# VideoKen: Automatic Video Summarization and Course Curation to Support Learning

Debabrata Mahapatra
Videoken
Bangalore, India
debabrata.mahapatra@videoken.com

Ragunathan Mariappan
School of Computing
National University of Singapore
mragunathan@nus.edu.sg

Vaibhav Rajan
School of Computing
National University of Singapore
vaibhav.rajan@nus.edu.sg

Kuldeep Yadav
Videoken
Bangalore, India
kuldeep@videoken.com

Seby A
Videoken
Bangalore, India
seby@videoken.com

Sudeshna Roy
Videoken
Bangalore, India
sudeshna.roy@videoken.com

## ABSTRACT

The number of high quality online videos is increasing rapidly. Online courses as well as universities do not fully leverage the content due to several open challenges in video search, indexing, summarization and customization requirements for specific courses, instructors or learners. We present a new web-based social learning platform called Videoken. Using novel video summarization algorithms, Videoken automatically creates Table of Contents for videos. This allows a textbook-like facility for non-linear search and navigation through the video, enables extraction of semantically coherent clips from within a video and improves video search through better semantic indexing. The platform also allows new ways of course creation and sharing of learning modules; and can be both integrated with existing Learning Management Systems and used independently.

## CCS CONCEPTS

• **Computing methodologies → Information extraction**; **Video summarization**; • **Applied computing** → *Education*;

## KEYWORDS

Videoken; Video Summarization; Course Curation; Table of Contents; Phrase Cloud

**ACM Reference Format:**
Debabrata Mahapatra, Ragunathan Mariappan, Vaibhav Rajan, Kuldeep Yadav, Seby A, and Sudeshna Roy. 2018. VideoKen: Automatic Video Summarization and Course Curation to Support Learning. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3184558.3186988

## 1 INTRODUCTION

Massive Open Online Courses (MOOC) and other online resources provide many high quality educational videos on the Internet. The overwhelming number of videos presents challenges of finding the most appropriate video and consuming the content effectively.

Consider the case of a learner trying to find a relevant video to learn any given topic. Online search would present thousands of videos that differ in content, presentation styles, duration, video quality etc. Finding the right video that is most suitable to the learner's background, learning goal and preferred learning style is non-trivial. Further, navigating through the search results usually involves skimming through the video contents by watching snippets of the video just to know whether the video is appropriate. This is time consuming even for a single video and definitely cannot practically scale beyond a few videos.

Although MOOC platforms provide well organized courses, according to the study conducted in [10] that investigates high dropout rates in MOOCs, the structure of a course designed by the course creators may not be helpful to all participants. Moreover, the authors suggest considering different design patterns for a course. A similar study in [3] reveals two important factors, among others, for high dropout rate: a) there exists a gap in the perceived expectation of the course content and what is actually delivered, b) the perceived difficulty level of a course vary among the participants. MOOCs are designed for large scale use. However, the studies in [3, 10] suggest that this very design may not have the flexibility that is necessary to reach a large audience. Instead, a customized approach for knowledge delivery can potentially lead to better learning outcomes.

Recently *Byte Sized Learning (BSL)* [9] has gained popularity in corporate training and instructors are (re)creating existing video contents to incorporate BSL in their courses. Thus, instructors cannot directly use the already available high-quality videos that are not short enough for BSL.

In this paper, we present an innovative web-based platform, called **VideoKen**, that addresses these challenges using novel multimedia technologies. In particular, we develop certain tools that, when used together, can provide efficient solutions to the problems of both the learners as well as instructors. Our platform brings the following novel additions to Learning Management Systems (LMS):

- We provide two different types of automatic video summarization: Multimedia Table of Content (*MMToC*) and Phrase Cloud, that enables the user to quickly perceive the content of a video without having to watch it. These features are

*WWW '18 Companion, April 23–27, 2018, Lyon, France*

ACM ISBN 978-1-4503-5640-4/18/04.
https://doi.org/10.1145/3184558.3186988

hyperlinked within the timeline of video, which facilitates non-linear navigation of a video. Further, the use of these summarization tools improves video search.

- By utilizing these video summarization capabilities, we build a tool that allows a user to quickly create a *Learning Object* from clips of videos, called *Kenlist*. This permits a learner to modularize the knowledge of interest in a customized manner and efficiently create BSL-based courses with no extra content creation.
- Integrating the standard features provided by a typical LMS system with kenlists, we develop a novel Course Curation functionality. This would empower an instructor to create high quality, highly customizable course material by using either existing publicly available videos or their own content.

