DOCSOFT
Liberating Speech

case study

115 E. California Ave. Suite 415 Oklahoma City, OK 73104 405.236.2466 docsoft.com

Stanford Captioning Project

A Joint Project between Stanford University Online Accessibility Program and the Office of Accessible Education and Docsoft Inc.

Background

Located between San Francisco and San Jose, California, in the heart of Silicon Valley, Stanford University is recognized as one of the world's leading research and teaching institutions. As a leader in these fields, the University has developed an extensive network of multimedia web-based outlets which provide access to a wide range of Stanford-related digital audio and video content for students and faculty within the Stanford Community as well as the general public. Such outlets include ***Stanford on iTunesU***, which allows users to download courses, faculty lectures, interviews, music and sports for replay on an iPod or computer, various websites like **admissions.stanford.edu** which features different general interest videos and information promoting various aspects of Stanford life, and podcasts like the ***Academic Technology Report Podcast*** which is a production of the Academic Technology Lab.

Challenges

Working to provide accessible content for these outlets is a priority to the University and one of the primary goals of the Stanford Online Accessibility Program and the Office of Accessible Education. However, the challenges to creating captioned media for these outlets are complex and require specific skills, technologies and software that can truly stretch a university's budget to the limits to accomplish. Stanford needed a system to manage the workflow for captioning media that would address the university's needs without breaking the budget and that was simple and easy to use. The university identified the three biggest obstacles in creating captioned media on campus and entered into a pilot project with Docsoft developers to create a system that would overcome these obstacles:

1. **Converting media from various sources into a web-ready format that works with accessible web-based media players:**

   A web application that performs media conversion in a simple upload/download process is needed.

2. **Extracting text transcripts from the audio stream of a media asset:**

   Most speech to text solutions currently available do not provide the accuracy that captioning requires, therefore audio files must be outsourced to transcription providers and then managed appropriately. Flexible costs are a definite requirement as well.

3. **Synchronizing the text transcripts with the media asset:**

   Ensuring that the text is synched with the media file, the transcript needs to be time-stamped with start and end times throughout. This feature allows for search of specific time points within the video. An automated, efficient production process is needed for a large-scale production environment.

DOCSOFT
Liberating Speech

case study

115 E. California Ave. Suite 415 Oklahoma City, OK 73104 405.236.2466 docsoft.com

Solution

Docsoft developers worked to customize the workflow of the Docsoft:AV appliance. The Docsoft:AV appliance is a software and hardware solution designed to audio mine the spoken content in digital audio and video files. The customization added the ability to outsource and manage the transcription of media files and provide a programmatic means to convert media to standard media formats. In addition, a custom interface for users and administrators was created to allow web-based management of the system.

The customized system allows users to log into the system and upload their media files in most any format currently being produced. The system will convert the file into a suitable format for use in the customized Flash media player which is currently experiencing wide-spread use across campus. At the same time video files are being converted to a web delivery format, the file is also automatically converted to an audio format and forwarded to a pre-selected transcription company based upon the cost/delivery time selection chosen by the user. Upon transcription, the files are reloaded into the appliance and are automatically converted to an XML time stamped file. The time stamped transcript can now be integrated into the website along with the video which results in a closed captioned solution.

The development project resulted in a system that provides Stanford University with the capability for automatic video conversion while providing a standardized cross-platform media player. Additional capabilities of the system provide for searchable text transcripts with full indexing of media files and a means for providing affordable transcription solutions with accuracy and consistency.

Implementation

The project was initiated during the fall of 2008, with initial functional specifications determined with Docsoft. Between January 2009 and July 2009, multiple iterations of the system were deployed and tested, culminating in the launching of the service at a Tech Briefing on July 31, 2009. As part of the test phase, a number of legacy videos that experience high volume traffic at Stanford's YouTube channel were captioned and re-released. Plans to continue retrofitting older non-captioned videos are currently in discussion.

Results

Today, several outlets on campus are producing captioned media on a regular basis. The Academic Technology Report Podcast, delivered via the iTunes and iTunesU delivery channels, is a leader in captioned educational shows with its bi-monthly podcasts. The University's Entrepreneurship Corner (**ecorner.stanford.edu**) is pushing the edge of captioned media to new lengths by allowing users to toggle captions on or off and select videos to be translated into different languages. In addition, users have the ability to search for key words or phrases and simply click on the key word or phrase to be taken to that specific point in the lecture or video presentation.