

Humans and Machines Learning Together

Dr. Stuart Shulman ¹

¹ Texifter, LLC, USA; stu@texifter.com

Abstract. DiscoverText is multilingual-capable platform designed for collecting, cleaning up, and categorizing text data. It is free for academics. This highly interactive workshop teaches participants how to use DiscoverText to build custom machine classifiers for sifting Twitter data, free text, emails, survey responses, RSS feeds, or any other digitized text and metadata. Topics covered include how to:

- fetch Twitter datasets,
- construct precise or broad social data queries,
- clean data for relevance,
- respect the right to be forgotten,
- use Boolean search on raw data archives,
- filter on metadata or other project attributes,
- tabulate, explore, and set aside duplicates, cluster near-duplicates,
- crowd source human coding and other annotations by collaborators,
- measure inter-rater reliability,
- adjudicate coder disagreements to rank human coders, and
- quickly build word sense and topic disambiguation machine classifiers.

By the end of this workshop, participants will be able to present valid, legal, defensible Twitter research based on qualitative, quantitative, mixed, and machine-learning methods without having to learn software coding, Python, or R.

Keywords: Machine-learning, annotation science, qualitative research methods, collaborative text analytics, multilingual text mining.

Necessary resources: Laptop and internet access.

Presenter's biographical note:

Stuart Shulman is a methodologist with more than 30 years of experience putting words in categories, as well as an entrepreneur, recovering political science professor, U.S. Soccer National C licensed retired coach, CEO of Texifter, inventor of DiscoverText, and Taoist garlic grower.

Dr. Shulman's academic publications are here:

<https://www.researchgate.net/profile/Stuart-Shulman/research>

Academic mentions of DiscoverText are here:

<https://discovertext.com/mentions>

