

SUBSCRIBE TO NEW SCIENTIST

Select a country

Subscribe



DON'T MISS THE LARGEST CELEBRATION OF SCIENCE IN THE U.S. APRIL 28 & 29, 2012, WASHINGTON D.C.



FESTIVAL HOST:

LOCKHEED MARTIN

PROUD SPONSOR:

NewScientist

USASCIENCEFESTIVAL.ORG

NewScientist

Tech

search New Scientist

Go

Log in

My New Scientist

Home News In-Depth Articles Blogs Opinion TV Galleries Topic Guides Last Word Subscribe Dating

Look for Science Jobs

SPACE TECH ENVIRONMENT HEALTH LIFE PHYSICS&MATH SCIENCE IN SOCIETY

Home | Tech | Life | News

Gender-spotting tool could have rumbled fake blogger

10:00 17 June 2011 by Paul Marks

For similar stories, visit the [The Human Brain](#) Topic Guide

Software that guesses a writer's gender could have prevented the world being duped into believing a blog that opposed the Syrian government and was striking out for gay rights was written by a young lesbian living in the country.

It turned out the author of the blog, "Gay Girl in Damascus", was a man – something the online gender checker would have picked up on. When *New Scientist* fed the text of the last blog post into the software, it said that the author was 63.2 per cent likely to be male.

Developed by Na Cheng and colleagues at the Stevens Institute of Technology in Hoboken, New Jersey, the ever-improving software could soon be revealing the gender of online writers – whether they are blogging, emailing, writing on Facebook or tweeting. The team say the software could help protect children from grooming by predators who conceal their gender online.

The fake blog highlights the problem of people masking their identity online. The truth about [Amina Abdullah](#) only emerged when the blogger disappeared, supposedly snatched by militiamen.

Online contacts realised that none of them had ever met Amina, and it turned out her blog photo had been stolen from a Facebook page. Then a 40-year-old American, Tom MacMaster living in Edinburgh, UK, confessed that he had been writing the blog all along.

Gender analysis

To determine the gender of a writer or blogger, Cheng and her colleagues Rajarathnam Chandramouli and Koduvayur Subbalakshmi wrote software that allows users to either [upload a text file](#) or [paste in a paragraph](#) of 50 words or more for gender analysis (update: since we posted this article, the servers have enjoyed more traffic than they can handle. If you can't get a response from clicking the links in this paragraph, we suggest trying again at a different time of day).

After a few moments, the program spits out a gender judgement: male, female or neutral. The neutral option points to how much of the text has been stripped of any indication of gender. This is something particularly prevalent in scientific texts, the researchers say.

To write their program, the team first turned to vast tranches of bylined text from a Reuters news archive and the massive email database of the bankrupt energy firm Enron. They trawled these documents for "psycho-linguistic" factors that had been identified by previous research groups, such as specific words and punctuation styles.

In total they found 545 of these factors, says Chandramouli, which they then honed down to 157 gender-significant ones. These included differences in punctuation style or paragraph lengths between men and women.

Other gender-significant factors included the use of words that indicate the mood or sentiment of the author and the degree to which they use "emotionally intensive adverbs and affective adjectives such as really, charming or lovely" which were used more often by women, says Chandramouli. Men were more likely to use the word "I", for example, whereas women used questions marks more often.

PRINT SEND SHARE



Who's behind the blogs you read? (Image: Plainpicture)

ADVERTISEMENT: USA SCIENCE & ENGINEERING FESTIVAL. DON'T MISS THE LARGEST CELEBRATION OF SCIENCE IN THE U.S. EXPO: APRIL 28 & 29, 2012 IN WASHINGTON, D.C. FESTIVAL HOST: LOCKHEED MARTIN. PROUD SPONSOR: NewScientist. USASCIENCEFESTIVAL.ORG

This week's issue

Subscribe



07 April 2012

ADVERTISEMENT

DON'T MISS THE LARGEST CELEBRATION OF SCIENCE IN THE U.S.

GRAND FINALE EXPO & BOOK FAIR APRIL 28 & 29, 2012 WASHINGTON, D.C.



