

Videolyzer: Quality Analysis of Online Informational Video for Bloggers and Journalists

Nicholas Diakopoulos

Georgia Institute of Technology
School of Interactive Computing
nad@cc.gatech.edu

Sergio Goldenberg

Georgia Institute of Technology
School of Literature,
Communication, and Culture
sergio.goldenberg@gatech.edu

Irfan Essa

Georgia Institute of Technology
School of Interactive Computing
irfan@cc.gatech.edu

ABSTRACT

Tools to aid people in making sense of the information quality of online informational video are essential for media consumers seeking to be well informed. Our application, Videolyzer, addresses the information quality problem in video by allowing politically motivated bloggers or journalists to analyze, collect, and share criticisms of the information quality of online political videos. Our interface innovates by providing a fine-grained and tightly coupled interaction paradigm between the timeline, the time-synced transcript, and annotations. We also incorporate automatic textual and video content analysis to suggest areas of interest for further assessment by a person. We present an evaluation of Videolyzer looking at the user experience, usefulness, and behavior around the novel features of the UI as well as report on the collaborative dynamic of the discourse generated with the tool.

Author Keywords: video analysis, information quality

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces – Graphical user interfaces.

INTRODUCTION

The validity, reliability, and ultimately the quality of information in online media is a growing problem as more and more people post information directly to the web without the oversight of an editorial process. While true for all forms of media, this especially resonates for video, as technology for disseminating and sharing videos becomes more pervasive. Consider for example video advertisement and video podcasts for topics ranging from politics to commerce and marketing. With such a growth in online video, it is apparent that more and more consumers will be getting their information from online videos. Topics like health, finance, and politics are already common place. We feel that understanding the quality of that information becomes paramount for a consumer wanting to make decisions based on it.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

Copyright 2009 ACM 978-1-60558-246-7/09/04...\$5.00.

Consider a specific example in the domain of health from user generated videos. A recent study [14] found that of 153 videos referring to vaccination or immunization on YouTube, 32% were negative and 20% more were ambiguous with reference to Canadian governmental health guidelines on the topic; a frightening proposition considering many may believe this contradictory information seen online. In the current political season in the U.S., we have seen a large number of political advertisements making claims that could benefit from close examination and a better understanding of their context.

In this work we begin to address the larger issue of information quality by building and studying a tool, Videolyzer, specifically designed to aid *political bloggers and journalists* in the activity of *watchdog journalism*, the process of combing through and evaluating the veracity of claims in the media. Videolyzer allows users to collectively analyze the quality of online political video and then aggregate and share these analyses with others. Bloggers and journalists can assess aspects of quality in the video, its transcript and annotations including bias, accuracy, and relevancy that can then be backed up with sources and reasons. The advantage of focusing on these users and with political content is that bloggers and journalists are already highly motivated to critique online videos, and political content is rife with claims and issues that would benefit from careful scrutiny and additional contextualization.

We have developed Videolyzer as a visual and discourse analytic sensemaking tool that innovates on issues of (1) *granularity*, which considers the level of detail of ratings and how different pieces or segments of video can be rated separately, (2) *multimedia*, by tightly integrating interactions between the video, its transcript, and annotations, and (3) *automation*, by incorporating elements of automatic content analysis to help draw attention to areas of the video and transcript potentially worthy of further human analysis. Interactions between transcript and video are tightly coupled, the result of acquiring high precision per word time-stamp information [8]. The whole system leverages an information quality ontology which lends structure and organization to the interface, but that is semi-formal to allow for analysis and expression of concepts outside the ontology

We also present an initial evaluation looking at the properties of the discourse generated with the tool and the behaviors and perceptions of users regarding the novel features we provide. Users saw benefit from Videolyzer through the collaborative evaluation of different facets of information quality. Their perception and understanding of various perspectives and the comprehensiveness of complex issues was enhanced using the tool. Furthermore, having time-anchored annotations in a semi-structured format is expected to lead to an overall benefit by allowing for deep indexing and more precise search of video as well as visualizations that can cue end consumers to the information quality of sections of video as they watch.

RELATED WORK

There is a wealth of related research on video annotation, visual analytics, argumentation support, and the semantic web that has informed the design and development of Videolyzer. However, much of the research on video annotation systems has largely focused on tagging or commenting on videos without an element of *discourse* or *collaborative sensemaking*. Videolyzer extends and synthesizes past work to marry video annotation interfaces with a notion of helping the user discuss and understand the quality of the information in the video using a semi-structured approach.

One of the earliest systems in the literature to approach video annotation was Mackay's EVA [19]. EVA was a video analysis application which allowed for tagging of events in real-time or in post-processing. The user could add free text notes which could be linked with tags in the audio track. The minimalist interface of EVA facilitated simple and low-cost annotation, at the expense of providing the kind of sophisticated multimedia analytic activity which Videolyzer is designed to support.

The Family Video Archive (FVA) [3] was a video annotation system which explored the symbiosis between automated and manual techniques for tagging *collections* of videos. FVA used suggestive annotation to provide potential tags based on free-form text that had already been entered. Whereas FVA defaulted to applying a tag to an entire video, Videolyzer is geared toward specifying annotations anchored to time intervals in the video.

Suggestive annotation systems have also been explored in the text domain. For example, the ClaimSpotter tool was oriented toward sensemaking of scholarly papers and suggested areas for annotation based on a document's "important" sentences and by detecting pertinent relations from an ontology [21]. Videolyzer also implements and evaluates suggestive annotation, but is geared toward concepts of interest to media studies. This is done by incorporating automatic content analysis to draw attention to potentially interesting parts of the video and transcript.

