Qualitative and Quantitative Methods in Libraries (QQML) 1:245 –252, 2014

A Mixed Method Framework for Evaluating Video Digital Libraries

Dan Albertson

School of Library and Information Studies, University of Alabama, Tuscaloosa, AL, USA

Abstract. This paper generalizes the methodology of a comprehensive interactive video retrieval study in order to propose a framework that can be applicable for evaluating digital libraries. The presented framework depicts a set of evaluations and other methodological consideration deemed to be significant to digital libraries. Quantitative methods described here measure outcomes and factors related to users and their action, perceptions, and experiences. Factors for which to compare and contrast these measures across the framework are also outlined. Qualitative measures, which are not common to larger interactive retrieval studies, are also proposed including technique and pertinent themes relevant to evaluations aimed at obtaining deeper understanding of users and digital libraries.

Keywords: video digital libraries, user-centered, evaluation framework

1. Introduction

From a user-centered perspective, digital libraries are retrieval tools that allow users to interact "in the loop" with centralized collections through user interfaces. Video digital libraries are retrieval systems that enable access to digitized video. Video digital libraries have been developed to make accessible collections, resembling special digital collections of video, from a variety of different domains, including education, medicine, nursing, law, entertainment, archived news, cultural heritage, and others.

The evaluation of video digital libraries should entail thorough user-centered analyses. Automatic video processing, on the other hand, has advanced through standardized systematic evaluations, like the Text Retrieval Conference's Video Retrieval Evaluation (TRECVID), which facilitate common system-centric evaluations. TRECVID has served as a testbed for evaluating new user interfaces and interaction techniques, e.g. collaborative video search and visualization interfaces, through systematic retrieval measures, but not

Received: 19.4.2013 / Accepted: 26.1.2014 ISSN 2241-1925

necessarily using standardized user-centered evaluations (Adcock, et al., 2008; Snoek, et al., 2010). The logical next step in video digital library research is to facilitate interactive research through foundations and/or frameworks for user-centered evaluation. Foundations for evaluating interactive video digital libraries will support research and ultimately lead to better interface designs. The analysis presented here strives to complement foundations in human-computer interaction (HCI), digital library, and user interface research.

2. Objectives and Need

This paper presents a framework to be used for developing methodologies for evaluating interactive video digital libraries; the objectives include:

- Develop a generalized framework for specifically conceiving and designing evaluation methodologies for video digital libraries.
- Illustrate the significant and influential experimental factors warranting evaluation, as part of one video digital library evaluation framework.
- Provide a foundation for user-centered video digital library research.

Digital library research needs foundations from which to build, frame, and guide future evaluations. This analysis will help researchers better understand how digital libraries and specific interactive components, like user interfaces, can be evaluated efficiently and effectively in a methodical and user-centric manner.

The need for such a framework is rather straightforward: there is currently not enough research, understanding, or generalized models of digital library evaluations being developed and applied. This gap contrasts from what has been developed for other areas of research, such as web usability, where standardized assessments and criteria have been in place for nearly two decades. The current state of digital library research presents opportunities for proposing foundations for user-centered evaluations. Video digital library projects are being developed for a wide range of collections, and digital library programs are conducting "in-house" analyses. So, this brings up the question: where and how do new video digital libraries begin evaluation? Furthermore, what are the strategies for developing initial methods and what are decisions based on? User-centered video digital library research is hindered without existing examples of evaluation frameworks; methods employed for user-centered video digital library studies are typically specific to a collection and/or user group, which, in turn, limits the ability to generalize and apply for other research projects.

Foundations are important and warranted for video research, specifically, separate from the evaluation of digital libraries built for other types of collections, due to the complex and multidimensional makeup of video. Video, unlike any other information resource, enables the automatic extraction of text, image, moving image, and audio, and, in addition, is a time-based format. As a result, video, in the context of digital library research, warrants independent analysis and thus separate evaluation frameworks and guidelines, as its structure affects users differently than text, image, and speech do alone.

3. Literature Review

HCI research has noted difficulties in generalizing methodologies and phenomenon corresponding to video digital libraries, as variability among different methodological considerations, e.g. user groups, domain, and system (i.e. interactive) components, will have an effect on the final experimental design (Christel, 2008a). However, it is warranted to initially attempt or openly discuss conceptual depictions of video digital library studies in order to support future research, keeping in mind the importance not to overgeneralize.

