Supporting Video Library Exploratory Search: When Storyboards are not Enough*

Michael G. Christel
Computer Science Department and Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
1-412-268-7799

christel@cs.cmu.edu

ABSTRACT
Storyboards, a grid layout of thumbnail images as surrogates representing video, have received much attention in video retrieval interfaces and published studies through the years, and work quite well as navigation aids and as facilitators for shot-based information retrieval. When the information need is tied less to shots and requires inspection of stories and across stories, other interfaces into the video data have been demonstrated to be quite useful. These interfaces include scatterplots for timelines, choropleth maps, dynamic query preview histograms, and named entity relation diagrams representing sets of hundreds or thousands of video stories. One challenge for interactive video search is to move beyond support for fact-finding and also address broader, longer term search activities of learning, analysis, synthesis, and discovery. Examples are shown for broadcast news and life oral histories, drawing from empirically collected data showing how such interfaces can promote improved exploratory search. This paper surveys and reflects on a body of Informedia interface work dealing with news, folding in for the first time an examination of exploratory transactions with an oral history corpus.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *video*

General Terms
Human Factors

Keywords
Digital video library, interactive video retrieval, exploratory search, video information views

1. INTRODUCTION
The overall market for online video is growing at a rapid clip, with YouTube the prime example accounting for over 28% of all videos watched online in November 2007: during that month, three quarters of Internet users in the United States watched on average 3.25 hours of online video [15]. As more users flock to online video and as online video collections dramatically expand in size, how do users interact with the video and find video items of interest? One primary way is through author-defined tags and user community-defined tags and recommendations, short text descriptors that grow with the community of users. Such "folksonomy" indexing dominates online multimedia repositories like YouTube, del.icio.us, and Flickr today, with one challenge for the video research community being to leverage this effort and enhance it through automatic information extraction and indexing methods. Another challenge is to establish benefits of such automated indexing methods beyond what is freely authored and established in video folksonomies. It may be that for fact-finding and lookup of specific topics, folksonomies and web portals that produce ranked lists of scrollable results are sufficient for the task. Rather than focus on fact-finding, this paper looks at a different sort of information search, *exploratory search*, and discusses why ranked lists of results or even traditional storyboards are not sufficient. Folksonomies can obviously improve interactive exploratory search interfaces as well, but this paper looks at what is possible without such social community recruitment and involvement in tagging and recommendations. Even without such additional descriptors, automated processing for video can extract and populate a number of information facets in support of exploratory search, as illustrated here with Carnegie Mellon University (CMU) Informedia processing of two very different archives: broadcast news, and life oral history interviews.

Marchionini breaks down three kinds of search activities: lookup, learn, and investigate, noting exploratory search as especially pertinent to learning and investigation [19]. He further deconstructs these activities as follows [19]:

- Learn: knowledge acquisition, comprehension/interpretation, comparison, aggregation/integration, socialization
- Investigate: accretion, analysis, exclusion/negation, synthesis, evaluation, discovery, planning/forecasting, transformation

Not surprisingly, high school and college teachers and professors working with the Informedia research group through the years have been quite interested in the use of the digital video library for these latter activities both for themselves and for their students. Likewise, the intelligence analyst community has been interested in these activities that go beyond satisfying a stated need through fact-finding and direct lookup [8, 10]. However, many traditional information retrieval studies, and forums like

***Proc. of the 2008 International Conference on Image and Video Retrieval CIVR'08,* July 7–9, 2008, Niagara Falls, Canada.**
http://doi.acm.org/10.1145/1386352.1386405

NIST's TRECVID, remain rooted in directed search and measures of precision and recall because there is a clear path to assessment given these quantitative metrics. This paper presents many uses of interactive video search, starting with TRECVID work and fact-finding but moving to exploratory search, which also moves the interface requirements beyond storyboards.

This paper shows the current state of Informedia storyboards first, as storyboards are the most frequently employed interface into video libraries seen today. That does not mean that they are sufficient. On the contrary, a 2007 workshop involving the BBC [1] witnessed discussion over the shortcomings of storyboards and the need for playable, temporal summaries and other forms of video surrogates for review and interactive interfaces for control. A BBC participant stated that we have had storyboards for over ten years now (more, if we go back to Mills' work with storyboards and QuickTime interfaces [20]), and that industry is looking to the multimedia research community for the latest advances. Playable video surrogates are reported in detail in that workshop [1] and so will not be emphasized here. This paper will focus on three challenges in interactive video search:

(1) Moving beyond fact-finding to exploratory search.

(2) Moving beyond broadcast news to interfaces supporting diverse video corpora.

(3) Evaluating such video interfaces for exploratory search.

The discussion will move from storyboards to numerous other ways of accessing information from video libraries as implemented in the CMU Informedia interface.

2. STORYBOARDS AND TRECVID

Storyboards work well for shot-based directed search information retrieval. This task has been used for TRECVID evaluations, and for such evaluations, storyboards consistently and overwhelmingly produce the best performance [4, 6, 11, 26].

Consider the TRECVID 2006 task to "find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible." The user when given this stated need might issue the text or-query "downtown city" returning 519 segments with 959 shots matching one or both text terms in the Informedia interface, with the resulting storyboard shown in part in Figure 1.

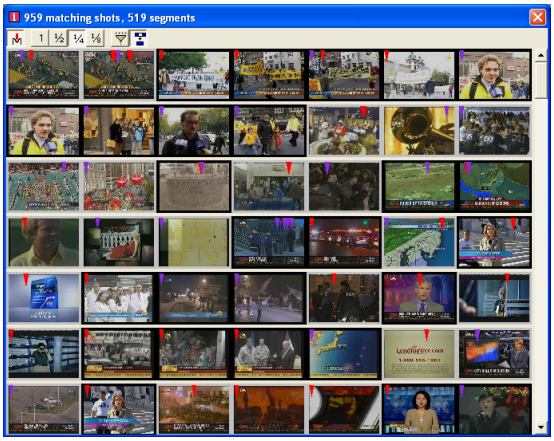

Figure 1. Informedia storyboard, TRECVID 2006 corpus.

Motivated users have been shown to navigate through thousands of thumbnails representing shots in such storyboard layouts as witnessed in the Video Olympics demonstrations at *CIVR 2007*. With the Informedia storyboards as shown, users on average can navigate through 2487 shots in the 15-minute timed period per TRECVID 2006 topic [7]. Clearly, the packed visual representation of thumbnails in this view allow for many shots to be reviewed efficiently. When the topic is very visual, as this one is regarding tall buildings, the user can scan the thumbnails and quickly isolate those potential candidates, for example the sixth shot on the top row. By adding trivial interactive control, here the ability to blow up the thumbnail pointed to by the mouse to full screen resolution by pressing the "Shift" key on the keyboard, the image can be verified for relevancy. A different shortcut key can cycle through a short burst of imagery from the shot to let motion and temporal information be reviewed quickly. The storyboard provides the overview, with details on demand available through further drill-down actions by the user. This "overview first, zoom and filter, details on demand" cycle, the Information Seeking Mantra [24], is often witnessed in novice user actions (users unfamiliar with TRECVID, *CIVR*, and Informedia) when addressing TRECVID topics [4, 6, 8]. Experienced users highly motivated to succeed on TRECVID tasks primarily stay with storyboard overviews, eschewing zoom and detailing actions for greater shot visual review efficiency of thousands of shots within 15 minutes [4, 6, 8].