More Latest news

Tweet your own highlight reel with #EpicPlay



13:00 09 April 2012

Sports fans who tweet while watching the game could help create more exciting highlights than those compiled by professional broadcasters

Online searches for future linked to economic success



17:50 05 April 2012

People in countries with a higher GDP are more likely to search for information about the year ahead rather than the previous year

Bruisable gadgets heal themselves in the sun

12:47 05 April 2012

ADVERTISEMENT: USA SCIENCE & ENGINEERING FESTIVAL. DON'T MISS THE LARGEST CELEBRATION OF SCIENCE IN THE U.S. EXPO: APRIL 28 & 29, 2012 IN WASHINGTON, D.C. FESTIVAL HOST: LOCKHEED MARTIN. PROUD SPONSOR: NewScientist. USASCIENCEFESTIVAL.ORG

Bayesian algorithms

Finally, the software combined these cues using a Bayesian algorithm, which guesses gender based on the balance of probabilities suggested by the telltale factors. The work will appear in an upcoming edition of the journal Digital Investigation.

It doesn't always work, however. When the software is fed text, its judgement on a male or female writer is only accurate 85 per cent of the time – but that will improve as more people use it. That's because users get the chance to tell the system when it has guessed incorrectly, helping the algorithm learn. The next version will analyse tweets and Facebook updates.

Bernie Hogan, a specialist in social network technology at the Oxford Internet Institute in the UK, thinks there is a useful role for such technology. "Being able to provide some extra cues as to the gender of a writer is a good thing – it can only help."

Even a "neutral" decision might indicate that someone is trying to write in a gender voice that does not come naturally to them, he says. "It could be quite telling."

Testing the gender software

What did the gender identifier make of three well-known authors? We fed it some sample text to find out.

V. S. Naipaul, a winner of the Nobel prize for literature, claims he can tell a woman's writing by reading just two paragraphs of text, and controversially thinks female authors are no match for his writing. The software's verdict on this extract from his book The Enigma of Arrival: 88.4 per cent male.

Mary Evans was a female novelist who famously wrote under the male nom de plume George Eliot. The software has the measure of her, though. Its analysis of the writer's gender from the first paragraphs of Middlemarch: 94.6 per cent female.

More than 14,000 of Sarah Palin's emails were released by the state of Alaska last week after a lengthy campaign by various media organisations to obtain access to them. One email from the archive was put through the system, but the software got it wrong: 70.77 per cent male.

0 tweets

Like 2



If you would like to reuse any content from New Scientist, either in print or online, please contact the syndication department first for permission. New Scientist does not own rights to photos, but there are a variety of licensing options available for use of articles and graphics we own the copyright to.

Have your say

Only subscribers may leave comments on this article. Please log in.

email: [input field]

password: [input field]

Remember me [checkbox]

Log in

Only personal subscribers may leave comments on this article

Subscribe now to comment.

Maybe Todd Ghostwrote Sarah Palin's Emails

Fri Jun 17 19:05:54 BST 2011 by ullrich fischer

login and reply report this comment



A new type of plastic turns red when it is damaged – and then heals itself when exposed to light

Stripping CO2 from air requires largest industry ever



15:30 04 April 2012 Geoengineering schemes to remove CO2 from the atmosphere would involve creating an industry 1000 times larger than any seen before

see all related stories

Most read Most commented

Bruisable gadgets heal themselves in the sun

Tides turn some habitable planets hellish

Marathon mouse space mission boosts bone protection

Galactic laser light show might lead us to aliens

Increased autism rates crush vaccine hypothesis

TWITTER

New Scientist is on Twitter



Get the latest from New Scientist: sign up to our Twitter feed

LATEST JOBS

Henry M Jackson Foundation: Linux System Administrator (207174)

Henry M Jackson Foundation: Research Assistant (207325)

Henry M Jackson Foundation: Bioinformatics Scientist (207329)

Paramount Recruitment: Medical Writer - Senior Medical Writer - Milan

Hays: Senior Scientific Advisor - Vaccines Middlesex