Cockburn and Dale developed a synchronous collaborative video annotation system, CEVA [6], which had the goal of enriching the video annotation process through either

synergistic or parallel analysis. Synergistic analysis describes the additional insight gained from having multiple video analysts share their different perspectives. Parallel analysis is the idea of using multiple analysts to speed up annotation by distributing the workload. We focus on the notion of synergistic analysis which expands the views and perspectives on the video.

Recently there have been several video applications in the commercial space such as Viddler and Asterpix, which address the issue of granularity in annotation (i.e. they allow for comments to be anchored to segments of the video). One feature that these systems lack, however, is the ability to combine granularity of annotation with discussion beyond the level of "neat" or "that's cool." In some sense these systems are designed to be widely applicable to *any* video content, whereas Videolyzer is targeted toward *informational* content with a concomitant need to understand that information's quality through an argumentation support system.

Argumentation support and visualization tools are in and of themselves a rich area of prior work (see [15] for a review). Compendium [1] is a widely used argumentation support system which builds off of ideas stemming from the Issue Based Information System (IBIS) methodology of Horst Rittel. The dialogue mapping and conversation modeling approach in Compendium has seen a wide range of applications including design, collaborative meetings, and more recently integration with video meetings in the Memetic system [23]. Memetic combines support for identifying and structuring video conferences based on issues, questions, and arguments / dialogue in the meetings. The highly structured argumentation functionality is integrated with a multileveled video timeline allowing for semantically enhanced navigation. Users can navigate the video based on the argument structure and vice versa.

Trellis is an argumentation system which combines free text descriptions with formal connections between statements [11]. Early versions of Trellis focused on comprehensively allowing for information source analysis in conjunction with connectives between statements drawn from those sources. Sources could be annotated with their credibility and reliability allowing the system to make inferences about statement support. One of the significant points levied against fully featured argumentation support systems such as Compendium and Trellis is the cognitive effort involved with manipulating a rich ontology [22]. Follow up work on Trellis acknowledged the substantial user effort required to use a formal reasoning representation and explored a range of tradeoffs in designing semi-formal representations which mitigate user effort but which still allow for some machine processing using the structure [5].

Collaboration is a significant component of many argumentation support systems. In the POLESTAR system [20], which was designed for intelligence analysts, collaborative review of intelligence reports was facilitated

through a semi structured mechanism for feedback including agreement / rebuttal, request for clarification, and request for supporting evidence. Videolyzer is similar to POLESTAR in its effort to enhance the quality and validity of inferences through peer-review.

The sense.us system [12], is oriented toward asynchronous discussion and annotation of information visualizations. Most interesting here is their notion of “embedded discussion” which anchors sub-discussions to particular view states of the visualization. We also utilize this design objective in Videolyzer to help ensure users have focused discussion on different parts of the video and transcript.

Videolyzer draws on a wide variety of related research ranging from video annotation, to argumentation support, the semantic web, and collaborative visualization systems. In this work we have synthesized across these various domains and have expanded them to provide a tightly integration video and transcript analysis system, which incorporates a semi-formal representation of information quality to aid users in collaboratively organizing the discourse around and making sense of informational videos.

VIDEOLYZER SYSTEM

Annotation and Interaction Model

Videolyzer is built on top of an ontology which facilitates the organization of different facets of information quality [7]. The term *information quality* indicates a degree of excellence in communicating knowledge or intelligence and is the focus of the analyses produced with Videolyzer. The model of information quality we use is derived primarily from studies of journalism and information science and is reified as a pragmatic set of tags that cover most of the space of information quality concerns including notions of *accuracy, reliability, validity, comprehensiveness, transparency, and bias*.

A central design goal of Videolyzer was to maintain expressivity for users’ communication goals while simultaneously leveraging structure to provide additional organization and sensemaking support. Thus the model incorporates free text comments and rationale as well as typed annotations. As suggested by Shipman in [22] we

Annotation Type	Actions on Own Annotations	Actions on Others’ Annotations
Video Claim	Add Source, Add Tags, Add Reaction	
Comment	Add Source, Edit Text, Delete	Add Tags, Add Reaction
Tag	Add Reason, Delete	Add Reaction
Reaction	Add Reason, Delete	
Reason	Add Source, Edit Text, Delete	Add Reaction, Add Tags
Source	Specify type, Edit Title, link, Rate expertise, trust	Edit Title and link, Rate expertise, trust

Table 1. Annotation types and associated actions depending on annotation ownership

took measures in our design to base our formalism on a model derived from studies of the domain familiar to our intended users (journalism) [7], as well as having non-formal ways to use the system while still providing additional benefit if the formalism was used. Finally, we made sure to provide adequate time for participants in our user study to gain familiarity with the application and lexicon. Part of our evaluation examines users’ ability to cope with this semi-formal structure.

The interface of Videolyzer translates the information quality model into a useful set of actions that users can take to interact with information in the video and in other annotations in a semi-structured way. The design goal of these interactions is to allow for an overall enhancement of the understanding of information quality (1) by allowing users to provide corroborating or refuting evidence for claims and comments thereby giving rise to triangulation or chains of evidence, (2) by facilitating fact checking and accuracy evaluation using alternate sources, and (3) by allowing for the expansion of different perspectives (i.e. comprehensiveness, fairness, and balance), and sources (transparency) around an issue.

The primary elements of the ontology exposed in the interface are *claims, comments, tags, reactions, reasons, and sources*. The two annotations initially available in the interface are claims and comments. After these are added, the other types of annotations become possible depending on ownership (see Table 1).