VideoKen is accessible on the web at www.videoken.com. In the following sections we briefly describe the key features of Videoken that we intend to demonstrate.

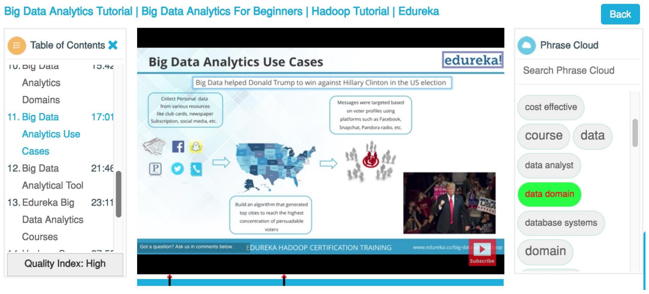
## 2 MMTOC: MULTI-MODAL TABLE OF CONTENTS FOR VIDEOS

In a textbook, the Table of Contents (**ToC**) helps a reader to find and navigate to a particular topic. However, in case of a video, especially if it is long, it is difficult to locate where a particular topic is covered in its time line. For typical educational videos that contain slides, we have developed an algorithm to automatically generate the ToC. Figure 1 and 2 illustrate the user interface for ToC of a video, to the left of the video. For this, we use information from different modalities, visual, audio and text (obtained from the transcript); hence the name Multi-modal ToC or **MMToC**. From a user experience point of view, MMToC provides several advantages.

First, it becomes as simple as navigating a book; because the titles of MMToC are linked to their corresponding time of occurrence in the video. So by clicking a particular ToC entry one can directly go to the start time of its corresponding frame in the video, much like the beginning of a chapter beginning in a book. According to the user study done in [1], having watched a video earlier, the user can locate a particular concept in a video faster if MMToC is available.

Second, for long lecture videos, the user can get a quick summary of topics covered by seeing the MMToC, even without watching the full video. This enables the user to find a better match for their requirements among many options in a video search result.

In order to generate MMToC automatically, we first segment the video temporally into scenes. Each scene is then represented by some of the key-frames in it. Using Optical Character Recognition (OCR), we extract the text information from the key-frames. In this work, we use a commercially available OCR engine, offered by the *Cognitive Services* of Microsoft Azure cloud computing. To prioritize among all the extracted texts, saliency scores of these texts are computed using information from multiple modalities. Visual saliency scores of a text include stroke width [2] of the characters, font size, geometrical location of the text etc. Occurrence of the text in the transcript of the video is also taken into account. If the transcript is not available, then we obtain it through automatic speech recognition (ASR) (Kaldi [11]). If the instructor has uttered the text then pitch information of the corresponding audio signal is also used in the saliency score. Using these scores, the titles are chosen among the extracted texts that form the MMToC entries.


Big Data Analytics Tutorial | Big Data Analytics For Beginners | Hadoop Tutorial | Edureka
Back
Table of Contents
10. Big Data Analytics Domains 15:42
11. Big Data Analytics Use Cases 17:01
12. Big Data Analytical Tool 21:46
13. Edureka Big Data Analytics Courses 23:11
Quality Index: High
Big Data Analytics Use Cases
edureka!
EDUREKA HADOOP CERTIFICATION TRAINING
www.edureka.co/big-data-and-hadoop
Subscribe
Phrase Cloud
Search Phrase Cloud
cost effective
course
data
data analyst
data domain
database systems
domain

Big Data helped Donald Trump to win against Hillary Clinton in the US election

Collect Personal data from various resources like club cards, newspaper Subscription, social media, etc.

Messages were targeted based on voter profiles using platforms such as Facebook, Snapchat, Pandora radio, etc.

Build an algorithm that generated top cities to reach the highest concentration of persuadable voters

Got a question? Ask us in comments below.

**Figure 1: User interface of MMToC and phrase cloud. Both kinds of summaries can be scrolled through. Upon clicking a phrase the (here: "data domain"), markers appear in the blue line parallel to the video timeline representing the time points where this phrase is present (in the video frame or audio).**

In general, many educational videos are produced with slides to facilitate instruction. However, there are also videos wherein information is conveyed without using slides such as classroom lectures delivered orally with or without using a black board, panel discussions, skill development videos from various domains, such as sports, cooking, code walk-through etc. In principle, for any instructional video, MMToC can improve the effectiveness of video consumption.

