Thorsten Ruprechter, Bsc

Increasing Efficiency In Video Transcription

Final Research Paper

submitted to

Austrian Marshall Plan Foundation

Supervisor
Assoc.Prof. Dipl.-Ing. Dr.techn. Christian Gütl
Graz University of Technology

Supervisor
Assoc.Prof. MSc Ph.D. Foaad Khosmood
California Polytechnic State University

Graz, May 2018

Abstract

Legislative and parliamentary proceedings present a rich source of multi-dimensional information that is crucial to citizens and journalists in a democratic system. At present no fully automated solution exists that is capable of capturing all the necessary information during such proceedings. Even if professional-quality automated transcriptions existed, other tasks such as speaker identification, entity disambiguation, and rhetorical position identifications are not fully automatable. While many governments rely on expensive, manually-produced transcriptions and annotations, others are left entirely without digital transcriptions. In this paper, we present a study using the Digital Democracy transcription tool. Human transcribers work to up-level and annotate California state legislative proceedings using the tool. Four phases of UI and functionality improvements are introduced and for each phase, the resulting change in efficiency is measured and presented. We work with a set of 7304 individual transcription sessions (1290 hours of video) where each session is the record of one bill discussion. We further concentrate on a set of 2800 sessions belonging to a single cohort of 20 editors who have experienced four versions of our transcription tool. We find that through introduction of features in the transcription tool, we can improve human assisted transcription efficiency by 10.7 percent over 3 phases. Our analysis regarding phase 4 remains inconclusive at submission time, due to not enough completed tasks being available.

Acknowledgements

Some of the contents and results of this paper were submitted to the 19th Annual International Conference on Digital Government Research organized within the European commission program on Interoperability solutions for public administrations, businesses and citizens (2018) as a research paper titled "Gaining Efficiency in Human Assisted Transcription". I want to thank Christian Gütl (Graz University of Technology) as well as Foaad Khosmood, Toshihiro Kuboi, and Alexander Dekhtyar (California Polytechnic State University) for their contributions and help during the creation of this paper. Although the initial submission was declined, a poster version was accepted to the conference.

The authors thank the Institute for Advanced Technology and Public Policy for providing the data used in this paper.

Contents

List of Figures

1 Introduction

1.1 Digital Democracy

The main goal of the *Digital Democracy* initiative is to provide full insight and access into US state legislative processes. For this, videos of legislative committee hearings are combined together with auxiliary information such as searchable transcriptions, bills discussed in hearings, identification of participating speakers like legislators, lobbyist, witnesses, and members of the general public. Furthermore, for legislators, their full legislative service history as well as campaign contributions and gift data are tabulated and presented to the user. The Digital Democracy initiative created and maintains technologies which link all of this information together. It provides an online platform which brings transparency into state legislative hearings and government proceedings that would otherwise not be accessible to the broader public. Citizens, journalists, and researchers would not be able to easily gain information about the content of those hearings and debates. In addition to giving insight into the proposal of new bills and laws, it also presents a chance for the general public to monitor lobbyists, lawmakers, and advocates. All information can be searched and queried, and the results are high quality thanks to human-assisted transcription. Digital Democracy started out being a platform only for the California legislature, but has since evolved to also cover the states of New York, Florida, and Texas. Practical usage of the aforementioned platform can be examined by visiting www.DigitalDemocracy.org. Figure 1.1 shows an example of how Digital Democracy links transcription and metadata to the actual hearing video.

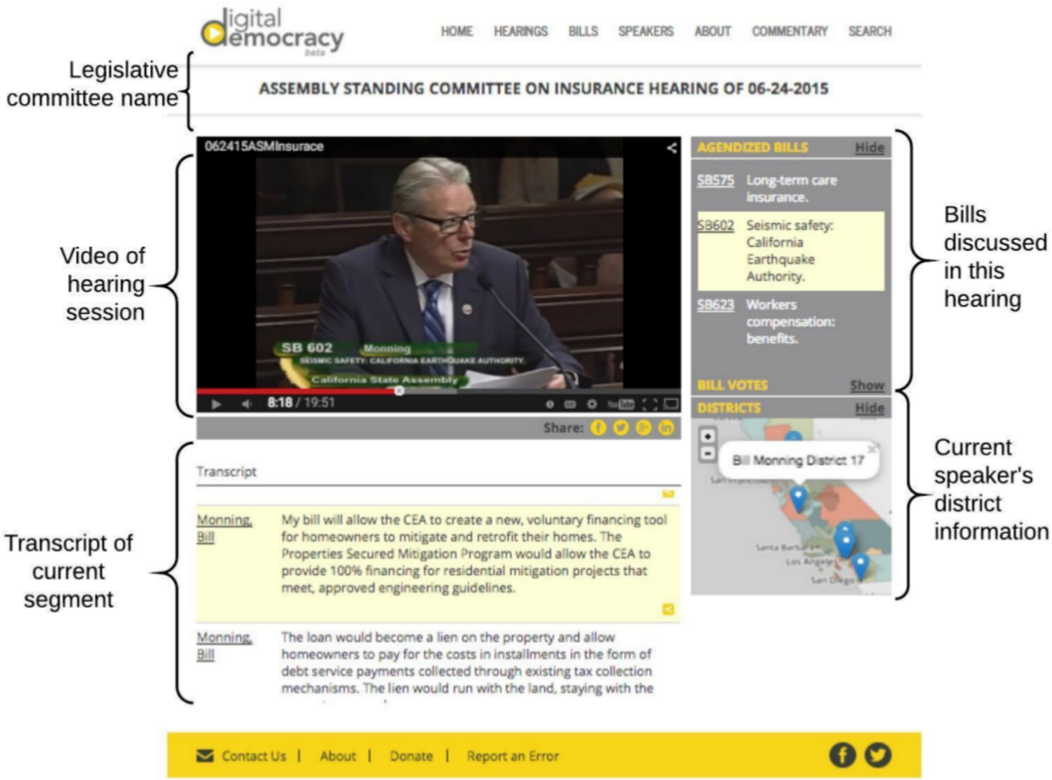

Figure 1.1: Transcript Display on the Digital Democracy Website

1.2 Transcription Process

Digital Democracy uses a human-assisted approach for generating transcription texts and metadata. While automatic transcription might be sufficient in other areas, a legislative setting requires professional and correct transcripts. This is achieved by human editors manually up-leveling transcription text and performing annotations such as speaker and position identification. The human editor use *Transcription Tool*, a software developed by the Digital Democracy project team to enhance transcripts.

1.3 Contributions and Research Questions

The main contribution made by the author of this research paper is the improvement of the tool used by human transcribers as well as measuring the effectiveness of these improvements. To scientifically tackle the problem of improving Transcription Tool and measuring efficiency, three research questions were posed.

First, an evaluation of the current interface has to be conducted. Before starting development, it is important to know which new features would speed up transcription progress. Another research question is to what extent do the improvements to the transcription tool increase efficiency in human assisted transcription independent of the task, state, or the particular people doing the work. Lastly, it is up to debate which aspects of this work yields the largest efficiency gain. To answer the third question, we must break down the various aspects of human assisted transcription and measure their output independently to the largest extent possible.

1.4 Overview

The remainder of this paper is organized as follows: First, Chapter 2 introduces the Digital Democracy initiative further. Chapter 3 provides related works both for similar legislative transparency efforts around the world and the state of tools comparable to our own. Chapter 4 describes our own tool and its capabilities. Chapter 5 reiterates the research problem and questions introduced above. Chapter 6 discusses our experimental design. Chapter 7 presents the results and finally, Chapter 8 provides conclusions and future work.

2 Digital Democracy Initiative

2.1 General Information

In the USA, state legislatures hold a lot of power. Unfortunately, monitoring and therefore holding them to account is difficult. It can be a time-consuming task for the general public to gain insight into the workings of state government. Official records about the content of state legislative hearings are only sparsely available.

Politically speaking, states are situated between federal and local government. State governments enforce policy and spending mandates on municipalities and local government while deciding on how federal budget is used. Even in times when federal government comes to a halt due to elections or political conflict, states still resolve policy issues and enact new bills and resolutions. Due to the massive amount of laws passed and altered each year, lobbyists are interested in influencing state legislators. Without a proper platform offering insight into the content of legislative hearings and the discussions happening during them, the public is shut out of the process. This is especially a problem for the news media, which also has problems finding information about simple facts such as bill text changes or votes. Without proper records being created, it is difficult to discover details about debates and negotiations about a bill. The position of people discussing an issue might be hard to guess without knowing details about the ongoings inside state house or senate.

As already mentioned, the focus of the Digital Democracy initiative is to increase transparency in government. The corresponding website provided by Digital Democracy (www.DigitalDemocracy.org) can be described as a statehouse accountability platform (DigitalDemocracy, 2017). It creates a searchable, verbatim record of all statements, whether they were made by lawmakers or witnesses during hearings in legislative committees and floor sessions in statehouses. The website opens up legislative proceedings to everyone and provides them with a simple interface to query for specific information in the full-text of the transcript, as seen in Figure 2.1.