When given a stated need, a short period of time to fulfill that need, many answer candidates, and an average precision metric to measure success, storyboards produce the best results [8, 11, 26]. Consider the same data set as used for Figure 1, though, and a precise need to find the one shot of downtown Chicago with a fire truck and people in the foreground. Consider a more open-ended need to contrast vehicular flow in different European cities. Consider a need to report on pedestrian traffic in Hong Kong, or the role of weather in urban/rural residency patterns in Africa. Such different needs may make use of the visual overview provided by a storyboard of 959 shots matching "downtown city" but so much more could be offered in the interface to support these actions. Other supporting views could be provided that leverage from more diverse, automatically derived metadata for video collections and an emphasis on other data facets besides visual thumbnails for shots.

3. THERE IS MORE TO SEARCH THAN FACT-FINDING

An exploratory search "may be characterized by the presence of some search technology and information objects that are inherently meaningful to users …often motivated by a complex information problem, and a poor understanding of terminology and information space structure" [27]. True end users of video collections – not the video indexing researchers themselves, but the end user community – often have poor understanding of what dimensions are available for searching and exploring in a video collection: e.g., face detection seems to work, so people detection should work just as well, or specific face recognition; automated speech recognition produces fine transcripts of talk in the studio, so it should work just as well for field reporting. Of course, the *CIVR* publishing community knows there are many pitfalls to automated multimedia content-based indexing and that such end user assumptions often result in disappointment. A challenge in

building search interfaces is to account for variability in the correctness of the underlying metadata and let the user explore through it, e.g., deciding whether to increase precision by dropping out low-confidence automated measures or increase recall by including them.

Gersh et al. discuss how exploring a set of information can help an analyst synthesize, understand, and present a coherent explanation of what it tells us about the world [10]. Marchionini and Geisler discuss agile views of files in the Open Video Digital Library [18, 19]: overviews of collections of video segments and previews of specific video objects, supporting Shneiderman's Information Seeking Mantra with dynamic interactive querying. The CMU Informedia research group has emphasized a multiplicity of views for information seeking, encouraging exploration, and providing insight from interactions with a large video corpus. Numerous published studies have empirically tested the usability of particular aspects of the interfaces [4, 6, 7, 8]. This paper looks broadly at the capabilities added through the years to the Informedia interface to support exploring video, using 2 collections as examples:

1. Broadcast news from CNN (English) and AZN (Mandarin) from January-May 2006: 240 hours, 11,191 story segments, 183,654 shots
2. Life oral history interviews from The HistoryMakers recorded from 1993 through August 2005: 913 hours, 399 interviewees, 18,254 story segments (1 shot per segment)

Experiences of CMU and University of Pittsburgh students and staff through the years are reflected in the comments made here. Direct quotes come from two pools of users. For broadcast news, six government intelligence analysts used the views reported here over a two-day period [8]. For HistoryMakers, over 50 students, including many from the Carnegie Mellon University History Department, made use of the views during the Fall 2007 semester; their transactions and reports are published here for the first time.

4. STORYBOARDS AS EXPLORATORY INTERFACES

Storyboards have been used for interactive search in TRECVID experiments by numerous research groups, with the blocky dense layout of Figure 1 being the most consistent presentation. Storyboards emphasize temporal flow within a story, with shots from the same video segment being displayed in the storyboard in time order. Recently, many researchers have looked to other ways of displaying shot thumbnail imagery to foster exploration along different dimensions than just temporal story flow. Two examples are Imperial College's Lateral Browser whereby a key thumbnail is plotted in the center with temporally adjacent shots represented in traditional linear storyboard fashion, but with a circular plot of thumbnails added to show shots related to the key by structure, color, texture, or other features [11]. A user may start with temporal shot navigation through storyboards but then jump to one of the circular plotted thumbnails to explore laterally along a dimension.

The MediaMill team achieved great success with interactive search of TRECVID topics using the CrossBrowser, [26, 28] which reduces the thumbnail set from the storyboard grid of Figure 1 to horizontal and vertical intersecting strips that repopulate as the user scrolls in either the horizontal (temporal, the time thread) or vertical (selected concept, e.g., "building") direction. Additional MediaMill layouts showing more thumbnails at once include the SphereBrowser and GalaxyBrowser [28].

Both the CrossBrowser and Lateral Browser layout strategies sacrifice thumbnail density for clarity. In extending Informedia storyboards we wished to retain the ability to pack many shots simultaneously before the user's eyes, to accommodate tasks where such an approach works well, e.g., identifying visually distinct shots like green soccer fields. We also wished to allow users to interactively control when there was a need for greater resolution over greater simultaneous display, so we trivially added the toolbar shown at the top of Figure 1 and Figure 2 to adjust thumbnail resolution. Most importantly, we added an ability to explore within the storyboard space through dynamic query sliders, interface widgets having proven success for supporting exploratory search [2, 19]. Consider a user wishing to browse the 959 shots of Figure 1 rather than linearly scan them. By using two filters to keep shots automatically tagged with medium or greater confidence as being both *outdoor* and *building* shots, the user sees the 20 shots shown in Figure 2.

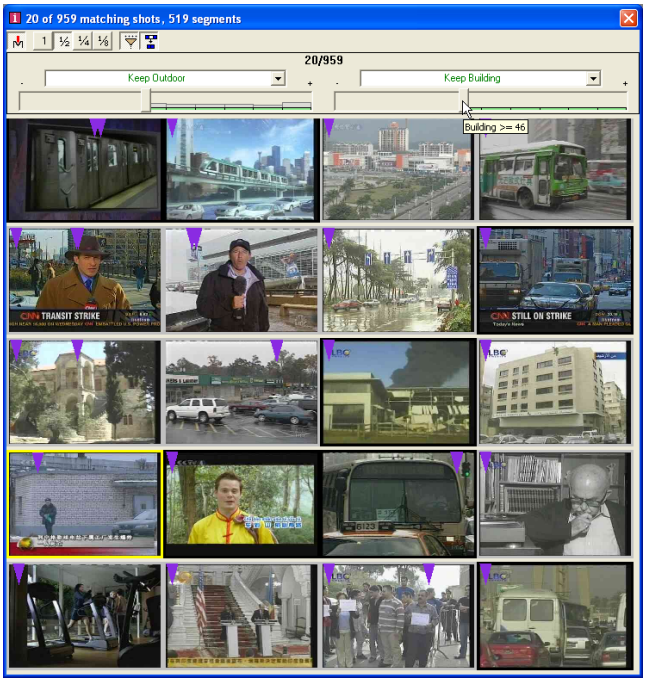

Figure 2. Using outdoor and building filters to reduce complexity of Figure 1 down to manageable size (959 to 20).

