

# When journalists and technologists monkey around

A computer scientist with a Medill education and a sports reporter with a physics background team up to develop software that instantly writes baseball articles.

By Angela Dee Kwan (MSJ09)

“BRADENTON, Fla.— Northwestern held off a late comeback bid by Georgetown to defeat the Hoyas 5-3 Friday. Trevor Stevens led the Wildcats with two hits and one run scored...” Fans who finish reading this story would know that Eric Jokisch pitched a complete game for the Wildcats and struck out nine batters—but they’d never suspect that a machine wrote it.

StatsMonkey, a computer program that automatically generates sports stories based on readily available data, was produced by graduate journalism students in the spring 2009 Interactive Innovation Project ([www.medill.northwestern.edu/innovation/interactive\\_archive.aspx](http://www.medill.northwestern.edu/innovation/interactive_archive.aspx)).

But unlike most projects limited to an academic and theoretical shelf life, StatsMonkey has lived on to become one of the first prototypes developed by the Medill-McCormick Center for Innovation in Technology, Media and Journalism.

← Recent graduates John Templon (back left) and Nick Allen were hired by McCormick Professors Kristian Hammond and Larry Birnbaum (front right) to build StatsMonkey, a computer program that translates sports data into a narrative.  
*Photo by Tommy Giglio.*

The Center explores the intersection between journalism and technology, a space that faces tremendous challenges but offers infinite opportunities to create critical solutions that will benefit the public. Scheduled to formally launch in the fall, The Center hired two summer interns, John Templon (MSJ09) and Nick Allen (MSJ09), to continue developing StatsMonkey.

"We want these classes and projects to be things that last beyond the classroom and become something that goes out into

up with on their own," Gordon says.

If the whole is greater than the sum of its parts, then technologists need journalists as much as journalists need technologists. McCormick students wrote the code for the five interactive tools and taught Medill students which problems were easy or hard to solve, and the journalists offered their knowledge of editorial judgment and audience needs.

Though all five tools were well received at the final presentation to representa-



Susan Dosemagen (left) and Lizz Kannenberg (both MSJ09) rehearse their script to introduce Tweedia, a widget that aggregates and displays current Twitter posts alongside a related news story.

"People care passionately about these games, but there's no existing way to serve them."

the world and makes an impact," says Kristian Hammond, director of The Center and McCormick professor of electrical engineering and computer science. "It's not the innovation [that matters], it's the impact."

For an industry historically rooted in words, innovation is a difficult word to define. Owen Youngman, associate director of The Center and Knight professor of digital media strategy, describes innovation as, "creating new ways to solve old problems, creating new products to serve old markets better or emerging markets for the first time, or finding new ways to serve audiences with products that are less expensive or more convenient."

Originally intended to reach the underserved community of college baseball fans, StatsMonkey was created by Templon, Allen, Tian Huang (BSJ09, MSJ09) and Thu Cung (McC09). It's just one of five new tools produced in the spring Interactive Innovation Project to engage audiences and more efficiently gather and analyze news. Hammond co-taught the course, the first inter-disciplinary one offered to Medill graduate and McCormick undergraduate students, with three instructors: Medill Assistant Professor Jeremy Gilbert, Medill Director of Digital Innovation Rich Gordon and McCormick Associate Professor of Electrical Engineering and Computer Science Larry Birnbaum.

"If you put journalists together to collaborate with technologists, they will develop something more novel, influential and innovative than either would have come

tives from media organizations, such as Hearst Corporation, Tribune Company, and ESPN, StatsMonkey was the most successful from an actual marketability and deployment standpoint, Gordon says. He attributes this success to the team's fluency in both disciplines—the lead software developer, Allen, is a computer scientist turned journalist, while Templon is a sports writer with a physics background and basic programming knowledge. The recent journalism graduates have since migrated across campus to McCormick's Ford Motor Company Engineering Design

Center to work under the supervision of Hammond and Birnbaum.

A sports reporter who designs a program that eliminates the need for a human writer may seem counterintuitive, but StatsMonkey benefits both the audience and the journalist. Because newspapers lack the manpower to cover all sports, StatsMonkey can meet the needs of underserved audiences, such as Division I fans or even little league dads. And any computer-generated story good enough to replace a newspaper article written by a human would give reporters additional time to cover other more interesting stories, Templon says.

But the beauty of StatsMonkey (formerly



Director of Digital Innovation Rich Gordon talks with Owen Youngman (left), Knight professor of digital media strategy, before students demo the five interactive media tools they created.



Assistant Professor Jeremy Gilbert (left), who led the spring 2009 Interactive Innovation Project, and Ivan Meyers, broadcast production manager (center), prepare students for the final presentation to media organizations, such as Hearst Corporation, Tribune Company and ESPN.

named Machine Generated Sports Stories) lies in its potential to easily accommodate any type of story that is data-driven and recurring, such as an economic indicator or quarterly earnings report.

"If your stories are repetitive, then they are probably boring, and I don't know why you are writing them," Templon says.

When the project began in April, Templon and Allen wanted to experiment with a data-rich story in high demand for its content, not its writing style. Baseball games—rife with statistics, yet restricted to a finite number of play-by-play permutations—surfaced as the ideal story archetype for the first version of StatsMonkey.

"People care passionately about these games, but there's no [existing] way to serve them," Birnbaum says. "Will it be the greatest story ever written? No. Well

crafted? Yes."

By graduation day the program could write a lede, a headline and a five-paragraph story from data scraped off CSTV. "The Machine," as it is aptly named in the byline, is also smart enough to construct specific play-by-plays, to select the star player's photo and to identify a perfect game.

But StatsMonkey has a long way to go. The program cannot describe a player's injury, add historical context or compare innings, which is why Templon and Allen are working to develop a more dynamic version. The duo of technologically advanced journalists are striving to craft machine-generated anecdotal ledes and detailed sub-narratives that convey the level of sophistication tantamount to stories produced by a human.

Allen says he is excited about the opportunity to combine his two passions and to develop a product that has real world viability. Several organizations have already expressed interest in funding StatsMonkey, which brings the team one step closer to the ultimate goal: license the tool to media companies to reach a wider audience with cheaper and faster reporting. For Allen these inter-disciplinary collaborations—such as StatsMonkey and The Center—embody the reason why he came to Medill. He knows that The Center is an important step in evaluating the future of journalism.

"It's critical for Medill to evolve as the industry evolves," Allen says. "In the long run, showing leadership in the area of journalism technology is going to help Medill stay relevant." ■