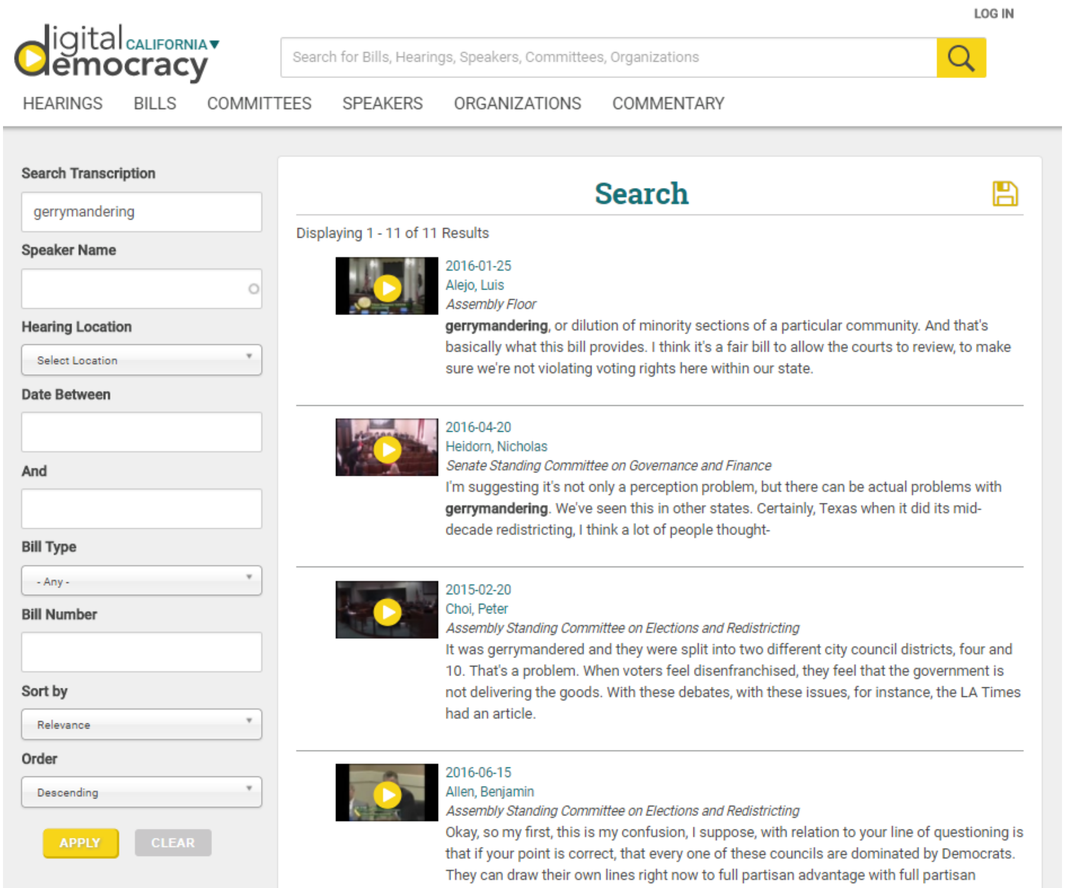

Figure 2.1: Search Functionality of the Digital Democracy Website

After a successful search request, the website provides a list of results. Each entry in the result list holds the part of the transcription text in which the search term occurred as well as a direct link to the hearing where it was mentioned. The link takes the user to the web page of the specific hearing and sets the video to the exact position where the search term was mentioned. This interface was already introduced in Chapter 1 and can be seen in Figure 1.1.

While this simple search interface is the centerpiece of the Digital Democracy website, one is also able to find additional information about state government there. Users can browse through information about hearings, bills, committees, speakers, as well as organizations and lobbyists.

2.2 History

The Digital Democracy online platform is accessible to the public since early 2015. In an official press release issued by the Public Affairs Office of California Polytechnic State University (2015), former state senator Sam Blakeslee mentioned that Digital Democracy was developed to "open up government". Also, he stated that there are no transcripts produced by the California Legislature during state legislative hearings. He further explained that due to this fact the public would have no way of seeing what really happens in hearings.

Blakeslee's claim was backed by a poll conducted by the Institute for Advanced Technology & Public Policy (IATPP) at Cal Poly San Luis Obispo, which he directs. The results of this survey show that a large amount of Californians would highly approve of changes to the Legislature's transparency. Regardless of party or ideology, citizens demand more insight into politics. Most people find it is especially important to have further access to information about budgeting. In addition to that, the interviewees were also very interested in being able to access documents and searchable information online. (Myers, 2015)

During explanations of his improvements to the Digital Democracy platform, Rovin (2016) mentions the F grade given to California by Davis and Baxandall (2014) in a comparison of access to online information regarding spending information. In this report published by the CALPRIG Education Fund, California placed last. In 2016, another investigation mandated by the CALPRIG Education Fund was carried out by Surka and Ridlington (2016). While other states improved their transparency policies and websites, California again came in last out off all states and received an F grade. No improvements to transparency policies were attempted by the state. This further shows the still existing need for a publicly accessible platform providing information about political proceedings, which the Digital Democracy initiative provides.

Although Digital Democracy was initially supposed to only focus on

California, other states are also lacking transparency regarding government proceedings. It soon became apparent that the ideas which led to founding this project are also applicable to other states in the USA. Therefore, the initiative integrated hearings held in New York into the system in early 2017. As of January 2018, the most recent additions to the selection of states for which Digital Democracy offers searchable information online are Florida and Texas. There have already been talks with stakeholders from Colorado, Michigan, Nebraska, and North Carolina to further expand the system into these states. However, due to an ever-growing amount of data and the need to keep costs as low as possible, this is a difficult task. At the moment, scaling the system to cover these states would be an immense challenge.

2.3 System Design and Transcription Process Phases

Presentation of information on the Digital Democracy website is only a fraction of what the project actually achieves. Before data can be queried and browsed properly, many preprocessing steps must be taken to provide correct information to people using the service. Some of these steps will be elaborated in more detail below.

The Digital Democracy initiative, while also influencing the political landscape, has technical issues to tackle and tasks that need to be solved. For example, correctly transcribing videos of hearings and detecting which person is currently speaking is one of these challenging tasks. Because of the difficulties caused by this process, speech transcription in the Digital Democracy Initiative is performed in two separate phases.

Firstly, an automatic speech recognition service transcribes the video of a hearing. However, this process of converting audio into text is error-prone. In addition to that, the speaker making the current utterance cannot be easily identified automatically. Therefore, a second phase is needed in which a human manually corrects errors of the speech recognition service, and adds annotations such as information about the current speaker. To ease this manual correction a web application, called the Transcription Tool, was introduced. While this tool is helpful in assisting transcribers, it still has some flaws and needs improvement. Transcription Tool will be explained in more detail in Chapter 4. Due to limited resources, both costs and time required to complete automatic processing and manual correction must be reduced as much as possible.

In addition to the before-mentioned automatic speech recognition service, other preprocessing steps are taken. Depending on state, year, and committee specific scripts pull data from third party sites. That metadata is then incorporated into the Digital Democracy Database and connected to the rest of the hearing data. Since this research paper mainly focuses on the improvements made to the Transcription Tool, it would be beyond the scope of this document to go into detail about the exact data sources. For example, additional information for California hearings and the people speaking in these discussions is extracted from MapLight, "Cal-Access," and The California Channel.

Most of the above-mentioned steps are part of an automatic (albeit partially human-assisted) pipeline. This preprocessing pipeline will be explained in more detail in Chapter 4.2.

3 Background and Related Work

In this section, literature search findings of mainly two areas are presented: Other government transparency efforts that includes recording and exposing legislative proceedings, and a survey of major tools available for similar transcription and annotation purposes comparable to Digital Democracy.

3.1 State of Government Transparency Around the World

In contrast to the problematic situation regarding government transparency in USA state governments, policies differ at federal level as well as in other parts of the world. Below we investigate some of the transcription systems which allow monitoring of lawmakers in the congress of the USA, the Commonwealth of Nations, and Europe.

Although American state governments do not provide searchable video or any transcripts to the general public, resources are available for congressional proceedings of the federal government. C-SPAN broadcasts congressional hearings and creates searchable transcripts which are linked to the video source (C-SPAN, 2018). In addition to that, citizens can access congressional records online (Library of Congress, 2018). However, search functionality is limited for these proceedings.

The British parliament provides official transcripts of parliamentary debates in a searchable manner. These records, officially called "Hansard", are accessible to the public via the parliament's online presence (Parliament of the United Kingdom, 2018). During a sitting day, an online version of this day's proceedings is published gradually, with the full Hansard being available the next morning. Similar to Digital Democracy, the website provides a search interface which allows to query transcripts for specific terms. It directly links the search results to the exact position in both textual transcripts and video recordings. Some of the Commonwealth Nations also implement such searchable Hansards. For example, the Parliament of Australia (2018), the New Zealand Parliament (2018), and the Parliament of Canada (2018) all offer Hansards on their websites while creating links between plain text records and corresponding videos. Figure 3.1 shows an example of the comprehensible search interface provided by the British government.

The European Union also provides online access to debate videos and verbatim texts (The European Parliament, 2018). Although agendas, reports, adopted texts, and audio recordings of debates are made available in all official languages of the EU, segments of speech in verbatim reports are not translated to all languages. This effectively means that transcripts only contain speaker utterances in the original spoken language, which is mostly not English. Although speeches contained in the report are linked to video sources and therefore translated audio, no full text search of all plenary sittings is available.

Similar to the EU, the parliament of Switzerland only creates textual transcripts in the original language of the speaker. However, a searchable official bulletin is made available which allows users to issue full-text queries on statements of council members (in their original language) and provides links to transcripts and videos (The Swiss Parliament, 2018).

In Austria, stenographic protocols of plenary sittings of the parliament are made available online (The Austrian Parliament, 2018b). The full-text of these records can be searched using the advanced search functionality provided on the website of the Austrian parliament, but there is no direct link from found text to video recordings (The Austrian Parliament, 2018a). As of now, there also exists no permanent video archive of past parliamentary sittings. However, sittings are streamed live on the parliamentary website and are stored for seven day on demand access in the archive of the Austrian national public service broadcaster (ORF, 2018). Most of this functionality and information is only available in German. Since the official services of the parliament in Austria do not keep track of votes, the website Addendum.org provides this service to the public by manually counting the physical votes of the representatives attending a sitting (Quo Vadis Veritas Redaktions GmbH, 2018).