Claims are statements or interpretations stemming from the video itself, whereas *comments* are free text remarks made by users of the system. Either can have *sources* associated with them and those sources (whether they be documents or entities) can be rated according to their *trustworthiness* and *expertise*. Tags (Figure 1) are used to evaluate the quality and bias of other annotations as well as to communicate and collaborate with other users. Quality tags are shown as a 3 point scale with the midpoint representing uncertainty. Users may also specify free text tags to add concepts outside the ontology. Annotations can be responded to by users in terms of their reactivity (agree, disagree, or hedge) as well as by adding tags. Finally, tag and reaction annotations can be buttressed by free text *reasons*.

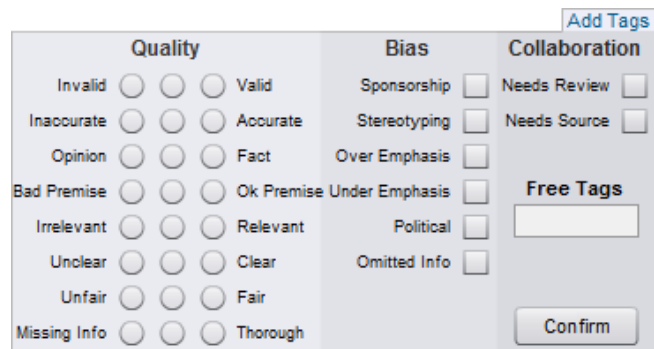


Figure 1. The tagging panel UI with quality, bias, and collaboration tags that users are presented with.

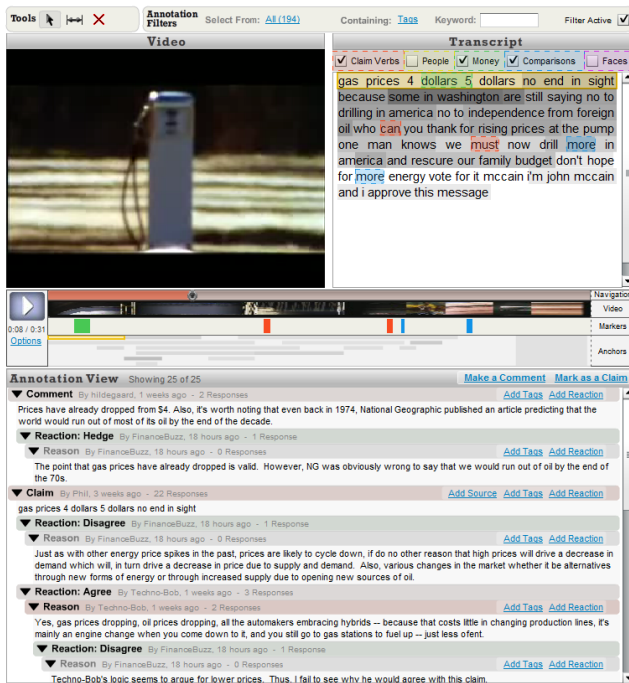


Figure 2. Overview of the Videolyzer interface showing (clockwise from upper left) the tool bar, filter options, transcript, timeline, annotation view, and video playback area.

Automatic Content Analysis

Videolyzer is primarily focused on helping users do information quality evaluations of video. Automatic methods of information quality annotation are largely impractical because many aspects of information quality have a degree of subjectivity, context, and connotation in their interpretation and evaluation. However, we have incorporated some automatic methods insofar as we want to examine if they can help *guide* human attention to areas of the video or transcript that may deserve more critical attention. As such, we use content analysis algorithms in Videolyzer to detect faces in the video, as well as references to people, money, comparisons, and modal verbs in the transcript. These are rendered as colored markers over the transcript (Figure 3) and timeline (Figure 4).

Face detection is important because many sources of information in video are people and a face can be a proxy for helping the annotator find these sources. We use the OpenCV face detector to preprocess the video. It is run on each frame separately and the temporal stream of detections is median filtered to remove any spurious detections.

We chose to extract references to people, money, comparisons, and modal verbs from the transcript because (1) media studies suggests that these features may lead to interesting analytic insights about the quality of information presented [10, 13] and (2) it is possible to automatically extract these features with a fairly high degree of reliability. References to people and money are extracted using the Open Calais web API from Thomson Reuters [2]. Comparisons in the text are detected by running a Part Of

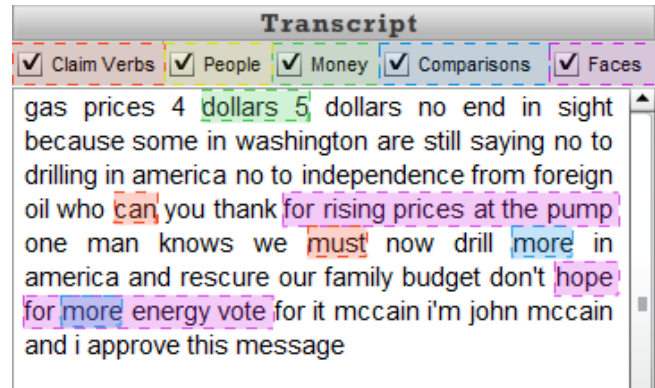


Figure 3. Color coded, automatically detected markers shown on the transcript view

Speech (POS) tagger on the transcript and then noting words detected as either comparisons or superlatives [4].