Through such sliders supporting confidence-based filtering across LSCOM-Lite concepts [21], the user is left in control as to whether precision is more important (e.g., reducing Figure 2 to only tall building shots), or whether recall is more important (e.g., allowing less confident outdoor-classified shots to show in Figure 2 so that perhaps 100 must be reviewed, but most tall buildings are then retained in the display). This storyboard display gives the user maximal control in terms of layout, thumbnail resolution, and thumbnail inclusion criteria, supporting exploration of the visual material in shots.

For many corpora, e.g., the HistoryMakers corpus of chiefly head-and-shoulders framed interviews, there are other dimensions worth exploring beyond visual detail. The next section presents interfaces appropriate to this corpus as well as for news.

5. A MULTIPLICITY OF VIEWS FOR EXPLORATORY SEARCH

When the information need is non-visual, thumbnail layouts like storyboards are not sufficient. For real-world users like government intelligence analysts and History students, their needs were often not addressed by storyboards at all. A common framework across both these user groups, and across both featured collections used in this paper (news and oral histories), is that time (when), location (where), and people (who) are important information dimensions. These dimensions can be addressed in part through storyboards coupled with dynamic query sliders, e.g., location as in Figure 2, or emphasizing people by filtering storyboards to include faces but exclude television anchorpeople. Hence, storyboards still play an important role for exploratory search in video, as they offer a rich window into the visuals. However, there are more facets with which to interact, and Informedia work through the past 13 years has led us to the ones discussed in this section.

A representative screen shot of the Informedia interface with HistoryMakers data is shown in Figure 3, with the tabs holding sets of video produced by query and browsing actions. For example, clicking on the purple term "Vietnam War" in the View Controller window would open a new tab filtered from 130 segments in the shown tab to just those 87 matching the phrase "Vietnam War." The View Controller in the upper left of every tab allows the video set to be rendered in different ways, i.e., checking a view "on" displays an additional window into the video set held in the tab. These windows each specialize in highlighting particular attributes of the video set, and lets users filter down to a subset in specialized ways. The advantage of the different views is to let the user explore the set of video data in varied means, rather than be restricted to a thumbnail grid as shown in the storyboard of the Shot Thumbnails view. Figure 3 shows three views:

- Segment Grid view emphasizing video story segments rather than shots, along with additional overlays of color-coded relevance score and term contributions.
- Nested Lists view, a text hierarchy of story segment titles organized by interviewee (discussed in Section 5.4).
- Named Entity view showing people, places, organizations, and their temporal associations (discussed in Section 5.5).

With the HistoryMakers, most shots show a talking head. With this corpus, a Segment Grid view, showing one thumbnail per interview story segment, is visually the same as showing a Shot Thumbnails storyboard view, one thumbnail per shot, as each segment holds but one shot.

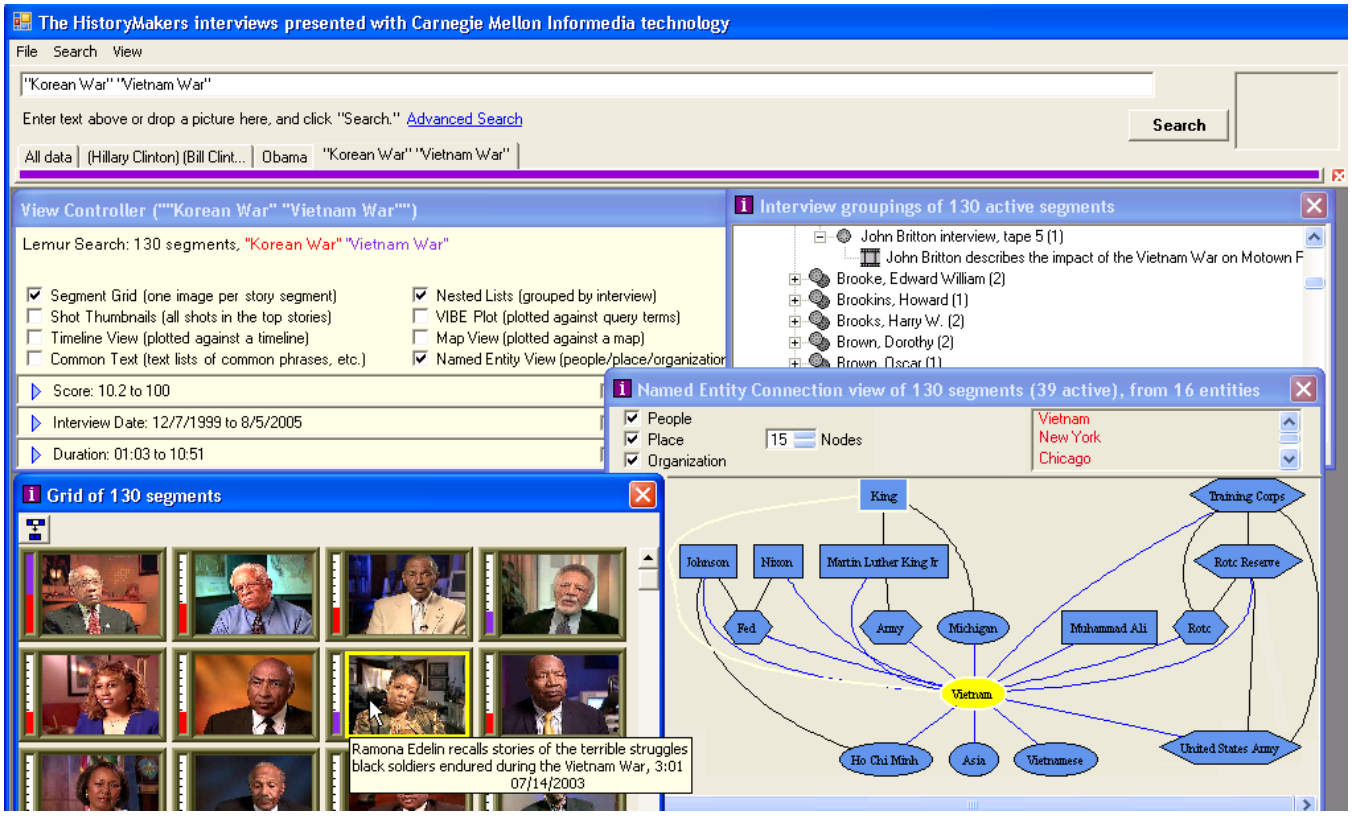

Figure 3. Informedia interface: query/browse action creates a video set shown in a tab; each set supports multiple views; views integrated through techniques from information visualization like brushing (here, mouse cursor over Ramona Edelin story in Segment Grid pops up text title there and highlights Vietnam-King named entity link from that story in the NE View).

