

Resource Options for Text Analysis

Much of today's interesting data come from social media platforms such as Twitter, Facebook, Google+, and Instagram. You may want to explore the public online conversation about a particular topic and gather that data for analysis. There are several tools that you can use or purchase to gather this data. Table 1 provides a few examples of different software and tools that can be used to access and/or analyze text (non-numeric) data.

In general, open source tools that require programming backgrounds are more powerful, but necessitate a skilled programmer who understands and can implement data analysis and social media API calls. Proprietary tools, though more expensive, already have the data agreements with each of the social media platforms and have simple, user-friendly interfaces from which one can query and visualize data.

Table 1. Software/Tools: Tradeoff between ease of use and price

	Software Packages(Open)	Open Source Languages	Software As A Service (SAAS) Options
Software/Tool (examples)	Voyant	Python, R	Sysomos, Synthesio, Crimson Hexagon
Software Price	Open Source (Free)	Open Source (Free)	~\$25,000 - \$50,000+/yr ¹
Skills Needed for Use	Basic	Intermediate/Advanced	Basic/Intermediate
Environment	Web Browser	PC/Cloud	Web Browser
Data Access	Provide your own ²	Limited Access using API calls ²	Built-in (Twitter, News, Blogs, Forums)

¹Prices will vary depending on the length of your license, number of users you have, and suite of tools you want to purchase. See, e.g., <http://www.pcmag.com/article2/0,2817,2494517,00.asp>; <http://www.pcmag.com/article2/0,2817,2494521,00.asp>; <http://www.pcmag.com/review/342729/sysomos>

² Open source tools can be incredibly useful in gathering and analyzing data. However, access to social data such as Twitter can be limited. Table 2 displays different mechanisms to gather Twitter data.

Table 2. Twitter Data Access Types

	Twitter REST API	Twitter Streaming API	Twitter Firehose
Tweet Limits	Depends	Depends on current tweet matching your criteria – sends only a sample of current activity	None
Call Limits	150-350 per hour	This is a push of Tweets from Twitter to you in near real-time	Unclear
Access Level	User/Developer Token needed	User/Developer Token needed	High – Agreement needed
Price	Free	Free	\$\$\$ Depends on partnership

Example of Staff/Skills needed³:

- Skill/Staff
 - Intermediate/advanced programmers 2-5 yrs experience using one of the following languages, or similar
 - Python, R, Java, Julia
- Scope of Work
 - Library setup: 2-4 hours (a fair amount of trouble shooting and understanding of the computing environment is sometimes needed)
 - Initial data scrape: 16-30 hours (depends on skill level and familiarity of programmer)
 - Data analysis: 40-120 hours (40 hours for setting up the data tables and doing basic descriptive analysis; 120 for more advanced natural language processing)
 - Data visualization/Report output: 40 hours (for initial setup) and maybe 4-8 hours for each subsequent report (assuming a periodic report with the same information, graphs, and structure)

³ The staff/skills needed are rough estimates based on basic analysis and visualizations. The staffing needed will vary greatly depending on the complexity of the project, analysis, data, and objectives.