Text modality, which is expressed through modal verbs (e.g. can, ought to, could, shall, may, should, might, will, must, would) together with their negations and adverbs is an indicator of the degree to which a speaker or communicator is committed to their claim [10]. The choice of modal verbs can indicate the attitudes, judgments, and political beliefs of a writer or speaker. Verb modals can vary from weak to strong (i.e. possibility to necessity). For instance, a weak claim might involve the phrase “ought to” whereas a strong claim would use the verb “must.” We developed a routine to detect modal verbs and their strength by filtering the textual transcript against a list of modal verbs identified by linguists [17].

Videolyzer Interface

The videolyzer interface is designed to tightly couple the interactions occurring across three multimedia indices: the transcript, timeline, and annotation views (Figure 2).

Transcript View

A precisely time-stamped transcript is included in the Videolyzer interface to facilitate fine-grained commenting (Figures 2 and 3). The video transcript is gathered using a human-computation method based on a game called Audio Puzzler [8]. In the process of game play, the video is transcribed and each word in the transcript time-stamped. Automatic methods for acquiring the transcripts could also have been used although usability and utility become problems for automatic transcripts with lower accuracy.

Time-stamp information allows for a very tight coupling of interactions between the transcript and the timeline. Actions such as hovering and selection are mirrored on the other view, which we expect to aid navigation and anchor creation on the video. For instance, users can select a piece of text as they normally would on the transcript and an anchor is added to the transcript view as well as the timeline view (Figure 4). Anchors are where further annotations are attached to the transcript or timeline. They are rendered over the transcript as gray boxes, the darkness



Figure 4. Stacked timeline view showing levels for navigation, stripe image, automatically extracted markers, and anchors.

of which is modulated by the number of annotations on the anchor normalized by the maximum number of annotations on any anchor for the video. These anchors are also semi-transparent so that areas of the transcript with more annotations appear darker. This is a form of social navigation allowing users to orient and navigate to anchors containing more annotations [9].

Timeline View

Navigation through the video is accomplished using standard video controls such as play and pause as well as by using a stacked timeline representation which shows four types of information (Figure 4). The top level of the timeline, labelled *Navigation*, is a standard video thumb which allows for non-linear seeking within the video. The second level down is a stripe image indicating the pixel colors of the video at that point in time. The space efficient stripe image clearly shows boundaries between shots and is meant as an information scent to aid the user in navigating between segments of the video. The *Markers* level shows the same automatically extracted markers as are shown over the transcript (Figure 3) but synced to the time of the video.

The bottom most level of the timeline shows all of the anchors that have been added to the video. Anchors on the timeline are rendered with the same social navigation information as those shown over the transcript. Anchors can be rendered as overlapping, semi-transparent areas, or as a “stacked” view. The stacked view algorithmically determines how many sub-levels of the timeline are needed to avoid overlaps between anchors and then sub divides the height of the anchors level accordingly.

Hovering over the timeline brushes the other views. It shows the frame at that point in time in the video player area of the interface (Figure 2 upper left), highlights the word using a transparent gray box in the transcript area at that time and shows a tracking popup above the timeline with the three words in the temporal neighborhood of that frame. To our knowledge this type of tightly coupled navigation and browsing between a video and its transcript has never been attempted in this way.

Annotation View

The annotation view (Figure 5) shows the set of filtered annotations attached to the currently selected anchor on the timeline or transcript. If no anchor is selected, the annotation view shows the global annotations which apply to the entire video. The view is structured as a set of hierarchically composed annotations which can be expanded or contracted by clicking the arrow at the left

hand side of each annotation. Each annotation shows its type, the username of whoever added the annotation, how long ago the annotation was made, and the number of responses to that annotation.

At the right hand side of each annotation is a list of action buttons for interacting with the information. As described earlier, the actions available depend on the type of annotation and whether or not the user has added the annotation or someone else has added it. The view is designed this way to minimize memory and cognitive issues associated with learning the ontology. All initial actions are visible in the interface, however some of them expand. For instance, if the user clicks “add reaction” a drop down will show the options of “agree”, “disagree” etc. Also, clicking “add tags” pops up the tagging panel shown in Figure 1. At the top of the annotation view are the two primary actions available when any anchor is created or selected “make a comment” and “mark as a claim.”

The coloring of each annotation indicates the degree of support of that particular annotation in the system with shades of green representing more support and shades of red indicating less support. For example, in Figure 5 there is a claim that has been identified at the top which is colored slightly red because someone reacted in disagreement to the claim and the reason for this disagreement was corroborated by other users (i.e. it is slightly green).

Degree of support values in the annotation view are computed bottom-up from the annotation hierarchy. Each type of annotation is assigned an initial quality weight (QW) in [-1, 1]. For instance, negative valence tags such as “invalid” start with a QW of -0.5 whereas positive tags such as “accurate” have a QW of 0.5. For each node in the hierarchy its quality score (QS) is computed by summing the QWs of its children. The sum is divided by a constant



Figure 5. The Annotation View showing a claim that has lost support based on the series of annotations responding to it.

and capped at -1 and 1 so that a child does not overly impact its parent. Annotations need corroboration from other similarly valenced siblings to substantially effect the parent's QS. The QS is then mapped to the color red if less than 0 and to green if greater than 0.

The quality weight of the parent node is also modulated by the normalized child QW sum so that the support that node transmits to its parent in turn takes into account the values of its children. We also consider changes in sign of the parent child relationships. So, for instance, if an initially negative node has accumulated a negative quality score based on its child annotations, the value it transmits (its QW) will be less negative. Intuitively, a negative annotation that many disagree with or have rated down has less negative impact on its own parent.

