

Twitter tweets chart the social whirl

November 8 2011

Data mining Twitter "tweets" may produce a gold mine for two University of Cincinnati computer science students.

William Clifton and Alex Padgett have developed a web-based application called The Tweetographer that allows users to learn about events in their cities or neighborhoods. The app works by collecting tweets sent by [large numbers](#) of Twitter users and extracting information about events – parties, concerts, games, etc. – happening nearby. It's like a real-time events guide.

The Tweetographer was the senior project for the pair, who are graduating during the 2011-12 academic year, Padgett in December and Clifton in June.

"We wanted to explore [data mining](#), which is an important area of research in Computer Science, in the context of social media," Padgett said. "Although the concept will work with many social media platforms, Twitter was the most accessible. Everything is out there in public domain, a giant pool of untapped data, tagged with latitude and longitude. It's very precise and lends itself to so many uses."

That broad utility created some difficulty for the developers as they tried to formulate a focused project.

"We realized that we could do all sorts of things with this data. We could add all sorts of functions, but we worked really hard to avoid 'feature creep' and decided to focus on events," Clifton said.

The Tweetographer, in practice, answers a common question for socially active people: "What's happening?" Since people who use Twitter often tweet about where they are going or what they want to do, The Tweetographer answers that question by listening in to the chatter. A user can get a sense of not only what is going on, but how popular various events are.

The application is so effective that it was initially overwhelmed the volume of data streaming in through millions of tweets in some large cities.

"Eventually we were able to come up with a solution for this with a kind of queuing system that let us handle a stream of that magnitude," Clifton said.

Another obstacle was making sense of all the available data. Although Twitter offers upwards of 140 million [tweets](#) a day, they are not posted in a uniform format.

"So many people type in their own shorthand," Padgett said.

The solution, according to Clifton, was to create a "thesaurus" of multiple [Twitter](#) synonyms.

"Do you know how many ways people type 'Tuesday'?" Clifton said.

All of the technical obstacles needed to be overcome on a tight deadline – just six months from assignment to presentation.

"If we had a couple of years, we could come up with something a lot more sophisticated," Padgett said. "Everyone is their own worst critic, and we had very high standards. We wanted to show an elegant, simple solution."

The Tweetographer got an enthusiastic reception at its unveiling.

"It blew people's minds," Clifton said. "One skeptic, in particular, wanted to test us. He said, 'If I tweet right now, it will show up,' and we said yes. He tweeted, and it popped up onscreen right away."

The future of The Tweetographer is yet to be written. Padgett and Clifton are making plans beyond graduation, yet still actively working on improving and evolving this project. Clifton thinks the "engine" developed for The Tweetographer has other useful applications, such as predicting election outcomes, or compiling product reviews.

"So much is out there," Clifton said.

Provided by University of Cincinnati

Citation: Twitter tweets chart the social whirl (2011, November 8) retrieved 14 May 2024 from <https://phys.org/news/2011-11-twitter-tweets-social.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.