5.1 Timeline View

The Timeline view emphasizes time, with the vertical axis by default the search action relevance score but also supporting other attributes like segment duration if selected from the combo box shown in the upper left of the view. Each green box (plot-box) represents at least one but possibly many video segments. Users can drill down in the timeline by rubber-band drawing a selection rectangle in the plot area, or by clicking on an x-axis button. Ease of use and satisfaction with such interactive control have been confirmed through numerous user studies with college students dating back to multimedia collage work in 2002 [3]. Figure 4 shows an interaction sequence against the 2006 news set following a query on "earthquake tornado volcano" that produces 128 segments. The plot shows interesting patterns in April so the user clicks "Apr" to drill into that month, and then bounds a stack of stories that map to April 3, 2006. Through a shortcut key these "9 of 128" active segments can populate their own video set (tab) and with the Common Text view and filtered storyboard Shot Thumbnails view, the user gets a clear picture and details of the tornadoes from that day.

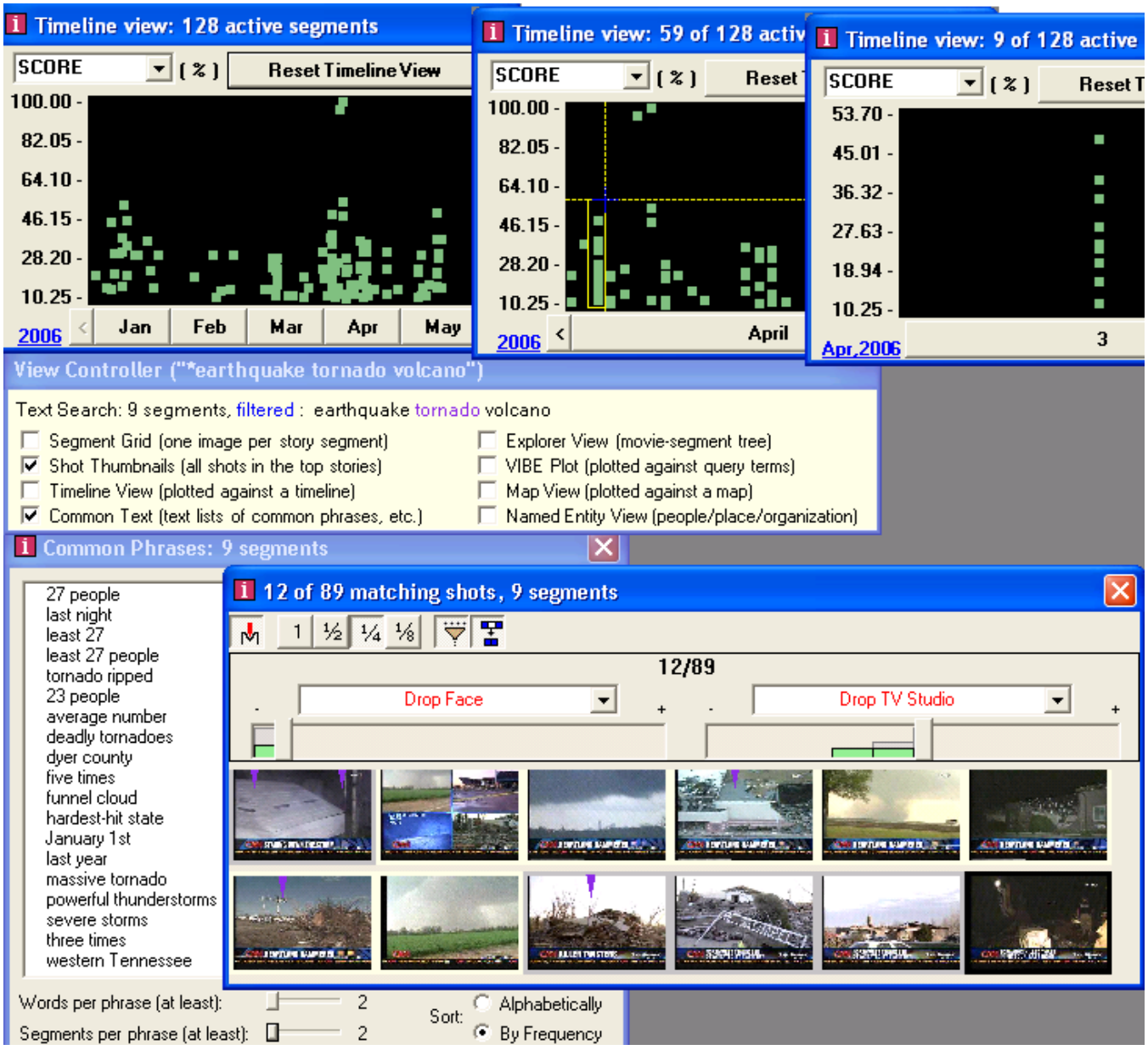

Figure 4. Timeline interaction: starting with 128 segments from "earthquake tornado volcano" query (upper left), interacting with timeline (middle upper) leads to nine segments on April 3, 2006 (upper right), which when loaded as its own video set shows only matches to "tornado", and a list of common phrases and non-face imagery that illustrate tornadoes and resulting damage.

For broadcast news, time plotting has been simplified to this point, using the very accurate broadcast date to plot each story to the day along the x-axis, but without additional parsing of time references within the news stories. With oral histories, the date of the interview is not nearly as helpful as the date of the broadcast for news. A steady stream of daily news allows for interactions like that shown in Figure 4, but oral histories are not recorded daily. However, oral histories contain a rich network of memories and time references, and these timeframes are what are plotted, rather than the recording date [5]. For example, a text search on *Korea Vietnam* produced 245 segments in the video set. The Timeline View notes 132 active segments as shown in Figure 5. The other 113 segments in the video set either do not mention a timeframe, or the archivists coupled with automated processing missed the timeframe, so these segments are not plotted within this view. The metadata emphasizes precision over recall, so tagged time references are very accurate, but some segments where the timeframe maybe could be inferred are not given the perhaps ambiguous time reference and so not plotted. Moving the mouse over a plot-box shows a tooltip with informative text, in the shown case indicating that this interview story with William M. Taylor (born 1930) contains a time reference to 1968 (indicated as its own line "1968") responsible for the green box that brushes yellow when the mouse arrow cursor moves over it. In addition, all other time references from the segment(s) under the mouse are brushed yellow. Brushing works across views, so the segment(s) under the mouse and their associations are highlighted yellow in other views in this video set as well, e.g., in Figure 3 the named entities from the pointed-to segment are brushed yellow. In this case, Taylor's story also has references to Nov. 1978; 1980; March 24, 1989; and 1991; showing up as additional yellow-colored plots in the timeline.