Filtering

Filtering in Videolyzer is provided as a way to narrow down the visible annotations in the annotation view, timeline, and transcript (Figure 2 upper right). When particular annotations are filtered out they are still rendered at reduced saturation and brightness in the annotation view so that users can see the context of the remaining annotations. Anchors with no annotations in the filter are not rendered, thus reducing visual clutter.

The filters were designed to facilitate users finding "interesting" annotations to respond to. The filtering mechanism is a logical AND of three components: "select from", "tags", and "keywords." Specifying tags allows the user to search for annotations tagged with any of the quality, bias, or collaboration tags detailed earlier. Keyword filtering applies to any annotation with free text to search.

The "Select From" component of the filter allows users to specify the set of annotations being shown. Users can select from "All", "My", "Highly Rated", "Contentious", and "Feeling Lucky." To detect highly rated annotations, the system thresholds the QS of any given annotation. "Contentious" annotations are detected when the variance of QWs on an annotation is above a threshold. The "Feeling Lucky" filter randomly selects a set of annotations.

USER STUDY

We conducted a user study to help gain an initial understanding of the user experience of the application, whether it was usable, and if it was useful to political bloggers and journalists. At the interface level specifically, we were interested in looking at the usage and behavior around the transcript, timeline, and automated suggestion markers. And from the content, we wanted to explore the collaborative dynamic of the discourse generated with the tool, including people's information quality evaluations. We collected data from questionnaires, interaction logs, screen recordings, interviews, and annotation content.

Procedure

The study was conducted in a usability lab on a dual monitor computer so that participants could have one

monitor for using the application and one monitor for looking up information in a browser. Each participant first filled out a background questionnaire soliciting information about their experience and education in journalism, their online news and video consumption and commenting behavior, and their interest in a variety of political issues. Then the participant had ~15 minutes to read through and watch videos from a tutorial explaining all of the features of Videolyzer. The participant was then given a warm-up period where they used the interface to accomplish a list of tasks to ensure their familiarity with the different features.

Finally, the participant was given two videos (one at a time) to analyze with the interface. Their task was not specified precisely, but instructions were given to analyze the quality of the videos and to have a discussion with other users. The experimenter left the room and checked back on the participant in 15-20 minute intervals until the participant felt comfortable that they had "finished" their analysis. Afterwards the participant filled out a questionnaire and answered some questions in an interview. The questionnaire included ratings for various aspects of their experience on a 7 point Likert scale (1 negative and 7 positive). Subjects were run in a sequential order in which each participant could see the annotations of all prior participants. In addition to the initial and final questionnaires and interview we logged all interactions that the user had with the interface including the type and quantity of all annotations made. We also screen recorded sessions to look for interaction patterns and usability issues.

Content

The videos chosen for the study were taken from YouTube and were selected for their multivalent properties. We focused on produced content since it tends to be more information dense. We chose videos which related to the energy issue in the United States. Energy policy is a timely and relevant topic of interest to many Americans so we thought participants would already have some familiarity with the topic. Also, energy issues lend themselves to both objective analysis through statistics and subjective analysis via different perspectives on the issue.

The warm-up video was a 30 second advertisement from the American Petroleum Industry. The first video in the study proper was a 30 second political advertisement from the Republican nominee for President, John McCain, called "Pump." This ad had garnered attention in the media as containing many weak claims so it was a natural candidate for generating discussion between participants. For this video the commentary was seeded with information gleaned from a FactCheck.org¹ expert analysis of the video as well as with comments from three pilot participants. The final video in the study was a 2.5 minute mini-documentary from

¹ FactCheck is a nonprofit advocate "that aims to reduce the level of deception and confusion in U.S. politics." <http://www.factcheck.org>

the Heritage Foundation about U.S. energy policy as it relates to ethanol production. This video was not seeded with any annotations.

Participants

We recruited 10 participants from both local political blogs and news outlets through email and via subsequent snowball sampling. Six had a degree in journalism or had on the job experience in journalism and four did not. All but one participant were male and ages ranged from mid 20s to late 50s. All of our participants could be considered “news junkies” in that they all read online news several times per day and watched online political videos at least 2 times a week (and most watched much more than that). Everyone was a frequent reader of comments on blogs and videos, and all but one person had done extra research to support a comment “often” or “sometimes.” Also 8 of the 10 participants said they had spent extra time to verify facts presented in news and had cited original sources of information in comments “often” or “sometimes.”

Results

Interface Usage

The Transcript was *strongly* preferred for both creation of new anchors and selection of existing ones. In fact no new anchors were created on the timeline, and 43 were created on the transcripts across all participants and both study videos. The transcript was also used much more heavily for navigation and selection of anchors with 82% (151 / 184) of the selections. Part of the reliance on the transcript could be associated with the fact that the videos used were primarily audio information centric [18]; other types of content, such as very visually oriented video, could elicit different interaction strategies.

We asked participants about their use of the transcript in the interview. They thought it was much easier to mark areas of interest on the transcript, that the anchors were larger and easier to select, and that it was more suited to navigation in general. Users had a bias toward interacting with textual information, so the textualization of the video through the tight coupling of the transcript with the timeline seemed to be of great benefit. Several participants said that it helped them find interesting areas of the video to comment on: “*For finding what I was interested in commenting on in the video the transcript was best,*” and “*Having the transcript there really allows you to get directly to what they’re saying without having to go in there and listen.*”

Participants thought that having the integrated transcript allowed for an enhanced discussion of the video. One participant indicated it was easier to stay focused on what was actually said and to be commenting on the same thing as others by looking at the highlighted anchor on the transcript. Another liked the ability to highlight specific phrases where the meat of the content was. Not everyone thought the transcript was unequivocally great though. “*Most people watch videos because they don’t want to deal with text,*” remarked one participant. The transcript also

lacked punctuation such as capitals, periods, and paragraphs and could be difficult to parse. Sometimes the textualization of the video (including all the annotations) was distracting from the visual of the video.

