Monday, November 26, 2007

Searching Video Lectures

A tool from MIT finds keywords so that students can efficiently review lectures.

By Kate Greene

Researchers at MIT have released a video and audio search tool that solves one of the most challenging problems in the field: how to break up a lengthy academic lecture into manageable chunks, pinpoint the location of keywords, and direct the user to them. Announced last month, the MIT Lecture Browser (http://web.sls.csail.mit.edu/lectures/) website gives the general public detailed access to more than 200 lectures publicly available though the university's OpenCourseWare (http://ocw.mit.edu/OcwWeb/web/home/home/index.htm) initiative. The search engine leverages decades' worth of speech-recognition research at MIT and other institutions to convert audio into text and make it searchable.

The Lecture Browser arrives at a time when more and more universities, including Carnegie Mellon University and the University of California, Berkeley, are posting videos and podcasts of lectures online. While this content is useful, locating specific information within lectures can be difficult, frustrating students who are accustomed to finding what they need in less than a second with Google.

"This is a growing issue for universities around the country as it becomes easier to record classroom lectures," says Jim Glass, (http://people.csail.mit.edu/jrg/) research scientist at MIT. "It's a real challenge to know how to disseminate them and make it easier for students to get access to parts of the lecture they might be interested in. It's like finding a needle in a haystack."

The fundamental elements of the Lecture Browser have been kicking around research labs at MIT and places such as BBN Technologies (http://www.bbn.com/) in Boston, Carnegie Mellon, SRI International (http://www.sri.com/) in Palo Alto, CA, and the University of Southern California for more than 30 years. Their efforts have produced software that's finally good enough to find its way to the average person, says Premkumar Natarajan, scientist at BBN. "There's about three decades of work where many fundamental problems were addressed," he says. "The technology is mature enough now that there's a growing sense in the community that it's time [to test applications in the real world]. We've done all we can in the lab."

A handful of companies, such as online audio and video search engines Blinkx (http://www.blinkx.com/) and EveryZing (http://www.everyzing.com/) (which has licensed

technology from BBN) are making use of software that converts audio speech into searchable text. (See "Surfing TV on the Internet (http://www.technologyreview.com/Infotech/18545/) " and "More-Accurate Video Search (http://www.technologyreview.com/Infotech/18545/) ".) But the MIT researchers faced particular challenges with academic lectures. For one, many lecturers are not native English speakers, which makes automatic transcription tricky for systems trained on American English accents. Second, the words favored in science lectures can be rather obscure. Finally, says Regina Barzilay (http://people.csail.mit.edu/regina/) , professor of computer Science at MIT, lectures have very little discernable structure, making them difficult to break up and organize for easy searching. "Topical transitions are very subtle," she says. "Lectures aren't organized like normal text."

To tackle these problems, the researchers first configured the software that converts the audio to text. They trained the software to understand particular accents using accurate transcriptions of short snippets of recorded speech. To help the software identify uncommon words--anything from "drosophila" to "closed-loop integrals"--the researchers provided it with additional data, such as text from books and lecture notes, which assists the software in accurately transcribing as many as four out of five words. If the system is used with a nonnative English speaker whose accent and vocabulary it hasn't been trained to recognize, the accuracy can drop to 50 percent. (Such a low accuracy would not be useful for direct transcription but can still be useful for keyword searches.)

The next step, explains Barzilay, is to add structure to the transcribed words. Software was already available that could break up long strings of sentences into high-level concepts, but she found that it didn't do the trick with the lectures. So her group designed its own. "One of the key distinctions," she says, "is that, during a lecture, you speak freely; you ramble and mumble."

To organize the transcribed text, her group created software that breaks the text into chunks that often correspond with individual sentences. The software places these chunks in a network structure; chunks that have similar words or were spoken closely together in time are placed closer together in the network. The relative distance of the chunks in the network lets the software decide which sentences belong with each topic or subtopic in the lecture.

The result, she says, is a coherent transcription. When a person searches for a keyword, the browser offers results in the form of a video or audio timeline that is partitioned into sections. The section of the lecture that contains the keyword is highlighted; below it are snippets of text that surround each instance of the keyword. When a video is playing, the browser shows the transcribed text below it.

Barzilay says that the browser currently receives an average of 21,000 hits a day, and while it's proving popular, there is still work to be done. Within the next few months, her team will add a feature that automatically attaches a text outline to lectures so users can jump to a desired section. Further ahead, the researchers will give users the ability to make corrections to the transcript in the same way that people contribute to Wikipedia. (http://wikipedia.org/) While such improvements seem straightforward, they pose technical challenges, Barzilay says. "It's not a trivial matter, because you want an interface that's not tedious, and you need to propagate the correction throughout the lecture and to other lectures." She says that bringing people into the transcription loop could improve the accuracy of the system by a couple percentage points, making user experience even better.