Social Video: A Platform for Collaborative Discoverability and Annotation of Disconnected Media

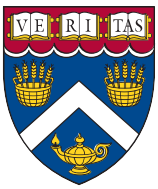
---

David Killeffer

May 21, 2016

Version: Proposal v0.1

Harvard University Extension School



HARVARD
Extension School

Department of Information Technology

Proposal

Social Video: A Platform for Collaborative Discoverability and Annotation of Disconnected Media

David Killeffer

1. Thesis Director Susan Buck
Department of Information Technology
Harvard University Extension School

2. Research Advisor Dr. Jeff Parker
Department of Information Technology
Harvard University Extension School

Supervisors Susan Buck and Dr. Jeff Parker

May 21, 2016

David Killeffer

Social Video: A Platform for Collaborative Discoverability and Annotation of Disconnected Media

Proposal, May 21, 2016

Reviewers: Susan Buck and Dr. Jeff Parker

Supervisors: Susan Buck and Dr. Jeff Parker

Harvard University Extension School

Department of Information Technology

51 Brattle Street, Cambridge, Massachusetts 02138

Abstract

There are very few good tools available that allow people to digitally categorize recorded videos. This is a problem because with the rise of digitization of recordings previously created on magnetic media and other media as well as the profileration of smartphones, there is an ever-growing volume of videos being made available to people, but a severe lack of ways to organize and search those videos. The absence of good tooling often means that many people amass large collections of self-recorded videos (both digital as well as older, analogue format recordings such as VHS, Mini-DV, etc.) which are largely inaccessible and make it extremely difficult to locate past recordings due to the absence of meaningful metadata. With the rise of cheap, large online storage resources and very powerful handheld recording devices, but few and inadequate tools to catalog them, there is a looming problem that many videos will go unwatched and uncared for, and the importance of their content go forgotten. By utilizing an easy to use web interface, this project attempts to bring the power of crowds to aid in adding metadata and meaningful annotations to online video for the benefit of other viewers.

Even if an individuals video recordings were somehow properly "tagged" with perfectly accurate metadata describing all the pertinent aspects of the video, without a rich, well-designed and easy to use search interface, the effort expended on such annotations would be wasted and the future utility of the recordings still in question. This project attempts to prototype a fully-featured, easy-to-use, faceted search interface which will help make locating and sharing recordings of precious memories easy and enjoyable. Additionally, once videos are properly tagged and annotated, users can explore further avenues to enrich sets of recordings by making "playlists" of segments of videos; for example, creating a "highlight reel" which is a playlist of all the recordings in a set that are tagged with the word "funny".

Acknowledgements

I would like to ackowledge the excellent help, guidance, and wisdom of both Susan Buck and Dr. Jeff Parker. They provided much needed tips and feedback over the course of the development of this thesis proposal which helped make it successful.

I want to also ackowledge and thank my wife Sarah and kids; thank you for being patient with me, putting up with my moods as I struggled through this process, and for all the tea.

Contents

1 Motivation and Problem Statement

1.1 Motivation

Today people are recording more videos than at any other time in history. Most people have very well equipped video recording devices that they carry in their pockets (smartphones) which have capabilities that dwarf even the most advanced handheld camcorders of just a few years ago. YouTube reports that they receive 300 hours of recorded video submissions PER MINUTE (see: https://www.youtube.com/yt/press/statistics.html, accessed 2015-05-22, http://www.tubefilter.com/2014/12/01/youtube-300-hours-video-per-minute/, accessed 2016-02-25, and http://expandedramblings.com/index.php/youtube-statistics/, accessed 2015-02-25). In addition to the stratospheric proliferation of newly recorded digital video, in recent years an entirely new industry of video preservation companies has sprung up which offer a variety of services to help both professional and consumer customers to digitize their recordings made on film, magnetic, or optical formats; these services will accept all manner of defunct formats and digitize the recordings captured on them in as high resolution as possible and create digital copies of the originals. With the future viability of nearly all physical media formats in doubt at best and al but assuredly over at worst, it seems the writing is on the wall for the future primacy of digital video as the format of the future. However, with the massive increase in the amount of recorded video content created, how should content be organized for sharing and annotated for posterity?