Suggestive Annotation and Use of Automatic Markers

Automatically extracted markers were only used by 4 of the 10 participants. Among the reasons others gave for their disuse were that the videos were too short and that looking at the text was sufficient for finding what they wanted to comment on. One user mentioned that the markers didn’t correspond to things that he cared to comment on. A few others thought that the markers and their colors were overwhelming and were more than was needed to use the system effectively.

Several participants thought however that for longer videos the markers could be helpful for highlighting different types of arguments, or even as simple visual metrics for comparing videos. One participant acknowledged their potential usefulness saying, “*People equate facts to mentions of money, and feeling to how many times you mention those verbs.*” Indeed, one user told us in the interview that he had used the money marker to hone in on an area of the transcript that he ultimately wanted to comment on. He also tried the claim marker to jog his memory about where they had said something in the video, but ended up deciding not to comment based on it. Two other participants were observed to have used the highlighted markers in the process of navigation to hone in on anchors in a vicinity where they were interested in exploring the discussion.

Annotations and Annotation Behaviors

Across all 10 participants there were 236 annotations created including comments, claims, reactions, reasons, sources, and tags. The breakdown and distribution of their usage is shown in Figure 6. At the top level, comments were used almost twice as often as claims were marked. At the response level, reactions and reasons were the most heavily used aspects of the ontology representing half of all annotations. Tags were used primarily for criticism (not support) of other annotations including some use of the collaboration tags, “needs review” and “needs source.” Tags were only used by 6 of the 10 participants however.

The ability to source comments and provide reasons was a powerful feature and one used by all participants. In the user experience questionnaire users rated their ability to see and access the sources on which others had based their

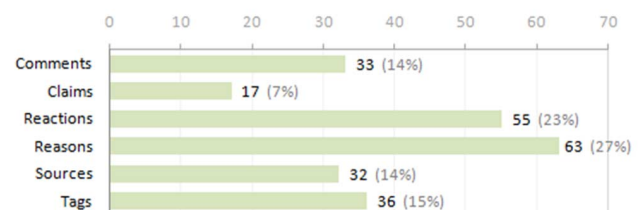


Figure 6. Annotation Type Usage and Distribution.

annotations as 6.60 / 7.0 with a stdev of 0.7. This is perhaps best summarized in one participant's words, *"I felt like my entry was not complete without one [a source] ... I needed to support my claims with more expert arguments."* There was also a good amount of corroboration and refutation that occurred between participants using sources. One person used the citation that another had previously added to refute that person's own comment, an activity that seemed to delight him. Another participant reflected on his use of sources, *"It turned out the biggest thing I was going after, the articles I easily found weren't supporting the point. It forces you to think about what you're doing."* The process of backing up arguments with reasoning and sources made participants more careful and reflective about the information they were using in the system.

One of our concerns with implementing a formal argumentation system was that the expressivity of the user could be compromised [22]. What we found though was that the semi-formalism that we designed was successful in letting people express and discuss the various aspects of information quality. The mean rating for, *"I felt I could express myself adequately with this tool."* was a 5.89 / 7.0 with a stdev of 0.93. The formal representations weren't used by all however. In fact there were two participants that only used comments and sources, but not reactions, reasons, or tags. One participant recognized that he had agency over the degree of formality of usage, saying, *"If you just want to make a comment, you make a comment, you don't have to put sources. Maybe you just want to get other people talking."* Some further observations about the use of the ontology are made in the following section.

Granularity and the specificity of annotation was another feature that many users appreciated. Several users mentioned that they liked being able to focus on a particular part of the video or to highlight specific statements and comments. It helped to organize the commentary into more focused and relevant threads. One participant noted, *"You're all launching off of an individual phrase or selection or image and so you're all on the same page."* Participants thought this helped at least somewhat to keep the discussion on topic.

Asynchronous interactions and collaborations among participants were observed in usage and by looking at the content produced in the annotations. Users were clearly influencing each other and building on each other's arguments. One participant remarked, *"As I come into this I see myself not just as a person that's gonna give my 2 cents, but my 2 cents are going to build up a potentially more interesting argument over time."* There were instances of joking, emoticons, and socializing and since we recruited from a local pool of participants several participants acknowledged that they recognized others by screen name. They referred to each other by name and one participant admitted that he let his knowledge of these other people influence how he responded to them in the system.

Participants responded to others with questions like, *"Where's your source?"* and in some cases replied to others' questions with answers or links to outside sources. Another way that users could interact with each other asynchronously was by reacting to another's annotation in terms of agreement, disagreement, or hedging. Of the 55 reactions recorded, 41 of those (75%) were reactions to other people's annotations. The reasons behind these reactions served to request additional information, sources, or parameters, to expand the issue, to add a caveat, to clarify a remark, or to add an alternate explanation. In general, reactions and their reasons expanded the comprehensiveness of the issues being discussed as participants brought individual knowledge, sources, and perspectives into the conversation.

