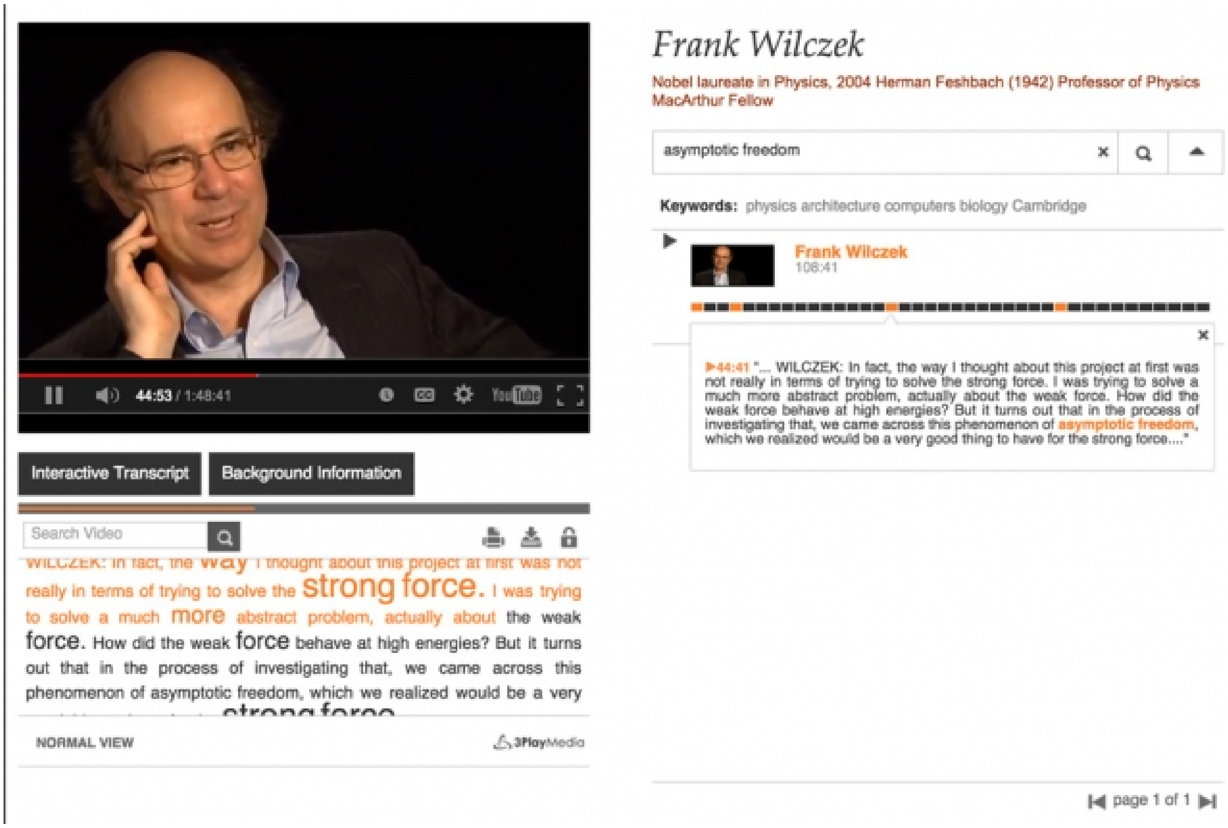
Automated captioning system processes hundreds of video-hours per day

April 9 2015, by Rob Matheson



WILCZEK: In fact, the way I thought about this project at first was not really in terms of trying to solve the strong force. I was trying to solve a much more abstract problem, actually about the weak force. How did the weak force behave at high energies? But it turns out that in the process of investigating that, we came across this phenomenon of asymptotic freedom, which we realized would be a very

▶44:41 "... WILCZEK: In fact, the way I thought about this project at first was not really in terms of trying to solve the strong force. I was trying to solve a much more abstract problem, actually about the weak force. How did the weak force behave at high energies? But it turns out that in the process of investigating that, we came across this phenomenon of asymptotic freedom, which we realized would be a very good thing to have for the strong force...."

An interactive transcript created by 3Play Media appears under and to the right of an online video talk by Nobel laureate Frank Wilczek. These transcripts scroll along with video, highlighting text that's spoken, and let users click words to bring them to that exact moment in the video. Credit: 3Play Media

In 2008, four students at the MIT Sloan School of Management developed a system for captioning online video that was far more efficient than traditional methods, which involve pausing a video frequently to write text and mark time codes.

The system used automated speech-recognition software to produce "rough-draft" transcripts, displayed on a simple interface, that could easily be edited. Landing a gig to caption videos from five MIT OpenCourseWare (OCW) classes, the students were able to caption 100 hours of content in a fraction the time of manual captioning.

This marked the beginning of captioning-service company 3Play Media, which now boasts more than 1,000 clients and an equal number of contracted editors processing hundreds of hours of content per day. Clients include academic institutions, government agencies, and big-name companies—such as Netflix, Viacom, and Time Warner Cable—as well as many users of video-sharing websites.

Today, 3Play's system works much as it did at MIT, but on a grander scale: Customers upload videos to 3Play's site, where automatic speech-recognition software produces transcripts and captions, which are then pushed to the cloud. Then, any of the contracted editors can choose which transcripts to edit. Finally, managers give each job a final look, before pushing it back to the cloud for customers to access.