Christel (2006, 2007, 2008a, 2008b, 2009) reflects further on the evaluation of interactive video retrieval systems. Findings from these studies, which inform the development of generalized methodologies, include:

- The assessment of various evaluation methods for video digital library experiments ranged from the positive benefits of discount usability studies, focusing on heuristics with a smaller user pools, to the expressed need for long-term observational studies (Christel, 2006, 2007)
- The importance of mixed methods, quantitative and qualitative, as each informs the other in interactive video research, and the emphasis on HCI metrics of efficiency, satisfaction, and effectiveness (Christel, 2006)
- Perceived limitations of typical video retrieval experiments, such as that evaluation of video shots depths at 1,000 is not realistic of actual use of users, who typically browse at depths of 100 to 200 results. Also, users seek other video segments beyond a shot, the standard for TRECVID evaluations, including clips, stories, and/or scenes (Christel, 2007)
- The importance of "ecological validity"; e.g. not overly relying on college students for experimentation, but assembling a broader and more representative user sample, collecting robust experimental datasets, and developing accurate experimental tasks (Christel, 2006, 2007)
- Context and domain are both significant in video digital library studies; interactive components like user interfaces, results displays, and video representations will vary accordingly, so researchers should be cautious not to overgeneralize (Christel, 2006, 2008a).

4. Framework

The framework is presented in Figure 1; specific measures and analyses of the developed framework are listed in Table 1. Figure 1 presents the primary components of this evaluation framework, *user*, *interaction*, *system*, and *domain*, which will be discussed individually in this section, along with how each is specific to video digital libraries. While Figure 1 presents a general depiction of digital library evaluation, which is to be expected in any high-level framework, additional details of each are presented in Table 1. Figure 1 also depicts various overlaps within the framework to be taken into account when evaluating video digital libraries.

In no particular order, there is the *user* of the video digital library, as shown in Figure 1 and Table 1. Since digital libraries are interactive retrieval tools, thorough user-centered analysis is needed. It may be significant to gauge users' level of expertise with certain types technologies, either generally speaking or

even perhaps with video-based tools specifically. Considering video digital libraries are many times domain-centric, researchers may also find it important to assess users' judgments on experimental instruments, such as search topics developed around a domain-specific collection. Examining users will be important, not only for ensuring the evaluation involves a representative sample, but also for measuring other user-centric factors, such as domain familiarity and/or prior knowledge, which have been shown to influence other factors within the interactive video retrieval process. A natural overlap between the user, their interaction, and the retrieval system is depicted in the framework. While evaluation of the user is applicable to other types of digital library studies, the framework here suggests that there are considerations of the user that are specific to video and video digital libraries. For example, if gauging prior knowledge with a domain-centric collection, users may have varying levels of familiarity with the different types of information comprising one information item (i.e. video), which can suggest more-detailed information needs among users of the collection and thus additional assessments that may be needed.

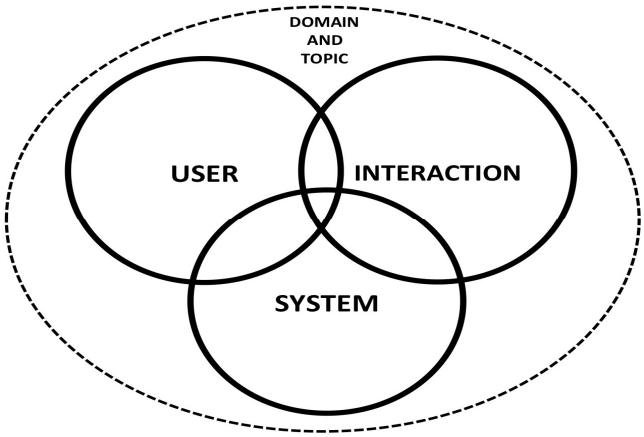

Figure 1: Framework for video digital library evaluation.

Interaction with the video digital library is next primary component of the framework. Interactivity is key to digital libraries; therefore, it is important that evaluations take into account how systems accommodate user actions with the system. Evaluating interaction should comprise assessments of what actions are being performed by the user, including types of search, browse, and other requests through the system. Factors related specifically to interaction, such as the number of steps and/or time to complete a given search topic, will be interrelated with the system (or digital library) and how it was designed. Interaction will also be associated with factors about the user, as they are the ones interacting with the digital library and may find certain types of interactions more effective than others. This component of evaluation is also relevant to video digital libraries specifically, as, due to video's composition,

more types or styles of user interaction will warrant evaluation, including multiple types of search, e.g. textual, visual, hybrid, and browse, e.g. title, inner-video / timeline, features.