User Experience

Participants rated their "overall satisfaction with the application and what it can do" as 5.50 / 7.0 with a stdev of 1.18. All other ratings are shown in Figure 7. For the most part users indicated that it was easy to navigate, read comments, and to annotate things. The last participant felt that anchors and annotations (almost 200 annotations at that point) were becoming dense enough to make navigation somewhat harder. A few users felt that the interface was over featured and felt overwhelmed with everything that it could do. The interface is clearly not suited for the average casual user. However, participants also acknowledged that they could fairly easily learn the core features and lexicon of the system within the 25-30 minute training and warm-up period. The rating for *"I would be able to produce a detailed analysis of the quality of a video using this tool."* was 5.89 / 7.0 with a stdev of 0.93, so users had indeed recognized the power of the tool and had gotten comfortable with the interface by the end of the study.

The learnability of the ontology used to drive the semi-formal interactions was somewhat of an issue for some users. Observations of the screen recordings of users showed that 6 of them were seen to hesitate and visibly vacillate their cursor between options in the ontology as they made a decision about what facet was most appropriate. One user was observed to have checked tags for "unclear" and "missing info" but then went back and

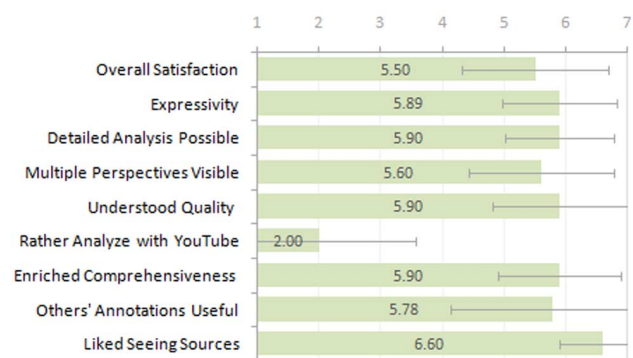


Figure 7. User Experience Ratings.

unchecked them in favor of “invalid.” It’s clear that there is additional cognitive overhead for using the semi-formal structure. At the same time, this structure had clear benefits for some. One participant remarked that, *“It is my reaction that I’m responding to this claim and then I have a reason that I’m providing and then I add the source to the reason. And that’s a nice structure ... I liked that.”*

A key design goal of Videolyzer was to enhance the understanding of the quality of the information people are consuming by providing the ability to see comprehensive information as well as diverse and varied perspectives on an issue. Users rated “I could see multiple perspectives on an issue by reading the annotations.” a 5.60 / 7.0 with stdev of 1.17, and “The annotations made the information in the video more comprehensive / complete.” a 5.90 / 7.0 with stdev 0.99. Also they rated “I felt I understood the quality of the information present in the video and its annotations better using this tool than I would have with a standard tool like YouTube.” as a 5.90 / 7.0 with stdev 1.10. Of course we can’t make any claim about an objective change in the information quality of the video, but these ratings indicate that participants’ *perceptions* of the quality of the information in the video were sharpened. Participants’ ability to see multiple and diverse perspectives on complex issues was enhanced. One remarked during the interview that, *“There were some thoughtful comments that made one consider different angles on the material.”*

DISCUSSION

The semi-formalism for quality analysis saw significant usage and did not substantially detract from people’s ability to express their evaluations. There were adequate opportunities for free text explanation including a familiar “comments” mechanism. People could choose to use the elements of the formalism that made most sense to them, could express evaluations of quality through free form text, and were not penalized for not using the formalism.

The ability to source and see sources was by far the most widely liked feature of the system. We suspect this is in large part due to the characteristics of our users; media savvy bloggers and journalists that know the value of buttressing and guarding arguments with thoroughly researched information. In fact one journalist went so far as to suggest she didn’t want to waste time looking at any comments that weren’t sourced. She suggested that the application would work best as a shared application in a single newsroom, allowing reporters to leverage each other’s sources and expertise to do research and analysis. Being able to add sources gave the interface a “serious” side that appealed to journalism professionals, and seemed to add to their acceptance of the technology. Designers of such interfaces for journalists or serious bloggers may want to consider this and other values central to the practice of journalism in their design [16].

Automatically extracted and highlighted markers such as money or comparisons were of some benefit for users but

may have added to an already overwhelming interface. While there were some interesting uses for navigation and comment creation their overall usefulness may be marginal compared to other more important features such as having the transcript tightly integrated with the video. Additional work is needed to assess whether the markers would be helpful for the analysis of longer videos or if there are other markers that are automatically extractable and better correspond to the types of things people are interested in.

Limitations

One limitation of Videolyzer in its current form is that transcripts must be obtained through the computer game before the video can be used in the application. This limits the ability of a user to use an arbitrary video, and means that either (1) a centralized repository of pre-processed, transcribed videos needs to be assembled for use or (2) we create a degenerate interface for videos without transcripts and provide a link to the transcription game. If a user is interested enough, they may choose to play the game before looking at the video.

Another limitation is the scalability of our interface representations. We designed our views for use with short (upto ~10 minute) internet videos, but in reality Videolyzer could also be useful for longer documentary type videos. One difficulty here is the scalability of the timeline representation for much longer videos, since for simplicity we currently don’t provide for zooming on the timeline.

The annotation view and the way in which hundreds of comments can be aggregated and represented is another difficulty with moving toward an internet scale application. The division of discussion across many anchors serves to keep commentary more focused and reduces the number of threads in any one anchor, but there is a limit to this. All tags and reactions (38% of annotations) can be directly aggregated using the ontology, but comments and reasons (41% of annotations) are plain text and need further processing before aggregation. This can be done using content analysis to compute derivative facets such as positive or negative sentiment in the text. Using both the structure of the ontology as well as derived tags via content analysis we expect to be able to develop alternate views and aggregate representations which reduce complexity further.