One popular method people used to use for creating "playlists", "favorite" recordings, and generally sharing of media in the past was to create so-called "mix-tapes" of favorite music and videos. People would typically use two "decks" to create the compilations; one would be used to play back the song or video they wanted to record, an the other deck would be used to create the new recording. Today this practice is an all-but forgotten media artifact of the 1980s and 1990s (rarely ever seen since the early 2000s). At the same time as people are no longer creating and sharing "mix-tapes" of either songs or videos for friends and family, the world is seeing the largest exponential growth of recorded media in its history; clearly the manual labor and time investment required in the now lost art of creating "mix-tapes" does not scale, and better solutions are needed for organizing, categorizing, and

sharing important digital video works. Without the benefit of carefully crafted and curated metadata to catalog the output of this new explosion of recordings, the future usefulness and viability of new recordings is in serious question. As video recording has shifted away from being created on physical media formats to digital formats, new digital recordings do not have the advantage of their physical media forebears which could be easily and simply labeled with a pen or marker to describe their contents.

At the same time as society has seen explosive growth in the proliferation of recordings due to technological advances, we have also seen the rise of the "social" web. People are sharing all aspects of their lives with friends, family, even perfect strangers online. Participants in the "social web" allow others to add metadata, comments, and add to their own digitally shared pictures, music, videos, etc. This has proven to be a very effective way to apply meaningful metadata to digital artifacts, and adds to the future longevity and viability of such digital artifacts (presuming that the metadata applied to individual files can be guaranteed to survive alongside the digital files themselves well into the future).

1.2 Personal Story

Several years ago my siblings and I lost our two remaining grandparents; my maternal grandmother, and my paternal grandfather, both within a couple years of each other. These were difficult losses to take, and I was left thinking of them often and the times we had together. At the same time, I was recently married and had started a family of my own, and enjoying all the highs, lows, and excitement of being a part of a young family with children. With sentimentality creeping into my mind more and more, we would often videotape our young son, and then his younger brother, and making memories with our children and saving them to video for posterity. Several years later I found myself with a very large collection of both VHS and Mini-DV tapes (well over a hundred tapes, at least), and I realized that my wife, kids, and I had not viewed most of these recordings, not even since they were originally taped. At an extended family dinner I inquired about all the old VHS tapes my parents had recorded of us kids growing up, including several important family milestones and a few select events that my grandparents would have been a part of; I was told that I was free to take any videotapes I could find. I did the same thing with my in-laws and gathered up all their old videotapes as well; soon I found myself in the posession of a virtual mountain of videotapes, some up to 25 years old. I knew that over time videotapes degrade and are subject to a process of "vinegarization" (see http://www.clir.org/pubs/reports/pub54/2what_wrong.html and http://www.emeraldartservices.com/visual-media-deterioration/), and I knew that if I wanted to preserve all

the precious memories that were captured on those tapes that I would need to digitize this collection.

Thus began a process of over a year's worth of work whereby I slowly digitized nearly 200 analogue tapes of various formats; VHS, Mini-DV, VHS-C, etc. Initially I was tempted to edit and cleanup the recordings as I imported them, but soon I found out just how much time and effort is involved in doing high-quality video editing and cleanup, and I realized I would never finish digitizing the tape collection if I stopped digitizing to edit each tape. Eventually I was able to successfully digitize about 99% of the videotape collection, and I was left with an enormous set of video files (one per tape). Some statistics of the collection:

- over 3000 different video files of various formats
- over 2.33 TB of disk space used
- over 200 total hours worth of recordings

While I certainly loved being able to go back and re-live many funny childhood moments that were now immortalized on an external hard drive, I quickly came to realize several things about my new massive video file collection:

1. I did not know ***who*** was in several of the videos, but I knew that my mom/my dad/my father-in-law/my uncle/etc. would know
2. I did not know ***when*** many of the recordings took place
3. I did not know ***where*** some of the recordings were made
4. It was unclear ***why*** some recordings were made, and not easy to ***decipher the purpose*** without viewing the video in its entirety (something I did not have the time to do when digitizing the entire old tape collection)
5. I did not know ***where*** many of the recordings were made
6. It was ***extremely difficult and time-consuming*** to be able to properly ***share*** old historical family moments from the video collection with the family because the videos were only labeled by type (VHS, Mini-DV, etc.) and tape number

The last point is perhaps the most poignant. In my eagerness and excitement to share my newfound digital video treasure trove with my family, I fumbled at several family dinners and gatherings when I was requested to play the "*funny barbecue video from 1987 where Nate accidentally hit cousin Al in the head*", and several other classic family gems; I simply couldn't find videos that I was looking for without basically brute-forcing my way and playing each video, fast-forwarding through the video until either I found (or did not find) what I was looking for, and then wrote down what video and at what time "that moment" that I was looking for was found in. The problem was that despite having invested over a year's worth of time and effort into preserving all these old family videos, it wasn't worth much to everybody

else (or me!) in their current form ***because nobody could find what they actually wanted to see without watching the entire collection***. And this revelation was the genesis of my thesis idea; to create a platform where I could leverage the knowledge and memories of my family to help me build up a set of rich metadata to annotate the family video collection, and then to reward them for their help in annotating the videos by building out a rich, faceted search interface to the video collection, as well as creating the ability to create "playlists" where users can make their own "highlight reels" of special moments, people, places, etc.