	Primary Measure	***Instrument***	***Analysis***
User	Background Familiarity	Pre-survey Pre-survey	Quantitative Quantitative
Interaction	Actions (of users)	System log Observation	Quantitative Quantitative
System	Search results	Results analysis Post-search survey	Quantitative Quantitative
User-Interaction	Complexity	Post-search survey	Quantitative
User-System	Interface and video library design and organization	Post-survey Interview	Quantitative Qualitative
Interaction-System	Time Steps	System log Observation	Quantitative Quantitative
User-Interaction-System	Topic completion Performance satisfaction Satisfaction with design for topic completion	Observation Post-search survey Interview Interview	Quantitative Quantitative Qualitative Qualitative
Domain and Topics	Topic-Representativeness Topic influences	Post-search survey Interview Observation	Quantitative Qualitative Quantitative

Table 1: Specific evaluations as comprising the framework in Figure 1.

The next primary component of the evaluation framework is the *system*. The system should be developed in order to return sufficient results, e.g. video clips, based on submitted user queries. The system together with considerations of the user and their interactions are all necessary and interrelated for the purposes of evaluating system performance when subjectively gauged by the researcher and the user, in terms of satisfaction with the interface's ability to accommodate search topic completion. The system can and should be evaluated specifically to video, as a video digital library will likely require more features needing evaluation than would text or even image systems due to its multi-dimensional (text, audio, image) and time-based structure.

Finally, the *domain* and search *topic* form, in a sense, overarching factors affecting other primary components of digital library evaluation. The user, interaction, and system are all either directly or indirectly influenced by factors related to domain and the representativeness of the search topics, particularly in a digital library study or context. Topics can comprise various types of needs and structures; video search topics may include low-level visual needs (e.g. "red circles"), textual needs, visual needs, which are semantic in nature (e.g. "Samford Stadium at night"), abstract visual needs (e.g. "clip with visuals that represent happiness"), and others. Subsequently, researchers will find it useful

to examine their particular video digital library based on the search topics that are most applicable for the domain. Video search topics have shown to be influential in terms of effects on topic complexity, performance, and subsequently how users interact with and retrieve digital video (Albertson, 2012; Albertson & Meadow III, 2011; Albertson, 2010). Thus, different categories of search topics can be designed based on the number of steps or requests (i.e. single versus multiple), semantic versus abstract, and the inclusion of different characteristics, such as visual, textual, or hybrid needs (or any combination of any of the above). The various types of information embedded within video will enable researchers to evaluate a video digital library across a wider range of search topics that may not be suitable for other kinds (textual, image, audio) of digital libraries. Creation of experimental search topics needs to be carefully planned, so it is possible to evaluate the appropriate factors of the user and video digital library. Domain can be influential as well as the nature of the collection will have direct impacts on the designs of system features (like user interfaces) and thus the system needs for evaluation.

5. Discussion

Both positive implications and limitations accompany generalized frameworks. This section includes a discussion of the further benefits for video digital libraries and, more specifically, evaluation, and various limitations of the framework. The field of digital libraries needs foundations to support research.

The broader positive implication for this framework is that researchers and developers have a higher-level framework to consider and use for developing methodologies. The framework presented here enables researchers to draw upon and apply examples applicable to the evaluation of their particular video digital library. For example, a designer for a digital library that operates within a specialized domain, such as folklore, would likely consider that most users approaching the system would have some pre-existing knowledge of the field and/or some experience with that field's information resources. Application of the proposed framework would suggest the need to validate such and to assess users knowledge with search topics representative of the domain and the collection. Furthermore, evaluating certain types of system factors may be more relevant for those particular users, including semantic and/or advanced types of searches, using specialized search terms, and video title browse to support experts' navigation through the collection of familiar or known items. A second positive implication is, given the close proximity between video and image digital library research, a framework such as the one presented here can be applied, at least partially, downward to support the evaluation of other information resources, such as images, because of fundamental similarities.

