



Imagine searching for a concert and pulling up the usual web pages – plus untagged Flickr photographs, Twitter remarks, YouTube videos, and Facebook comments. Or asking when the band will perform again and getting back a table of dates and locations.

Associate Professor Luis Gravano is supercharging search engines to conduct exactly those types of searches. Often, that means tapping the chaos of social media. “It’s not so much about just returning a list of individual Web pages as it is about combining and making sense of all information on the Web to increase the effectiveness of a search,” he said.

For example, many online photos are tagged to refer to specific events. Others have time and GPS data that coincide with the time and location of an event. Sometimes photos are forwarded or linked to other people who have commented about an event.

“We analyze these tags, comments, and links, and automatically cluster them to correspond to real-world events,” Gravano said. His team has already shown that it can aggregate such information. It is now probing how to fit the data together to develop more powerful searches.

“If there is a concert or political demonstration, people take pictures, tweet, and form groups around these activities. We want to capture and associate this content with real-world events automatically. We’ll return results that correspond to a specific event at a certain time on a particular street in New York City.”

Gravano also wants to improve our ability to extract structured information, such as tables, from the Internet. Today, he explained, anyone who wants to analyze the characteristics of past infectious disease outbreaks would have to sift through hundreds or thousands of search engine results.

Gravano’s extraction technology searches for pages that are likely to contain the desired structured information, which is often embedded in natural language text. It then extracts, analyzes, and puts the information into a table automatically. Unfortunately, the process is prone to errors. Information is sometimes out of date or wrong. Writing is often ambiguous.

Gravano hopes to reduce errors by using such trusted sources as government documents, university archives, newspapers, and specialized websites, as well as by analyzing the frequency and context of the extracted information.

He also taps crowd wisdom to assess the reliability of popular sources. “Popularity is a step in the right direction – if you trust people to go to trustworthy sources,” Gravano said.

B.S., ESLAI (Argentina), 1991; M.S., Stanford, 1994; Ph.D., 1997

*Supercharging
Search Engines*

LUIS GRAVANO

Associate Professor of
Computer Science