Dean's Innovation Grants

Final Report

Date: 9/30/2019   Project Title: Making Digital Collections A/V Materials Accessible

Project Team Members: William Hicks

Project Purpose

This project sought to learn about and initially address accessibility compliance issues, specifically related to A/V materials in the UNT digital collections. It consisted of a review of applicable law, technical standards, and best practices, as well as an audit of digital collections' content. It also included the testing, evaluation of services, creation and review of methods for accessible media alternatives (e.g. caption files), and the development of code necessary to deliver said alternatives to users. A whitepaper will be available in the UNT Digital Library by October 31, 2019 presenting an assessment of costs, benefits, challenges, and opportunities, as well as practical recommendations for policies, procedures, and technical tools that will improve compliance and access issues.

Project Activities

- Attended the Conference and Workshops: "Accessing Higher Ground" in Denver, CO. Introduction to legal concepts, practical demonstrations and hands-on work with captioning tools. Discussions with experts on realistic expectations of user needs.
- Review relevant U.S. law, WAI standards.
- Write/Deploy code to Aubrey Content Delivery Systems (see outputs below)
- Generate and evaluate vendor supplied machine and human created caption files.
- Supervise student employees in the correction of machine-generated caption files.
- Contract with vendor for sample Audio Description alternative media files.
- A/V content Audit (see outputs below)
- Whitepaper (in progress)

Budget

- $1,734.79 – Accessing Higher Ground Conference (AHG), 5 days. Workshops and presentations on varying topics.

- $138.00 – Access to two archived International Association of Accessibility Professionals (IAAP) workshops (3 hours).
- $500.00 – temi.com – approximately 83 hours of automated Transcriptions/Captions.
- $500.00 – trint.com – approximately 33 hours of automated Transcriptions/Captions.
- $750.00 – Undergraduate wages – Correction of machine-generated transcriptions.
- $99.99 – Software License for "Closed Caption Creator."
- $660.00 – Video Caption Corporation – Creation of two audio-described videos (approx. 6 minutes of media)
- $1,200.00 – rev.com (split with Digital Projects operational account for budget overage). Human generated transcriptions/captions.

Project Outputs/Outcomes

Regardless of issues related to their legal necessity, responsibilities for creation, etc. prior to this project the digital collections were unable to store or display captions/transcripts of A/V materials. During the project period Software Development Unit personnel made necessary infrastructure modifications and I configured our video player to display closed captions from webvtt files where extant. Additionally, I wrote code to parse caption files and transform them into searchable, readable, and clickable-to-timestamp offset HTML transcripts. We now support multiple language transcripts, subtitle/translations, and 'chapters' for better discovery.

Examples:

- https://digital.library.unt.edu/ark:/67531/metadc1115874/m1/
- https://digital.library.unt.edu/ark:/67531/metadc489982/m1/
- https://texashistory.unt.edu/ark:/67531/metadc977400/m1/
- https://texashistory.unt.edu/ark:/67531/metapth779501/m1/

Unique to any digital repository (as far as we can tell), I added a "request captions" link to all uncaptioned media objects, allowing users a clear communication channel to make such requests. Our goal is to honor requests within 48 hours for UNT community members. We have received six such requests so far.

We generated over 100 hours of transcription files of varying quality. Some of this material has been made live, others have not (see anecdotal information below).

This grant allowed us to perform a content audit, our first near-comprehensive picture of A/V materials available in the digital collections. An audit is important in understanding how the Libraries should think about budgeting, resource allocation, and return on investment. The audit was a one-time snapshot of the collections generated through a programmatic parsing/analysis of a (mostly) undocumented API available for each published item in the

system. In short, I queried the SOLR index to return a list of identifiers for objects with a resource type of either "video" or "sound" and then downloaded a copy of each item's "resource object" as JSON via a shell script and wget. Using Python, I evaluated each file and wrote relevant data (metadata fields like dates, collections, durations, etc.) to CSV which could be analyzed in Excel. Since we do not make resource objects available for unpublished items a less comprehensive (and less exact) analysis was performed by reviewing the output of a SOLR query on hidden A/V materials, returning only physical descriptions. In many cases this contained duration and the presence of sound or silence. From this we learned a number of things, a few highlights of which are:

- We currently have about 62,000 published objects and a backlog of 31,000 unpublished items in the digital collections. The bulk of unpublished items come from the KXAS-NBC 5 Collection; several hundred move from an unpublished to published state every month.
- Prior to the grant we had not calculated how many minutes of A/V we make available online. Since transcription services are priced on a per-minute basis, this is an important figure to be aware of. The audit finds that we have approximately 900,000 published minutes (620 days' worth) of material available through the digital collections. The lengths of each recording vary considerably with files as short as 15 seconds or less, and others exceeding 3 hours. Most files fall into a range of 45 seconds to 15 minutes in length.
- The audit helps us understand our average yearly output, with a total period average of approx. 75,000 minutes added per year, generally trending more recently towards 115,000 MPY.
- Roughly 99% of our objects have sound, but there are over 42 hours of silent video contained in the collections. Detecting silence is an important area of exploration since we should try to exclude those from submission to vendor transcription systems for cost savings, but also because they pose a difficult research medium for the blind who have no meaningful way of evaluating the material. Though a somewhat contentious assertion, these materials, along with other non "talking head' videos in the digital collections may need to be considered for a much more thorough metadata descriptive record, descriptive webvtt alternatives or audio- or extended-audio descriptive alternative media, all at a much greater expense than simple captioning services.
- Based on values in metadata records over half of all audio/video minutes contain music of some kind, with over 100,000 minutes or recordings that contain singing or other vocal performance. Instrumental music, of course, does not need to be transcribed, but chapter/section markers would benefit many users; similarly, choral and other sung works would prove extraordinarily difficult/impossible to transcribe via automated services though they are arguably required to exist.

Anecdotal Information

An important part of the project involved evaluating the quality of machine generated and human-interceded transcripts. This is an important area of consideration since machine generated transcripts are becoming increasingly accurate and generally cost a fraction of what human-generated ones do.

I evaluated two vendor services that provided automated transcriptions and one where the vendor employees generate or correct machine transcriptions. Machine transcriptions tended to have success rates of perhaps 85-95% depending on a variety of factors. Generally slow, clearly enunciated, well-miked English speech performed well. Speakers with accents, regionalisms, rapid pacing, or who sung, performed less well. One service was overly aggressive at capturing stutter-starts ("ums") while the other took a less literal approach, cutting and cleaning poor speech habits. Both had trouble with complex sentence structures and punctuation. Inconsistencies in numbering and unique/special terms and proper nouns also made it hard to consider these services as serious options for research-grade materials where we would likely hope to achieve a 99% accuracy rate, but at 1/10 to 1/4 the price of a human generated service, it may be worth considering for some collections. Both services generated transcriptions within only a few minutes of their upload time, provided APIs, methods for batch imports, file renaming and management, and both had capable, modern, through-the-web editors for corrections and file export to webvtt standards. Human interceded services by contrast returned files usually within a few days and with the exception where the source media contained cross-talk, poor audio-quality, or contained non-English languages, returned near-perfect results.

Part of my evaluation involved employing student workers to correct machine generated content. In some ways the workflow is similar to metadata authoring practices we already use in the Digital Projects Unit, but using different tools. Once corrected, the outputs appear to be of a similar quality of those by vendors who provide human generated/corrected transcripts of their own. A number of modern browser-based editors exist to generate transcription files, either from scratch or from existing source files, and in the end a number of factors will likely affect the price; in the end, I believe that the costs of employing student workers would be slightly more expensive in the long run, but would come with certain advantages in limited cases:

- First, at the present time, most transcription services are English-only or offer foreign transcriptions at either a much higher rate or through machine-only generation with varying quality. When considering foreign language, or as is the case with a number of our oral histories, multi-lingual content, it might be wise to employ a bilingual Spanish-speaking student who can transcribe and potentially translate content.
- Second, musical transcriptions are virtually impossible today; but some measure of interpretation would be possible by experienced music library students/staff with access to scores, libretti, etc. Local experts would be well suited to generate either lyric-based transcriptions or add timestamped markers for movements, sections within pieces, etc. Such a service would be largely novel in the world of digital libraries, and however

ethically desirable, or valuable such enhancements might be to discovery, the legal and practical necessity is somewhat unclear.

I found that while the content audit was an important first step in understanding the scope and shape of our collections, it was far from perfect. Metadata records with errors or omissions were the most common problem encountered that likely effected statistical analysis, but many issues related to A/V accessibility are quite subjective and, in the end, a full analysis would require some level of human evaluation of many more files than I was able to observe. Some 'silent' videos for example, contained no digital audio track, others contained a single test tone in the opening few seconds of the footage, while still others have an audio track which contains no audible sound. This would naturally have effects on automated testing tools looking at decibel levels or for the presence of an embedded track. To remedy this we might look to visualization tools to create waveforms from audio signals, but again this would take some concerted level of inspection and staff time. Similarly, issues abound as we look to other tangentially related issues: some videos had poorly miked speakers, leading to larger questions of responsibility for source file integrity, and the appropriateness of pre-processing audio before ingest. If a file is known to have poor audio quality before adding it to the repository and we have tools to correct them and make the materials accessible to a wider audience, should we take the time or bear the cost to do so? Similarly, no current methods are in place to evaluate strobing or flashing effects or warn users of their occurrence. While a rare occurrence, user with epilepsy should be alerted to potential hazards should we know of their existence.

Finally, though it featured as a less significant part of my overall development time and tests, one area of pressing concern was exploring media alternatives for blind users, typically provided as an audio-described file. I contracted with a vendor to create an "audio description" and an "extended audio description" of two files totally 6 minutes. The cost per minute was extraordinary, and reveled problems we (and others) face with such content.