Due to recognition errors in OCR and ASR, there may be errors in the titles of MMToC entry. We build a moderation tool that can be used to modify the existing ToC of a video or to create a ToC in case it is not already generated automatically. To create such a manual ToC for a video, where the topic titles are not apparent, the user has to understand the content and come up with suitable titles. This exercise encourages the user to reflect up on his/her learning, hence leading to better comprehension.

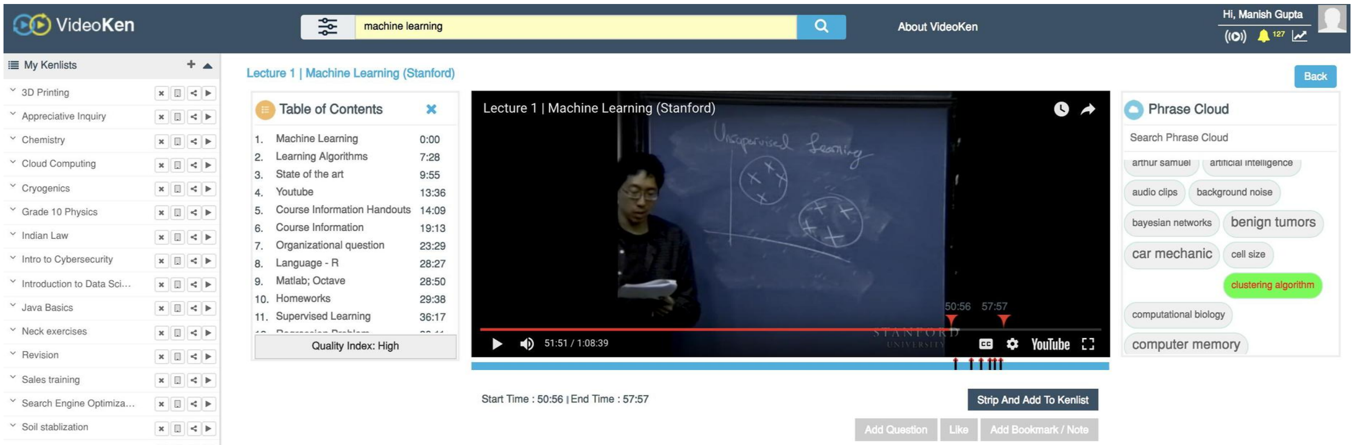
## 3 PHRASE CLOUD

A textbook also contains a set of index words. While the book's ToC covers the major topics, the index points to many specialized concepts, presented in the book. In our platform, we bring similar capabilities for a video. Instead of limiting to words, we automatically find the relevant phrases used in the video that might be of interest to users. We call this set of extracted phrases the *phrase cloud*.

The user interface, illustrated in Figure 1, has a parallel time-line to the video player. Upon clicking an item in the phrase cloud (which is on the right of the video), the occurrences of the selectedd phrase in the time-line is shown with markers. This allows the user to navigate a video in non-linear manner, and quickly reach to the point of interest that discusses the selected phrase.

Like MMToC, the user can scroll through the entire phrase cloud and get a quick summary of the major concepts discussed in the video. Thus MMToC and phrase cloud provide two different forms of hyperlinked video summarization.

The relative significance of phrases (measured by the frequency of occurrence) is depicted by its font size in the cloud: higher the



| | | |
|---|---|---|
| 1. | Machine Learning | 0:00 |
| 2. | Learning Algorithms | 7:28 |
| 3. | State of the art | 9:55 |
| 4. | Youtube | 13:36 |
| 5. | Course Information Handouts | 14:09 |
| 6. | Course Information | 19:13 |
| 7. | Organizational question | 23:29 |
| 8. | Language - R | 28:27 |
| 9. | Matlab; Octave | 28:50 |
| 10. | Homeworks | 29:38 |
| 11. | Supervised Learning | 36:17 |

Start Time : 50:56 | End Time : 57:57

**Figure 2: Illustration of stripping a video to a clip and adding to a kenlist. The red markers on the video timeline can be adjusted to select the portion of the video that will be added to a kenlist. All kenlists are listed on the leftmost panel.**

font size more the significance. Furthermore, with the 2D space to show these phrases, we segregate them topic-wise.

The phrase cloud is generated in multiple steps. First, we obtain the transcript of the video. If not available, then we perform Automatic Speech Recognition on the audio. For this, we use Kaldi [11] ASR engine. Then the most important concepts are extracted using *Text Rank* [6] for this. As a result, we get a set of key words/phrases. Then, we represent the extracted words in the form of continuous vectors by using Word2Vec [7] for further processing. Using a dictionary of concepts from various domains, we compute the significance value of each extracted phrase. The word mover's distance [5] metric is used to find the distance of a particular phrase with respect to the concept dictionary; lesser the distance higher the significance. Finally, we use DBSCAN [4] to cluster these phrases and put them in disjoint sets that represent separate topics.

