The New York Times

In Case You Wondered, a Real Human Wrote This Column

By STEVE LOHR   SEPT. 10, 2011

"WISCONSIN appears to be in the driver's seat en route to a win, as it leads 51-10 after the third quarter. Wisconsin added to its lead when Russell Wilson found Jacob Pedersen for an eight-yard touchdown to make the score 44-3 ... ."

Those words began a news brief written within 60 seconds of the end of the third quarter of the Wisconsin-U.N.L.V. football game earlier this month. They may not seem like much — but they were written by a computer.

The clever code is the handiwork of Narrative Science, a start-up in Evanston, Ill., that offers proof of the progress of artificial intelligence — the ability of computers to mimic human reasoning.

The company's software takes data, like that from sports statistics, company financial reports and housing starts and sales, and turns it into articles. For years, programmers have experimented with software that wrote such articles, typically for sports events, but these efforts had a formulaic, fill-in-the-blank style. They read as if a machine wrote them.

But Narrative Science is based on more than a decade of research, led by two of the company's founders, Kris Hammond and Larry Birnbaum, co-directors of the Intelligent Information Laboratory at Northwestern University, which holds a stake in the company. And the articles produced by Narrative Science are different.

"I thought it was magic," says Roger Lee, a general partner of Battery Ventures, which led a $6 million investment in the company earlier this year. "It's as if a human wrote it."

Experts in artificial intelligence and language are also impressed, if less enthralled. Oren Etzioni, a computer scientist at the University of Washington, says, "The quality of the narrative produced was quite good," as if written by a human, if not an accomplished wordsmith. Narrative Science, Mr. Etzioni says, points to a larger trend in computing of "the increasing sophistication in automatic language understanding and, now, language generation."

The innovative work at Narrative Science raises the broader issue of whether such applications of artificial intelligence will mainly assist human workers or replace them. Technology is already undermining the economics of traditional journalism. Online advertising, while on the rise, has not offset the decline in print advertising. But will "robot journalists" replace flesh-and-blood journalists in newsrooms?

The leaders of Narrative Science emphasized that their technology would be primarily a low-cost tool for publications to expand and enrich coverage when editorial budgets are under pressure. The company, founded last year, has 20 customers so far. Several are still experimenting with the technology, and Stuart Frankel, the chief executive of Narrative Science, wouldn't name them. They include newspaper chains seeking to offer automated summary articles for more extensive coverage of local youth sports and to generate articles about the quarterly financial results of local public companies.

"Mostly, we're doing things that are not being done otherwise," Mr. Frankel says.

The Narrative Science customers that are willing to talk do fit that model. The Big Ten Network, a joint venture of the Big Ten Conference and Fox Networks, began using the technology in the spring of 2010 for short recaps of baseball and softball games. They

were posted on the network's Web site within a minute or two of the end of each game; box scores and play-by-play data were used to generate the brief articles. (Previously, the network relied on online summaries provided by university sports offices.)

As the spring sports season progressed, the computer-generated articles improved, helped by suggestions from editors on the network's staff, says Michael Calderon, vice president for digital and interactive media at the Big Ten Network.

The Narrative Science software can make inferences based on the historical data it collects and the sequence and outcomes of past games. To generate story "angles," explains Mr. Hammond of Narrative Science, the software learns concepts for articles like "individual effort," "team effort," "come from behind," "back and forth," "season high," "player's streak" and "rankings for team." Then the software decides what element is most important for that game, and it becomes the lead of the article, he said. The data also determines vocabulary selection. A lopsided score may well be termed a "rout" rather than a "win."

"Composition is the key concept," Mr. Hammond says. "This is not just taking data and spilling it over into text."

Last fall, the Big Ten Network began using Narrative Science for updates of football and basketball games. Those reports helped drive a surge in referrals to the Web site from Google's search algorithm, which highly ranks new content on popular subjects, Mr. Calderon says. The network's Web traffic for football games last season was 40 percent higher than in 2009.

Hanley Wood, a trade publisher for the construction industry, began using the program in August to provide monthly reports on more than 350 local housing markets, posted on its site, builderonline.com. The company had long collected the data, but hiring people to write trend articles would have been too costly, says Andrew Reid, president of Hanley Wood's digital media and market intelligence unit.

Mr. Reid says Hanley Wood worked with Narrative Science for months to fine-tune the software for construction. A former executive at Thomson Reuters, he says he was struck by the high quality of the articles.

"They got over a big linguistic hurdle," he observes. "The stories are not duplicates by any means."

He was also impressed by the cost. Hanley Wood pays Narrative Science less than $10 for each article of about 500 words — and the price will very likely decline over time. Even at $10, the cost is far less, by industry estimates, than the average cost per article of local online news ventures like AOL's Patch or answer sites, like those run by Demand Media.

NARRATIVE SCIENCE'S ambitions include moving further up the ladder of quality. Both Mr. Birnbaum and Mr. Hammond are professors of journalism as well as computer science. The company itself is an outgrowth of collaboration between the two schools.

"This kind of technology can deepen journalism," says John Lavine, dean of the Medill School of Journalism at Northwestern.

Mr. Hammond says the combination of advances in its writing engine and data mining can open new horizons for computer journalism, exploring "correlations that you did not expect" — conceptually similar to "Freakonomics," by two humans, the economist Steven D. Levitt and the author Stephen J. Dubner.

Mr. Hammond cited a media maven's prediction that a computer program might win a Pulitzer Prize in journalism in 20 years — and he begged to differ.

"In five years," he says, "a computer program will win a Pulitzer Prize — and I'll be damned if it's not our technology."

Should it happen, the prize, of course, would not be awarded to abstract code, but to its human creators.