1.3 Problem Statement

The problems that my thesis thus attempts to address are:

- how to ***effectively annotate*** and add valuable metadata to a ***shared collection of video recordings*** via a loosely affiliated group of friends and family
- how to best ***organize and present valuable user-added metadata*** from video recordings in ***a faceted search interface*** that provides an easy way for people to find and play back videos they might otherwise not even know exist
- enabling users to ***explore libraries of richly annotated videos*** and ***make "playlists" of their favorite videos/segments*** based on the metadata that the videos have been tagged with

My thesis project will attempt to answer these questions by architecting a web application prototype where users can add videos, annotate who is in those videos, what the content of those videos is, where they take place, when they take place, and more. This metadata will be aggregated and used to power a faceted search, which will expose the video collection in a powerful new way to users and allow them to explore and reengage with the video collection in new and compelling ways.

2 Prior Work

The idea of annotating videos is not entirely new or novel, but such tools have not become commonplace in the same way that high quality image editing software and facial recognition algorithms have brought new dimensions to digital photography. YouTube has empowered people to share their recordings with the world in new ways and given rise to entirely new forms of entertainment. In the realm of education with the rise in online education and increasing pressure to make class lectures available to students both on-campus and remote, many classes are now recorded and distributed online, and there has been some scholarly research work done to enable students to annotate and share notes on classes. In addition to work on video annotations in academia, there has also been some work in industry as well, although these tools have largely not been in the hand of consumers and end-users.

Some prior works in the area of video annotation include:

2.1 OpinionCity

by Daniel P. Coffey: danielpcoffey@gmail.com , Spring 2015

OpinionCity is a website that was created by Daniel P. Coffey for a Digital Media Capstone project at Harvard Extension School. It is a tool for real-time group feedback on videos that have been uploaded to YouTube and allows for collaborative, time-code based annotation of videos as well as whole-video annotations. Users may select a video to "upload" to OpinionCity where the video from YouTube will play in an embedded player, and then users can add comments to the video at specific timecodes or apply their remarks to the entire video. Users can also invite others to join in and comment on the video as well.

While OpinionCity does allow for users to add annotations to specific parts of videos, it does not appear to allow for very granular annotations; specifically, a user may add an annotation at a specific part of the video denoting the "beginning" of an annotation, but cannot mark the "end" of the annotation. This means that the annotation functionality is somewhat limited in that users are not able to define during what specific timeframe in a video a person is in, or where the video was shot, etc. For annotating relatively shorter-length videos such as are often found on

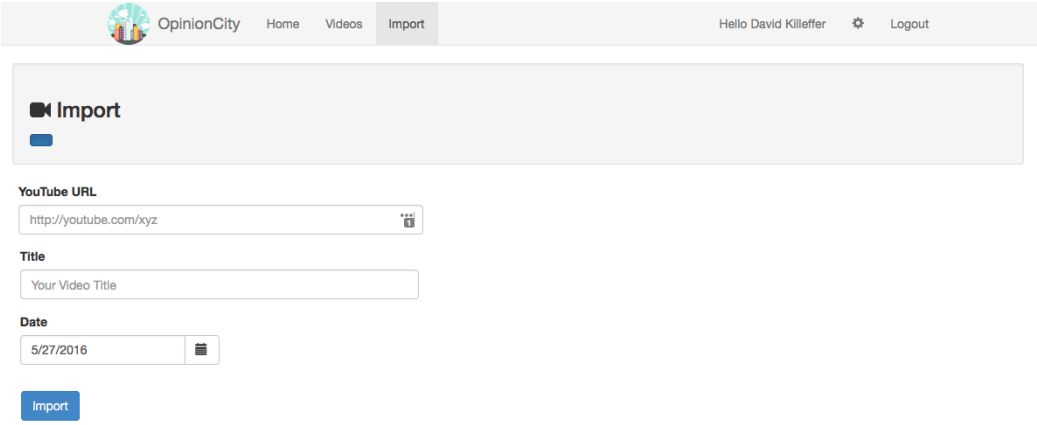


Fig. 2.1: *(OpinionCity)* importing a video from YouTube into OpinionCity

YouTube, this may be fine, but for annotating entire tape-length (typically 60-120 minutes or occasionally longer), this mechanism would be insufficient.