## 4 KENLIST: A LEARNING OBJECT

Often times, one would like to refer back to previously watched videos that are associated with a particular topic or knowledge domain. The *playlist* feature of YouTube enables users to do that. However, in general, whatever knowledge we are searching for may not be readily available in a single video. The user's requirement might be distributed across different videos; combining the useful portions of these videos would make a complete package of knowledge that is specific to the user's requirements. For example, while learning various sorting algorithms, one student may find that merge sort is best explained in first 10 minutes of an hour long lecture, quick sort is best explained somewhere in the middle of another video for 20 minutes, and so on. In Videoken we provide a functionality, namely **Kenlist**, that allows the user to efficiently aggregate portions of video material from different sources.

Kenlists can be considered as a learning *object* for a particular knowledge domain that is constructed from a set of video clips. The embedded video player in our platform provides clipping options within a video. Note that neither the video nor the clip is stored in Videoken, only relevant time pointers are stored and the video is played directly from Youtube or the selected channel.

Features like MMToC and phrase cloud are complementary to this clipping action. Because, using these hyperlinked summaries one can precisely locate the portion of interest and customize the amount of information that needs to be registered for future references. This is illustrated in Figure 2. Besides the semantic content exploration, we also provide relevant search results (discussed in section 6) for user query that can further help in building useful kenlists.

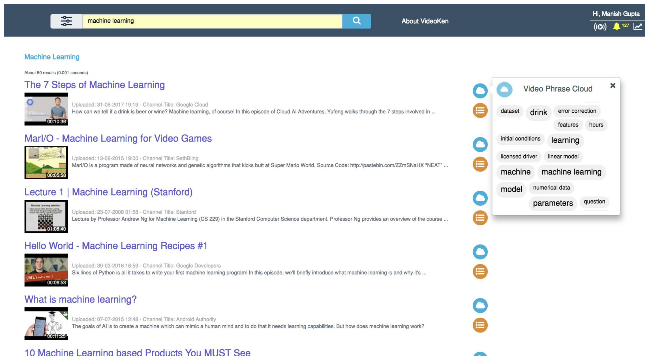
In addition, Videoken provides features found in other Learning Management Systems (LMS). Users can add bookmarks to the videos at specific points in time and annotate it by adding notes. To encourage peer-to-peer learning, we enable sharing of kenlists. While accessing a shared kenlist, a user can provide her feedback to the creator of that kenlist as well.

## 5 COURSE CURATION

While a learning object like Kenlist helps a learner put together videos for later references, from an instructor's point of view, it may not be directly used as a course. We provide some tools that allow an instructor to extend a kenlist to a full fledged course. A course is structured by adding sessions to it. The instructor can add kenlists to a session and supporting contents relevant to it; for example, documents for further readings and assignments.

We also provide LMS features: quizzes can be added at appropriate places in the course; to provide an engaging learning experience the instructor can add questions with multiple choice answers at different points in the kenlist. While watching the kenlist as a student, these questions automatically appear pausing the video until the question is answered. To promote social learning, students can start a discussion for sessions, kenlists, videos, documents and quizzes as well. These courses can also be shared among instructors.

To track the learning outcomes of students we provide an analytics dashboard for the instructor that shows information such as course performance summaries, leaderboards and how much time is being spent for a video by the students. The course creation features allows an instructor to leverage thousands of publicly accessible educational videos and build a customized course. Therefore, it avoids recreating some of the already available videos.



Machine Learning

About 50 results (0.001 seconds)

The 7 Steps of Machine Learning

00:10:36

Uploaded: 31-08-2017 19:19 - Channel Title: Google Cloud

How can we tell if a drink is beer or wine? Machine learning, of course! In this episode of Cloud AI Adventures, Yufeng walks through the 7 steps involved in ...

MarI/O - Machine Learning for Video Games

00:05:58

Uploaded: 13-06-2015 19:00 - Channel Title: SethBling

MarI/O is a program made of neural networks and genetic algorithms that kicks butt at Super Mario World. Source Code: http://pastebin.com/ZZmSNaHX "NEAT" ...

Lecture 1 | Machine Learning (Stanford)

01:08:40

Uploaded: 23-07-2008 01:58 - Channel Title: Stanford

Lecture by Professor Andrew Ng for Machine Learning (CS 229) in the Stanford Computer Science department. Professor Ng provides an overview of the course ...

