Chapter 12
Real-Time Sentiment Analysis of Microblog Messages with the Maltego “Tweet Analyzer” Machine

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT
For social media to work as a “human sensor network” or a relevant source for “eventgraphing” for some fast-moving events, it is important to capture real-time and locational information. It may help to not only capture information from a particular social media platform but from across the Web. In such a context, Maltego Carbon 3.5.3/Chlorine 3.6.0’s Tweet Analyser “machine” (with AlchemyAPI built-in) and used in combination with other “transforms,” may serve the purpose—at least for initial and iterated sampling of the related messaging. This tool may be used to capture information from social media accounts, social media accounts, linked URLs, geolocational information, and other information of research value. Maltego is an open-access tool with a community version and a proprietary commercial version available by subscription. Maltego Chlorine’s Tweet Analyser has a built-in sentiment analysis feature.

INTRODUCTION
Whether explicit or implicit, part of the social media “bargain” that some social media companies have made with the general public is to not only ingest people’s data, but to share the public-facing information. Such sharing involves corporate citizenship and good public relations. It enables datamining research that is value-added to the collected data. These companies enable open-access to some of their data through application programming interfaces (API) instead of merely leveraging the data for advertising and in-company research purposes. Content that is publicly released by their respective creators (the social media users) is available to the masses, and the privately held information is available based on

DOI: 10.4018/978-1-5225-0846-5.ch012
authenticated access (usually those of the original content creator). For most social media platforms, access is enabled by registering with the target social media platform service (usually with email-verified accounts) and the use of access keys. Those using software applications to access social media platform databases often use proper OAUTH (an open standard used to enable resource owners to authorize third parties to access server resources, without revealing who is making the query). A majority of social media holdings are open-access, with very low costs of entry. (Command-line access through high-level programming languages is enabled through the Web and Internet; third-party software tools—many of them open-source and free—enable easier data access and data visualizations.)

Twitter, one of the world’s foremost microblogging sites, has long been a favorite with datamining researchers and the broad public. Twitter has enabled access via its application programming interfaces (APIs), enabling two broad types of data capture: the extraction of captured recent historical data (static) and also real-time (or near real-time) and continuous streaming data (dynamic). The static API enables the download of public streams (multi-user streams and topic streams), user streams (single users), and site streams (multiple stream extractions simultaneously, used for servers) (“The Streaming APIs,” 2015). The streaming and iterated API enables the near-real-time capture of emergent communications. This feature may be used as a perpetual or continuously-running dashboard.

While developers have tapped into Twitter’s APIs and scraped data (in Javascript Object Notation / JSON; comma separated values / .csv formats; hypertext markup language / HTML, and extensible markup language / XML), the general public itself has often had limited access—often through software tools, web interfaces, or limited applications. This lack of access is especially the case for continuous or dynamic data extractions. Also, most research require close human readings of the captured messaging and trained coding for sensemaking. Most researchers do not have the access to sentiment analysis tools or content network creation tools based on the extracted contents.

A new “machine” in Maltego Carbon 3.5.3 / Chlorine 3.6.0, though, enables near real-time topic-based microblogging extractions from Twitter and near real-time sentiment analysis (with the built-in AlchemyAPI for English language text, not the Unicode characters). In addition, the other affordances of this software tool—extracting data from the Surface Web—means that researchers that are interested in using this to capture a sense of the fast-changing collective mindset (as expressed on both Twitter and on the Surface Web) may be able to achieve a fairly broad and accurate sampling of unfolding events and/or socio-cultural phenomena. Researchers can reach out beyond a (somewhat porously) bounded system and into the larger Web and Internet ecosystems. The affordances of this “machine” (a designed macro) are tailor-made for using Twitter data for a “human sensor network” (the usage of on-ground data shared from social media accounts) and for “eventgraphing” (mapping out unfolding events). Knowing real-world sentiments enable individuals to both react and pro-act, such as to correct misperceptions, particularly when there is a level of broad-scale predictivity about the direction of public opinion based on a particular topic or issue (Nguyen, Wu, Chan, Peng, & Zhang, 2012, p. 1).

In the literature, both approaches are important use cases for human awareness of in-world events. This chapter introduces the Tweet Analyser™ “machine” with a few real-world examples. The data extractions and data visualizations for this chapter were conducted using the proprietary versions of Maltego Carbon 3.5.3 and Chlorine 3.6.0; there is also a community and open-source version of this tool but with a cap on the amount of nodes extracted (the community version tops out at a dozen vertices; the proprietary version enables extractions of over 10,000+). Even with 10,000 nodes, that may not actually be at-scale of what is available.