3.2 Transcription Editing And Annotation Tools

Although automatic algorithms nowadays manage to generate reasonable transcripts as well as additional contextual data from digital sources, human-assistance is still mostly necessary to ensure correctness on a professional level. Due to this, tools for both research and commercial use emerged in the past years. However, transcription software differs strongly from field to field.

Entity tagging or annotation tools such as introduced by Stenetorp et al. (2012), Papazian, Bossy, and Nédellec (2012), or Widlöcher and Mathet (2012) focus on linking metadata to plain text. They allow users to create links between existing or newly created entities as well as their text occurrences. In addition to that, further details such as description of relationships can be specified. Although such annotation tools fulfill a different purpose as transcription tools do overall, both create metadata for plain text records. Entity information created by transcription tools can be used to derive more in-depth information about interactions and relationships existing in the current setting. Such information could for example be represented by speaker identification and speaker alignment regarding a currently discussed issue.

Audio transcription tools must allow the user to pause, rewind, or in any other way manipulate the currently investigated files while editing transcripts. Even though audio tracks are a different media type than video, many audio transcription systems are similar to those handling video in that the initial text presented to the user is also created by automatic speech recognition (Luz, Masoodian, Rogers, & Deering, 2008; Burke, Amento, & Isenhour, 2006; Revuelta-Martìnez, Rodrìguez, & Garcìa-Varea, 2012; Whittaker & Amento, 2004; Basu, Bepari, Nandi, Khan, & Roy, 2013). Some tools provide the option to investigate waveform representation of played audio, such as the one implemented by Luz et al. (2008). The main difference between transcription tools developed for audio and video is that the latter could be used to derive additional information. An example for such information is that it might be easier to identify speakers based on their physical on-screen appearance than by voice or text alone.

Software which assists users in transcription of handwriting or ancient texts has to properly handle display and navigation of texts or still images, so single words and characters can be properly deciphered (Toselli, Vidal, & Casacuberta, 2011; Castro-Bleda et al., 2017). These tools use raw transcripts produced by optical character recognition technologies as their primary source of text. Image analysis was also incorporated in the preprocessing pipeline of Transcription Tool, where results of automatic face recognition help transcribers to identify speakers.

The subset of tools most similar to Transcription Tool are those used to create and edit video subtitles, captions, or transcripts. Although subtitle and caption editors do not fall into the exact same category as those focusing on transcription texts, they still provide valid input for user interface decisions and features implemented in a transcription tool. One of the earliest approaches of providing a software package for creating and synchronizing video transcripts was introduced by Nivre et al. (1998). The authors chose to split up the main components into separate tools. *TransTool* is used to generate the transcription texts while *SyncTool* synchronizes transcripts with video recordings by enabling the user to manually set time codes. Seps (2013) created *NanoTrans*, a tool which allows for creation of both textual and phonetic transcriptions. Besides the usual approach of providing UI panels for both transcription and video, *NanoTrans* also includes a panel visualizing the audio track in waveform. In addition to that, a button panel is available to insert non-speech event tags into the transcript. A different approach was taken by Deshpande, Tuna, Subhlok, and Barker (2014), who developed *ICS Caption Editor*, a crowdsourced caption generator which enables students to collaborate on correction of pre-processed captions generated from lecture videos. To improve quality of texts, users can request a review on problematic or complicated sections directly through the editing interface.

Besides these research-oriented transcription tools, commercial software exists which provides assistance in editing video transcripts, captions, or subtitles. While most of these tools share the same functionality, some have unique features or user interfaces. NowTranscribe Ltd (2017) offers an audio transcription tool which shows the automatically generated transcripts in light gray. Pressing the tab key accepts the currently displayed word while pressing any other key will allow the user to modify the text. The automatic transcription services provided by cielo24 (2018) come with a sophisticated transcription editor. This tool is the one most similar to Transcription Tool functionality wise. Some of the features of this editor are: allowing users to navigate between utterances using a button panel, jumping to specific parts of a video by entering a timestamp, adding speaker information to utterances, the option of auto-pausing the video player while a user is typing, hotkeys for navigating the transcript, and shortcuts for adding sound tags.

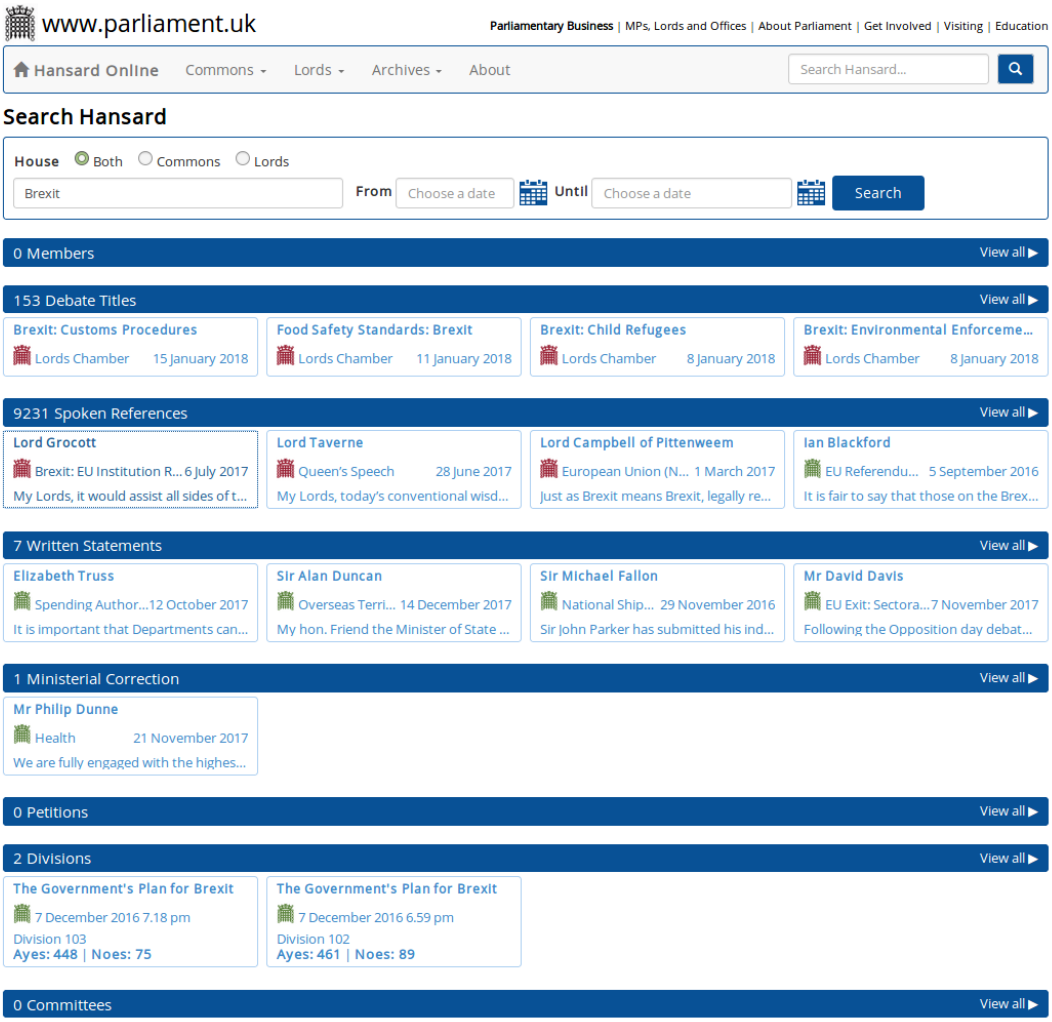

Figure 3.1: Screenshot of an Example Search Using the Interactive Hansard Provided by the British Parliament Website

4 Digital Democracy Transcription Tool

4.1 General Functionality

The Digital Democracy Transcription Tool serves as the main source of semi-structured data for most of the information and content provided by the Digital Democracy initiative and its website. One of the tool's purposes is to handle administrative tasks such as importing hearing videos, updating hearing metadata, or supervising the transcription process. Besides this administrative usage, Transcription Tool is used by human transcribers to edit automatically generated transcripts. Due to different usage scenarios, two user roles exist: "Admin" and "Editor". Admins handle administrative tasks while also being able to edit any transcript they want. All transcribers are assigned the "Editor" role which limits them to only access transcripts which were specifically assigned to them.

Transcribers enhance the previously automatically generated transcripts by using the tool's transcription user interface. Part of the transcription screen can be seen in Figure 4.1.

The necessity for human transcribers stems from the fact that the textual transcripts produced by automatic systems are not high quality enough for professional and government purposes where slight alterations could be important. Especially legislative bill information and personal names are not always correct. To fulfill the professional requirements demanded for proceedings of legislative hearings, these mistakes have to be corrected. Additionally, editors must identify speakers and decide on their alignment regarding the current issue discussed in the hearing. Lastly, editors work to standardize utterance length by merging or splitting utterances. This mainly serves the purpose of allowing proper presentation of transcripts on the Digital Democracy website and making sure no utterances exist that are too short to stand alone. In the context of Digital Democracy, an utterance can be defined as one to many sentences spoken by the same speaker containing a line of thought.

Before transcripts are given to editors, a preprocessing pipeline is executed to enhance the results produced by automated transcription services and annotate additional metadata. This pipeline takes official hearing videos and subtitle files which were produced automatically as an input. At the end of the pipeline, Editors are then assigned segments of a hearing video for which automatically generated transcripts are already available, which they then correct and enhance. These segments of video are called transcription *tasks* and are automatically generated by the tool, making use of the previously entered hearing information. A more detailed overview of the task generation and preprocessing pipeline will be given in the following section.

