

MassMine: Your Access To Data

Nicholas M Van Horn¹ and Aaron Beveridge²

¹Capital University

²University of Florida

15 August 2016

Paper DOI: <http://dx.doi.org/10.21105/joss.00050>

Software Repository: <https://github.com/n3mo/massmine>

Software Archive: <http://dx.doi.org/10.5281/zenodo.193078>

Summary

MassMine (Van Horn and Beveridge 2016b) is a research tool designed to simplify the collection and use of data from social media and news websites, such as Twitter, Tumblr, Google Trends, Wikipedia, as well as web URL scraping. MassMine runs on personal computers and in “headless” environments (i.e., servers/clusters). In addition to searching, downloading, parsing, and storing data, MassMine handles any authentication credentials on behalf of the user, allowing users to focus on their research rather than programming code to negotiate with online APIs. Further, MassMine automatically monitors and respects rate limits commonly imposed by online services, ensuring that a data collection project does not violate usage rights set in terms of service agreements. Custom data requests are made simple, limited only by the restrictions of the data providers, allowing for flexible deployment for individualized research needs. All data collected is managed automatically for the user. This means that all tweets, trends, posts, etc. related to users’ research questions are saved permanently on local hard drives. Built to accommodate short and long-lived data collection projects, MassMine makes it easy to create a database of social media content spanning days, months, or years. Visit the official online website (Van Horn and Beveridge 2016a) for downloads, as well as installation and usage instructions.

References

- Van Horn, Nicholas M., and Aaron Beveridge. 2016a. “MassMine Official Website.” <http://www.massmine.org>.
———. 2016b. “MassMine: Social Media and Online Web Content Mining Software.” <https://github.com/n3mo/massmine>.