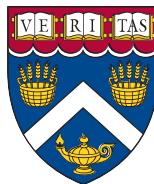


Social Video: A Platform for Collaborative Discoverability and Annotation of Disconnected Media

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 proliferation of smartphones, there is an ever-growing volume of videos being made available to people, but a severe lack of ways to categorize 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 individual's 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 acknowledge 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 acknowledge 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
1.1	Motivation	1
1.2	Personal Story	2
1.3	Problem Statement	4
2	Prior Work	5
2.1	OpinionCity	5
2.2	Media-rich Video Annotation Tool (MVAT)	7
2.3	Amazon X-Ray	9
2.4	Conclusion	11
3	Requirements	12
3.1	Overview	12
3.2	Details	12
4	Design	15
4.1	Overview	15
4.2	Technologies Used	15
5	Work Plan	17
6	Risks and Alternatives	18
7	Preliminary Schedule	19
	Bibliography	21

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* [[@You15](#); [@Tub14](#); [@DMR16](#)]. 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" [[@LR95](#); [@Eme16](#)], 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.

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 [OpinionCity](#), [Media-rich Video Annotation Tool \(MVAT\)](#), and [Amazon X-Ray](#), which are discussed below:

2.1 OpinionCity

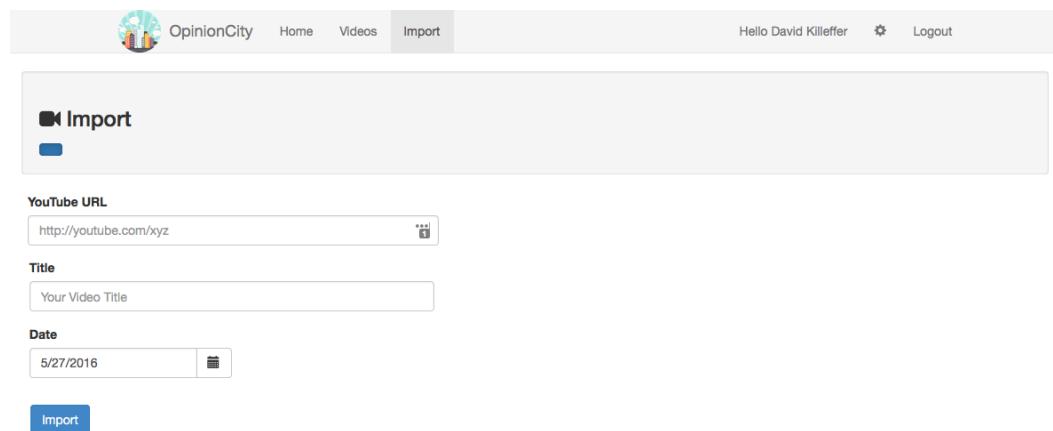


Fig. 2.1: (OpinionCity) importing a video from YouTube into OpinionCity

by Daniel P. Coffey: danielpcoffey@gmail.com, Spring 2015 [[@Cof15](#)]

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

Your Videos

Date	Title	Invite Collaborators	Remove
05/27/2016	2014 Audi R18 e-tron: Audi's Infamous Diesel-Hybrid Tested! - Ignition Ep. 125	email@gmail.com, otherperson@alta <button>Invite</button>	X
05/27/2016	2016 Audi R18 LMP1 Spied Testing At Monza Circuit!!	email@gmail.com, otherperson@alta <button>Invite</button>	X
05/27/2016	2015 Audi R18 E-Tron Quattro Le Mans Aero Kit - High Speed Fly Bys	email@gmail.com, otherperson@alta <button>Invite</button>	X

Collaborations

Date	Title	Remove
You haven't yet been invited to collaborate!		

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

The screenshot shows a video player displaying a race car on a track. The video has a timestamp of 00:00:08 and a playback speed of 1.00x. Below the video is a comment input field labeled "New Comment..." and a "Global" checkbox. A "Add New" button is also present.

Timecode	Commenter	Comment	Vote	Complete	Date	Remove
00:00:03	David Killeffer	love that sound as it rounds the corner!	0	<input type="checkbox"/>	03:11 PM	X
00:00:17	David Killeffer	where can I rent one??	0	<input type="checkbox"/>	03:11 PM	X

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

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 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)

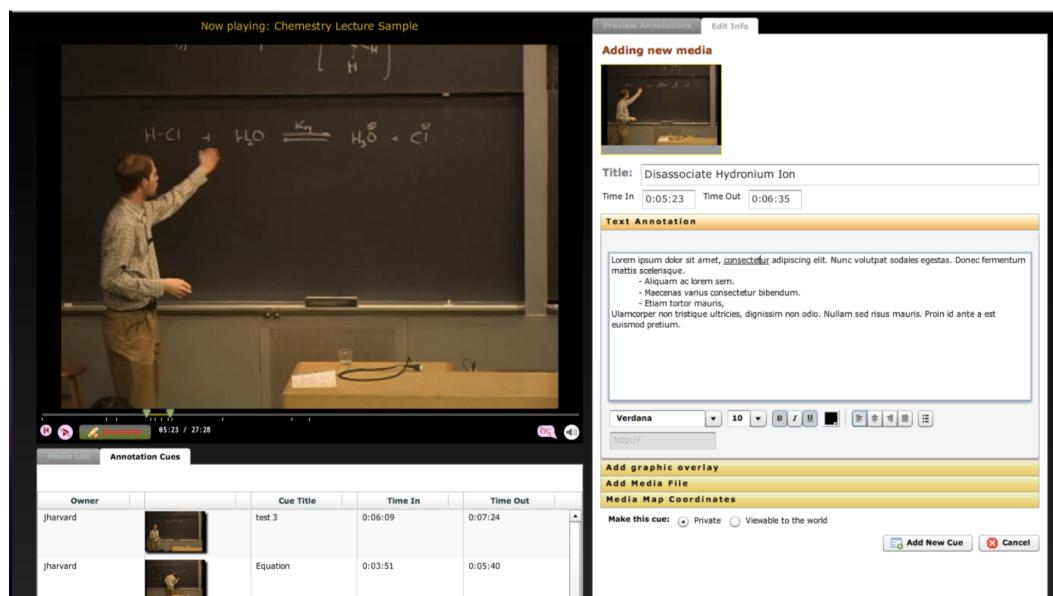


Fig. 2.4: (MVAT) video annotation edit view

by Philip Desenne: desenne@fas.harvard.edu, May 2012 [Des12]

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

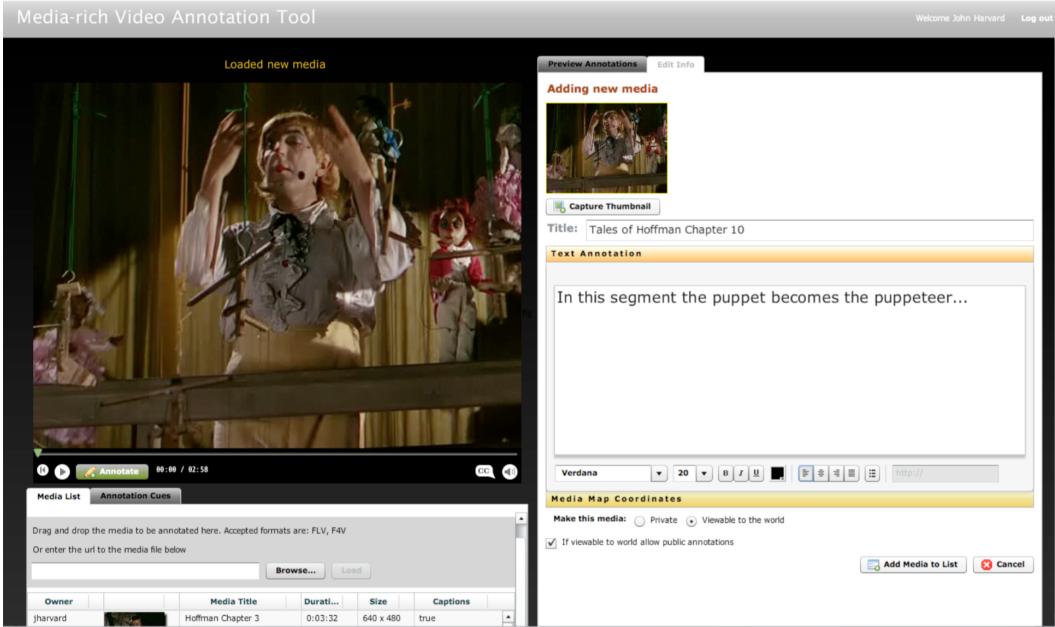


Fig. 2.5: (MVAT) adding video metadata view

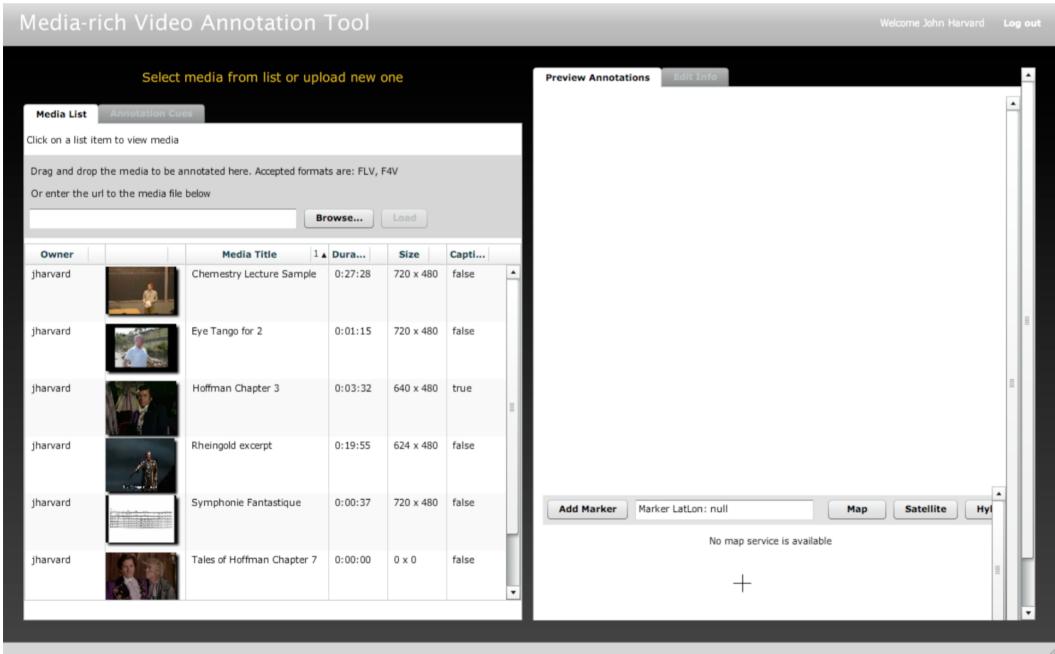


Fig. 2.6: (MVAT) video management view

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 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.

2.3 Amazon X-Ray

by Amazon.com, Inc.

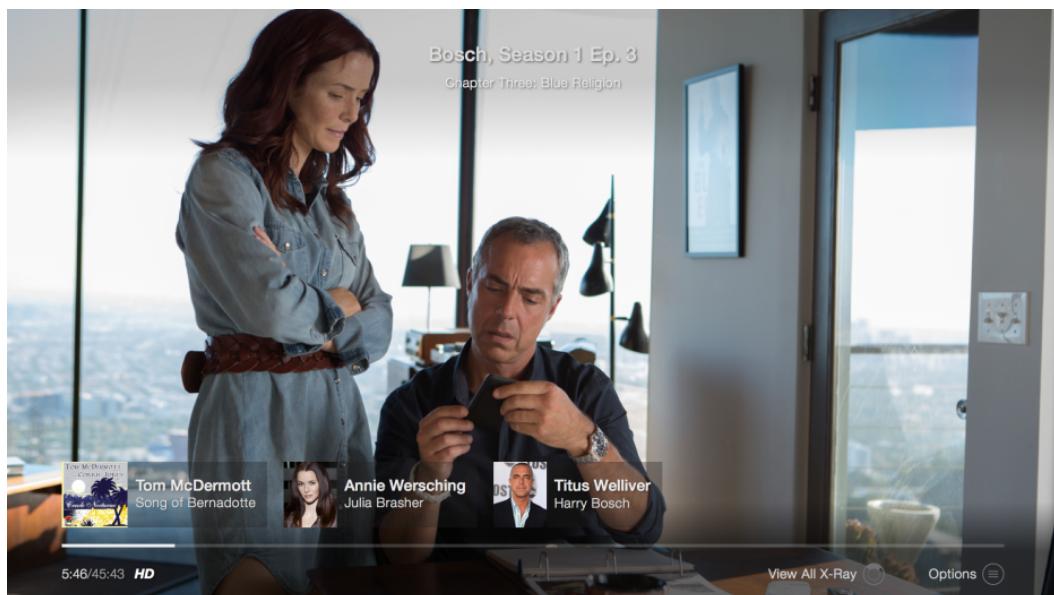


Fig. 2.7: (Amazon X-Ray) pausing a video displays relevant annotations and links to IMDB

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

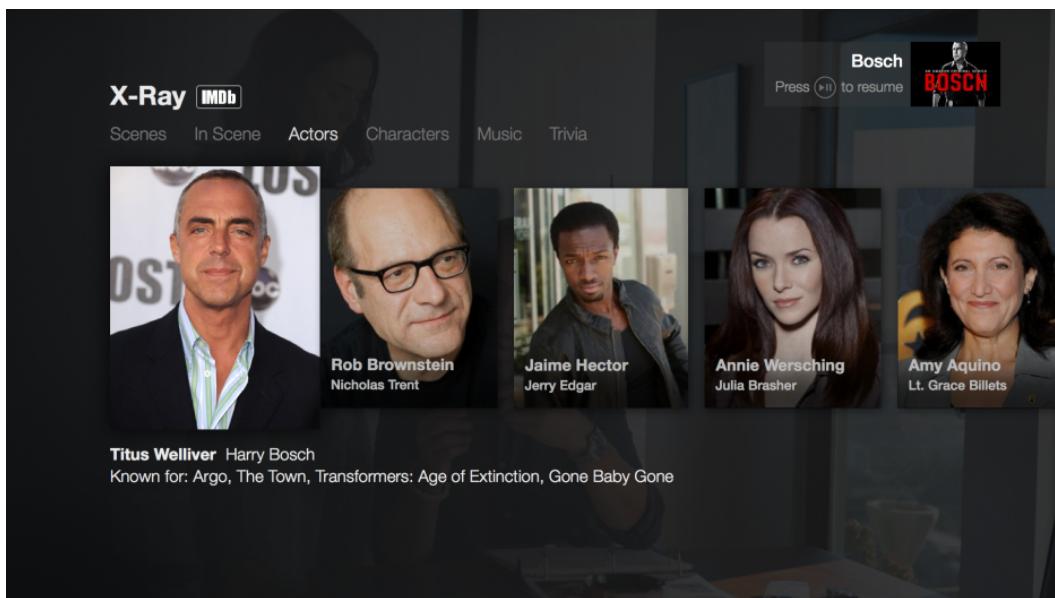


Fig. 2.8: (*Amazon X-Ray*) displaying an annotation of all main actors in a film, linking to individual actor profiles on [IMDB](#)

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 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 window, and the annotations change as characters move in and out of screen, when a new song begins and ends, etc.). While the user-facing interface to Amazon X-Ray is fairly polished and intuitive for viewers, there is no mechanism to add or edit video annotations for the public or viewers; the canonical nature of the original source recordings (being that they generally come from scripted performances, and recordings which have the benefit of having machine-parseable closed-captioning) obviates the need for such user-submitted annotations.

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

[@Wir15a; @Dat16; @Ama15; @Eng12; @Bea15; @Wir15b; @Giz12]

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.

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. *A primary motivating factor for this project is that the individual that uploads a video may not know all of the possible metadata that should be associated with that video, and so relies on the recollection and annotations of others to enrich the media collection.*

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**

Individual users may "invite" other users of the system (or invite people not yet registered on the website) to join a custom group. When a user invites another to join their group, the videos that the inviter has uploaded become viewable to the other members of the group, and the invitees may add metadata to the group. Currently I am weighing the pros and cons of having annotations that other users add to a video be moderated by the original uploader/owner of the video; part of the difficulty in determining whether or not to enforce moderation is that it seems a premature optimization, and solving a *potential* problem, but not necessarily a known one. One advantage of enforcing moderation of user-added annotations is that it would present an opportunity for the original uploader/owner of a video to

normalize the annotations and ensure correct and standardized spelling of names, places, dates, etc.

- **Video Upload**

Registered users will be able to upload their own videos to the website which they can then add annotations to, or invite other users to annotate. Common video formats such as .MP4, .MOV, .AVI will be supported, and other additional formats may be added as well if time permits.

- **Rich Annotations**

This is one of the core features of the project; users will be able to add annotations on a whole video basis (e.g., tagging a video as a particular family's recordings, for example), or more commonly, on segments of a video. [OpinionCity](#) allows for users to add a comment to an entire video, or to insert a comment at a specific time in the video, but this would not allow for knowing exactly from and to a particular annotation is valid (e.g., annotating when a person is first shown in frame and when they leave the frame); [Amazon X-Ray](#) works this way, but does not allow viewers to add or edit such annotations (and the original source for such annotations could potentially be machine generated from a timecoded script or algorithmic analysis of the video and not from human-generated annotations). The [Media-rich Video Annotation Tool \(MVAT\)](#) does support adding annotations that have a beginning and ending within a video (and also offers other types of useful annotation types for pedagogical purposes), but this prototype will focus purely on textual annotations which are used for categorization and tagging. The user interface will focus on giving users the ability to add fairly simple annotations of the following types, which will form the basis for faceted search capability:

- *People*: annotate a video as having discrete people present in the entire video, or in time-based segments of the video (e.g., "*Bob is in the video from 3 minutes, 24 seconds until 7 minutes, 3 seconds*")
- *Places*: annotate an entire video as taking place in a particular location (e.g., "*Paris, France*"), or mark a segments of the video as being in a particular place (similarly to how they can mark *people* as being in a segment of the video)
- *Dates*: annotate a video as taking place in a particular year, month, or day, as well as marking segments as taking place on particular dates
- "*Tags*": a more free-form type of annotation, *tags* will allow users to mark entire videos or segments with user-defined characteristics such as *funny*, *sad*, *scary*, *barbecue*, and so forth. Some experimentation may be required to see if a more limited set of potential tags is more desireable than allowing for free-form user entry of *tags*; this will be determined during user testing of the prototype.

- **Faceted Search**

The heart of the project will be the faceted search feature; this will enable users to quickly and easily find all the videos that a particular person is in, all the videos recorded at a particular place, in a particular year/month, etc. User supplied annotations will form the basis for the facets. Depending upon the breadth/depth of user-generated annotations, some curation of the facets may be necessary.

- **Video Playlists**

If time permits, the system should allow for users to create video "playlists" of their favorite moments/segments of videos; the interface for adding videos should be very simple (perhaps a drag and drop interface) to build up a playlist from a search result. Users can mark a playlist as private (only viewable to themselves), or public, which could be shared with others, enhancing the utility of the website and encouraging further user participation. A stretch goal would be to allow for not just discrete video playlists (e.g., arbitrarily adding entire videos or segments of videos to a playlist), but creating a playlist from search terms or facets (e.g., construct a playlist called "funny" combined with "vacation", which would have all of the video segments tagged with the keywords "funny" and "vacation"). This would bring new levels of engagement to the site by incentivizing user participation because they would be able to quickly see how their annotation activities influenced already established public video playlists.

Design

4.1 Overview

The prototype will take the form of a website, and will be created using [Node.js](#), [Express.js](#), HTML, CSS, and JavaScript for the backend. The frontend will make use of [Bootstrap](#) for styling and responsive layout, as well as [React.js](#) for dynamic user interface management, particularly for managing video annotations. Additional libraries may be used to aid with the video player display and management. My selection of an almost exclusively JavaScript based technology stack is based on my desire to gain further experience with JavaScript as a server-side technology, as well as to minimize the "context-switching" costs for myself that I would otherwise have to undergo if using a more mixed-mode technology stack (for more background on context switching and the costs involved, see: <http://www.newvoicemedia.com/blog/agile-development-context-switching-comes-at-the-price-of-delivery/>; <https://psygrammer.com/2011/03/28/unnecessary-context-switches-the-myth-of-multitasking/>; <http://www.petrikainulainen.net/software-development/processes/the-cost-of-context-switching/>; <http://www.joelonsoftware.com/articles/fog0000000022.html>; <http://www.bryanbraun.com/2012/06/25/multitasking-and-context-switching/>; <http://blog.ninlabs.com/2013/01/programmer-interrupted/>).

[@Cow12; @Psy11; @Kai14; @Spo01; @Bra12; @Res13]

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. Another core requirement of the system is that a faceted search be supported to allow for easy discovery of videos of interest to users. To support the faceted searching of videos, [Elasticsearch](#) will be used as a document repository of all video annotation data. When users add annotations to a video, that metadata will be added to Elasticsearch and made available for searches on videos, as well as be used to display top categories of videos.

Annotations will be stored in a document-based repository, [Elasticsearch](#). Document based repositories such as [Apache Solr](#) and [Elasticsearch](#) are commonly used to

power search engines because of their capabilities such as fast search indexing, flexible document schema support, and rich support of a wide variety of client libraries. Interactions with the Elasticsearch document repository are executed via a [RESTful API](#) and [JSON](#) messages are exchanged to read/write documents to the repository.

Work Plan

My plan for the development of this project will be to follow a loosely based "Scrum" development methodology, where I attempt to adhere to 2-week "sprints" wherein I successfully complete discrete functional parts of the project, and break out larger functionality into small chunks that can be built and completed in the shorter sprint timeframe. Before investing in much of the actual coding portion of the project, I will be developing wireframes that visually indicate the proposed "look and feel" of a particular webpage.

6

Risks and Alternatives

I will primarily be using Apple MacBook Pro laptops for the development of this project, currently running OS X 10.11.x El Capitan. With respect to tooling, I will be using WebStorm as my IDE of choice because it offers a rich development environment for JavaScript projects and I am already very familiar with it. Source code will be kept up to date in a personal GitHub repository so that I can synchronize my work between the two primary laptops I will be using, as well as to keep myself accountable for continual progression (by checking when code is committed and milestones are hit). For hosting, I currently have a web hosting account through A Small Orange which I will be evaluating to see if that is sufficient for running the website on; due to the application I am planning to build being a Node.js application which also requires an Elasticsearch document repository for the annotations, I may need to search out other hosting options. During much of development though, I can run the entire application locally on my laptop.

In terms of risks, there is a risk to the project related to video hosting; the large filesize of many of the videos I wish to upload could become very expensive to pay for hosting for, and so free/alternative hosting solutions should be investigated. As the prototype is intended primarily to be a "proof-of-concept" and not necessarily a commercially viable product for the market the potential problem of where to host the videos is not a grave risk, but is an important consideration. Currently I intent to research and assess the suitability of a few different video hosting platforms that could be leveraged for the project that would involve minimal or no cost, such as YouTube. There may also be a scope risk in that adding some of the nicer features of the Requirements, such as giving users the ability to create video playlists. As I embark on the development phase of the project I will need to re-assess the viability of completing some of more advanced features in the context of the overall goal of making a rich video annotation prototype.

Preliminary Schedule

Month/Year	Description
July 2016	<ul style="list-style-type: none"> Continue to develop thesis proposal and get approval to proceed System design and architecture diagramming Research on Node.js, Express.js, and Elasticsearch Initial wireframing exercises of the front-end user interface
August 2016	<ul style="list-style-type: none"> Summer vacation
September 2016	<ul style="list-style-type: none"> Tooling setup; source code repository, establish workflow for development Determine server setup and configuration; install Node.js, Elasticsearch Develop basic Node.js / Express.js application allowing for user account creation, sign-in, users creating groups
October 2016	<ul style="list-style-type: none"> Develop initial version of video upload utility Develop administrative toolkit where administrators may invite new users to website and approve video annotations submitted by users (annotation mechanism not yet built) Develop initial basic video annotation mechanism on videos (people, places, dates, tags)
November 2016	<ul style="list-style-type: none"> Further refinement of video annotations mechanism (people, places, dates, tags) Develop faceted search based on user-submitted annotations (Elasticsearch repository)
December 2016	<ul style="list-style-type: none"> User testing of platform, solicit user feedback Develop plan for writing thesis in conjunction with Susan Buck Time permitting, create functionality allowing users to create video "playlists"
January 2017	<ul style="list-style-type: none"> Final bugfixes, code cleanup, review codebase and improve documentation Continue thesis draft, work with Susan Buck on document revisions

Month/Year	Description
February 2017	<ul style="list-style-type: none"> Continued user testing, feature tweaks, bug fixes Update documentation on how project has changed/deviated from original plan since inception Begin writing final thesis draft for approval
March 2017	<ul style="list-style-type: none"> Final bugfixes, code cleanup, review codebase and improve documentation Code freeze Continue thesis draft, work with Susan Buck on document revisions
April 2017	<ul style="list-style-type: none"> Final thesis document approval and review with Susan Buck Receive overall grade for thesis project Prepare presentation for HES IT Symposium
May 2017	<ul style="list-style-type: none"> Present thesis at HES IT Symposium HES Graduation

Tab. 7.1: Preliminary Project Schedule

Bibliography

- [Des12] Desenne, Philip. „Prototype for a Media-rich Video Annotation Tool“. Harvard University Extension School, 2012 (cit. on p. 7).

Websites

- [@Ama15] Inc. Amazon.com. *Press Release: For the First Time Ever, X-Ray for Movies and TV Shows Now Available Directly on Your TV - Answer the Classic Movie-Watching Question "Who's That Guy?" with Your Amazon Fire TV.* 2015. URL: <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=2034369> (visited on Apr. 13, 2015) (cit. on p. 11).
- [@Bea15] Venture Beat. *Amazon launches X-Ray on Fire TV to bring IMDb movie and TV data to the big screen.* Ed. by Paul Sawers. 2015. URL: <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/> (visited on Apr. 13, 2015) (cit. on p. 11).
- [@Bra12] Bryan Braun. *Multitasking and Context Switching.* 2012. URL: <http://www.bryanbraun.com/2012/06/25/multitasking-and-context-switching> (visited on June 25, 2012) (cit. on p. 15).
- [@Cof15] Daniel Coffey. *OpinionCity.* 2015. URL: <http://opinion.city> (visited on June 10, 2016) (cit. on p. 5).
- [@Cow12] Charlie Cowan. *Agile Development: Context switching comes at the price of delivery.* 2012. URL: <http://www.newvoicemedia.com/blog/agile-development-context-switching-comes-at-the-price-of-delivery/> (visited on June 12, 2012) (cit. on p. 15).
- [@Dat16] Internet Movie Database. *X-Ray for Movies & TV by IMDb.* 2016. URL: <http://www.businesswire.com/multimedia/home/20150413005383/en/> (visited on July 26, 2016) (cit. on p. 11).
- [@DMR16] DMR. *By The Numbers: 125+ Amazing YouTube Statistics.* Ed. by Craig Smith. 2016. URL: <http://expandedramblings.com/index.php/youtube-statistics/> (visited on Feb. 25, 2015) (cit. on p. 1).
- [@Eme16] LLC Emerald Art Services. *Visual Media Deterioration.* 2016. URL: <http://www.emeraldartservices.com/visual-media-deterioration/> (visited on Feb. 25, 2015) (cit. on p. 2).

- [@Eng12] Engadget. *Amazon announces X-Ray for Movies, a Kindle feature that uses IMDB to name the actors for you*. Ed. by Dana Wollman. 2012. URL: <https://www.engadget.com/2012/09/06/amazon-announces-x-ray-for-movies-a-kindle-feature-that-uses-im/> (visited on Sept. 6, 2012) (cit. on p. 11).
- [@Giz12] Gizmodo. *Amazon's X-Ray for Movies Knows What You're Watching—And Who's In It*. Ed. by Brian Barrett. 2012. URL: <http://gizmodo.com/5941067/amazons-x-ray-for-movies-knows-what-youre-watchingand-whos-in-it> (visited on Sept. 6, 2012) (cit. on p. 11).
- [@Kai14] Petri Kainulainen. *The Cost of Context Switching*. 2014. URL: <http://www.petrikainulainen.net/software-development/processes/the-cost-of-context-switching/> (visited on Apr. 6, 2014) (cit. on p. 15).
- [@LR95] Council on Library and Information Resources. *What Can Go Wrong With Magnetic Media?* 1995. URL: http://www.clir.org/pubs/reports/pub54/2what_wrong.html (visited on Feb. 25, 2015) (cit. on p. 2).
- [@Psy11] Psygrammer. *Unnecessary Context Switches & the Myth of Multitasking*. 2011. URL: <https://psygrammer.com/2011/03/28/unnecessary-context-switches-the-myth-of-multitasking/> (visited on Mar. 28, 2011) (cit. on p. 15).
- [@Res13] Ninlabs Research. *Programmer Interrupted*. 2013. URL: <http://blog.ninlabs.com/2013/01/programmer-interrupted/> (visited on Jan. 19, 2013) (cit. on p. 15).
- [@Spo01] Joel Spolsky. *Human Task Switches Considered Harmful*. 2001. URL: <http://www.joelonsoftware.com/articles/fog000000022.html> (visited on Feb. 12, 2001) (cit. on p. 15).
- [@Tub14] Tubefilter, Inc. *YouTube Now Sees 300 Hours of Video Uploaded Every Minute*. 2014. URL: <http://www.tubefilter.com/2014/12/01/youtube-300-hours-video-per-minute/> (visited on Feb. 25, 2015) (cit. on p. 1).
- [@Wir15a] Business Wire. *For the First Time Ever, X-Ray for Movies and TV Shows Now Available Directly on Your TV — Answer the Classic Movie-Watching Question "Who's That Guy?" with Your Amazon Fire TV*. 2015. URL: <http://www.businesswire.com/multimedia/home/20150413005383/en/> (visited on Apr. 13, 2015) (cit. on p. 11).
- [@Wir15b] Wired. *Amazon's X-Ray Shows Movie Info Whenever You Hit Pause*. Ed. by David Pierce. 2015. URL: <http://www.wired.com/2015/04/amazon-xray-fire-tv/> (visited on Apr. 13, 2015) (cit. on p. 11).
- [@You15] YouTube, LLC. *YouTube.com Statistics*. 2015. URL: <https://web.archive.org/web/20150520171813/https://www.youtube.com/yt/press/statistics.html> (visited on May 22, 2015) (cit. on p. 1).