4.2 Human-Assisted Transcription Pipeline

Figure 4.2 visualizes the human-assisted transcription process and the task generation process as an activity diagram. Most of the automated parts of the pipeline visualized in this figure are beyond the scope of this thesis. However, it may be important for the reader to understand where and when the information and metadata the editors use as a starting point for their transcription work were generated. Therefore, the following few paragraphs will be used to describe the transcription pipeline in more detail. Please note that this diagram is modeled after the state of the preprocessing pipeline in spring 2018. As already mentioned, Digital Democracy is constantly evolving. This leads to parts of the project such as Transcription Tool and its task generation logic, the preprocessing pipeline, or auto-correction scripts constantly being updated.

The transcription tool processing pipeline uses hearing videos which were either recorded from a live stream or downloaded from a video archive as an input. The actual video source is dependent on the state. After the video was successfully stored and indexed in the Digital Democracy video archive, automatic trimming and cutting is performed. This shortening and partitioning of videos into clips serves multiple purposes. First, recordings of hearings might contain silent periods at the beginning and end of a

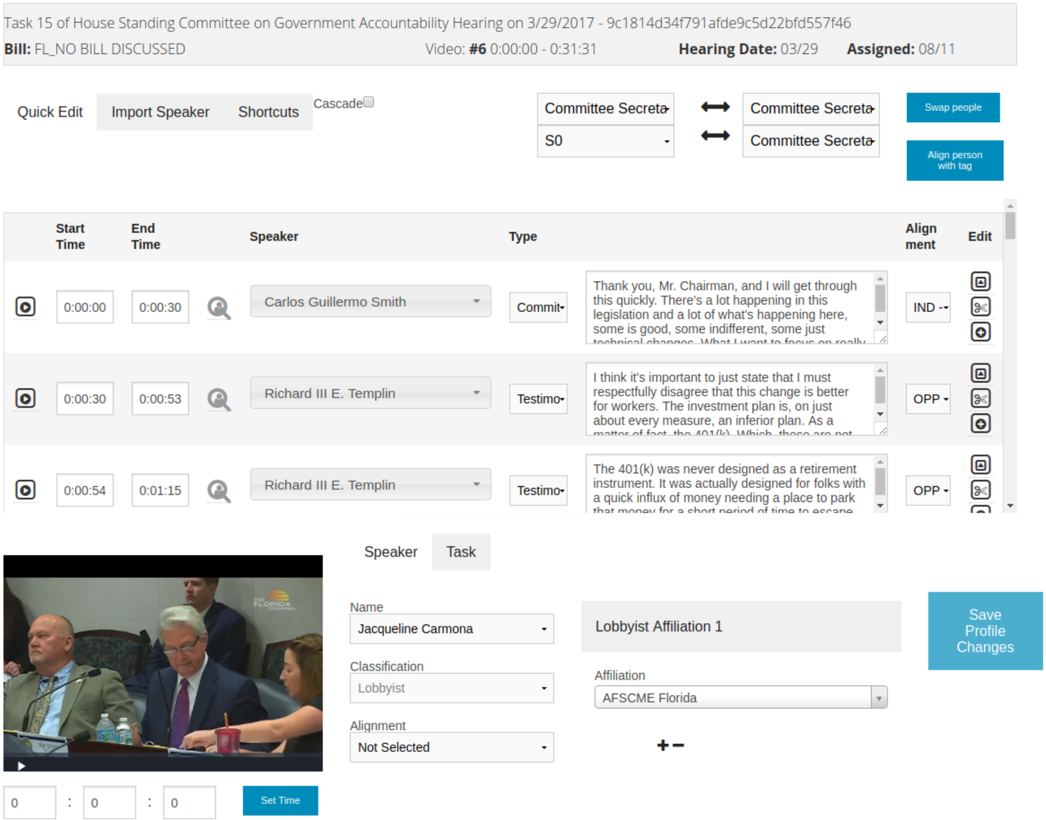

Figure 4.1: Main Screen of Transcription Tool's Transcript Editing Interface

video, which have to be removed. Second, videos are cut into separate clips with a length of about thirty minutes. This guarantees faster generation of automated transcripts by the external transcription service.

When trimming and cutting is finished, the hearing shows up in the admin interface of the Transcription Tool. Admins can then review the clips and make adjustments by manually trimming or cutting it. If everything is in order, they send the video to an external transcription service via the user interface. As mentioned earlier, Cielo24 (cielo24, 2018) is the currently used service to generate automated transcripts. In the near future, Digital Democracy plans to also enable usage of other transcription engines such as Watson (IBM, 2018) as a possible transcription service. The video transcripts returned by the external services are stored in SubRip subtitle (SRT) files.

Each file holds the full transcript of a single video, fragmented into short intervals of speech (utterances).

Submitting a hearing video to an external transcription service also triggers a first diarization using only audio. For this diarization of utterances a toolbox introduced by Rouvier, Gay, Khoury, Merlin, and Meignier (2013), called LIUM_SpkDiarization, is currently used. During this process, speaker tags are assigned to separate segments of the video to determine when speaker changes occur and which person spoke when. Each one of the generated tags represents a different, unidentified speaker. Producing this information is one of the most crucial steps, since work done by human transcribers is much easier (and therefore faster) with correct speaker tags.

After sending the video to the external service, Admins add more information to the hearing. For example, they annotate the hearing with committee and bill discussion information. Adding this information is called "Bill Tagging". In every hearing, multiple bills could be discussed over different or overlapping periods of time. Creating this data is essential to provide proper bill information for the Digital Democracy initiative as a whole. It also mandates how the Transcription Tool generates tasks for this hearing video at a later point in time.

Admins can also choose to use in-house voice, face, and text (VFT) analysis services to generate more sophisticated diarization using audio and textual transcripts. However, the main benefit of the VFT system is its speaker prediction algorithm. This prediction is only carried out if a speaker model applicable to the currently processed legislative hearing is existent. VFT-analysis attempts to link each of the formerly created diarization tags, which represent unidentified speakers, to an actual person. Each tag is aligned with several person suggestions also containing a confidence value. If processing the video using VFT was successful, the resulting diarization and speaker predictions are saved to the database and used as a base for further preprocessing. In case the VFT service reported an error or the Admin chose to not use VFT processing for a hearing video, the earlier mentioned basic diarization is used.

As the final step of the preprocessing pipeline, automatic text correction is executed on the raw transcript. Python scripts remove consecutively occurring white spaces, capitalize proper nouns, and convert lexical representations of numbers to numerical ones. If needed, utterances which are too short to stand alone are merged into longer ones. After this automatic preprocessing of transcripts has finished, TT automatically generates tasks which refer to a specific time interval in a hearing video. A task represents a short work package which will be assigned to a single editor. Administrators can then choose to either manually assign editors to important tasks, or let the tool automatically distribute the workload among the editors.

Finally, as already mentioned beforehand, human editors start to enhance the preprocessed transcripts (up-leveling) by working on the tasks assigned to them.

The time span and content of tasks generated out of a hearing depend on the nature of the hearing as well as the aforementioned information added by an Admin. For the sake of completeness the correlation between a hearing, hearing videos, bill discussions, and tasks will be shortly explained. A hearing is made up of one to many hearing videos. Over the course of the hearing, multiple bills can be discussed. Bill discussions can continue over different videos, while multiple bills can also be discussed at the same time. Tasks will then be generated in such a way that there is one full bill discussion per task. If bill discussions overlap multiple videos or multiple bills are discussed at once, a more sophisticated approach is used. However, describing this approach here would be too much of a technical detail.

4.3 Technical Details

Transcription Tool is being developed as a web application. The backend uses the Java based web framework Spring MVC (2018), while also utilizing Spring Boot. For the frontend, presentation of content is mainly provided by the template-driven JavaScript library Ractive.js (2017), while also making use of native JavaScript and JQuery (2018). All data is saved to the Digital Democracy Database (DDDB), which uses MySQL (2018) and is stored on an Amazon Web Services cloud server (2018). TT is still under active development, and new features are added periodically to improve transcriber efficiency.