The framework's primary limitations can be summed up by the idea that it cannot be exact, only approximate, which again is typical of other models. Whether evaluating user interaction or interface design for digital libraries, due to variations among experimental design, user group, etc., methods will vary or need customizing from the exact depiction of any generalized framework. For example, independent of a user's expertise in a particular field or with a

domain's information sources, high search tool experience of the user may result in deviation of a standard or typical evaluation of a digital library. Furthermore, findings presented in this particular framework may vary as users prolong their experience with a certain video digital library. In other words, this framework may not necessarily be the most applicable for longitudinal evaluation, which would be valuable to evaluate use in more natural of settings or garner qualitative trends over longer periods of time, in addition to individual users trying to satisfy an information need in an experimental setting. The framework strives to present findings using a suitable balance between generalizable methods and application in actual digital library evaluations.

The domain, independent of its individual users, can also cause variation from the framework. For example, art history, traditionally considered a visually-oriented domain, would demonstrate potential need for evaluating visual search features for video digital libraries. However, it should not be implied that the evaluation of visual features for art history would be warranted for other domains. For example, collections that pertain to video oral histories, where a full-length video can essentially contain one visual, i.e. one person's talking head, may not warrant the same evaluation of interface features as would art history, as emphasis would be placed on the story being told, not necessarily a range of visuals within the videos and clips.

Limitations also include that additional system or interface features, such as different types of video surrogates, may warrant evaluation. This study emphasizes search and browse, independent of how the video is being represented to users through the user interface. Moreover, while textual retrieval is typically accompanied with text-based surrogates, video surrogates can be enhanced with different combinations of text, image, audio or moving image information, all for supporting user decisions. Surrogate evaluation presents interesting considerations for future analyses of the framework.

References

Adcock, J., Pickens, J., Cooper, M., Anthony, L., Chen, F., and Qvarfordt, P. (2008), "FXPAL interactive search experiments for TRECVID 2007", in Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2007), November 5-6, 2007, NIST, Gaithersburg, MD.

Albertson, D. (2012). Examining feedback in interactive video retrieval. Journal of Information Science, 38(6), 501-511.

Albertson, D. & Meadows III, C. (2011). Situated topic complexity in interactive video retrieval. Journal of the American Society for Information Science and Technology, 62(9), 1676-1695.

Albertson, D. (2010). Influences of users' familiarity with visual search topics on interactive video digital libraries. Journal of the American Society for Information Science and Technology, 61(12), 2490-2502.

Albertson, D. (2010). Analyzing user interaction with the ViewFinder video retrieval system. Journal of the American Society for Information Science and Technology, 61(2), 238-252.

Christel, M.G. (2006), "Evaluation and user studies with respect to video summarization and browsing", in Society for Imaging Science and Technology (ISandT) and International Society for Optical Engineering (SPIE) Symposium on Electronic

Imaging, Multimedia Content Analysis, Management, and Retrieval in San Jose, CA, January 17-19, 2007, Society of Photo-Optical Instrumentation Engineers, Bellingham, WA.

Christel, M.G. (2007), "Examining user interactions with video retrieval systems", in Hanjalic, A., Schettini, R. and Sebe, N. (Eds.), Proceedings of International Society for Optical Engineering Conference (SPIE), Vol. 6506, Multimedia Content Access: Algorithms and Systems in San Jose, CA, February, 2007, Society of Photo-Optical Instrumentation Engineers, Bellingham, WA.

Christel, M.G. (2008a), "Amplifying video information-seeking success through rich, exploratory interfaces", in Tsihrintzis, G., Virvou, M., Howlett, R. and Jain, L. (Eds.), New Directions in Intelligent Interactive Multimedia, Vol. 142, Springer, Berlin and Heidelberg, pp. 21-30.

Christel, M. G. (2008b), "Supporting video library exploratory search: When storyboards are not enough", in Proceedings of the 2008 international conference on Content-based image and video retrieval (CIVR '08) in Niagara Falls, Ontario, July 7-9, 2008, ACM, New York, NY, USA, pp. 447-456.

Christel, M.G. (2009), "Assessing the usability of video browsing and summarization techniques", in Multimedia Content Analysis, Signals and Communication Technology, Springer, USA, pp. 1-34.

Snoek, C.G.M., van de Sande, K.E.A., de Rooij, O., Huurnink, B., Uijlings, J.R.R., et al. (2010), "The MediaMill TRECVID 2009 semantic video search engine", in Proceedings of the Text REtrieval Conference Video Retrieval Evaluation (TRECVID-2009), November 16-17, 2009, NIST, Gaithersburg, MD.