder synchronous and asynchronous annotation; shared memos; and RTF, CSV, PDF, and XML format reporting.

DiscoverText imports Federal Docket Management System (FDMS) public comment archives from Regulations.gov, as well as spreadsheets, large email collections, blog and wiki content, plain text files, HTML, and XML tagged data. Using free public APIs and Gnip-enabled premium PowerTrack capabilities, DiscoverText provides direct access to content from Facebook, Twitter, Tumblr, WordPress, Disqus, Google+ and other live social or RSS data feeds. Texifter builds custom API connections to private discussion forums and third party databases. DiscoverText connects directly to SurveyMonkey via an API for easy survey data transfer.

This platform brings topic modeling, sentiment detection, machine-learning, advanced measurement as well as other information retrieval and natural language technologies into an "ActiveLearning" loop. User-created choices improve the performance of our text classification algorithm over time. Texifter provides training and support services complimenting the enterprise license model, as well as a trained "CoderVerse" to code and annotate documents or manage projects in a timely and cost-efficient manner.

Our expertise in qualitative methods, text analysis, linguistics, statistics, and computer science allowed us to create a comprehensive solution. DiscoverText delivers benefits to both public and private sector users. Texifter has lowered the barriers to entry that prevent many people from utilizing powerful text analytic software. DiscoverText shortens the time it takes to create accurate, time-saving, custom machine text classifiers via our "ActiveLearning" method featuring a patent pending approach to "CoderRank" for enhanced machine-learning.

## Company Designations

SINs:	132-51 & 132-52
CAGE Code:	5T2L2
DUNS #:	832343086
Federal Tax ID:	27-0850393
NAICS:	541511 – Custom Computer Programming Services
SIC:	7371 – Computer Programming Services
PSC:	R540 – Study/Building Technology
FSC:	R – Professional, administrative, and management support services
NAGC:	3022717
SBA ID:	P1161031

## Key Personnel

- Dr. Stuart W. Shulman, Founder & CEO
- Mark J. Hoy, Chief Technical Adviser
- Keith Payne, Senior Architect

## Key Capabilities

- Multilingual Text Analytics & Machine-Learning
- Crowd Source Human Annotation & Validation
- Survey Analysis (SurveyMonkey Integration)
- Social Data Capture & Cleaning
  - PowerTrack Filtering by Gnip
  - Twitter, Facebook, Tumblr, WordPress, Disqus
  - Every Undeleted Tweet in History
  - Tencent Weibo (Chinese Microblog)
- Public Comment Analysis (FDMS/Regs.Gov)
- Document Redaction, FOIA & eDiscovery Tools
- Dirty Word Detection and Custom Redaction Lists
- Duplicate Detection & Near-Duplicate Clustering

## Past and Current Federal Government Customers

- Department of Agriculture
- Federal Communications Commission
- Fish & Wildlife Service
- National Labor Relations Board
- Comptroller of the Currency
- Small Business Administration Office of Advocacy
- National Oceanic and Atmospheric Administration

## Contact Information

Dr. Stuart Shulman

[stu@texifter.com](mailto:stu@texifter.com)

413-992-8513