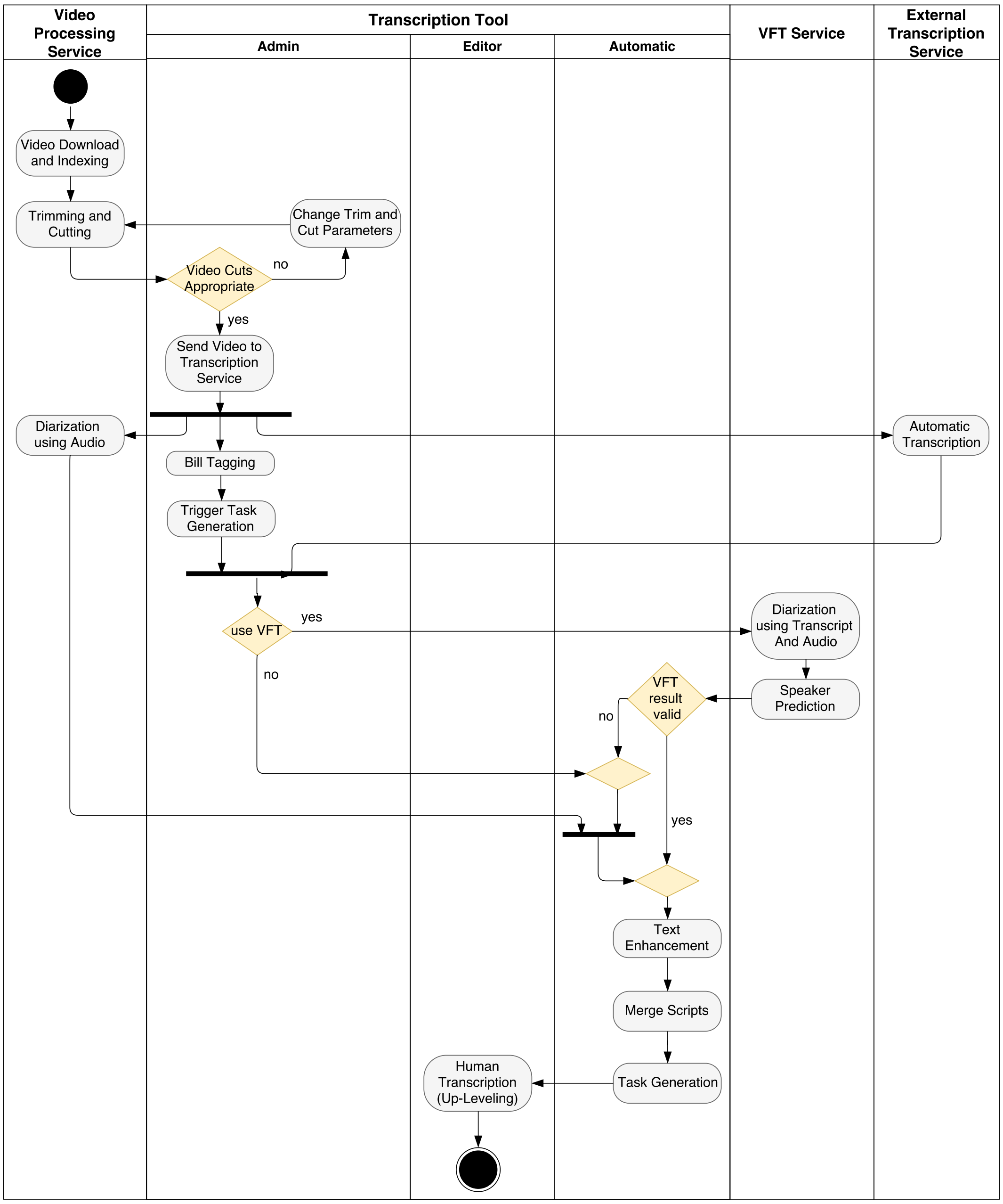

Figure 4.2: Human-Assisted Transcription Pipeline

5 Research Problem

The bottleneck created by the manual transcription process leads to a considerable obstacle for Digital Democracy, both in regards of monetary cost as well as time delay before final transcripts are available. For this reason, improvements to Transcription Tool were necessary. Therefore, the first part of the research problem is to make the tool more efficient.

Secondly, there has been no reliable way for the Digital Democracy initiative to measure performance changes that new tool releases bring with them up until now. The question if newly introduced versions lead to improvement could not be answered. Another problem was the lack of clarity regarding the concrete composition of transcription time. In order to further improve efficiency, the crucial question of which general actions transcribers take and how these interactions attribute to transcription time must be answered.

5.1 Data Sources and User Types

As already mentioned in Chapter 4.1, there are two types of users working with the transcription tool: Administrators and Transcribers. Administrators are mostly staff working for the IATPP. Transcribers are student workers which are employed on a short-term basis. Most students are initially not familiar with the terminology used in a legislative setting as well as the transcription interface itself. Due to the relatively high turnover of student staff and the cost of training new transcribers, it is necessary for the tool to work efficiently and provide an easy and straightforward interface. In spring 2018, 36 transcribers and 4 admin users were interacting with the tool on a regular basis.

While some students are always working from a dedicated workspace in an office, others work remotely. Because of this fact and the students' unregulated working hours due to them also being occupied with their studies, it is difficult to keep track of the current progress of transcription tasks. Besides missing reports of progress, there is no simple way of properly quantifying transcriber productivity and therefore tool efficiency without keeping track of tool interactions. Without detailed insight into these parameters, there is no easy way of estimating workload produced by incoming transcription tasks.

5.2 Research Questions

As already mentioned shortly in Chapter 1.3, three major questions occur.

5.2.1 Transcription Tool Improvements

Before improvements to the tool can be made and new features can be introduced, research is necessary to determine which parts of the software need improvement. The general interface must be reevaluated and requirements for new features have to be gathered. For this, transcribers must be interviewed to determine current problems with the tool, and find solutions how these could be solved. Afterwards, developments for new features can begin. New features and improvements will be released in multiple release phases. In the end, Transcription Tool's requirements must be reevaluated to determine if the improvements solved some of the tool's issues.

5.2.2 Transcription Efficiency

Efficiency of the tool and its different versions has to be analyzed and quantified. It is important to ensure the introduced changes did not disturb the current workflow or lead to any disruptions of the transcription process. In addition to that, measuring transcription time is important to allow proper planning and estimation of the duration and cost of future transcription operations.

5.2.3 Transcriber Interactions

Besides transcription tool efficiency, identifying general interaction patterns and how much time they take in general has to be specified. Until now, only assumptions could be made about the amount of time specific processes such as correcting text errors in the textual transcript would take. So far, supervisors of the Digital Democracy initiative assumed the following about the manual editing process of transcripts:

- Identifying speakers and their affiliation takes up most of the transcription time
- Only minor text corrections are necessary, most of the text only has to be proofread
- Splitting up utterances is a lot more time-consuming than merging

6 Experimental Design

6.1 Recording User Interaction

In order to answer the questions and address the problems introduced in Chapter 5, actions taken by the editors had to be investigated. To achieve this, interactions between editors and the transcription tool were recorded and saved into logs. This not only allows the system to keep track of the transcription progress of separate videos, but also enables performance measurements and identification of interaction patterns in a live environment. It also provides additional information for debugging. To prevent the system from saving log records too excessively, only specific interactions with HTML elements and interaction patterns relevant to the transcription workflow are taken into consideration. The following JavaScript events are recorded: *focusin*, *focusout*, *click*, *change*, and *keydown* (Mozilla, 2018). In addition to standardized browser interactions, custom events interesting to our use case such as minimizing or leaving the browser window and usage of keyboard shortcuts were also included in the recordings. Those named custom events include: "window", "visible", and "keypressed".

For each event, up to eight values are recorded:

- *event*: String value describing the triggered event. This can either be a standardized JavaScript event, or a custom event defined for our purpose.
- *element*: Either the tag name of the HTML element the event was recorded on, or additional information for custom events.
- *tagId*: Unique identifier given to this element. This id might contain information such as the database id of specific entities to enable further analysis.
- *value*: Text content of the targeted HTML element when the event was triggered. In case of a custom event, this field is used for providing additional information about the specific interaction.

- *keyPressed*: Integer value describing the button pressed for triggering this event. For keyboard interactions, ASCII decimals are used. Left, middle, and right mouse buttons are represented by values 0, 1, and 2.
- *timestamp*: UNIX timestamp in milliseconds (UTC).
- *editorId*: Unique identifier for the current editor.
- *ip*: IP address, only used to detect if person was working remotely or from the project lab

Logs are stored in a JSON format. A separate log file is created for each transcription task and contains JSON representations of all events triggered by the editor during completion of this task. A concrete example for a record of a single event can be seen in listing 6.1.

Listing 6.1: Example of a Single JSON-Entry Describing an Event in a Log File


```
{
  "event":       "click",
  "element":     "TEXTAREA",
  "tagId":       "textarea-20959669",
  "value":       "Mr. President, I withdraw
                 my point of order.",
  "keyPressed":  1,
  "timestamp":   1512408954042,
  "editorId":    107,
  "ip":          "123.45.6.789"
}
```


Before these logs are analyzed, they are saved to the Digital Democracy MySQL database. This allows for faster processing and easier data access. Nightly executed Python scripts insert log information produced during the last day into the database.

6.2 Metrics

To measure tool efficiency, a statistical analysis using specific metrics was performed. These metrics are coined by terms which have a distinctive meaning within the scope of this paper. They will be described over the following few paragraphs.

Firstly, it has to be mentioned that hearing video duration does not represent a reliable base value for calculation of performance metrics. In some states, videos are uploaded untrimmed and contain silent periods which do not require any transcription work. Therefore, pure speaking time, or *video speech time* (VS_t) of a task was chosen to determine its real length. As seen in Table 6.1, VS_t accounts for only 83.68 % of the overall video time in a task on average. A task is a short work package which is assigned to a single transcriber. VS_t was derived from utterance timestamps returned by the automatic transcription process.

Transcription time (T_t) is the time needed to complete the transcription for a specific task. To evaluate the assumptions made in section 5.2.3, T_t was split up into separate components. It can be stated that transcription time includes interactions such as startup time (time from loading the page until first interaction), speaker identification, text correction, as well as splitting and merging utterances. However, there are operations which can not be reliably measured, such as as the transcriber proofreading text and just watching the video or not doing any work at all (being idle). These activities were combined into $T_{passive}$. Equations 6.1, and 6.2 show a summarization of how T_t was defined for our analysis.

$$T_t = Startup + TextCorrection + SpeakerId + Split + Merge + T_{passive} \tag{6.1}$$

$$T_{passive} = Proofread + Idle \tag{6.2}$$

Transcription ratio per task (TR_t) describes the time in minutes it takes an editor to work on a minute of video speech in a specific task (see Equation 6.3).

Editor Productivity (EP) is the average TR_t over all tasks of an editor, while N_{te} describes the number of tasks completed by a specific editor (see Equation 6.4).

The main metric used to measure total efficiency of the tool and its separate versions is called transcription tool efficiency (TTE). As seen in Equation 6.5, TTE represents the average editor productivity over the amount of all editors (N_e).

Table 6.1: Comparison of Video Duration, Video Speech Time (VS_t), and Their Relation Over All States

State	Tasks	Duration (h)		% of VS_t in Video	
		Video	Speech	per Task	Overall
CA	346	125.51	114.65	86.48	91.34
FL	1890	350.31	330.43	94.21	94.33
NY	415	150.5	138.1	84.18	91.76
TX	4653	663.35	531.25	80.38	80.09
All	**7304**	**1289.67**	**1114.43**	**84.47**	**86.41**

$$TR_t = \frac{T_t}{VS_t} \quad (6.3) \qquad EP = \frac{\sum_{i=1}^{N_{et}} TE_{t_i}}{N_{et}} \quad (6.4) \qquad TTE = \frac{\sum_{i=1}^{N_e} EP_i}{N_e} \quad (6.5)$$

All metrics as well as the formulas depicted in Equations 6.3, 6.4, and 6.5 can be calculated over an arbitrary subset of transcription tasks (e. g. only specific tool versions) and editors.

7 Results

7.1 Improvements and Features

To properly evaluate Transcription Tool, interviews with transcribers were conducted both before and after feature development. First, transcribers were asked which part of working on a transcription poses the most effort for them. Second, they were questioned if they could think of any improvements for the tool. Results of the interviews conducted in the beginning were combined with suggestions made by the project leaders as well as feature ideas which came up during brainstorming sessions to conclude improvement requirements.

7.1.1 Finding Tool Requirements

In September 2017, five transcribers worked full time for Digital Democracy. An interview with each of them was conducted, evaluating the baseline version of the Transcription Tool.

Effort necessary to identify people was named to be one of the main factors increasing transcription time. In addition to that, splitting up utterances to achieve correct length and speaker assignments was identified as an expensive factor. In many cases, a textual utterance which the diarization scripts determined to be spoken by one person actually contained sentences by another speaker. In general, utterances seemed to be too long. Also, the video player was lacking options such as enlarging the video or other more convenient ways to interact with it. Another request mentioned by the transcribers was that names of speakers in the database could not be changed via Transcription Tool. Lastly, transcribers criticized the long loading time of the tool.

7.1.2 Feature Development

Several changes aiming at improving stability, usability, and efficiency of Transcription Tool were developed. These changes were gradually introduced from September 2017 until March 2018 as four different tool versions. The following improvements to the baseline tool version were implemented and released incrementally as separate feature versions:

- *Profile Preview*: A speaker profile picture preview next to speaker names was added (see Figure 7.1). This is especially helpful when searching for people using the search interface of Transcription Tool.
- *Video Features*: Better video player functionalities such as full-screen mode as well as slowing down and speeding up video were implemented. Also, functionality to edit people's names was added.
- *Utterance Navigation*: Due to existing difficulties when navigating a transcript and the corresponding video, additional interface buttons were added which set video time to the beginning of utterances. Also, jumping to the previous or next utterance from the currently played one was enabled using additional UI elements (see Figure 7.2).
- *VFT*: Several other student projects and theses focus on providing better speaker recognition while combining voice, face, and text analysis. This version introduces an interface for linking results of those speaker identification algorithms to the tool. A list of speaker suggestions for utterances is made accessible to transcribers upon clicking an icon (see Figure 7.3).

Table 7.1 gives an overview of each version's functionality and their naming.

7.1.3 Evaluating Tool Improvements

After the last feature release was in use for more than a month in February 2018, another round of interviews was performed. This time, ten transcribers answered the same questions about effort and improvements as in September 2017.

First, it has to be said that transcribers still mentioned splitting utterances as a time consuming matter. Therefore, the problem of automatically splitting utterances has to be further addressed in the future. Changing settings

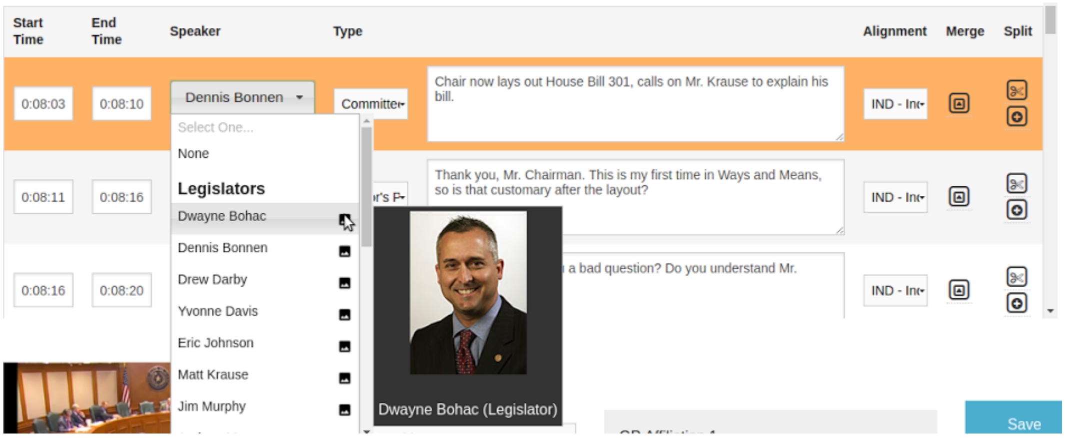

Figure 7.1: Profile Picture Preview on a Person Selection List, As Introduced in Version 1. Hovering the Preview Icon Shows the Profile Picture Overlay.

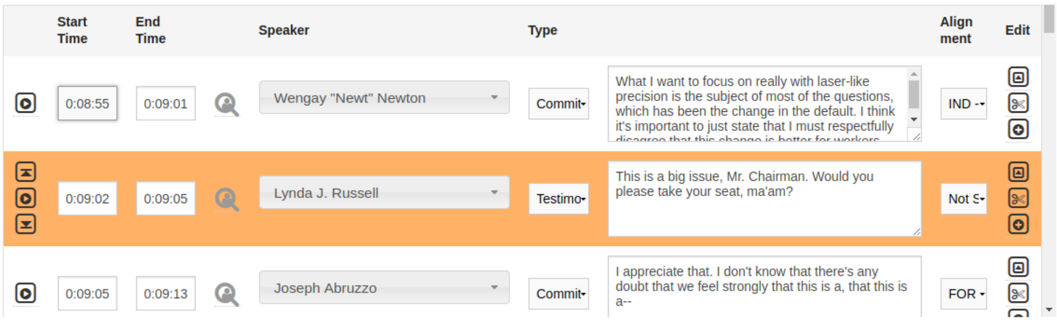

Figure 7.2: Updated Utterance Element with Navigation Buttons Introduced in Version 3 on the Left.

of the automatic speech recognition and diarization system has also not solved this problem. However, due to the improvements made to the video player and video navigation in general, no more complaints about the player were brought up. Also, no issues about loading times and speaker name corrections were mentioned anymore, meaning the newly implemented improvements also solved these problems. Although transcribers stated that the profile picture preview implemented in version 1 is very helpful for identifying legislators and lobbyist, recognizing speakers of the general public is still problematic due to them lacking profile pictures. Overall, all transcribers stated that improvements and new features implemented over

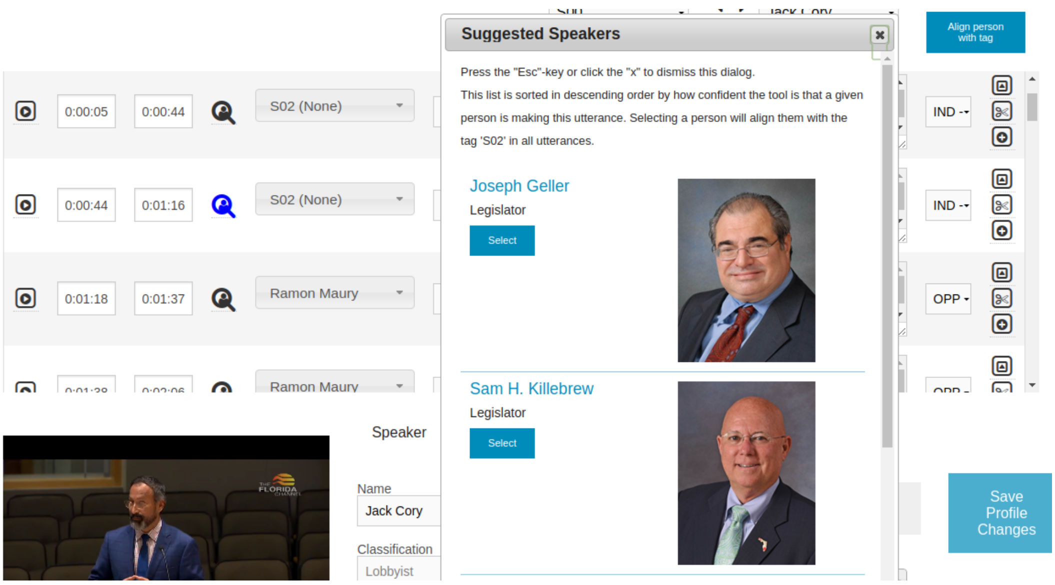

Figure 7.3: Screenshot of Version 4, Displaying the Speaker Suggestion Dialog. The Blue Icon Indicates the Utterance for which the Suggestion is Currently Displayed.

the past few months increased tool efficiency and stability. For a scientific analysis of tool efficiency see Chapter 7.2.

7.2 Tool Efficiency Analysis

An efficiency analysis for a cohort of 20 editors, who worked with all versions of the tool, was performed. Table 7.2 shows an efficiency improvement of 10.7% from version 0 to version 3. However, version 4 seems to produce an decrease in efficiency. Due to not having enough usage data to further investigate version 4, a more detailed analysis will be performed at a later time when more data is available. Figure 7.4 visualizes the tool efficiency by comparing the time necessary to transcribe one minute of video speech over each version for the given editor cohort. In addition to that, the performance change each version brought with it on average as well as for single editors can be seen in Figure 7.6.