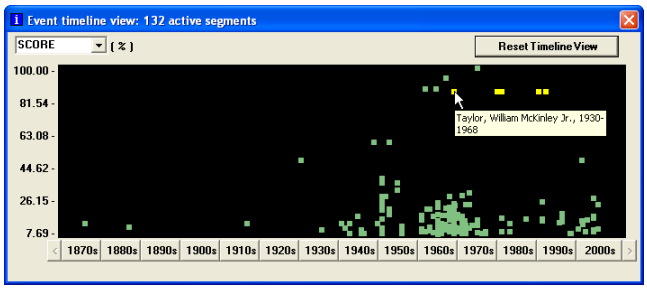

Figure 5. Timeline view following "Korea Vietnam" query in The HistoryMakers corpus.

5.2 Map View

The Map View emphasizes geographic distribution of the locations mentioned by the segments in the video set. Automatic named entity processing is run against the text metadata for video segments, e.g., the spoken transcript. A confidence measure is added to every tagged location reference based on the disambiguation provided in the sentence, segment, and full video broadcast/interview, along with heuristics regarding commonly mentioned cities, states, and countries. For example, a mention of London is taken with high confidence to be London, Ontario if that is how it is referenced in the transcript, and middle confidence as London England if there is no state or country qualifier mentioned in the story. A location Springfield remains at low confidence unless additional disambiguating container locations qualify it to a city within a particular state or country. The user has control over whether to consider just high confidence locations (favoring precision) or consider all tagged locations (favoring recall) just like the storyboard filters provide interactive control in Figure 2.

The Map View presents an overview for exploring locations represented in a video set, and for filtering that video set to a smaller active subset via locations. Consider for example a query on Hillary/Bill/President Clinton, which returns 204 segments as shown in Figure 6. Gray states are not mentioned, and by default the mentioned states are colored green, just like the default plot color in timeline and VIBE plots are green. If a View Controller slider "Color-code" is checked, the user can see that slider's relationship to plotted states, e.g., the lowest scoring segments are from the north central, northeast, and northwest.

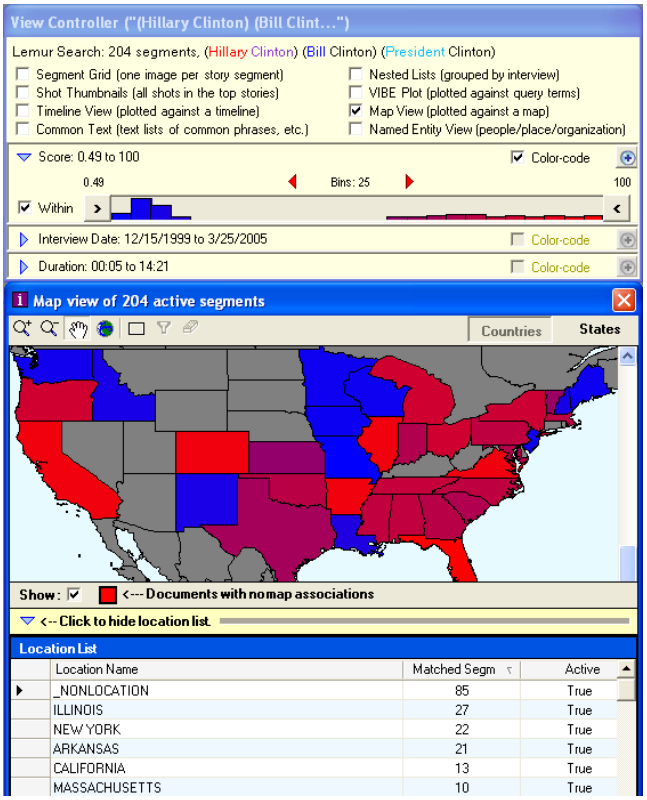

Figure 6. Map view of query results color-coding states based on query engine relevance score.

As an example of views allowing exploration, the user could open the timeline for the set of Figure 6 and click "2000s" to drill down just to the segments with 2000 decade references, filtering the set from 204 to 24. Opening this subset in its own tab lets the user see the map plot shown in Figure 7. The difference in map plot shows the U.S. Southeast remains an emphasis in stories discussing Clinton with 2000 time references. Such exploration could be the basis for follow-up research investigations, and such views as shown in these figures can form illustrations used in analysis reports. The African American oral history corpus included many U.S. state references along with country references, so for that domain, the support of "Countries" or "States" as the basis for the Map View works well. Other corpora of course may require different or additional layers of detail.

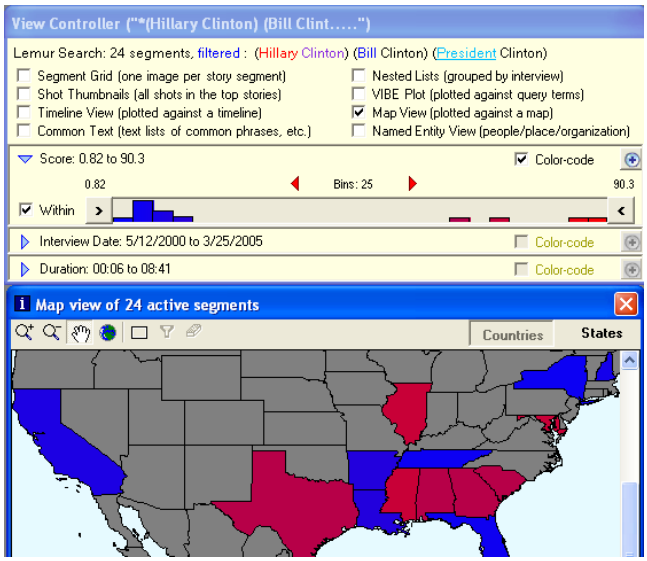

Figure 7. Video from Figure 6, filtered down to just 2000-2005 time references, shows shift in geographical pattern.

5.3 VIBE View

VIBE stands for "Visualization By Example" and is a presentation technique first developed at the University of Pittsburgh School of Library and Information Science [22]. It uses the query terms (words for text queries, matching cities/states/countries for geographic searches) as anchors for the plot, and then distributes the video segments on the plot based on their relative score from each of the anchors. It is most useful for exploring relationships between multiple matching query terms.

As an example, consider a map search on the "Four Corners" U.S. states against the HistoryMakers corpus, producing 362 segments in the resulting video set with 4 matching terms: Utah, Colorado, New Mexico, and Arizona. The VIBE View for this video set is shown in Figure 8, with the 4 states serving as plot anchors.

