

Computer translator reads between the tweets



Peter Macdiarmid/Getty Images

A shop in Tahrir Square is spray painted with the word Twitter after the government shut off internet access on February 4, 2011 in Cairo, Egypt. Anti-government protesters have called today 'The day of departure'. Thousands have again gathered in Tahrir Square calling for Egyptian President Hosni Mubarak to step down.

April 3, 2011 | Christopher Joyce | National Public Radio

Share: [f](#) [t](#) [+](#) [Mail Print Comment](#)

Facebook and Twitter have been touted as tools for North African and Middle Eastern revolutionaries. They're also sources of intelligence on politics and culture. But sifting through all those Tweets is a Herculean task. Now scientists are developing a computer program that doesn't just translate — it interprets nuance as well.

There are already computer programs to translate text from one language to another. But languages like Arabic or Urdu are tough — the script and the grammar are hugely different from European languages, and digitized dictionaries and grammatical algorithms for those languages are still in the early stages.

But these languages are politically important. Urdu, for example, is the language of Pakistan and of many Muslims in India. It's a mix of Hindi and Persian and uses Arabic script.

Mining For Sentiment And Opinion

Computer scientist Rohini Srihari says existing computer translators for Urdu are often too literal.

"What I want is to determine who are the people, places and things being talked about," she says. "Is there an opinion being expressed? Is it a positive or negative opinion being expressed?"

At the University of Buffalo, Srihari has developed a natural language program that she says can do that. The computer has "learned" the nuances of written Urdu. Some of this is fairly mechanical — Urdu doesn't have the same kind of clear breaks between one word and the next the way English does. Some things are subtle, as in characters and words whose placement may connote sentiment.

"And when you are able to figure out what the topic of the conversation is," she says, "what kind of sentiment is being expressed around that, that's the goal of what we are trying to do."

On the screen, you can mouse over a section of script and if it carries a "negative" connotation — it will highlight red. If it has a "positive" sentiment, it glows green.

Srihari says the computer allows her to mine the Internet. Her research company, called Janya, Inc., gets funding from the Pentagon for the project.

"So in Twitter posts and tweets and so on, if there's specific factual information that's being mentioned — they want that extracted," says Srihari. "There's also definitely an interest in sentiment and opinion mining."

Getting More Voices Into The Conversation

Social media and the opinions that get slung around digitally around the world's hot spots are also of interest to political scientists and historians.

SPONSOR

VALLEY PERFORMING ARTS CENTER
CALIFORNIA STATE UNIVERSITY, NORTHRIDGE
SHIRLEY MACLAINE
MAY 6

ADVERTISE WITH US

Contribute Now

support the unbiased, comprehensive news you rely on from KPCC



Become a member today!

- | | |
|-----------------------------|-----------------------------|
| LISTEN | FIND US ON |
| Live Stream | Facebook |
| Podcasts | Twitter |
| iPhone App | Newsletters |

» Find us on social media

SPONSOR

HOOLOGAN Theatre Company presents
SeUSSICAL
TWO WEEKENDS ONLY
APRIL 1ST-10TH

ADVERTISE WITH US

Most Viewed | Most Commented

- \$10,000 reward offered for information on suspects in Dodger Stadium attack
- Dodger Stadium victim is paramedic, father of two
- Entitlement cuts are on the table!
- The latest in medical marijuana paraphernalia on display at HempCon 2011
- Crowd boos Sheen on stage during his 1st tour stop

SPONSOR

Ernest Tucker, a history professor with the Center for Middle East and Islamic Studies at the U.S Naval Academy, struggles with Urdu himself. But Tucker does speak and read Persian, which is close, and he regularly reads publications from the region.

He argues that history is best told not by what the Napoleans say, but by the foot-soldiers, or, in this case, the tweeters.

"And that's the goal of all historians anywhere," he says, "to try to get the voices of more and more people into the conversation, and anything that can do that, particularly this kind of thing, is a wonderful gift."

Tucker says he's skeptical about how well a computer is going to identify sentiment — he says you'll still need a human linguist to fine-tune any translation. For example, he says it's common in Middle Eastern languages to employ couplets from traditional poetry to convey feelings — symbolic language that could confuse a computer program.

"For the Iranians, for the Pakistanis, for the Indians," he says, "it's still part of the living connection to the cultures of the past."

Rohini Srihari acknowledges the program isn't perfect. It gets flummoxed by things like Urdish, a mashup language for text messaging that's part Urdu, and part English. But it has given her insight into what Urdu speakers have been talking about lately.

"A lot of the conversation, believe it or not, was about cricket — that seems to be on everyone's mind all the time," she says.

Last week would have been a good time to tune into Urdu cyberspace. Pakistan played India in the world cricket semi-finals. Pakistan lost; no doubt the web traffic was full of strong sentiments. Copyright 2011 National Public Radio. To see more, visit <http://www.npr.org/>.

Related Stories

- [View this story at NPR](#)

World News

- [Search for radiation leak turns desperate in Japan](#)
- [Deadly Suicide Blast Hits Bus Stand In Pakistan](#)
- [Libyan Rebels Advance On Brega; U.S. Role To Shift](#)
- [Computer translator reads between the tweets](#)

» [MORE WORLD NEWS](#)

Leave a comment

We welcome all comments on our website and encourage active, intelligent debate. However, comments/posts including hate speech, profanity, spam or personal attacks will be removed. Thanks for being part of our community.

Name

Email address

Comment



[ADVERTISE WITH US](#)

Tourists hunt photo ops along the L.A. River

By Olga Khazan | Art | Los Angeles



On a cloudy Sunday afternoon, two dozen photography enthusiasts trudge along the muddy banks of the L.A. River. Though not always known for its beauty, the river provides no

shortage of photo ops as it churns with chocolate-brown water from the recent rains. As they pause to capture everything from ... [More](#)

[More from About Town](#) | [Sign up for the A&C Newsletter](#)

DAY IN A PAGE Radio/Story Archive

April 4 2011 GO

89.3 KPCC | Southern California Public Radio

News | Programs | Blogs | Events | About | Support Us
Pledge Now | Contact Us | Search
Terms & Conditions | Privacy | Website Feedback

© 2011 Southern California Public Radio
474 South Raymond Avenue, Pasadena, CA 91105
626-583-5100

Foundations:



Partners:

