

# Semi-Automated Rediscovery of Lost YouTube Music Videos

Daniel Sebastian  
Harding University  
Computer Science Dept  
Searcy, Arkansas, USA 72149  
dsebasti@harding.edu

Frank McCown  
Harding University  
Computer Science Dept  
Searcy, Arkansas, USA 72149  
fmccown@harding.edu

Michael L. Nelson  
Old Dominion University  
Computer Science Dept  
Norfolk, Virginia, USA 23529  
mln@cs.odu.edu

## ABSTRACT

Users frequently post popular material to YouTube, and in response, others link to these videos from social media, blogs, forums, and email. However, this content may be removed for numerous reasons, only to resurface again at another URL. This continuous movement and breaking of the web graph makes it difficult for users to relocate content that has moved in YouTube. We present Volitrax, an add-on for FireFox which redirects users to YouTube music videos that have moved to a different URL within YouTube. Volitrax acts as an intermediary that corrects the web graph transparently so YouTube links continue to work even after the content has changed locations.

## Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries – Systems Issues.

## General Terms

Design, Experimentation.

## Keywords

Music video, web graph, linkrot, web browser add-on.

## 1. INTRODUCTION

YouTube is one of the world's largest and most popular video digital libraries, and it is home to millions of music videos. Many of the videos posted to YouTube garner a great number of links by music fans in social media, blogs, forums, and email. However, many of the music videos are subsequently inaccessible from YouTube for a number of reasons: users delete videos or close their accounts, or copyright owners petition for the removal of copyrighted work. Copies of the removed music videos can often be found at other URLs or resurface later as music lovers and copyright owners engage in a game of cat-and-mouse. This continuous movement and breaking of the web graph requires users to manually relocate content that has moved in YouTube.

In order to extend the life of links to YouTube music videos, we have created Volitrax, a FireFox add-on that redirects users to YouTube music videos that have moved locations within YouTube. (Volitrax is a combination of “volatile” and “tracks”,

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

JCDL'13, July 22–26, 2013, Indianapolis, Indiana, USA.  
ACM 978-1-4503-2077-1/13/07.

as in music tracks.) Volitrax acts as an intermediary that corrects the web graph transparently so YouTube links continue to work even after the content has changed locations.

## 2. BACKGROUND & RELATED WORK

YouTube has been the source of study in a number of contexts including characterization [2], video sharing [1], video discovery [3], and preservation [5][8]. In previous work [7], we have shown that the half-life for a music video in YouTube ranges from 9 to 18 months, depending on the publication date of the video, and there are frequently hundreds if not thousands of copies of the same music video within YouTube at any one time. This work is an extension of [7] where we make use of the fact that multiple copies of music videos are available at other URLs, and we have built a tool to aid in the discovery of these new URLs. This work fits into a larger framework of rediscovering lost web resources [4] and correcting broken links in the web graph [6].

## 3. FINDMOBILE TOOL

### 3.1 Overview

Volitrax works in a client-server environment where clients (those with the Firefox add-on installed) make use of central server which stores music metadata locally and pushes the data out to the Web Infrastructure where it may outlive the lifetime of the service. Figure 1 shows the Volitrax server interfacing with the Volitrax add-ons and various services like Blogger, tumblr, and delicious. These services all offer APIs which make pushing and pulling information straightforward. Data is supplied to Volitrax from volunteer user input and from the last.fm API.

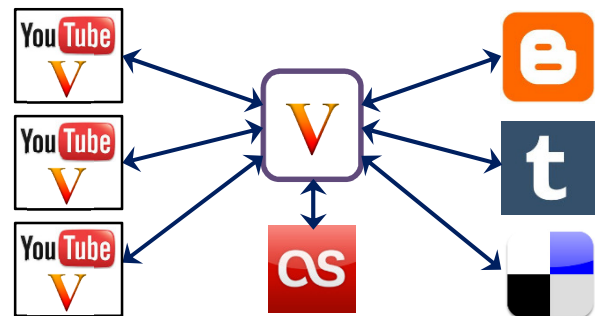


Figure 1. Volitrax clients interact with server which stores long-term data with Blogger, tumblr, & delicious

The Volitrax add-on examines the HTTP traffic when the user is visiting URLs on youtube.com. The add-on uses the YouTube API to obtain information about the video being viewed, and when it suspects the user is viewing a music video (by examining

the categorization and video title), it will use the last.fm service to find additional metadata about the song and perform spell corrections. The user may also be asked to verify the metadata (see Figure 2) although this option can be turned off easily. Metadata about the music video is stored on the server and is pushed to long-term data storage (the Web Infrastructure).

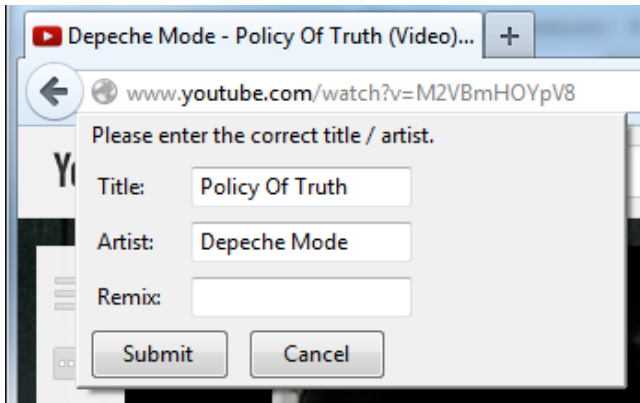


Figure 2. Volitrax Asking the user for metadata about the music video

When the user accesses a YouTube music video that has been removed (as shown in Figure 3), the YouTube API will return a 40x http status code. When this happens, Volitrax will look for an alternate location in its database of music videos so it can automatically redirect the user to another copy. If this is the first time the URL has been seen by Volitrax, it will prompt the user for information about the missing video. The user is able to provide the title and artist if they know it and a new URL if they can discover a copy themselves.

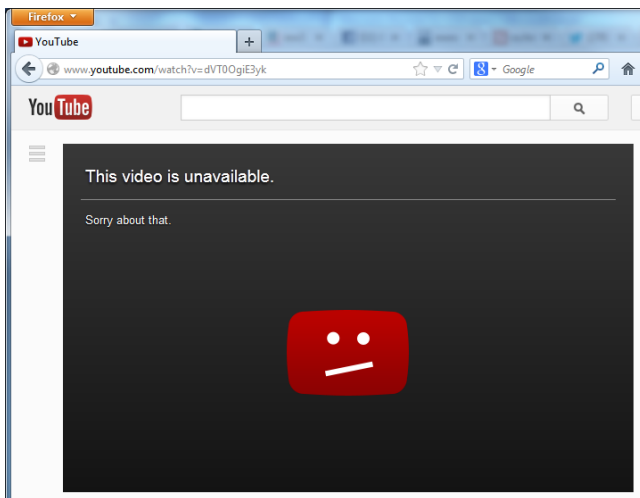


Figure 3. Music video has been removed from YouTube with no hint where to find a replacement

### 3.2 Voting on User Input

Volitrax makes use of information supplied by users to improve the redirection experience for all users. Any system that relies on the collective anonymous input of the public will ultimately require separating out good input from bad input. Volitrax uses a voting system to determine which inputs are better than others. For example, if more users agree that a song's title is A instead of

B, then A will be the selected title. Users are also given the chance to confirm if an automated redirect is resulting in a "good enough" copy of the requested video. Of course this voting method is not full-proof, but it is an adequate light-weight solution that does not require user authentication. In the future, other trust metrics could be employed such as giving new users less weight when issuing votes.

### 3.3 Long-term Storage

As mentioned earlier, Volitrax stores data in a number of commercial services that are not originally designed for such purposes. However, storing music video information in these services will significantly increase the chance of such information surviving well past the lifetime of the Volitrax application. Because these services were not intended to be used as a database, their APIs are throttled to at best one request per second. This makes starting a new server with no local information very slow.

## 4. ACKNOWLEDGMENTS

Thank you to Vivens Ndatinya who laid some of the groundwork for the Volitrax add-on. This research was supported by the National Science Foundation (IIS 1008492 and 1009392).

## 5. REFERENCES

- [1] Cheng, X., Dale, C., Liu, J. 2007. Understanding the characteristics of internet short video sharing: YouTube as a case study. Technical Report arXiv preprint arXiv:0707.3670.
- [2] Cheng, X., Dale, C., Liu, J. 2008. Statistics and social network of YouTube videos. In *2008 16th International Workshop on Quality of Service*, 229-238.
- [3] Cunningham, S.J., Nichols, D.M. 2008. How people find videos. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '08)*, ACM, New York, NY, USA, 201-210.
- [4] Klein, M., Nelson, M. L. 2010. Evaluating methods to rediscover missing web pages from the web infrastructure. In *Proceedings of the 10th annual international ACM/IEEE Joint Conference on Digital Libraries (JCDL '10)*. ACM, New York, NY, USA, 59-68.
- [5] Marchionini, G., Shah, C., Lee, C.A., Capra, R. 2009. Query parameters for harvesting digital video and associated contextual information. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '09)*. ACM, New York, NY, USA, 77-86.
- [6] Popitsch, N., Haslhofer, B. 2011. DSNotify – A solution for event detection and link maintenance in dynamic datasets. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(3):266-283, September 2011.
- [7] Prellwitz, M., Nelson, M. L. 2011. Music video redundancy and half-life in YouTube. In *Proceedings of the 15th international conference on Theory and Practice of Digital Libraries: research and advanced technology for digital libraries (TPDL '11)*. Springer-Verlag, Berlin, Heidelberg, 143-150.
- [8] Shah, C. Tubekit: a query-based YouTube crawling toolkit. 2008. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '08)*. ACM, New York, NY, USA, 433.