Because not every state which Digital Democracy processes had an active session year during the time new versions were released, a separate analysis

Table 7.1: Transcription Tool Versions with Feature Descriptions

#	Name	Description
0	Baseline	No changes
1	Profile Preview	Hovering icon near people's names shows preview of their profile picture
2	Video Features	Full-screen for video player, UI to change playback rate for video
3	Utterance Navigation	Buttons for to directly jumping to utterance in video and going to next or previous utterance, interactive error messages, UI for manually setting video time, preprocessing changes to utterance length
4	VFT	Incorporation of voice, face, and text analysis to identify speakers

is performed to compare tool efficiency per state. Figure 7.5 provides an overview over these results. Analysis of behaviors - bar/pie chart with what takes most time: speaker identification, text correction, splitting and merging utterance, general idle time.

7.3 Transcriber Interaction Time

Comparing percentages of the separate interactions leading to the final transcription time as posed in Equation 6.1 makes for a similar result across all tool versions. Figure 7.7 shows the percentage of time an editor spends on labeled interactions within a single task on average.

To quantify the concrete changes in timing for each version and how the interactions attribute to the changes, the vector α_{i_v} is added to each interaction feature. α_{i_v} acts as a series of multipliers denoting the increase or decrease in time interaction (i) produced in this version (v) compared to the base version (version 0), as seen in Equation 7.3. Generally, values less than 1 denote an increase in efficiency compared to the base case in version 0. In Equation 7.1, AIT_i stands for the duration of time spent for the given interaction type in a task in relation to the video speech time (VS_t). Averaging this value for all tasks in a version, as seen in Equation 7.2 leads

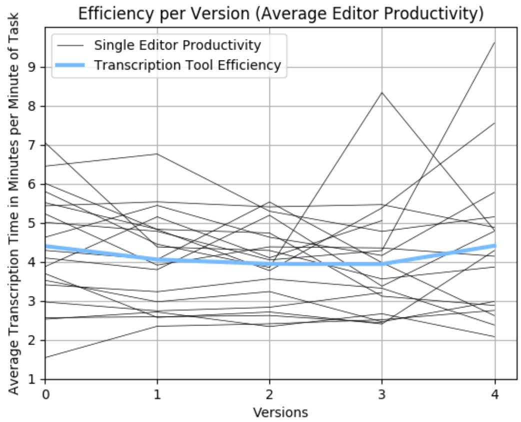

Figure 7.4: Transcription Tool Efficiency per Version

to a general representation of how much time consuming this interaction type was in this version (AIT_{i_v}). The changed formula for T_t can be seen in Equation 7.4. Consecutive performance changes over multiple versions illustrated by this formula can be best displayed in a matrix, with each cell containing a concrete value for α_{i_v} with rows representing different tool versions (v) and columns realizing the different interactions (i). Columns in both matrices can be labeled as follows from left to right: Startup time, text correction, speaker identification, splitting time, merging time and passive time.

$$AIT_i = \frac{TimeInteraction_i}{VS_t} \quad (7.1)$$

$$AIT_{i_v} = \frac{\sum_{t=1}^{N_{tasks_v}} AIT_i}{N_e} \quad (7.2)$$

$$\alpha_{i_v} = \frac{AIT_{i_v}}{AIT_{i_0}} \quad (7.3)$$

$$T_{t_v} = \alpha_{1_v} \cdot Startup + \alpha_{2_v} \cdot TextCorrection + \alpha_{3_v} \cdot SpeakerId + \alpha_{4_v} \cdot Split + \alpha_{5_v} \cdot Merge + \alpha_{6_v} \cdot T_{passive} \quad (7.4)$$

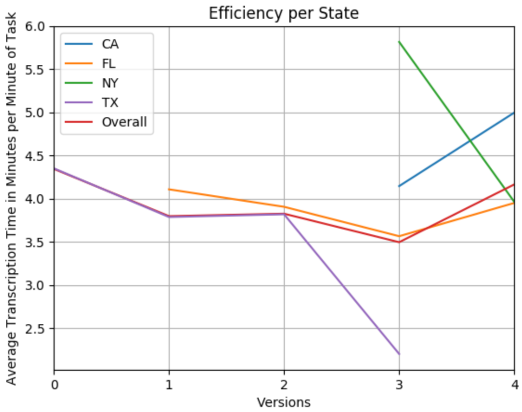

Figure 7.5: Transcription Tool Efficiency per State

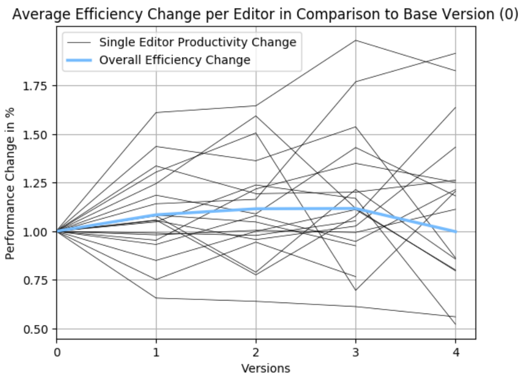

Figure 7.6: Performance Change per Version and Editor in Comparison to Base Version

$$TEC_{editors} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.91 & 0.75 & 1.02 & 0.86 & 1 & 1.42 \\ 1.06 & 0.81 & 0.88 & 0.82 & 1.26 & 1.73 \\ 0.73 & 0.67 & 0.90 & 0.24 & 4.49 & 1.90 \\ 1.02 & 0.47 & 0.87 & 0.26 & 3.65 & 1.98 \end{pmatrix} \tag{7.5}$$

$$TEC_{Texas} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.89 & 0.74 & 0.94 & 0.82 & 0.98 & 1.07 \\ 0.95 & 0.68 & 0.90 & 0.63 & 1.15 & 1.17 \\ 0.55 & 0.60 & 0.51 & 0.44 & 1.17 & 0.54 \end{pmatrix} \tag{7.6}$$

Table 7.2: Transcription Tool Efficiency Change Over All Investigated Versions

Version	*TTE* (minutes)	Efficiency Improvement (%)
0	4.401	-
1	4.014	8.793
2	3.943	1.768
3	3.938	0.129
4	4.407	-11.91

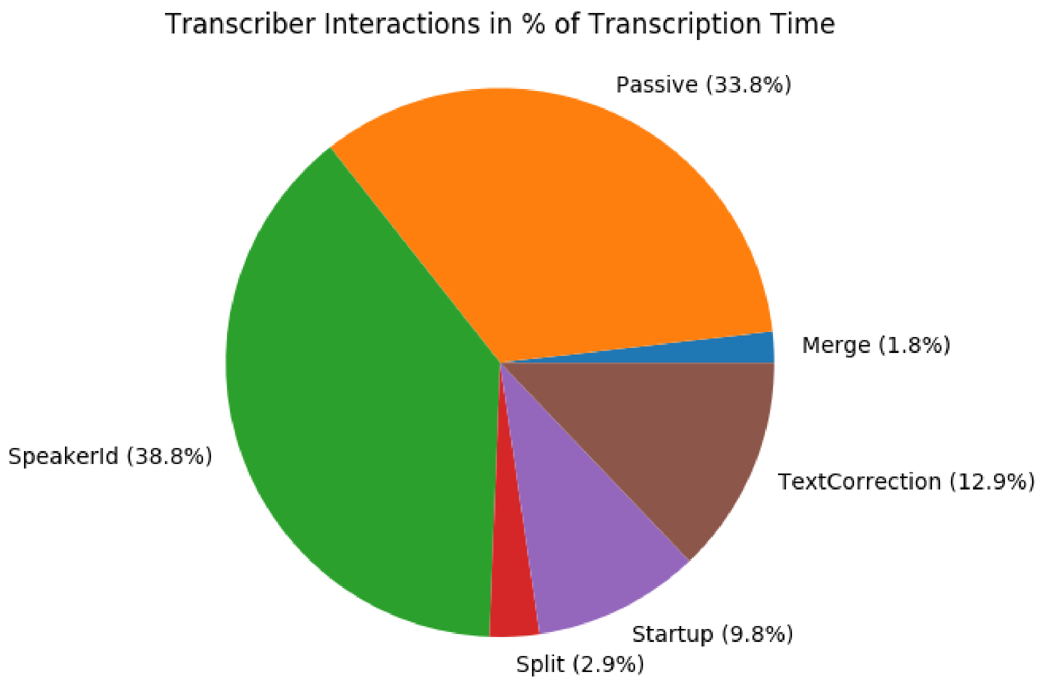

Figure 7.7: General Editor Interactions and Their Contribution to Overall Transcription Time On Average

8 Conclusions and Future Work

In the absence of fully automated, highly accurate transcription technology, human assisted transcription is at present a viable and cost effective option for capturing speech from legislative proceedings. We are interested in a deeper understanding of human assisted transcription systems, how to make them more efficient and finally how the human contribution is distributed across various functions expected from human transcribers.

In this paper, the Digital Democracy project and the transcription tool that is necessary for the bulk of the input data preparation for the project was introduced. Also, we introduce four sets of improvements to the tool and do a study of how these improvements affect the efficiency and cost of the entire operation. It is found that on the average a 10.7% increase in efficiency is realized by the first three phases of improvements. We did not have enough information to fully analyze the final phase of improvements, but the data we did have indicates a reversal of the trend and a decrease in efficiency in the final phase. Our working hypothesis is that unanticipated user interface problems led to the entire tool becoming significantly slower and thus the logs reflect that every aspect of the work took longer to complete. The results of research done about this issue will provide more certainty on this issue.