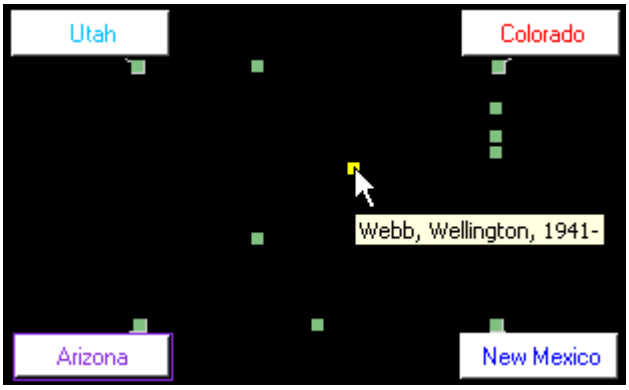

Figure 8. VIBE View for query term relations, here drilling down to see a story discussing both Colorado and Arizona.

Through user interaction, the VIBE View provides an understanding of relative contributions of multiple query terms. Pressing and moving the mouse while over an anchor drags the anchor to a new location. All the segments that match that anchor are likewise repositioned. Such interaction lets the user see from this plot that of 24 segments mentioning Utah, 23 only mention Utah with the other mentioning just Utah and Colorado, and 2 stories mention Arizona and New Mexico but neither of Utah or Colorado: the plot-box for these 2 stories is located between the Arizona and New Mexico anchors. Unchecking an anchor from the list (checklist not shown) removes it from the VIBE plot and deactivates segments that only match that anchor's term, e.g., unchecking Utah leaves "339 of 362 active segments." As with the timeline, users can brush over the points and get tooltip titles for the segments, right click on them to get a menu of actions, and can draw a rubber band zoom box on the plot area. If users right-click on a plot-box that represents a single segment, they can play its video, show its movie info, or navigate to its online biography, just like with thumbnail representations. If the plot-box represents more than one video, users can post the set of segments to a new tab with the "Show set in new tab…" menu item.

5.4 Text Views: Common Text, Nested Lists

Hearst discusses two popular methods for grouping: (1) clustering, grouping items according to some measure of similarity; and (2) hierarchical faceted categories, a set of category hierarchies each of which corresponds to a different facet relevant to the collection to be navigated [14]. The Common Text View supports clustering based on simple statistics for the text metadata associated with a video segment. This text metadata most often is the transcript of the narrative provided through closed captioning and human transcription or supplied through speech recognition, along with recognized overlay text and any additional text descriptors that might be provided through formal means like human archivist titling or informal means like social collaborative tagging. The Common Text View presents the most common phrases from a video set's metadata, organized and filtered by phrase length and frequency. An example is shown in Figure 4. The list as currently implemented could be trivially improved with some redundancy removal (e.g., Figure 4 shows "least 27 people" but also "least 27" and "27 people"), but even in its current form this view has received the most use by History students exploring the HistoryMakers corpus in the Fall of 2007.

The Nested Lists View organizes the segments according to a domain-specific hierarchy. For oral histories, the segments can be nested under the speaker, and speakers nested under their HistoryMaker category, e.g., "PoliticalMakers" or "BusinessMakers." For news, the segments are nested under the broadcast date and then the broadcaster and broadcast language. The facets to use depend both on the corpus and the anticipated needs of the users, with the Nested Lists View (e.g., Figure 3) supporting scrolling through hierarchical views of nested text.

5.5 Named Entity View

The maps in Section 5.3 are useful for showing geographic distribution, and the text list accompanying the map (shown in Figure 6) helps in promoting small regions that might get overlooked in a choropleth map display. Another way to show locations is to chart them as boxes, with lines connecting the boxes if the locations are mentioned together in N or more story segments, with N under control of the user. Similarly, the chart can show organizations and people named entities as well, using the automated named entity extraction method first employed by Informedia processing to generate collage interfaces [3].

An example is shown in Figure 1, with 39 segments connected to the user-selected center of focus, "Vietnam", through N=2

segments or more mentioning "Vietnam" and the plotted named entity. The difficulty with users seeing this view for the first time is that initial plots may be criss-crossing webs of complexity: too many links to too many boxes (i.e., nodes). The user can change the center of focus, and reduce the plot complexity with a maximum number of nodes to plot. As with the Common Text view, there should also be better redundancy removal, e.g., the use of canonical named entities so that both "King" and "Martin Luther King, Jr." are not plotted as shown in Figure 3.

Depending on user background, this view was either very promising or ignored. History students ignored it in favor of the linear lists of the Common Text view. The government intelligence analysts familiar with tools producing similar plots were eager to use it and drilled down with it to show video skims emphasizing how two named entities are related. For example, in Figure 3 there is a link between "Muhammad Ali" and "Vietnam" produced by 3 stories in the set of 130 mentioning these two entities in close temporal proximity. Through a pop-up menu, a video skim emphasizing the Ali-Vietnam connection is played, composed of relevant extracts from the 3 stories. This combination of named entities to provide an overview, user zooming to identify a neighborhood of interest, and use of that action to define a video summary, addresses the skim discussion of [1] in a different way: the playable video summary is not defined a priori but only through user interaction, in this case setting up Ali-Vietnam as the portions of interest.

6. FURTHER USAGE DATA

38 HistoryMakers corpus users opted in to surveys (mostly students, 15 female, average age 24) showing them as experienced web searchers but inexperienced digital video searchers. For "Do you search any web/online information systems?" (1=Not at all, 5=Very frequently (several times daily), the answer distribution was 1-1-5-13-18 while for "Do you use any digital video retrieval system (video stored and searchable on a computer)?" with the same scale, the distribution was 13-12-10-2-1. The latter two questions have been asked in numerous Informedia studies, e.g., see [6] with similar participant groups, and since 2002 the web search experience has grown from modest to frequent, and video search experience from not at all to modest. It is anticipated that as video search sites like YouTube maintain popularity that encourages numerous imitating sites [15], the college student of the future will have even greater experience in online video searching.

Six intelligence analysts (1 female; 2 older than 40, 3 in their 30s, 1 in 20s) used the stated news corpus (Section 3) as well as for TRECVID studies reported in [8]. These analysts compared to the university students participating in cited Informedia studies, were older, more familiar with TV news, just as experienced with web search systems and frequent web searchers, but less experienced digital video searchers. Their expertise was in mining text sources and text-based information retrieval rather than video search. Of course working with even more analysts would have been desirable to better represent the user pool of people mining open broadcast sources for information as their profession. Global political situations, demands on analysts' time, and logistics limited our access to six individuals over a two-day period.

For both groups, users commented on the potential of the various views of Section 5, with general praise in concluding survey remarks, e.g., "views (named entity, map, etc.) useful for exploratory topics." The Storyboard was noted for what is does best: "quick presentation of great volume in imagery." In general, the feedback was positive, e.g., one analyst remarked: "This is the fourth information retrieval system I have evaluated in my analyst role, and it is the easiest to learn and use, provides great functionality and features, and has great utility for the Intelligence Community."

