The New York Times
nytimes.com

October 4, 2006

Software Being Developed to Monitor Opinions of U.S.

By ERIC LIPTON

WASHINGTON, Oct. 3 — A consortium of major universities, using Homeland Security Department money, is developing software that would let the government monitor negative opinions of the United States or its leaders in newspapers and other publications overseas.

Such a "sentiment analysis" is intended to identify potential threats to the nation, security officials said.

Researchers at institutions including Cornell, the University of Pittsburgh and the University of Utah intend to test the system on hundreds of articles published in 2001 and 2002 on topics like President Bush's use of the term "axis of evil," the handling of detainees at Guantánamo Bay, the debate over global warming and the coup attempt against President Hugo Chávez of Venezuela.

A $2.4 million grant will finance the research over three years.

American officials have long relied on newspapers and other news sources to track events and opinions here and abroad, a goal that has included the routine translation of articles from many foreign publications and news services.

The new software would allow much more rapid and comprehensive monitoring of the global news media, as the Homeland Security Department and, perhaps, intelligence agencies look "to identify common patterns from numerous sources of information which might be indicative of potential threats to the nation," a statement by the department said.

It could take several years for such a monitoring system to be in place, said Joe Kielman, coordinator of the research effort. The monitoring would not extend to United States news, Mr. Kielman said.

"We want to understand the rhetoric that is being published and how intense it is, such as the difference between dislike and excoriate," he said.

Even the basic research has raised concern among journalism advocates and privacy groups, as well as representatives of the foreign news media.

"It is just creepy and Orwellian," said Lucy Dalglish, a lawyer and former editor who is executive director of the Reporters Committee for Freedom of the Press.

Andrei Sitov, Washington bureau chief of the Itar-Tass news agency of Russia, said he hoped that the objective did not go beyond simply identifying threats to efforts to stifle criticism about an American president or administration.

"This is what makes your country great, the open society where people can criticize their own government," Mr. Sitov said.

The researchers, using an grant provided by a research group once affiliated with the Central Intelligence Agency, have complied a database of hundreds of articles that it is being used to train a computer to recognize, rank and interpret statements.

The software would need to be able to distinguish between statements like "this spaghetti is good" and "this spaghetti is not very good — it's excellent," said Claire T. Cardie, a professor of computer science at Cornell.

Professor Cardie ranked the second statement as a more intense positive opinion than the first.

The articles in the database include work from many American newspapers and news wire services, including The Miami Herald and The New York Times, as well as foreign sources like Agence France-Presse and The Dawn, a newspaper in Pakistan.

One article discusses how a rabid fox bit a grazing cow in Romania, hardly a threat to the United States. Another item, an editorial in response to Mr. Bush's use in 2002 of "axis of evil" to describe Iraq, Iran and North Korea, said: "The U.S. is the first nation to have developed nuclear weapons. Moreover, the U.S. is the first and only nation ever to deploy such weapons."

The approach, called natural language processing, has been under development for decades. It is widely used to summarize basic facts in a text or to create abridged versions of articles.

But interpreting and rating expressions of opinion, without making too many errors, has been much more challenging, said Professor Cardie and Janyce M. Wiebe, an associate professor of computer science at the University of Pittsburgh. Their system would include a confidence rating for each "opinion" that it evaluates and would allow an official to refer quickly to the actual text that the computer indicates contains an intense anti-American statement.

Ultimately, the government could in a semiautomated way track a statement by specific individuals abroad or track reports by particular foreign news outlets or journalists, rating comments about American policies or officials.

Marc Rotenberg, executive director of the Electronic Privacy Information Center in Washington, said the effort recalled the aborted 2002 push by a Defense Department agency to develop a tracking system called Total Information Awareness that was intended to detect terrorists by analyzing troves of information.

"That is really chilling," Mr. Rotenberg said. "And it seems far afield from the mission of homeland security."

Federal law prohibits the Homeland Security Department or other intelligence agencies from building such a database on American citizens, and no effort would be made to do that, a spokesman for the department, Christopher Kelly, said. But there would be no such restrictions on using foreign news media, Mr. Kelly said.

Mr. Kielman, the project coordinator, said questions on using the software were premature because the department was just now financing the basic research necessary to set up an operating system.

Professors Cardie and Wiebe said they understood that there were legitimate questions about the ultimate use of their software.

"There has to be guidelines and restrictions on the use of this kind of technology by the government," Professor Wiebe said. "But it doesn't mean it is not useful. It can just as easily help the government understand what is going on in places around the world."