New reporter? Call him AI, for algorithm

July 11 2012, by Rob Lever

The media covers President Barack Obama's press conference in June 2012. The new reporter on the US media scene takes no coffee breaks, churns out articles at lightning speed, and has no pension plan.

The new reporter on the US media scene takes no coffee breaks, churns out articles at lightning speed, and has no pension plan.

That's because the reporter is not a person, but a computer algorithm, honed to translate raw data such as corporate earnings reports and previews or sports statistics into readable prose.

Algorithms are producing a growing number of articles for newspapers and websites, such as this one produced by Narrative Science:

"Wall Street is high on Wells Fargo, expecting it to report earnings that are up 15.7 percent from a year ago when it reports its second quarter earnings on Friday, July 13, 2012," said the article on Forbes.com.

While computers cannot parse the subtleties of each story, they can take vast amounts of raw data and turn it into what passes for news, analysts say.

"This can work for anything that is basic and formulaic," says Ken Doctor, an analyst with the media research firm Outsell.

And with media companies under intense financial pressure, the move to automate some news production "does speak directly to the rebuilding of the cost economics of journalism," said Doctor.

Stephen Doig, a journalism professor at Arizona State University who has used computer systems to sift through data which is then provided to reporters, said the new computer-generated writing is a logical next step.

"I don't have a philosophical objection to that kind of writing being outsourced to a computer, if the reporter who would have been writing it could use the time for something more interesting," Doig said.

Scott Frederick, chief operating officer of Automated Insights, another firm in the sector, said he sees this as "the next generation of content creation."

The company got its start in 2007 as StatSheet, which generates news stories from raw feeds of play-by-play data from major sports events.

The company generates advertising on its own website and is now beginning to sell its services to other organizations for sports and real estate news.

"Over the next 12 to 24 months, every media property will need some automation strategy," Frederick told AFP.

To mimic the effect of the hometown newspaper, the company generates articles with a different "tonality" depending on the reader's preference or location.

For the 2012 Super Bowl, the article for New York Giants' fans read like this: "Hakeem Nicks had a big night, paving the way to a victory for the Giants over the Patriots, 21-17 in Indianapolis. With the victory, New York is the champion of Super Bowl XLVI."

For New England fans, the story was different: "Behind an average day from Tom Brady, the Patriots lost to the Giants, 21-17 at home. With the loss, New England falls short of a Super Bowl ring."

"Data becomes the seeds of the content trees. When you can create an entire story out of raw data, that is technologically impressive," Frederick said.

Kristian Hammond, chief technology officer at Chicago-based Narrative Science, said he had been involved in computer content generation for more than a decade.

Hammond is on leave from Northwestern University, where he was on the computer science faculty and headed a joint project generating content with the university's journalism school.

The company formed in 2010 has 40 clients including Forbes, and some corporate clients which use the technology to take spreadsheets or other data for internal reports that are more readable.

"We're about two-thirds engineering and one-third journalism," he said.

"We knew there were places in traditional journalism where raw data was used as the driver for telling stories, and we wanted to take that

model and turn it into something a machine can do," he told AFP.

While some articles are reviewed by editors, others are automatically delivered without human intervention because of client preference or because the task is too voluminous: Narrative Science, he said, produced stories on 370,000 Little League baseball games in the past year.

The computers cannot pick up on certain things, such as if an injury or weather affects the game.

"If it's not in the data, we can't say anything about it. We're very aware of that, but more of what goes on is data-driven," Hammond said.

"The feedback has been very positive. We haven't done anything goofy or embarrassing so far."

One goof came from a company called Journatic, a partner of the Chicago Tribune, which uses a combination of human editors in the US and overseas and computer algorithms to generated "hyperlocal" news.

Some news organizations complained when they discovered the "bylines" generated were made-up names, not real journalists, in the Tribune, Houston Chronicle and San Francisco Chronicle, a violation of ethics policies for the dailies.

Journatic chief executive Brian Timpone said the flap stemmed from a misunderstanding with news clients and the fact that bylines were needed to be seen on Google News.

"We're taking them off," Timpone said, arguing that should not distract attention from the business model which can help media companies.

"The way news is produced has not changed in 50 years," he told AFP.

Timpone said his company can produce news more efficiently "with technology, lots of local news gathering, and a distributed writing team."

"It's not about algorithms. Algorithms only work if the data is structured. There's no way to automate everything."

(c) 2012 AFP

Citation: New reporter? Call him Al, for algorithm (2012, July 11) retrieved 26 April 2024 from https://phys.org/news/2012-07-al-algorithm.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.