



Engineering

[Back](#)

Crowdsourced data analysis with Clockwork Raven

By

[Chris Aniszczyk](#)

Thursday, 16 August 2012

Today, we're excited to open source [Clockwork Raven](http://twitter.github.com/clockworkraven/) (<http://twitter.github.com/clockworkraven/>), a web application that allows users to easily submit data to [Mechanical Turk](http://en.wikipedia.org/wiki/Amazon_Mechanical_Turk) (http://en.wikipedia.org/wiki/Amazon_Mechanical_Turk) for manual review and then analyze that data. Clockwork Raven steps in to do what algorithms cannot: it sends your data analysis tasks to real people and gets fast, cheap and accurate results. We use Clockwork Raven to gather tens of thousands of judgments from Mechanical Turk users every week.

Motivation

We're huge fans of human evaluation at Twitter and how it can aid data analysis. In the past, we've used systems like Mechanical Turk and CrowdFlower, as well as an internal system where we train dedicated reviewers and have them come in to our offices. However, as we scale up our usage of human evaluation, we needed a better system. This is why we built Clockwork Raven and designed it with several important goals in mind:

- Requires little technical skill to use: The current Mechanical Turk web interface requires knowledge of HTML to do anything beyond very basic tasks.
- Uniquely suited for our needs: Many of our evaluations require us to embed tweets and timelines in the task. We wanted to create reusable components that would allow us to easily add these widgets to our tasks.
- Scalable: Manually training reviews doesn't scale as well as a system that crowd sources the work through Mechanical Turk.
- Reliable: We wanted controls over who's allowed to complete our evaluations, so we can ensure we're getting top-notch results.
- Low barrier of entry: We wanted a tool that everyone in the company could use to launch evaluations.
- Integrated analysis: We wanted a tool that would analyze the data we gather, in addition to provide the option to export a JSON or CSV to import into tools like R or a simple spreadsheet.

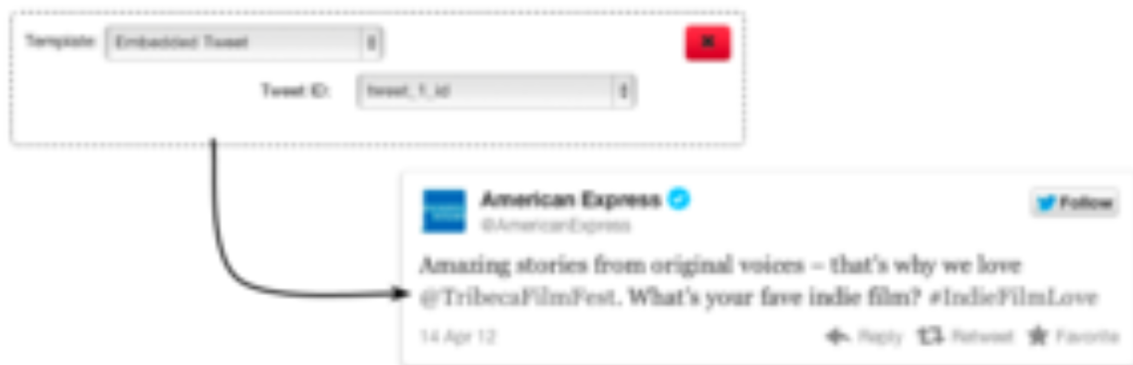
Features

In Clockwork Raven, you create an evaluation by submitting a table of data (CSV or JSON). Each row of this table corresponds to a task that a human will complete. We build a template for the tasks in the Template Builder, then submit them to Mechanical Turk and Clockwork Raven tracks how many responses we've gotten. Once all the tasks are complete, we can import the results into Clockwork Raven where they're presented in a configurable bar chart and can be exported to a number of data formats.

Here's the features we've built into Clockwork Raven to address the goals above:

- Clockwork Raven has a simple drag-and-drop builder not unlike the form builder in Google Docs. We can create headers and text sections, add multiple-choice and free-response questions, and insert data from a column in the uploaded data.

- The template builder has pre-built components for common items we need to put in our evaluations, like users and Tweets. It's easy to build new components, so you can design your own. In the template builder, we can pass parameters (like the identifier of the Tweet we're embedding) into the component. Here's how we insert a tweet:



- Clockwork Raven submits jobs to Mechanical Turk. We can get back thousands of judgements in an hour or less. And because Mechanical Turk workers come from all over the world, we get results whenever we want them.
- Clockwork Raven allows you to manage a list of Trusted Workers. We've found that having a hand-picked list of workers is the best way to get great results. We can expand our pool by opening up our tasks beyond our hand-picked set and choosing workers who are doing a great job with our tasks.
- Clockwork Raven authenticates against any LDAP directory (or you can manage user accounts manually). That means that you can give a particular LDAP group at your organization access to Clockwork Raven, and they can log in with their own username and password. No shared accounts, and full accountability for who's spending what. You can also give "unprivileged" access to some users, allowing them to try Clockwork Raven out and submit evaluations to the Mechanical Turk sandbox (which is free), but not allowing them to submit tasks that cost money without getting approval.

- Timeline SBS A/A Equalized across User Geo

Results Summary

Average value of 'What's new?' in the toolbar over time

Chart Data:

Month	Average Value
Jul	-0.02
Oct	-0.03
Dec	0.02
Jan	-0.08

Filter By

What's new? is most relevant to the toolbar user?

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

What's new? is most relevant to the toolbar user?

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

Control Trust has Most Relevant Engagement?

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

Test Trust has Most Relevant Engagement?

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

User Type

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

Display:

 - ☐ Overall Value
 - ☐ Trust it
 - ☐ Fear the worst
 - ☐ Trust it
 - ☐ Fear is irrelevant

We're actively developing Clockwork Raven and improving it over time. Our target for the next release is a comprehensive REST API that works with JSON (possibly Thrift as well). We're hoping this will allow us to build Clockwork Raven into our workflows, as well as enable its use for real-time human evaluation. We're also working on better ways of managing workers, by automatically managing the group of trusted workers through qualification tasks and automated analysis of untrusted users' work.

Acknowledgements

Follow @clockworkraven (<https://twitter.com/clockworkraven>) on Twitter to stay in touch!

Share:

