



Digging the Connections

Situated in the industrial steel and coal mining area of eastern Pennsylvania is the City of Bethlehem. During the fall of 2005, the city's police department conducted an investigation of graffiti that began popping up in various locations throughout the city. As part of the investigation, department personnel became themselves miners—not miners of coal, but rather miners of data. Their pick and shovel—a data-mining software program recently developed by researchers at Pennsylvania's Lehigh University and in beta testing by the department.

In the investigation the software was able to show links between the words, paint colors, and locations of the graffiti. The suspect responsible was identified along with his gang affiliation, and the investigation led to multiple arrests on a varied group of charges, including drug violations.

Leveraging funding from the Office of Justice Programs' National Institute of Justice and the National Science Foundation, Lehigh's Dr. William Pottenger and his team worked with the Bethlehem police to create the Bethlehem Police Department Information Extraction System. The system is a datamining tool that delves into unstructured narrative text and pulls out links between names, birthdates, addresses, vehicles, and the like, helping investigators quickly find any similarities between crimes.

When Pottenger originally came to the department with his proposal, Randy Miller, Bethlehem's deputy commissioner of police, told the professor that the department's historical data existed in a variety of formats, ranging from typewritten reports to microfilm to WordPerfect® and Microsoft® Word®.

"The problem was, we didn't have the funding or the time to update data from one system to the next," Miller says. "For example, our WordPerfect files still existed on an old server, but we didn't have any software that could access them. Using sophisticated data mining techniques and algorithms, the Lehigh team is now able to take the old records management systems, pick out key words, and make the information accessible.

"This is a common problem faced by police departments. The old information is still good, but it is lost. Fortunately for us, ours is not lost any longer. Now we might find that a string of robberies taking place right now shows the same modus operandi as a string that occurred 10 years ago. Data mining can find the link, and our investigators might learn that the man who committed the previous robberies was just released after serving 10 years in the State prison. They now have a solid suspect," Miller says.

Another example of how the data-mining software could help investigators, explains Miller, is searching records of cold homicide cases for the word "blood." Finding that word in a 20-year-old case report could indicate the possibility of uncovering DNA evidence that might generate new leads.

"We're getting data and relationships we thought were long lost," he says. "Because of this system, we can view and extract important pieces of data. It fills in a gap that I think all police departments have." He adds, "Investigators get promoted, retire, and move on, and the experience and knowledge about past cases moves on with them, but the data remains. Now, with data mining, it won't be forgotten."

Lehigh's Pottenger, who has been working in data-mining research for several years, became interested in helping law enforcement when he realized the amount of underutilized data collected by most police departments.

"There is a lot of power that can be found in this type of resource," Pottenger says, citing as an example a nationwide methamphetamine bust engineered by the U.S. Department of Justice that used manual collection to link records across the country. "Our goal is to make it possible for law enforcement to do this easily and automatically. Anything that was ever recorded, even if it was just a comment on a form, or even a receipt, could contain valuable information that might link cases."

Pottenger notes that although there are commercial products with similar data-mining capabilities, their licenses tend to be too expensive for many police departments to afford. The project focused on coming up with ways to reduce the “knowledge engineering” cost so as to produce technology that could be marketed for less money. “We’re in discussions now about licensing it,” Pottenger says. “We’ve had interest expressed by two or three different law enforcement projects.”

In developing the software, which runs on an ordinary personal computer, the Lehigh team consulted with several law enforcement personnel to ensure that the final product is easy to use. The system offers three different search modes: a Google™ type string search interface, a standard form with drop-down menus, and a new-incident search that compares results against the existing database.

With the string interface, Pottenger says, users simply type in the words they want to find and can also qualify them by using a field value such as “Age: 15.” The form with drop-down menus allows users to choose terms from one or more categories—for example, the term “father” from the “Relationship” menu. The new-incident search asks users to choose the data they want to match, such as weapon type. The goal of each search mode is to provide a list of closely matching previous incidents and a short list of suspects.

**The National Law Enforcement and
Corrections Technology Center System**
Your Technology Partner
www.justnet.org
800-248-2742

In a preliminary evaluation of the system, commonly occurring crimes such as theft, malicious mischief, and burglary were used to look for links among modus operandi. Pottenger and his team plan a further evaluation in Florida that will test the system in conjunction with the State’s FINDER system, which tracks pawn transactions. Pottenger plans to mine written narrative police reports for data on stolen items and match these data against the pawn transaction database.

For more information about the Bethlehem Police Department Information Extraction system, contact William Ford at the National Institute of Justice, 202-353-9768 or William.Ford@usdoj.gov; or Dr. William M. Pottenger, 610-758-3454 or DrWMPottenger@gmail.com.



This article was reprinted from the Summer 2006 edition of *TechBeat*, the award-winning quarterly newsmagazine of the National Law Enforcement and Corrections Technology Center system, a program of the National Institute of Justice under Cooperative Agreement #2005-MU-CX-K077, awarded by the U.S. Department of Justice.

Analyses of test results do not represent product approval or endorsement by the National Institute of Justice, U.S. Department of Justice; the National Institute of Standards and Technology, U.S. Department of Commerce; or Lockheed Martin Aspen Systems Corporation. Points of view or opinions contained within this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

The National Institute of Justice is a component of the Office of Justice Programs, which also includes the Bureau of Justice Assistance, Bureau of Justice Statistics, Office of Juvenile Justice and Delinquency Prevention, and Office for Victims of Crime.