Data-mining the Web for feelings, not facts

A new kind of search scours the Internet for users' emotions and translates them into hard data

BY **ALEX WRIGHT**
NY TIMES NEWS SERVICE, NEW YORK

Computers may be good at crunching numbers, but can they crunch feelings?

The rise of blogs and social networks has fueled a bull market in personal opinion: reviews, ratings, recommendations and other forms of online expression. For computer scientists, this fast-growing mountain of data is opening a tantalizing window onto the collective consciousness of Internet users.

An emerging field known as sentiment analysis is taking shape around one of the computer world's unexplored frontiers: translating the vagaries of human emotion into hard data.

This is more than just an interesting programming exercise. For many businesses, online opinion has turned into a kind of virtual currency that can make or break a product in the marketplace.

Yet many companies struggle to make sense of the caterwaul of complaints and compliments that swirl around their products online. As sentiment analysis tools begin to take shape, they could not only help businesses improve their bottom lines, but also revitally transform the experience of searching for information online.

Several new sentiment analysis companies are trying to tap into the growing business interest in what is being said online.

"Social media used to be this cute project for 25-year-old consultants," said Margaret Francis, vice president for product at Scout Labs in San Francisco. Now, she said, top executives "are recognizing it as an incredibly rich vein of market intelligence."

Scout Labs, which is backed by the venture capital firm started by the CNet founder Halsey Minor, introduced a subscription service that allows customers to monitor blogs, news articles, online forums and social networking sites for trends in opinions about products, services or topics in the news.

In early May, the ticket marketplace StubHub used Scout Labs' monitoring tool to identify a sudden surge of negative blog sentiment after rain delayed a Yankees-Red Sox game.

Stadium officials mistakenly told hundreds of fans that the game had been canceled, and StubHub denied fans' requests for refunds, on the grounds that the game had actually been played. But after spotting trouble brewing online, the company offered discounts and credits to the affected fans. It is revaluating its bad weather policy.

"This is a canary in a coal mine for us," said John Whelan, StubHub's director of customer service.

Jodange, based in Yonkers, New York, offers a service geared toward online publishers that lets them incorporate opinion data drawn from more than 450,000 sources, including mainstream news sources, blogs and Twitter.

Based on research by Claire Cardie, a former Cornell computer science professor, and Jan Wiebe of the University of Pittsburgh, the service uses a sophisticated algorithm that not only evaluates sentiments about particular topics, but also identifies the most influential opinion holders.

Jodange, whose early investors include the National Science Foundation, is working on a new algorithm that could use opinion data to predict future developments, like forecasting the effect of newspaper editorials on a company's stock price.

For casual Web surfers, simpler incarnations of sentiment analysis are sprouting up in the form of lightweight tools like Tweetfeel, Twendz and Twitrratr. These sites allow users to take the pulse of Twitter users about particular topics.

A quick search on Tweetfeel, for example, reveals that 77 percent of recent tweeters liked the movie *Julie and Julia*. But the same search on Twitrratr reveals a few misfires. The site assigned a negative score to a tweet reading "julie and julia was truly delightful!!" That same message ended with "we all felt very hungry afterwards" — and the system took the word "hungry" to indicate a negative sentiment.

While the more advanced algorithms used by Scout Labs, Jodange and Newssift employ advanced analytics to avoid such pitfalls, none of these services works perfectly. "Our algorithm is about 70 to 80 percent accurate," said Francis, who added that its users can reclassify inaccurate results so the system learns from its mistakes.

Translating the slippery stuff of human language into binary values will always be an imperfect science, however. "Sentiments are very different from conventional facts," said Seth Grimes, the founder of the suburban Maryland consulting firm Alta Plana, who points to the many cultural factors and linguistic nuances that make it difficult to turn a string of written text into a simple pro or con sentiment. "'Sinful' is a good thing when applied to chocolate cake," he said.

The simplest algorithms work by scanning keywords to categorize a statement as positive or negative, based on a simple binary analysis ("love" is good, "hate" is bad). But that approach fails to capture the subtleties that bring human language to life: irony, sarcasm, slang and other idiomatic expressions. Reliable sentiment analysis requires parsing many linguistic shades of gray.

"We are dealing with sentiment that can be expressed in subtle ways," said Bo Pang, a researcher at Yahoo who co-wrote *Opinion Mining and Sentiment Analysis*, one of the first academic books on sentiment analysis.

To get at the true intent of a statement, Pang developed software that looks at several different filters, including polarity (is the statement positive or negative?), intensity (what is the degree of emotion being expressed?) and subjectivity (how partial or impartial is the source?).

For example, a preponderance of adjectives often signals a high degree of subjectivity, while noun- and verb-heavy statements tend toward a more neutral point of view.

As sentiment analysis algorithms grow more sophisticated, they should begin to yield more accurate results that may eventually point the way to more sophisticated filtering mechanisms. They could become a part of everyday Web use.

"I see sentiment analysis becoming a standard feature of search engines," said Grimes, who suggests that such algorithms could begin to influence both general-purpose Web searching and more specialized searches in areas like e-commerce, travel reservations and movie reviews.

