

# Picking the President: Twindex, Twitris Track Social Media Electorate

 [dataversity.net/picking-the-president-twindex-twitris-track-social-media-electorate/](http://dataversity.net/picking-the-president-twindex-twitris-track-social-media-electorate/)

8/3/2012

The U.S. is a mere three months away from choosing who will be its leader for the next four years. The semantic web is a supporting player in the action.

This week, of course, saw the debut of the [Twitter Political Index](#) (Twindex), a joint effort between Twitter, Topsy, and the Mellman Group and NorthStar Opinion Research polling groups. Since the Semantic Web Blog last spoke with Topsy execs [here](#), the company has refined its sentiment analysis to the point where it could be released for the Twindex. The sentiment analytics engine ingests hundreds of millions of English-language tweets a day and computes sentiment for all terms in Twitter, though that's not publicly available yet.

In its Twindex incarnation, Topsy aggregates the underlying sentiment score minute by minute, and then that is rolled up into an hourly and daily score for each candidate, says Rishab Aiyer Ghosh, co-founder and chief scientist at Topsy Labs. Behind the scenes, "that score is normalized so that it is on 0 to 100 scale comparing to all the other terms people talk about," he says, which is important for keeping perspective on the candidates in context relative to whatever else may be on the mind of the collective social media conscience. It also is weighted to include the scores of the previous two days before its publication at the end of the day, and smoothed out so that it doesn't jump around in helter-skelter fashion.

The sentiment algorithm, Ghosh says, was designed specifically for dealing with short-form content, and in sample testing Topsy has found that the algorithms agree with human perceptions of positive or negative sentiment 90 percent of the time. Additionally, it has done work in the background to narrow things down based on geography, so it's pretty confident that the scores it comes up with are based on representative U.S. tweets. "We worked with the polling firms a lot to validate it, too, to see how the sentiment algorithm should be used," adds Ghosh. "This is definitely a skeptical audience, or one you want to be skeptical, and they were pretty positive about it," he says.

Though not to be published as part of the Twindex, Topsy itself will publish analysis of sentiment on the candidates for individual groups of states, such as swing states. "We will publish more detail based on geography," he says. "Only 1 percent of tweets actually have geo-tagging-enabled. But we have built technologies to expand from that. We use machine learning to process billions of tweets to come up with inferences of where the location is, and that has been pretty accurate." The Twindex site will link to that.

This approach has its advantages and disadvantages compared to traditional opinion polling. Questions and audience can't be controlled, for example. On the other hand, you do get insight into what hundreds and hundreds of thousands of people who are just out there expressing themselves think in an instant. "You can get an opinion right now and that is valuable information for everyone trying to understand what people think." Ghosh says.

## From Twindex to Twitris (And Wolfram Alpha Too)

That's not the only semantic way to see how we're feeling about the 2012 election, though. Wright State University and the Kno.e.sis Ohio Center of Excellence in Knowledge-enabled Computing, headed up by director and LexisNexis Ohio Eminent Scholar Dr. Amit P. Sheth, has as a piece of its semantic social web application project Twitris the [Election 2012 360° Social Media Analysis](#) site. Sheth, who discussed how Twitris handles social media analysis in our story [here](#), brings similar Search & Explore, Sentiment and Network (as in influential users and connectivity by topic) analysis capabilities to the site as it did for its Occupy Wall Street and India Against Corruption Twitris-based sites. Sheth told us previously that, for example, a candidate's team could use the Network Analysis

feature to see what influential users are talking positively about their candidate, perhaps to help target donations, or the Browse by Location tag to see what issues are being discussed in a certain area, to see if that information can be parlayed into speeches the candidate will make in that locale.

Sheth, [posting a Google Plus](#) message about the Election 2012 page, said this Twitris site is the way to follow through on the [comment](#) reportedly made by Adam Sharp, head of government, news and social innovation at Twitter, about how the Twitter Political Index could be used: “When the Twitter Political Index is giving a different indication than the polls as to where the winds of the electorate are shifting, that is a signal to perhaps dig deeper and gain a better understanding for the complexities of voting behavior.” Not only can Twitris Election 2012 let you do that now, Sheth said, with a nod to its Sentiment tab, and “soon you will be able to associate correlation between real-world events with its analysis here,” he noted. Click on Sentiment to currently view weekly charts of candidate sentiment by domestic, economic, international and social issues, the candidates themselves, or their parties.

And, if you just can’t get enough of the intersection between semantic technology and elections, over at the blog of the Wolfram Alpha knowledge engine – one of the sources for Siri that can provide results based on its structured data – there was a recent discussion of how it can help “to provide some useful context and analysis, particularly when it comes to understanding past election outcomes or predicting this year’s results.”

Writing for the [blog](#), C. Alan Joyce explains how users can leverage Wolfram Alpha to ask questions about the impact of various demographic groups on presidential election races. “You can use Wolfram|Alpha to try to understand where the [senior vote](#) might have the most impact in Florida, for example. Or you might look closer at the specific origin of the [Hispanic population in Florida](#)—increasingly split between Hispanics of Cuban origin ([who have traditionally voted Republican](#)) and other groups, such as Puerto Ricans, who tend to favor Democratic candidates,” he says.