Future Work

A central area of future work is in developing effective *aggregate* visualizations of annotations both in the annotation view and in the timeline view. For the watcher of informational video online it is important to convey as much ancillary information for sense-making as possible without being overwhelming. Our goal will be to develop visualizations, which give people the appropriate cues to credibility and trust in the information of the video as they watch it, but also allow them to drill down for the rationale when appropriate or desired.

Among the things that we would like to study in future deployments are content and context of usage. Currently we have focused the content and user context of Videolyzer on political video but the tool could just as well be used with other topics (e.g. health, science, finance) as well as with other genres of video (e.g. documentary, interview, lecture). One deployment context could be a team of journalists acting as watchdogs who would use the tool to do internal critiques that are then published. An alternative scenario would be to study the use of the tool among a wider range of online bloggers and journalists that opt in to the system.

CONCLUSIONS

Videolyzer innovates by providing a fine-grained and tightly integrated video, transcript, and annotation analysis tool which incorporates a semi-formal representation of information quality to aid users in collaboratively organizing the discourse around and making sense of informational video. In our evaluation we found that participants' understanding of the information in the video in terms of its comprehensiveness, context, and multiple perspectives was enhanced using the tool. Users were able to collaborate and build off of (or refute) each others' arguments as they added their own unique information to the fray. Specificity and granularity of annotation was quite useful to users as it focused their attention and helped keep discussion on track. This was augmented by the tight integration of the transcript which was the preferred modality for interacting with the video. Also, automated content analysis, used to highlight different sections of text was of some value to participants in navigation and in finding interesting sections to comment on.

ACKNOWLEDGMENTS

The first author would like to thank Kathryn Bergman and his media savvy family for making this work possible. Support was provided by NSF IGERT-0221600.

REFERENCES

1. Compendium. <http://compendium.open.ac.uk/>
2. Open Calais. <http://www.opencalais.com>
3. Abowd, G., Gauger, M. and Lachenmann, A., The Family Video Archive: an annotation and browsing environment for home movies. in *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, (2003), 1-8.
4. Brill, E., A Simple Rule-Based Part of Speech Tagger. in *Conference on Applied Computational Linguistics*, (1993), 112-116.
5. Chklovski, T., Ratnakar, V. and Gil, Y., User Interfaces with Semi-Formal Representation: a Study of Designing Argumentation Structures. in *Intelligent User Interfaces (IUI)*, (2005).
6. Cockburn, A. and Dale, T., CEVA: A Tool for Collaborative Video Analysis. in *GROUP*, (1997), 47-55.
7. Diakopoulos, N. and Essa, I., An Annotation Model for Making Sense of Information Quality in Online Video. in *International Conference on the Pragmatic Web*, (2008).
8. Diakopoulos, N., Luther, K. and Essa, I., Audio Puzzler: Piecing Together Time-Stamped Speech Transcripts with a Puzzle Game. in *ACM Multimedia*, (2008).
9. Dieberger, A., Dourish, P., Höök, K., Resnick, P. and Wexelblat, A. Social Navigation: Techniques for Building More Usable Systems *interactions*, 2000.
10. Franklin, B., Hamer, M., Hanna, M., Kinsey, M. and Richardson, J.E. *Key Concepts in Journalism Studies*. Sage Publications, 2005.
11. Gil, Y. and Ratnakar, V., Trusting Information Sources One Citizen at a Time. in *International Semantic Web Conference (ISWC)*, (2002).
12. Heer, J., Viegas, F. and Wattenberg, M., Voyagers and voyeurs: supporting asynchronous collaborative information visualization. in *Proceedings of CHI*, (2007), 1029-1038.
13. Jackson, B. and Jamieson, K.H. *unSpun: finding facts in a world of disinformation*. Random House Trade Paperbacks, 2007.
14. Keelan, J., Pavri-Garcia, V., Tomlinson, G. and Wilson, K. YouTube as a source of Information on Immunization: A Content Analysis. *Journal of the American Medical Association*, 298 (21). 2482-2284.
15. Kirschner, P., Buckingham Shum, S. and Carr, C. (eds.). *Visualization Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer-Verlag, 2003.
16. Kovach, B. and Rosenstiel, T. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. Three Rivers Press, 2007.
17. Leech, G. *Meaning and The English Verb*. Pearson Education Limited, 2004.
18. Li, F., Gupta, A., Sanocki, e., He, L.-w. and Rui, Y., Browsing Digital Video. in *Proceedings of CHI*, (2000).
19. Mackay, W.E. EVA: an experimental video annotator for symbolic analysis of video data. *SIGCHI Bull.*, 21 (2). 68-71.
20. Pioch, N. and Everett, J., POLESTAR: Collaborative Knowledge Management and Sensemaking Tools for Intelligence Analysts. in *Conference on Information and Knowledge Management*, (2006), 513-521.
21. Serano, B., Buckingham Shum, S. and Motta, E., ClaimSpotter: An Environment to Support Sensemaking with Knowledge Triples. in *Intelligent User Interfaces (IUI)*, (2005).
22. Shipman, F. and Marshall, C., Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. in *CSCW*, (1999), 333-352.
23. Simon Buckingham Shum, R.S., Michael Daw, Ben Juby, Andrew Rowley, Michelle Bachler, Clara Mancini, Danus Michaelides, Rob Procter, David De Roure, Tim Chown, Terry Hewitt. *Memetic: An Infrastructure for Meeting Memory*, 2006.