Pang envisions a search engine that fine-tunes results for users based on sentiment. For example, it might influence the ordering of search results for certain kinds of queries like "best hotel in San Antonio."

As search engines begin to incorporate more and more opinion data into their results, the distinction between fact and opinion may start blurring to the point where, as David Byrne once put it, "facts all come with points of view."

ILLUSTRATION: NY TIMES NEWS SERVICE

Joe Palia, 29, poses in his space suit behind his home in Holiday, Florida.
PHOTO: NY TIMES NEWS SERVICE

Mad for Mars

Joe Palia's mission in life is to go to Mars. He's got as far as the Arctic

BY **JOHN BARRY**
NY TIMES NEWS SERVICE, ST PETERSBURG

What brought a bright young man who sweated through electrical and nuclear engineering classes at MIT to this, a month on a frozen rock in the Arctic, with a fish bowl on his head, a Buzz Lightyear space suit, a shotgun to scare off polar bears and a busted "incinerator toilet"?

That's how bad Joe Palaia wants to get to Mars.

Joe is almost 30. His mission in life is Mars, but his Martian clock is ticking. Bailout-happy, cash-for-clunkering politicians are making it very difficult for the space program. Since graduating from MIT three years ago, Palaia has done all he can to keep hope alive.

To advance the cause of manned Mars missions, he left wife and home in Holiday, Florida, to spend last month with five other volunteers in a can-shaped shelter on top of the world just 1,448km from the Earth's North Pole. The place was Canada's Devon Island, which hasn't made much news outside musk oxen circles since a meteor fell on it 20 million years ago.

Their assignment: Pretend they're on Mars.

Palaia and mates wore fake space suits. They endured snow, rain and fog. They slept through hurricane-force winds and blazing sunshine at midnight. They were armed for bear but saw one mosquito and one rabbit.

They did accomplish something astronauts may one day attempt on the Red Planet. They drew water from rock.

By now in the story, your eyebrow may be up to your hairline. North Pole. Fake space suits. Polar bears vs polar bunnies. Incinerator toilets. Fourth rock from the sun. The fair question is: "Is Palaia nuts or what?"

He seems to be an intelligent, adventurous, focused young man. He just happens to be in a hurry to leave Earth. Draw your own conclusions.

The Mars Society has been sending volunteers on shoestring expeditions to the Arctic since 2000. It's as close to a Mars environment as you can find on Earth. Everything is like Mars except the polar bears, oxygen and Twitter.

Back in 2000, the Mars Society set up a tall fiberglass tube on the rock and furnished it with generators, stove, showers and cubby holes each just big enough for a bunk and a shelf.

Palaia's crew was the 12th to occupy the tube. He answered an open call for volunteers because going to Mars has been all he's thought about. After MIT, he helped start a company called 4Frontiers in New Port Richey, Florida, aimed at getting on the ground floor of Mars commercial opportunities. When he married, he told his wife, Melissa, he'd eventually have to leave for three years or so for a round trip.

The Mars Society made him chief engineer of the Devon Island expedition, meaning he had to keep the tube heated and the balky incinerator toilet working.

When they got to the island by bush plane it was snowing sideways.

The team was three men, three women. Besides engineer Palaia, it included a geologist, two NASA workers, a seismologist and a fifth-grade teacher. Each was encouraged to bring a personal research project. The geologist mined and cooked the mineral gypsum, which was all over the island, and also happens to exist in the polar regions of Mars. When heated to 149°C, it releases water.

Palaia brought an aircraft. He'd persuaded a Gainesville company called Prioria Robotics to loan him a small robotic plane rigged with surveillance cameras. Palaia's project was to show that he could fly the thing while encumbered in a spacesuit. He flew it six times.

For every outdoor mission, the crew was required to wear the space suits. They'd been sewn by a Denver costumemaker. Each was canvas, badly frayed, and had a button-up fly. The uniform included a backpack that was basically a Tupperware container. It contained a fan that blew air into the stifling bubble helmet.

The purpose of the suits was to test astronaut mobility. Besides, the canvas kept Palaia warm. They enjoyed only three or four sunny days. When they got there, the temperature was minus 15°C. It got up to about 7°C.

Ask Palaia how six people got along for a month in a tube, and he'll tell you that gypsum is hydrated calcium sulfate. He's scientific, not one for idle gossip. Mostly, it seems, he and his mates worked. It took two weeks to make the tube livable. They generally kept at it every day from 9:30am to midnight, working around bad weather.

They also had to practice a polar bear drill. It consisted of gathering behind the guy with the shotgun, stripping off space suits, and running as fast as possible. That may or may not be necessary on Mars.

As soon as Palaia's tired crew returned from Devon Island, a study committee appointed by US President Barack Obama issued a gloomy report on NASA's manned spaceflight program. Basically, it said, the money isn't there, even to send someone to the moon.

Palaia says private enterprise will do it if the government can't. "Look what we're accomplishing with a bunch of volunteers." One way or another, he's going where no man has gone before.