Additionally, OpinionCity does not appear to have any search functionality. Users may add annotations to videos, but there is no mechanism whereby they can then later search what annotations have been added. The project seems to have focused much more on the real-time aspects of commenting on a video, where a group of interested people may be watching a video at the same time and then making annotations and comments, and quickly viewing what others have likewise annotated and commented. Here are some screenshots of OpinionCity:

2.2 Media-rich Video Annotation Tool (MVAT)

by Philip Desenne: `desenne@fas.harvard.edu` *, May 2012*

The Media-rich Video Annotation Tool (MVAT) is a prototype tool developed by Philip Desenne as part of an A.L.M. in Information Technology thesis project at Harvard Extension School. Motivation for the development of the MVAT stemmed from Desenne's work as an Academic Technologies Product Manager to support learning and simplify the process of creating and sharing video annotations amongst

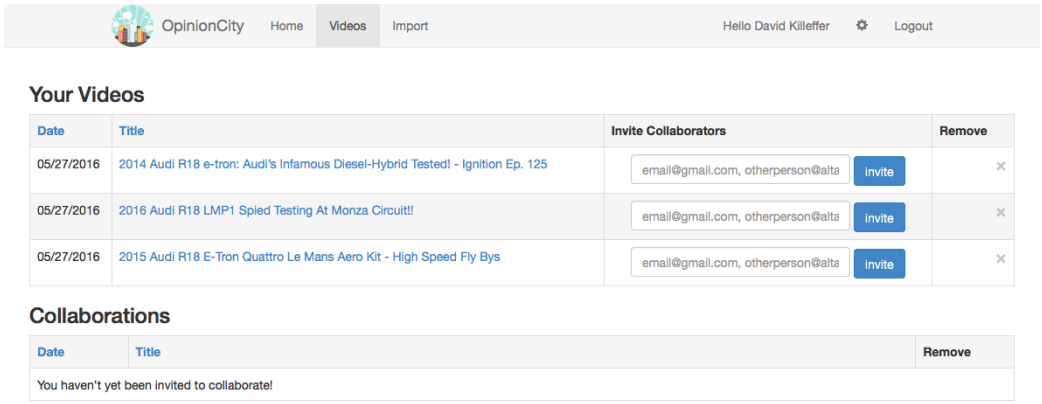


Fig. 2.2: *(OpinionCity)* listing of all videos associated with a user account

students in a pedagogical context. MVAT allows for a wide variety of media rich annotations, including adding text, HTML, pictures, actual vector drawings that users add, geographical notations, etc., all of which are very useful and support the educational aims of lecture videos.

The prototype focused on allowing users to create "media-rich" annotations so users could add much more than just plain text or image annotations, as well as link to outside supporting resources, and have a very simple, easy-to-use interface. MVAT was developed as an Adobe Air standalone application, and requires a data synchronization mechanism to upload annotations to an online SQL database from the embedded SQL-Lite database. Desenne acknowledged that while his selection of Adobe Air / Flex as a development platform enabled him to rapidly prototype the MVAT due to his experience with Adobe Air / Flex, it is a rather limiting choice long-term since Flex "was unleashed from Adobe" and Flash video usage has largely gone to the wayside in favor of open standards for video such as HTML5 video. Additionally, the MVAT prototype was limited to a single computer, and so other students were not able to benefit from, search for, or share the annotations made by one user with other classmates.

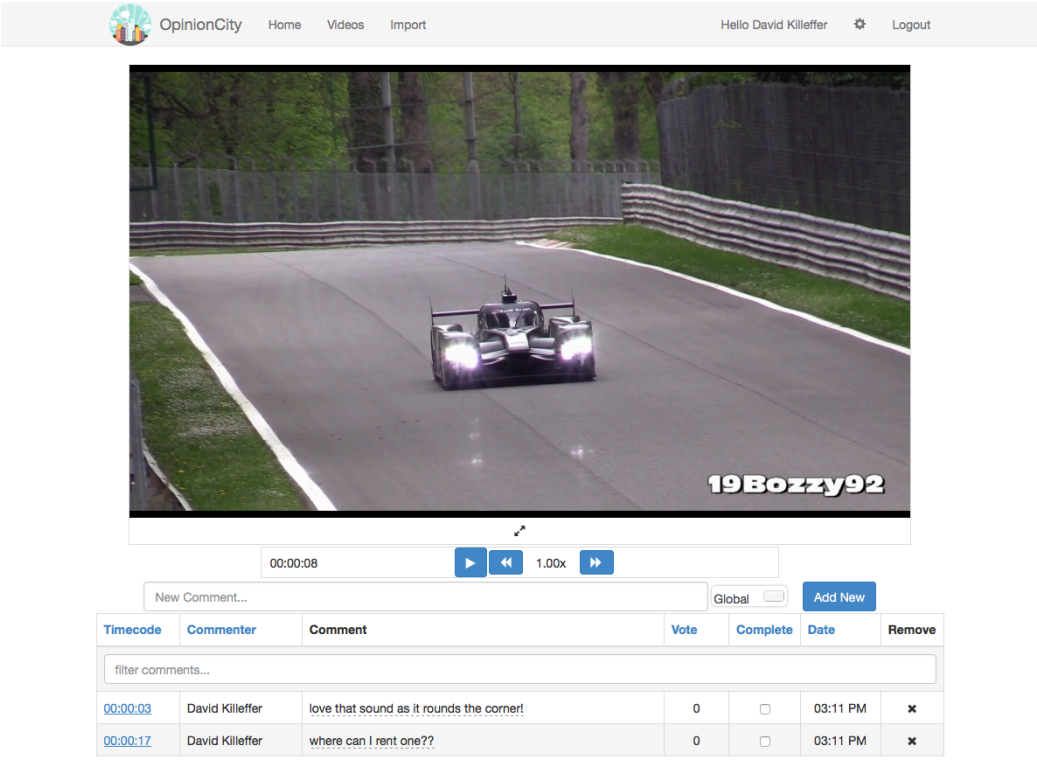


Fig. 2.3: *(OpinionCity)* another view of adding video annotations

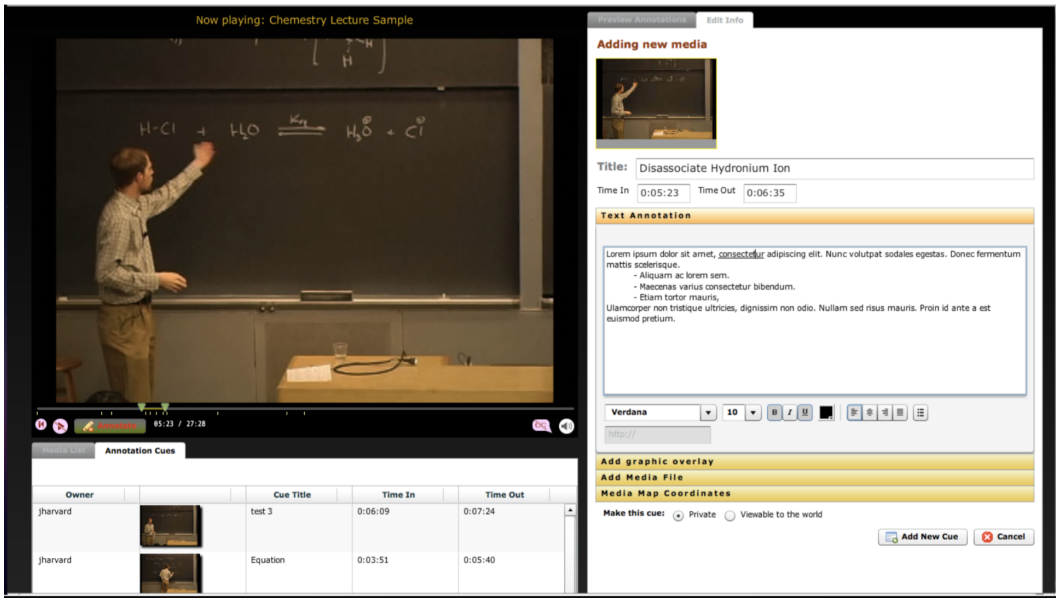


Fig. 2.4: *(MVAT)* video annotation edit view

2.3 Amazon X-Ray

by Amazon.com, Inc.

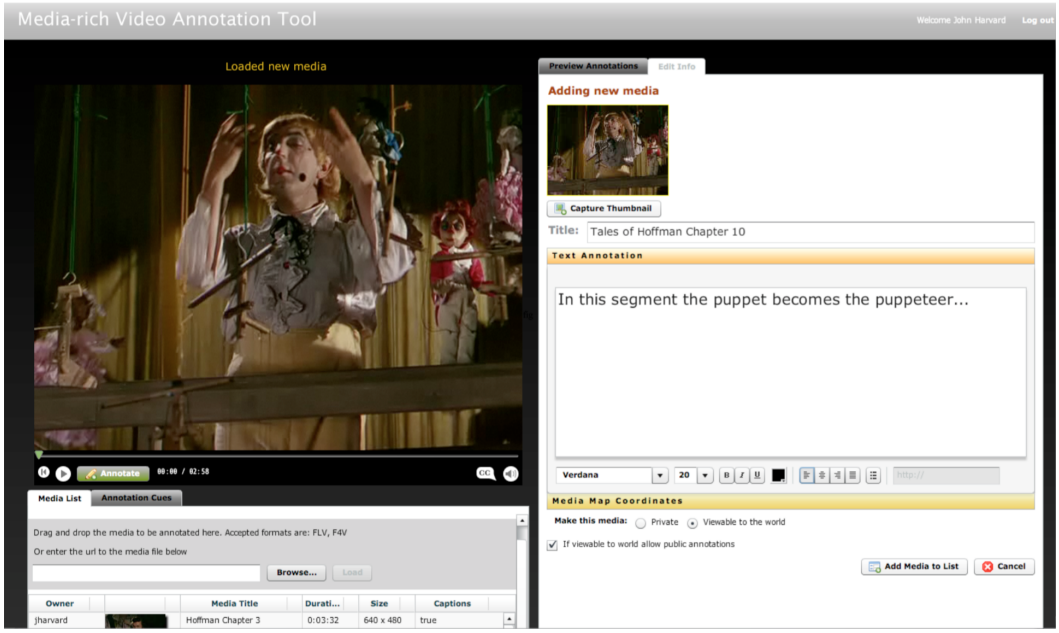


Fig. 2.5: *(MVAT)* adding video metadata view

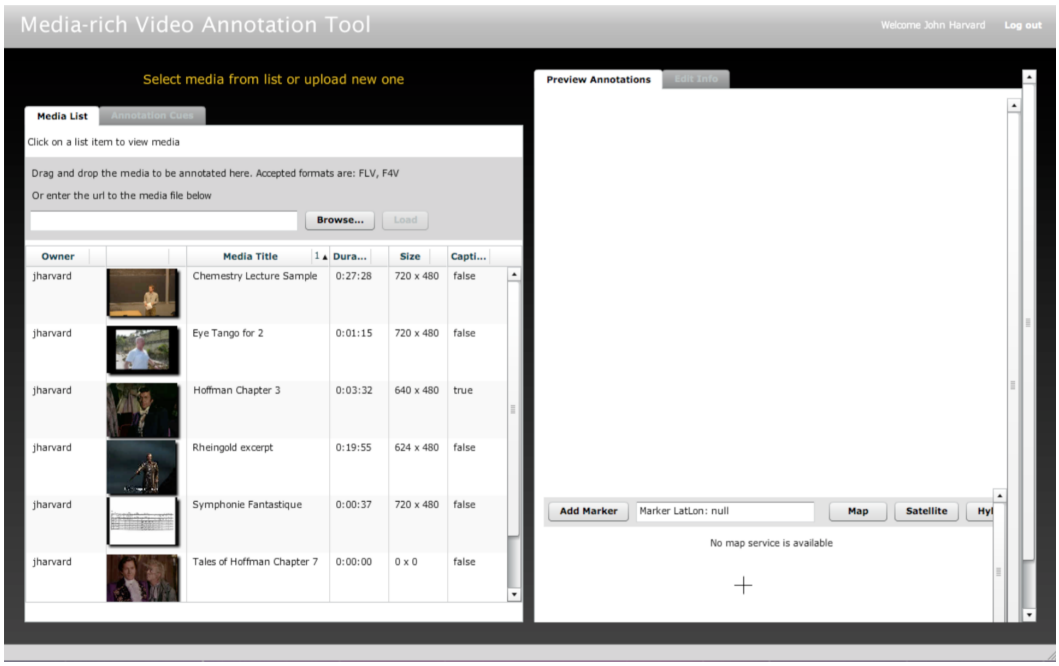


Fig. 2.6: *(MVAT)* video management view

Amazon has developed a technology that is now being used in several of their video players, from HTML5 enabled web browsers, to portable devices such as the Kindle Fire, and to the Amazon Fire TV sticks. X-Ray presents a video overlay on top of a video as it is playing (or on some devices, when a video is paused) and shows relevant metadata such as the actors that are currently appearing onscreen (and links to their IMDB webpages), the director(s), links to artists whose music is currently playing, etc. One of the motivating factors for the development of X-Ray