According to 3Play, the company can process captions in a few hours per video-hour—compared with traditional methods, which can take more than 10 hours per video-hour.

"It's about creating accurate captioning at scale," says 3Play co-founder and chief technology officer C.J. Johnson '02, MBA '08, who co-invented the system in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL). "The questions we asked were: How can we process one file faster, and how can we process thousands of files a day to scale up to meet the demands of the Internet?"

The transcripts created by the system also have time data behind each word. This has paved the way for "interactive transcripts" that accompany video content posted by MIT, and other universities, to online-learning platforms, including edX. These transcripts scroll along with video, highlighting text that's spoken, and let users click words to bring them to that exact moment in the video.

In addition to Johnson, 3Play's co-founders and system co-inventors are Josh Miller MBA '09, Chris Antunes MBA '08, and Jeremy Barron MBA '08.

Tools of the trade

Over the years, 3Play has also developed a number of tools aimed at easing workflow. One tool allows users to switch captioning formats with the click of a button; another lets users cut-and-paste text from the interactive transcript to create clip reels.

But they've also developed tools to meet the ever-changing rules and regulations regarding captioning: In January 2016, the Federal Communications Commission, for instance, will require online clips of full TV shows to have closed captions.

To support this requirement, 3Play has developed a video-clip captioner tool, being released this month, that generates closed captions for short clips by automatically extracting them from the full transcript.

Earlier this year, 3Play developed a return-on-investment calculator, so your average YouTube uploader can learn if captioning is worth the cost. To do so, the company drew on third-party data on thousands of YouTube videos that showed significant increases in viewership with the addition of captions.

When given a video link, the calculator crawls the user's channel to tally viewership of noncaptioned videos and, based on that data, estimates the boost in traffic and search-engine optimization, and how that could all add value with more advertising revenue, among other things.

"Everyone wants to know, 'If I invest money into anything, what's my return?'" Johnson says. "It's the same for captioning."

"Cutting the right corners"

3Play's system took shape at MIT, where the co-founders started "thinking about captioning from the viewpoint of manufacturing," Johnson says. "This meant cutting the right corners to make the captioning process more efficient, but not defective."

While doing some work for OCW in 2007, Johnson learned of the laborious, time-consuming process for captioning videos. "Quickly, it became obvious that this was something ripe for innovation, and technology could be applied to make this process a little easier," Johnson says.

Automatic speech-recognition technology seemed like the clear solution. But, as it turns out, the technology is only about 80 percent accurate, at best, because of errors caused by accents, complex vocabulary, and background noise, among other things.

For years, researchers had been trying to close that 20 percent gap, to no avail. So the real innovation needed to come "after the fact," Johnson says—decreasing the time it takes to edit an imperfect draft.

This led the students to the Spoken Language Systems Group, directed by James Glass, a senior research scientist at CSAIL. Working over the summer, they developed a prototype of the 3Play interface that, among other things, automated traditionally manual tasks, such as grouping words into frame sequences.

Today, that interface has become a key to the system's efficiency, Johnson says. Transcripts appear to editors as simple documents, with a video on the side. Incorrect or inaudible words in the interface are flagged and additional features make editing easier. Any edits made to the transcripts are reflected in the captions, and time is synchronized.

"For 'cutting corners,' the idea is to apply technology to improve the way people edit transcripts … to get as close as we can to perfection, while minimizing the time it takes a person to correct the errors," Johnson says.

Selling captioning

Launching in 2008—after prototyping their technology with OCW—3Play used MIT's Venture Mentoring Service to learn how to craft a business plan, attract customers, and earn funding. "VMS was the No. 1 thing that helped us launch," Johnson says.

And MIT Sloan's entrepreneurship and innovation program, in which Johnson and Miller were enrolled, "was vital for creating the foundation for us to start up out of MIT," Johnson says.

After 3Play's co-founders graduated from MIT Sloan, the company set up shop in a tiny apartment in Somerville, Massachusetts—where the four, without computer-science backgrounds, tried to grow a Web-based company. "We had 'JavaScript for Dummies' books on our desks," Johnson recalls. "We were figuring it all out on the fly."

At one point, they found a list of every college and university in the country, and started calling, one-by-one, to pitch their service. "We were trying to make money month-to-month by selling captioning," Johnson says.

Then one day, representatives from Yale University, which had used the service once before, called to say they were coming to 3Play's corporate headquarters—"our dumpy apartment," Johnson says. "So we light a few candles and we uncomfortably talk to them about how we're going to do their captioning," he says.

Yale went on to become 3Play's first big client outside MIT. Dozens of other educational institutions followed, including Princeton University, Boston University, Harvard Business School, Johns Hopkins University, and others.

Over the years, MIT has used 3Play to caption videos produced for its Infinite History project, MIT Sloan, and the Industrial Liaison Program—which was the first to use interactive transcripts in 2009.

More recently, the company entered the entertainment space, landing Netflix as a client. That was a product of the company's ability to process a lot of content in a short amount of time, Johnson says: "It's all about the ability to scale."

More information: www.3playmedia.com/

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Automated captioning system processes hundreds of video-hours per day (2015, April 9) retrieved 23 June 2024 from https://phys.org/news/2015-04-automated-captioning-hundreds-video-hours-day.html