1. Virtually no digital library provides such alternative files, even though WCAG require them of certain types of video footage. A concurrent survey of accessibility accommodations in digital repositories I am currently administering has over 100 respondents with only two claiming to provide audio descriptions for their video files.
2. Given our infrastructure, providing such files would require duplicate objects tied together by metadata reference fields. Though some providers like Netflix are able to supply audio description as a separate, toggleable concurrent track to the original, our player is unable to support that feature at this time.
3. Vendor supplied scripts were well voiced, but demonstrate that the content of the audio description is itself, an interpretive act, and that the individual needs of the user would likely dictate what should be described in the context of a research-grade video. Students of art, iconography, or fashion history might have different questions of a news clipping than someone interested in politics for instance.
4. The back-and-forth process of vendor-created audio described files took weeks to complete. Such delays would be highly problematic for researchers.

Best Practices

Aspects of this project could be repeated at other institutions and would likely have similar results; differences would likely occur in the following ways:

- Content audits are largely dependent on reporting, exporting, querying abilities of repository software. The scope of materials and audience would heavily influence the shape of the resulting data.
- Vendor-generated transcriptions will likely fluctuate in pricing and accuracy, in all likelihood becoming less expensive and more accurate with time. We would expect accelerated improvements in transcriptions of regionalisms and foreign language materials and the ability to transcribe multi-lingual content with greater ease than is currently possible.

Program Continuity

The white paper will make some targeted suggestions but generally speaking the following are important concepts that will need to be considered in the future:

1. The libraries should monitor pending litigation and ongoing appeals regarding accessibility in universities and other places of public accommodation as several major cases continue to work their way through the courts. Importantly, since the UNT Digital Collections contain materials created by both members of the UNT community and partnering institutions/organizations, and since it contains works that were generated for broadcast television, the libraries should contract, consult, or otherwise seek clarification and authoritative legal analysis regarding UNT's obligations under The Rehabilitation Act, The Americans with Disabilities Act, and the 21st Century Communications and Video Accessibility Act regarding the specific materials we make available to the public.
2. Subsequent to such clarification, the libraries should seek to publish a detailed explanation of both our responsibilities and obligations, and those of our partners on the digital collections sites through both our accessibility pages, policies, and a VPAT. Partners should be notified of their legal obligations, remediation options, and future responsibilities related to A/V deposits etc.
3. Within the libraries we will need to decide how to move forward with funding captioning and other accessibility-related remediation efforts in the digital collections; accessibility compliance should form part of the regular workflow.
4. Library personnel and administration will need to decide how funds are allocated within divisions and who is responsible for bearing the cost of captioning or other services. This requires hard questions to be asked and answered. Is it the Music Library and the Special Libraries Division who should fund content provided by the UNT School of Music? Who funds captioning for "State of the University" addresses by the President,

videos of student organizations, or external partners such as the Abilene Library Consortium? Is that cost only applicable to new deposits, retrospective, etc.? Should/could the Libraries offer or use grant funding to cover these costs in similar ways that we provide "Rescuing Texas History Mini Grants"

5. We will need to discuss if it is appropriate to provide machine generated transcripts at no cost, while providing methods for cleanup/editing or human-interceded versions as a value-added service. Could we provide partners with access to or training in caption file editing and allow them to fix automated transcripts themselves? This would closely align with how YouTube currently works.

6. Metadata records currently indicate the presence of "sound" or "silence" with abbreviations "1 film (2 min., 9 sec.) : sd., b&w ; 16 mm." or "1 film (2 min., 9 sec.) : si., b&w ; 16 mm." in a physical description field. Such abbreviations are unclear. Silent video, at least should be explicitly stated as such somewhere in the record.

7. For users with epilepsy or vestibular disorders, videos with strobing, pulsing, or other rapid visual effects should make note in their metadata record using a field that can be displayed on the video player page.

8. Metadata authoring guidelines on the library website and embedded in the metadata editing system should be updates to reflect best practices related to accessibility.

9. Videos identified as silent should provide a simple webvtt caption with a single declaration indicating "no sound."

10. Providing audio-described alternatives to video content is an extraordinarily expensive and potentially time-consuming process when outsourced to a vendor, but is likely required for compliance with some content we host. Aside from issues related to who would bear such costs, it would be wise to consider developing policies and procedures for this type of alternative media. Several organizations offer trainings for audio describers and the service is becoming more common in arts organizations, concert venues, and theaters. With access to tools such as Adobe Premier and FFMPEG a training manual and workflow could be created for inhouse production of these sorts of files. As an alternative, webvtt descriptions could be created to produce a lower-fidelity text format, but this would also rely on inhouse content generation.

11. Digital Projects Unit personnel put on occasional "metadata days" where staff and students are invited to learn how to create metadata records. A similar program should be devised to demonstrate how to create/edit webvtt transcriptions.

12. In coordination with other members of the Library staff a permanent workgroup or committee should be chartered to address these and related needs found throughout the organization. Significant overlap of issues can be found within Collection Development in the assessment of electronic resource acquisition, and likely also with facilities, circulation, and other library sites and systems.