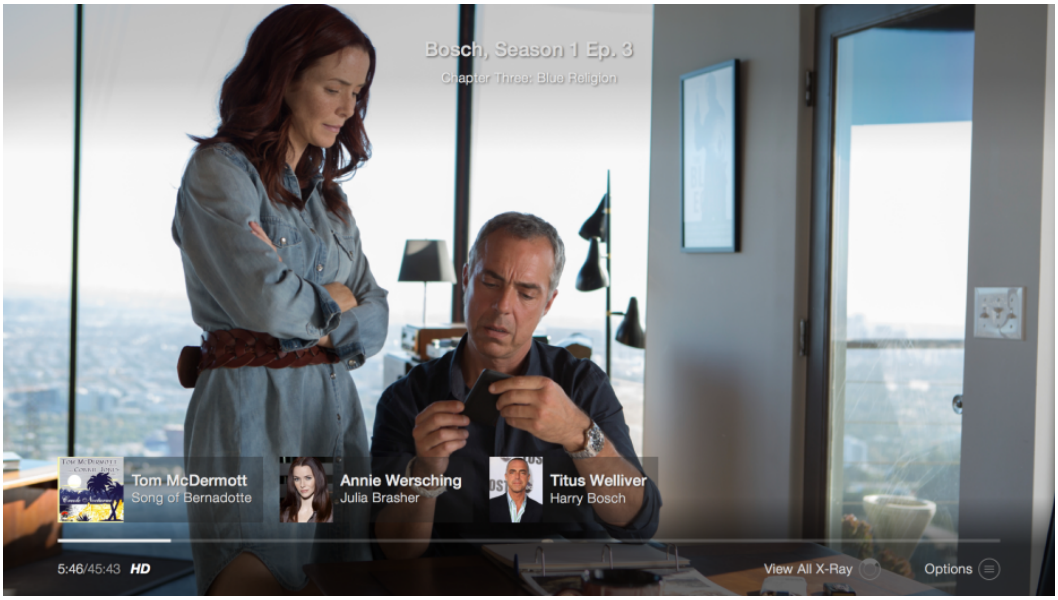


Fig. 2.7: *(Amazon X-Ray)* pausing a video displays relevant annotations and links to IMDB

for Amazon was to allow viewers to answer questions such as "*Who's that guy?*", "*What's she been in?*", *or* "*What is that song?*" (see http://www.businesswire.com/multimedia/home/20150413005383/en/).

The metadata that is used to power these real-time annotations comes from the Internet Movie Database, which is an Amazon owned property. Unfortunately there is not much in the way of technical details on the underlying technology or architecture of Amazon X-Ray, so it is difficult to find out how Amazon has created and implemented this technology, and how the metadata that powers the annotations is created (viewing an Amazon Prime hosted video in a web browser, for example, show you these annotations when you mouseover the player windown, and the annotations change as characters move in and out of screen, when a new song begins and ends, etc.).

For more information on Amazon X-Ray, see:

- http://www.imdb.com/x-ray/
- http://www.businesswire.com/multimedia/home/20150413005383/en/
- http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=2034369
- http://www.engadget.com/2012/09/06/amazon-announces-x-ray-for-movies-a-kindle-feature-that-uses-im/
- http://venturebeat.com/2015/04/13/amazons-x-ray-arrives-for-fire-tv-and-fire-tv-stick-bringing-context-to-instant-video-on-the-big-screen/
- http://www.wired.com/2015/04/amazon-xray-fire-tv/

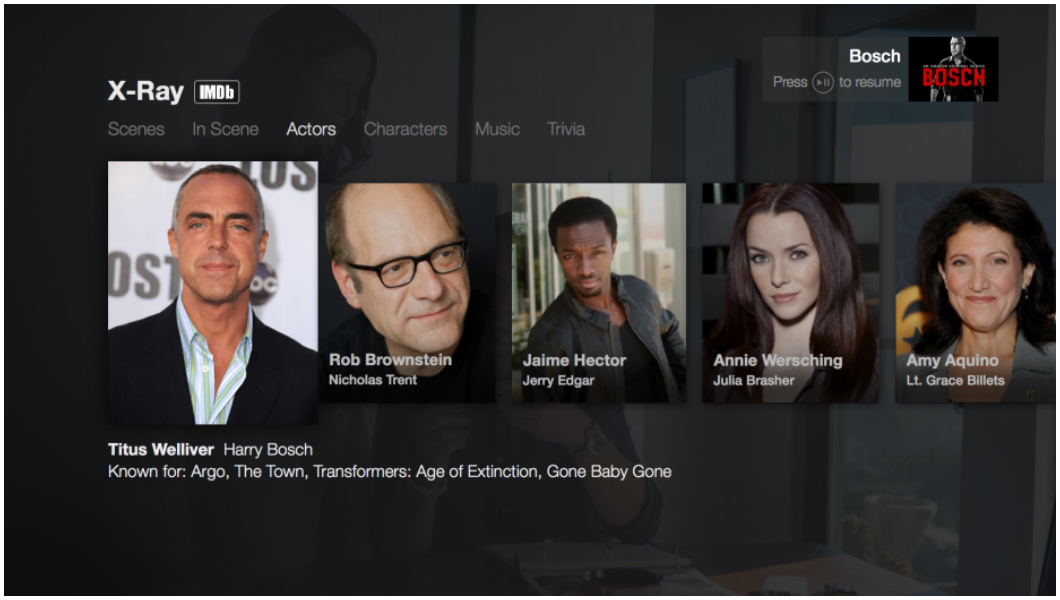


Fig. 2.8: *(Amazon X-Ray)* displaying an annotation of all main actors in a film, linking to individual actor profiles on IMDB

- http://gizmodo.com/5941067/amazons-x-ray-for-movies-knows-what-youre-watchingand-whos-in-it

2.4 Conclusion

There has been some very interesting and promising scholarly work done in the area of video annotations previously, but this thesis project has several unique aspects that it attempts to add and extend off of such prior work to make this project novel. Prior work does not appear to have focused too much on the search aspects of annotations that area created or the discoverability of videos that users might not have otherwise ever seen or known about were the annotations not present. Social sharing aspects of prior work in the area of video annoation seems to be a missing aspect of much prior work as well. A large part of the impetus for this project is the acknowledgement that the user in posession of a recording may themselves not know enough to properly annotate the video, but knows others (in this particular case, family members, but the same logic could easily apply to friends, colleagues, classmates, etc.) that would be able to add correct annotations.

3 Requirements

3.1 Overview

The Social Video platform prototype will allow for users to upload videos, define groups of individuals that are invited to annotate that video, create actual annotations on videos, enable administrators to "approve" annotations prior to those annotations being made public and searchable, and provide a rich, faceted search interface to find and locate videos of interest.

3.2 Details

Detailed requirements for the prototype are as follows:

- **User Authentication and Authorization**
  New users may create an account and request a particular level of access (**Regular User** or **Administrative User**). **Regular Users** are allowed to annotate videos, but **Administrative Users** can upload videos, create annotations, and "approve" annotations that **Regular Users** have added prior to those being made public and searchable.

- **User-Defined Groups**

- **Rich Annotations**

- **Video Upload**

- **Faceted Search**

- **Video Playlists**

3.3 Technologies Used

In contrast to some earlier work, a key goal of this prototype is for the tool to be a purely online system, allowing for web-based collaboration between users.

4 Design

4.1 Overview

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

4.2 Technologies Used

In contrast to some earlier work, a key goal of this prototype is for the tool to be a purely online system, allowing for web-based collaboration between users.

5 Work Plan

6 Risks and Alternatives

7 Preliminary Schedule

Month/Year	Description
July 2016	do some work
August 2016	do more work
September 2016	do more work
October 2016	do more work
November 2016	do more work
December 2016	do more work
January 2017	do more work
February 2017	do more work
March 2017	do more work
April 2017	do more work
May 2017	last bit of work

Tab. 7.1: Preliminary Project Schedule

List of Figures

List of Tables

Glossary

Here are some of the terms used throughout this proposal.

- **Amazon X-Ray**
  a reference tool incorporated into several video players that allows for the display and linking of actors, acresses, directors, and linked data when viewing a movie or television show hosted by Amazon. (see https://en.wikipedia.org/wiki/X-Ray_%28Amazon_Kindle%29, http://www.amazon.com/gp/help/customer/display.html?nodeId=201423010)

- **DVD**
  Digital Video Disc

- **Elasticsearch**
  TODO

- **faceted search**
  TODO

- **Node.js**
  TODO

- **Mini-DV**
  a successor to the wildly popular VHS home recording system, Mini-DV is a cassette tape format for video recording. It captures video in a resolution close to that of DVD. (see http://techterms.com/definition/minidv)

- **VHS**
  acronnym for "Video Home System", it is a widely-adopted videocassette recording (VCR) technology that was developed by Japan Victor Company (JVC) and put on the market in 1976. It uses magnetic tape 1/2 inch (1.27 cm) in width. It was extremely popular in the 1980s and 1990s, and declined in use during the early 2000s. (adapted from http://whatis.techtarget.com/definition/VHS-Video-Home-System and https://en.wikipedia.org/wiki/VHS)

Colophon

This thesis was typeset with LaTeX 2$_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at http://cleanthesis.der-ric.de/.

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

Cambridge, Massachusetts, May 21, 2016

David Killeffer