However, there was not much use of views as evidenced in collected transaction logs for both user communities. The exception was the Common Text view, which was used frequently even by these first-time users (both groups). Text titling for video segments and text transcripts were also heavily accessed. A prior eye-tracking empirical study looked at digital video surrogates comprised of text and thumbnail imagery to represent documents, and found that participants looked at and fixated on text far more than pictures. They used the text as an anchor from which to make judgments about the list of results [16]. Here, too, first-time system users accessing either oral history or news video began with and stayed primarily with text-based representations. There were exceptions, e.g., one History student gathering data for a term paper made repeated use of the Map View to investigate differences on a topic between the New York and Chicago areas. For the six analysts, transaction logs show collectively 1433 video plays and 433,031 shot scans on their own exploratory searches, with some probing of timelines and named entity charts. Overall, though, the analysts were cautious in their use of the system, staying with default settings as is typical of most usage studies of information visualization systems for the initial trial sessions.

7. CONCLUDING REMARKS: EVALUATION HURDLES

Empirically evaluating interfaces like the multiple views presented in Section 5 has proven to be difficult. Partly this is due to the interface complexity. If low-level simple tasks are used for evaluation, it is easier to attribute differences in task performance to the different visualization attributes, but the simple tasks may bear little resemblance to real-world tasks. If complex tasks that come closer to real-world tasks are used, then more factors may confound the observed outcomes [17]. Another difficulty is in determining the appropriate metrics to use. Measures for efficiency, satisfaction, and effectiveness are recommended in general [9], but these may be difficult to assess for visualization interfaces where browsing, querying, navigating, and scanning are all actions interwoven in the information access process [12, 13, 23]. For example, do users who spend more time with a visualization system act so because it promotes exploration of potentially relevant areas, or are they spending more time because of problems comprehending the interface? For simple fact-finding tasks, effectiveness can be easily assessed, but the task is not well suited for visualization. If the user is asked to solve a precise information need, then the statement of that need can obviate the use of a browsing, exploratory interface (hence, the reason why exploratory visualization interfaces are not necessary for TRECVID tasks where the topics are stated with adequate text and visual detail). The user could just enter that precise query itself into the system and check the top answers, or prepare for intense visual inspection and scrolling of storyboards of thousands of shots. However, if the information need is more

ambiguous and vague, then evaluation of effectiveness becomes tricky: was the need solved and to what degree? Good information visualization promotes a cycle of exploration and understanding that does not fit the traditional usability evaluation metrics of effectiveness and efficiency.

Plaisant suggests three "first steps" for improving interactive information visualization evaluation [23]: "the development of repositories of data and tasks, the gathering of case studies and success stories, and the strengthening of the role of toolkits." The first two are within the realm of new directions for TRECVID: to help in benchmark-based open evaluation of information visualization interfaces targeting large video corpora by providing a test repository and suitable exploratory tasks, where such tasks are motivated and defined based on gathered case studies of real-world exploratory video use.

Informedia storyboards were evaluated primarily through discount usability techniques, two of which were heuristic evaluation and think-aloud protocol [4]. Storyboards were found to be an ideal roadmap into a video possessing a number of shots, and very well suited to the TRECVID interactive search task emphasizing the retrieval of shots relevant to a stated task, producing excellent fact-finding performance [4, 6, 7, 11]. The evaluation of the non-storyboard views for exploration in video corpora is just beginning. In a first hour with the Informedia system these additional means are not utilized, perhaps because of inherent deficiencies but also very likely due to their being different from more traditional "issue text search, get ranked list of results" information lookup interactions. Only by working with a set of users over time might these additional views get noticed and employed in exploratory tasks, with quantitative usage data and qualitative feedback offering insights into their true utility. Plaisant and Shneiderman propose new research methods to deal with complex systems and changing use over time with "Multi-dimensional In-depth Long-term Case-studies (MILC)" [25]. Ideally, MILC research could be conducted with intelligence analysts, History students and faculty, and other representative user groups for different corpora over time, to see changing patterns of use and utility as those groups gain familiarity and experience with the system. In the term "Multi-dimensional In-depth Long-term Case studies" the multi-dimensional aspect refers to using observations, interviews, surveys, as well as automated logging to assess user performance and interface efficacy and utility. The in-depth aspect is the intense engagement of the researchers with the expert users to the point of becoming a partner or assistant. Long-term refers to longitudinal studies that begin with training in use of a specific tool through proficient usage that leads to strategy changes for the expert users. Case studies refer to the detailed reporting about a small number of individuals working on their own problems, in their normal environment. A HistoryMakers corpus-Informedia interface beta test is underway across a number of universities including CMU with plans to conduct such longitudinal work. In general, there are three significant hurdles that should be addressed in evaluating interactive, exploratory interfaces for video, each discussed in its own subsection.

7.1 Corpus Size

Users want and need a large corpus to explore. The nature of exploratory search is that users will rummage through peripheral material, discover some aspect of interest in a tangent, and in that tangent go probing for additional material without running into the boundaries of the test corpus. If the corpus is not large enough, there will not be a rich enough periphery for any new discoveries, and there will not be enough material to support various tangential directions of inquiry.

In their first hour interacting with the HistoryMakers system, 38 participants in a study considered the corpus as having just a bit too much content. On a 1 (too little content) to 5 (too much content) scale, the average rating was 3.43. It is anticipated that these users will desire more data as they spend more time with the system, with plans to eventually grow The HistoryMakers corpus from 400 interviewees to 5000. With the broadcast news corpus, the intelligence analysts were immediately impatient with corpus size. Within their two day workshop they could explore with the system, and attempted some investigations that were dead-ends because of small corpus coverage. One analyst tried to discover relationships between the Libyan leader's family and other countries for 2005 but found only a few stories on the leader, none on family members. Another analyst authored the information need "natural environments affected by global warming" which did not have much support in the corpus. For this topic for example, the same 3 "stock CNN topics" on bears, glaciers, and hurricanes dominated the found support materials. If the corpus covered more time and sources, one could envision materials on volcanoes, tsunamis, El Niño, etc., now being available to draw into support materials.

7.2 Task Definition

Users want to explore their own topic, not someone else's stated topic. The very nature of stating the topic already moves it toward more directed search. The motivation for finding relevant material is not as strong as when the user provides the task. For example, when given TRECVID tasks and procedure, the analysts did not feel compelled to find hundreds of relevant shots, even with the instructions to "find as many as possible in the 15 minutes." They felt that the tens of shots already collected were sufficient to fulfill the task and were satisfied with their relatively small answer sets [8]. The analysts were also not interested at all in sports topics, scoring relatively low on these topics but high on others compared with student performance in TRECVID experiments [8]. Some participants in a HistoryMakers experiment where the tasks were stated indicated frustration that they could not search their own information needs. Even when given an open-ended exploratory task, they wanted more control over the search topic. One student typed in an online survey form "I think this would be a fantastic resource to browse and get ideas ABOUT topics to research" (emphasis from student). Another typed "It was also hard to jump in and search on a topic that I hadn't thought about ahead of time" and a third typed "Trying to search for info that someone ELSE specifies, rather than coming up with your OWN topics, is frustrating. If I'd been searching for topics with which I had more of a personal stake and/or background knowledge, I might have been more satisfied/successful."