List of Figures

2.1	(<i>OpinionCity</i>) importing a video from YouTube into OpinionCity	5
2.2	(<i>OpinionCity</i>) listing of all videos associated with a user account	6
2.3	(<i>OpinionCity</i>) another view of adding video annotations	6
2.4	(<i>MVAT</i>) video annotation edit view	7
2.5	(<i>MVAT</i>) adding video metadata view	8
2.6	(<i>MVAT</i>) video management view	8
2.7	(<i>Amazon X-Ray</i>) pausing a video displays relevant annotations and links to IMDB	9
2.8	(<i>Amazon X-Ray</i>) displaying an annotation of all main actors in a film, linking to individual actor profiles on IMDB	10

List of Tables

7.1 Preliminary Project Schedule	20
---	----

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, actresses, 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>)

- **Apache Solr**

a popular, open-source fast document-oriented search engine similar to ElasticSearch, which is likewise built on top of the Apache Lucene engine (see <http://lucene.apache.org/solr/>)

- **Bootstrap**

A popular open-source framework used to speed up development of websites by supporting HTML, CSS, and JavaScript and supplying a responsive, mobile-ready layout mechanism. (see <http://getbootstrap.com/>).

- **DVD**

abbreviation of "Digital Video Disc", these are digital discs similar in size to CDs but can store much more information and were first used primarily for video content, but later gained popularity for storing data files as well. DVDs were first sold in 1997 and soon became wildly popular, largely supplanting VHS as the primary delivery mechanism for movies.

- **Elasticsearch**

a fast document-oriented search engine also based upon the open-source information retrieval library Lucene (similar to Apache Solr). ElasticSearch achieves high speed data retrieval due to searching indexed data fields rather than searching text directly, and is a competing product to Apache Solr. (see <https://en.wikipedia.org/wiki/Elasticsearch>. <http://www.elasticsearchtutorial.com/basic-elasticsearch-concepts.html>)

- **Express.js**

a minimal and flexible framework for building web applications and APIs under Node.js. (see <https://expressjs.com/>)

- **faceted search**

a search technique that categorizes items into "buckets" (e.g., facets) that allow for drill-down searches into distinct and more specific categories. An example of a faceted search would be the various distinct and increasingly specific product categories/specifications for a product item search on Amazon; for example, when searching for men's shoes, facets can be broken down by shoe size, color, style, manufacturer, purpose, etc.

- **JSON**

a human-readable, text-based format for documents that is also machine-parseable. JSON documents are structured as *key:value* pairs, where the *value* can be an array, a single text string, or another JSON document entirely.

- **Node.js**

a JavaScript runtime built on [Chrome's V8 JavaScript engine](#), Node.js allows you to run JavaScript on the server-side, and build a wide variety of applications using JavaScript as the primary language. (see <https://nodejs.org/en/>)

- **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>)

- **React.js**

an open-source JavaScript library for building user interfaces, originally developed and used by [Facebook](#). (see <https://facebook.github.io/react/>)

- **RESTful API**

"an application program interface that uses HTTP requests to GET, PUT, POST, and DELETE data" (see <http://searchcloudstorage.techtarget.com/definition/RESTful-API>)

- **VHS**

acronym 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 $\text{\LaTeX} 2_{\varepsilon}$. 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://cleantheesis.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