Another contribution in this paper is the derivation of a statistically informed model (Equation 7.4) of distribution of human transcription related functions. Figure 7.7 demonstrates the typical distribution of functions per unit of time. Equation 7.5 and Equation 7.6 are matrices that denote the improvement broken down by function, as measured by decrease in human assisted processing time per unit of video compared to the base version of the tool. Equation 7.5 does this across all states but with the same cohort of editors. Equation 7.6 does it just for the state of Texas but only with the first three phases of improvements. As shown in Figure 7.7, the vast majority of the time (85.5%) is spent on either text correction (Equations 7.5 and 7.6, column 2), speaker identification (column 3), or passive (column 6). Text correction shows a clear downward trend across the tool versions, especially when considering only Texas. The speaker identification function also shows clear improvement in efficiency. After a very slight up-tick with version 1 of the tool, speaker identification coefficients decrease with every subsequent version of the tool. Splitting functions also show a clear downward trend.

Startup and passive times appear not to follow the trend. In the case of passive for the editor cohort (Equation 7.5), the number seem to be increasing. The passive time calculation is slightly different in that it has a "catch all" idle component calculated by subtracting the sum of the time the editor spent in all the other categories, from the total task completion time. We assume editors are mostly just watching videos during this time, but it is not clear. Similarly startup time involves measuring unknown time elements when editors first access the tool but before they begin work.

Future work will first try to complete the analysis of phase 4 of the improvements, as well as paint a more comprehensive picture of each state (Figure 7.5). In addition to that, further future work will attempt a longer longitudinal study with a constant cohort in different states.

Bibliography

Amazon Web Services, Inc. (2018). Amazon web services - cloud computing services. Retrieved January 15, 2018, from aws.amazon.com

Basu, J., Bepari, M. S., Nandi, S., Khan, S., & Roy, R. (2013, November). Satt: semi-automatic transcription tool. In *2013 international conference oriental cocosda held jointly with 2013 conference on asian spoken language research and evaluation (o-cocosda/caslre)* (pp. 1–6). doi:10.1109/ICSDA.2013.6709903

Burke, M., Amento, B., & Isenhour, P. (2006). Error correction of voicemail transcripts in scanmail. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 339–348). CHI '06. Montreal, Quebec, Canada: ACM. doi:10.1145/1124772.1124823

California Secretary of State. (2017, July). Cal-access. Retrieved April 25, 2018, from cal-access.sos.ca.gov

Castro-Bleda, M. J., Vilar, J. M., Llorens, D., Marzal, A., Prat, F., & Zamora-Martinez, F. (2017). A system for assisted transcription and annotation of ancient documents. In *Proceedings of the 15th international workshop on content-based multimedia indexing* (37:1–37:5). CBMI '17. Florence, Italy: ACM. doi:10.1145/3095713.3095752

cielo24. (2018). Cielo24. Retrieved January 10, 2018, from cielo24.com

C-SPAN. (2018). Congressional chronicle - members of congress, hearings and more. Retrieved January 15, 2018, from c-span.org/congress

Davis, B. & Baxandall, P. (2014, April). Following the money 2014. Retrieved April 25, 2018, from calpirgedfund.org/reports/caf/following-money-2014

Deshpande, R., Tuna, T., Subhlok, J., & Barker, L. (2014, October). A crowd-sourcing caption editor for educational videos. In *2014 ieee frontiers in education conference (fie) proceedings* (pp. 1–8). doi:10.1109/FIE.2014.7044040

DigitalDemocracy. (2017). Digitaldemocracy homepage. Retrieved April 20, 2018, from digitaldemocracy.org

IBM. (2018). Watson Speech to Text. Retrieved February 10, 2018, from ibm.com/watson/services/speech-to-text

Interoperability solutions for public administrations, businesses and citizens. (2018, May). 19th annual international conference on digital government research. Retrieved May 14, 2018, from https://ec.europa.eu/isa2/events/19th-annual-international-conference-digital-government-research_en

Library of Congress. (2018). Congressional record. Retrieved January 15, 2018, from congress.gov/congressional-record

Luz, S., Masoodian, M., Rogers, B., & Deering, C. (2008). Interface design strategies for computer-assisted speech transcription. In *Proceedings of the 20th australasian conference on computer-human interaction: designing for habitus and habitat* (pp. 203–210). OZCHI '08. Cairns, Australia: ACM. doi:10.1145/1517744.1517812

MapLight. (2017, July). Maplight - revealing money's influence on politics. Retrieved April 25, 2018, from maplight.org

Mozilla. (2018). MDN web docs - event reference. Retrieved January 10, 2018, from developer.mozilla.org/en-US/docs/Web/Events

Myers, J. (2015, April). Big support in bipartisan poll for a more transparent california legislature. Retrieved April 25, 2018, from /ww2.kqed.org/news/2015/04/30/big-support-in-bipartisan-poll-for-a-more-transparent-california-legislature

New Zealand Parliament. (2018). Read hansard reports. Retrieved January 15, 2018, from parliament.nz/en/pb/hansard-debates/rhr

Nivre, J., Tullgren, K., Allwood, J., Ahlsén, E., Holm, J., Gronqvist, L., ... Sofkova, S. (1998). Towards multimodal spoken language corpora: transtool and synctool. In *Proceedings of the workshop on partially automated techniques for transcribing naturally occurring continuous speech* (pp. 11–23). Transcribe '98. Montreal, Quebec, Canada: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1628291.1628293

NowTranscribe Ltd. (2017). Nowtranscribe. Retrieved January 15, 2018, from nowtranscribe.com

Oracle Corporation. (2018). Mysql. Retrieved January 10, 2018, from mysql.com

ORF. (2018). ORF TVthek. Retrieved February 14, 2018, from tvthek.orf.at

Papazian, F., Bossy, R., & Nédellec, C. (2012). Alvisae: a collaborative web text annotation editor for knowledge acquisition. In *Proceedings of the sixth linguistic annotation workshop* (pp. 149–152). LAW VI '12. Jeju, Republic of Korea: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2392747.2392771

Parliament of Australia. (2018). Hansard. Retrieved January 15, 2018, from aph.gov.au/Parliamentary_Business/Hansard

Parliament of Canada. (2018). Publication search. Retrieved January 15, 2018, from ourcommons.ca/parliamentarians/en/publicationsearch

Parliament of the United Kingdom. (2018). Hansard online. Retrieved January 15, 2018, from hansard.parliament.uk

Pivotal Software. (2018). Spring. Retrieved January 10, 2018, from spring.io

Public Affairs Office of California Polytechnic State University. (2015, May). Lt. gov. newsom and former state sen. blakeslee trail-blaze digital platform to transform government transparency. Retrieved April 25, 2018, from digitalcommons.calpoly.edu/pao_pr/4596

Quo Vadis Veritas Redaktions GmbH. (2018). Addendum – Politometer. Retrieved February 14, 2018, from http://tvthek.orf.at

RactiveJS contributors. (2017). Ractive.js. Retrieved January 10, 2018, from ractive.js.org

Revuelta-Martìnez, A., Rodrìguez, L., & Garcìa-Varea, I. (2012). A computer assisted speech transcription system. In *Proceedings of the demonstrations at the 13th conference of the european chapter of the association for computational linguistics* (pp. 41–45). EACL '12. Avignon, France: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2380921.2380930

Rouvier, M., Gay, P., Khoury, E., Merlin, T., & Meignier, S. (2013). An open-source state-of-the-art toolbox for broadcast news diarization. In *In proc. of interspeech*.

Rovin, J. (2016, March). *Reducing costs in human assisted speech transcription*. Digital Commons @ Cal Poly. Retrieved April 20, 2018, from digitalcommons.calpoly.edu/theses/1538

Seps, L. (2013). Nanotrans—editor for orthographic and phonetic transcriptions. In *Telecommunications and signal processing (tsp), 2013 36th international conference on* (pp. 479–483). IEEE.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., & Tsujii, J. (2012). Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings* *of the demonstrations at the 13th conference of the european chapter of the association for computational linguistics* (pp. 102–107). EACL '12. Avignon, France: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2380921.2380942

Surka, M. & Ridlington, E. (2016, April). Following the money 2016. Retrieved April 25, 2018, from calpirg.org/reports/cap/following-money-2016

The Austrian Parliament. (2018a). Erweiterte suche. Retrieved February 14, 2018, from parlament.gv.at/SUCH

The Austrian Parliament. (2018b). Stenographische protokolle. Retrieved February 14, 2018, from parlament.gv.at/PAKT/STPROT

The California Channel. (2017, July). The california channel - politics and public affairs that shape california. Retrieved April 25, 2018, from calchannel.com

The European Parliament. (2018). Debates and videos | planary. Retrieved January 15, 2018, from europarl.europa.eu/plenary/en/debates-video.html

The jQuery Foundation. (2018). Jquery. Retrieved January 10, 2018, from jquery.com

The Swiss Parliament. (2018). Search official bulletin. Retrieved January 15, 2018, from parlament.ch/en/ratsbetrieb/suche-amtliches-bulletin

Toselli, A. H., Vidal, E., & Casacuberta, F. (2011). Computer assisted transcription of text images. In *Multimodal interactive pattern recognition and applications* (pp. 61–98). Springer.

Whittaker, S. & Amento, B. (2004). Semantic speech editing. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 527–534). CHI '04. Vienna, Austria: ACM. doi:10.1145/985692.985759

Widlöcher, A. & Mathet, Y. (2012). The glozz platform: a corpus annotation and mining tool. In *Proceedings of the 2012 acm symposium on document engineering* (pp. 171–180). DocEng '12. Paris, France: ACM. doi:10.1145/2361354.2361394