7.3 Time with the Users

Users will start simple and with what they know from prior experience. As discussed in Section 6, users coming to the Informedia system for the first time tend to be very experienced and comfortable with web-based search like Google text search, and less familiar with video search, but this is changing. So, first user trials with the Informedia system by History students and faculty stay with text search and a reliance on text-centric views like the Common Text view, as evidenced by transaction logs from Fall 2007.

Similarly, the analysts began working with the Informedia system by issuing numerous, often complex text queries. In interviews, the most common remark (by 4 of the 6) was that there was too little time to get into the other provided views, e.g., "Too little time to get a full grasp of the system – more days would be needed to know about the advanced features and to get practice in using the advanced features." Their transaction logs show that the most frequent operations performed were text search, with views like the named entities connection graph, map visualizer, timeline, and the visualization-by-example scatter plot touched upon but not used in detail. If evaluations are only done on "first impressions" with users new to an exploratory search system, the exploration will not be very deep and the users will stay with what they are familiar with from other systems and contexts. The most important challenge for evaluating exploratory video search systems is to conduct long-term investigations, the "L" from the MILC framework [25]. Then, as users' search strategies mature and their comfort levels and expertise with a system increase, the utility of other features like the many views overviewed in Section 5 can be truly assessed.

8. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation under Grant No. IIS-0205219 and Grant No. IIS-0705491. Special thanks to The HistoryMakers and its executive director Julieanna Richardson for enabling this video retrieval evaluation work, and to CMU History Professor Joe Trotter for his enthusiastic support.

9. REFERENCES

[1] *Proc. ACM Int'l Workshop on TRECVID Video Summarization* (Augsburg, Germany, in conjunction with *ACM Multimedia*, Sept. 28, 2007), ISBN: 978-1-59593-780-3.

[2] Ahlberg, C. and Shneiderman, B. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *Proc. ACM CHI* (Boston, MA, April 1994), 313-322.

[3] Christel, M., Hauptmann, A.G., Wactlar, H.D., and Ng, T.D. Collages as Dynamic Summaries for News Video. In *Proc. ACM Multimedia* (Juan-les-Pins, France, Dec. 2002), 561-569.

[4] Christel, M., and Moraveji, N. Finding the Right Shots: Assessing Usability and Performance of a Digital Video Library Interface. In *Proc. ACM Multimedia* (New York, NY, Oct. 2004), 732-739.

[5] Christel, M. Windowing Time in Digital Libraries. In *Proc. ACM/IEEE Joint Conf. Digital Libraries* (Chapel Hill, NC, June 2006), 190-191.

[6] Christel, M., and Conescu, R. Mining Novice User Activity with TRECVID Interactive Retrieval Tasks. In *LNCS 4071: Proc. Image and Video Retrieval (CIVR)* (Tempe, AZ, July 2006), Springer, Berlin, 2006, 21-30.

[7] Christel, M., and Yan, R. Merging Storyboard Strategies and Automatic Retrieval for Improving Interactive Video Search. In *Proc. CIVR* (Amsterdam, July 2007), 486-493.

[8] Christel, M. Establishing the Utility of Non-Text Search for News Video Retrieval with Real World Users. In *Proc. ACM Multimedia* (Augsburg, Germany, Sept. 2007), 706-717.

[9] Frøkjær, E., et al. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proc. ACM CHI* (The Hague, The Netherlands, April 2000), 345-352.

[10] Gersh, J., Lewis, B., et al. Supporting Insight-Based Information Exploration in Intelligence Analysis. *Comm. ACM* 49(4), 2006, 63-68.

[11] Hauptmann, A., and Christel, M. Successful Approaches in the TREC Video Retrieval Evaluations. In *Proc. ACM Multimedia* (New York, NY, October 2004), 668-675.

[12] Hearst M.A. 1999. User Interfaces and Visualization. In Baeza-Yates R, Ribeiro-Neto, B (eds.), *Modern Information Retrieval*, Addison Wesley/ACM Press, New York.

[13] Hearst, M.A., et al. Finding the flow in web site search. *Comm. ACM* 45(9), 2002, 42-49.

[14] Hearst, M.A. Clustering vs. Faceted Categories for Information Exploration. *Comm. ACM* 49(4), 2006, 59-61.

[15] Helft, M. YouTube's Traffic Continues to Snowball. *NY Times* (Jan. 17, 2008), bits.blogs.nytimes.com/2008/01/17/.

[16] Hughes, A., et al. Text or pictures? An eyetracking study of how people view digital video surrogates. In *Proc. Conf. Image and Video Retrieval (CIVR)* (Urbana-Champaign, IL, 2003), 271-280.

[17] Kobsa A. An Empirical Comparison of Three Commercial Information Visualization Systems. In *Proc. IEEE InfoVis* (San Diego, CA, Oct. 2001), 123-130.

[18] Marchionini, G., and Geisler, G. The Open Video Digital Library. *D-Lib Magazine* 8(12), Dec. 2002, www.dlib.org.

[19] Marchionini, G. Exploratory Search: From Finding to Understanding. *Comm. ACM* 49(4), 2006, 41-46.

[20] Mills, M., et al. A Magnifier Tool for Video Data. In *Proc. ACM CHI* (Monterey, CA, May 1992), 93-98.

[21] Naphade, M., et al. Large-Scale Concept Ontology for Multimedia. *IEEE MultiMedia* 13(3), 2006, 86-91.

[22] Olsen, K.A., et al. Visualization of a Document Collection: The VIBE System. *Info. Processing & Mgt* 29 (1993), 69-81.

[23] Plaisant, C. The Challenge of Information Visualization Evaluation. In *Proc. ACM Advanced Visual Interfaces* (Gallipoli, Italy, May 2004), 109-116.

[24] Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. IEEE Symposium on Visual Languages* (Boulder, CO, Sept. 1996), 336-343.

[25] Shneiderman, B., and Plaisant, C. Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. In *Proc. BELIV'06 Workshop, Advanced Visual Interfaces Conf.* (Venice, May 2006), 1-7.

[26] Snoek, C., Worring, M., Koelma, D., and Smeulders, A. A Learned Lexicon-Driven Paradigm for Interactive Video Retrieval. *IEEE Trans. Multimedia* 9(2), Feb. 2007, 280-292.

[27] White, R.W., et al. Supporting Exploratory Search, Introduction. *Comm. ACM* 49(4), 2006, 36-39.

[28] Worring, M., Snoek, C., et al. Mediamill: Advanced Browsing in News Video Archives. In *LNCS 4071: Proc. Image and Video Retrieval (CIVR)* (Tempe, AZ, July 2006), Springer, Berlin, 2006, 533-536.