

Semantic App Helps Researchers Understand Prescription Drug Abuse

 dataversity.net/semantic-app-helps-researchers-understand-prescription-drug-abuse/

6/11/2012

There's been a lot of attention given to the issue of prescription drug abuse, in the wake of violent crimes such as one last year that left four people dead in a pharmacy shooting in Suffolk County, New York. A recent study from the [Workers Compensation Research Institute](#) also shows that prescription drug abuse is the fastest growing drug problem in the United States, with over fifteen thousand people dying last year from an overdose. And, the U.S. Senate in late May approved an amendment to reclassify drugs that contain hydrocodone, a highly-addictive substance found in Vicodin and Lortabas, among other drugs, as Schedule II substances, while giving law enforcement more tools to monitor distribution of such drugs and also decreasing access to them for non-medical purposes.

What, you may ask, does any of this have to do with semantic technologies? Dr. Amit P. Sheth, Wright State University [Kno.e.sis Ohio Center of Excellence in Knowledge-enabled Computing](#) director and LexisNexis Ohio Eminent Scholar, and Dr. Raminta Daniulaityte of the school's [Center for Interventions, Treatment and Addictions Research \(CITAR\)](#), have a ready answer : PREDOSE, an application for understanding pain-killer drug abuse through the semantic analysis of social media conversations. More specifically, it's automated data collection and analysis tools to process web-based data to determine the knowledge, attitudes, and behaviors of addicts, related to buprenorphine, OxyContin and other pharmaceutical opioids. It's a [National Institutes of Health \(NIH\)](#)-funded project created by a partnership between Kno.e.sis and the CITAR.

In its role providing substance-abuse related services, academic research, and services research, CITAR engages in live interview projects and population surveys with subjects. But such approaches are costly and have a time lag to coordinate information and extract results, subjects might not fully disclose information, and research questions also can unintentionally have a bias. On the other hand, on "social media [people are] uninhibited about what they want to share," says Sheth, who also leads the [Twitris](#) social media analysis project (covered [here](#)). They represent a potentially valuable source of anonymous information on what people are doing with painkillers and other drugs that could be helpful in driving policies or other efforts for curbing abuse. Of course, on Twitter or sites like [Bluelight](#), which provides a forum for open information and discussion board about ecstasy and other drugs, people aren't necessarily adhering to proper language structure, and they're using a lot of abbreviations and slang names for drugs (oxy and multiple variants for OxyContin, for instance), too.

"It's very challenging text to analyze," says Sheth. The combination of domain knowledge at CITAR and semantic expertise at Kno.e.sis is a strong one-two punch to start to solve the issue. Work on PREDOSE, which has a three-stage architecture, has been underway for about a year now. The first part, which has largely been tackled, is collecting the user-generated data from social media sources; the project also involves creating semantic web data to correspond to the user-generated data and providing analysis tools to researchers. The work has included the development of the PREDOSE Drug Abuse Ontology (DAO) and the PREDOSE Annotator, to help capture and identify entities and represent the semantics of colloquial expressions in user-generated commentary.

"Unless you can establish a map between a drug and all its synonyms, you may miss some important posts," says Delroy Cameron, the key PhD student involved in the project. By establishing the mapping in content from the slang term to the standard name of the drug, the tool can facilitate analysis of the data by capturing additional relevant posts. As an example, Oxycontin also has a slang name of hillbilly heroin.

PREDOSE also already enables gauging sentiment, "to help us understand and automatically to extract information about substances people favor," says Daniulaityte. "In the case of a new drug, for example, we can very quickly

update how users are reacting to it or their opinions.”

With its annotation capabilities for entity extraction, sentiment and mood in hand, focus will be on doing more work to automatically extract relationships among entities. Already the team has made a discovery around loperamide, an over-the-counter drug used to control diarrhea, which is also known as Imodium. Having those terms mapped to each other was an important step in making it possible for the team’s analysis of extra-medical use of the term to discover that, whatever the name, it was being used to control opiate withdrawal symptoms from drugs like Oxycontin – something that had not been previously reported in the epidemiological literature, according to the project team. Other relationships to be explored could be around the methods of use of a drug, or side effects, or prices, notes Daniulaityte.

“The technology perspective is that we have done more than half the work, but ultimately our project will allow scientists to ask more complex questions,” says Sheth. For example, from all web posts with a mention of Oxycontin and also of withdrawal, what is the most common sentiment among users? Or, on the heels of the 2010 FDA-mandated Oxycontin reformulation to reduce its use for getting high, they can ask the tool to display the change in overall positive and negative mood/sentiment about that, to gauge if the policy had its intended effect and whether it might be something to extend to other prescription drugs. “That part is not done but the core technique of being able to collect all the data, annotate it and spot things for a particular purpose is there,” he says.

The demographic scale and real-time capability the tool enables potentially can be useful for helping inform policy decisions. “It adds another perspective to epidemiological data,” says Daniulaityte. “All policy decisions should be data-driven. And especially valuable about this type of research is that it provides a very rapid assessment of situations,” which is critical when drug abuse patterns change as rapidly as they do.

Down the road there also is more potential for active engagement, borrowing on some Twitris concepts, for example, to provide people with information on where to get help if they really want it, or here are some statistics about long-term abuse via a targeted message. “We would hope to make this tool public and also propose a larger follow-on that would allow other epidemiologists working in this area to also benefit from the web-based tool,” Sheth says.