Hello World - Machine Learning Recipes #1

00:06:53

Uploaded: 30-03-2016 16:59 - Channel Title: Google Developers

Six lines of Python is all it takes to write your first machine learning program! In this episode, we'll briefly introduce what machine learning is and why it's ...

What is machine learning?

00:11:25

Uploaded: 07-07-2015 12:48 - Channel Title: Android Authority

The goals of AI is to create a machine which can mimic a human mind and to do that it needs learning capabilities. But how does machine learning work?

10 Machine Learning based Products You MUST See

**Figure 3: Illustration of search results upon user query. Adjacent to each video entry in the results are icons for MMToC (orange) and Phrase Cloud (blue). Hovering on the corresponding icon shows the MMToC and Phrase Cloud thereby providing two kinds of video summarization, that can help the user find the most appropriate video.**

## 6 SEARCH

Given a search query from a user we first obtain the search results (list of videos) from YouTube. Then we re-rank this list to make it more relevant to the user's query and profile. We leverage the semantic metadata extracted from videos for re-ranking. It is achieved by using a recommendation system [8] that takes user profile and video features into account. User profiles are created by using their activity logs on our platform. Apart from the description of a video (provided by the content creator), which is already used in YouTube search, we use the MMToC, phrase cloud and its presence in a *KenList* as some of the extra video features.

In real time, the user query and the YouTube list of videos are given as input to the recommendation system. We provide the pre-computed user and video features to the trained recommendation model to score each of the video in the list, for a particular user and re-rank the results. A typical search result is illustrated in Figure 6. Upon hovering the cursor on either phrase cloud or MMToC icon, one can see the corresponding form of summary for the video. So, if the content is not relevant, one need not even view the video. In the final search result, we also present already existing courses relevant to the query. However, the course curated on our platform only appears if the creator makes it publicly available. Further, to narrow down the search results, we provide category and channel in filters while searching. These are primarily YouTube channels, such as Khan Academy, MIT OCW, NPTEL, Stanford, TED Ed etc.

## 7 CONCLUSIONS

Online education through the use of ever-increasing number of videos is still far from being successful, due to difficulties in effective video consumption and knowledge delivery through online courses. We develop VideoKen, a web-based social learning platform that provides tools to help both the learner as well as the instructor, through novel tools for finding relevant videos and curating effective courses. MMToC and phrase cloud are two video summarization tools that also aid in non-linear navigation and kenlist creation. Using kenlists, a user can quickly put together a coherent set of video clips. If utilized to its true potential by curating courses with the existing educational videos, this platform can expedite the process of customized knowledge creation and acquisition.

## 8 ACKNOWLEDGMENTS

We thank Kundan Shrivastava, Ranjeet Kumar, Abhinay Kumar, Onkar Hoysala, Astha Srivastava, Aakash Tyagi, Saahil Madaan, Harish Reddy Pittu and Shant Purushotam Kumar for their contributions in building the Videoken platform. We also thank Vishnu Raned, Ashish Vikram and Manish Gupta for useful discussions.

## REFERENCES

[1] Arijit Biswas, Ankit Gandhi, and Om Deshmukh. 2015. MMToC: A Multimodal Method for Table of Content Creation in Educational Videos. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM '15)*. ACM, New York, NY, USA, 621–630. https://doi.org/10.1145/2733373.2806253

[2] B. Epshtein, E. Ofek, and Y. Wexler. 2010. Detecting text in natural scenes with stroke width transform. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2963–2970. https://doi.org/10.1109/CVPR.2010.5540041

[3] Thommy Eriksson, Tom Adawi, and Christian Stöhr. 2017. "Time is the bottleneck": a qualitative study exploring why learners drop out of MOOCs. *Journal of Computing in Higher Education* 29, 1 (01 Apr 2017), 133–146. https://doi.org/10.1007/s12528-016-9127-8

[4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *Kdd*, Vol. 96. 226–231.

[5] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International Conference on Machine Learning*. 957–966.

[6] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

[7] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[8] Nagarajan Natarajan and Inderjit S Dhillon. 2014. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics* 30, 12 (2014), i60–i68.

[9] A. H. Omer. 2015. Is Bite Sized Learning The Future Of eLearning? https://elearningindustry.com/bite-sized-learning-future-of-elearning. (2015). Accessed: 2018-01-10.

[10] Daniel FO Onah, Jane Sinclair, and Russell Boyatt. 2014. Dropout rates of massive open online courses: behavioural patterns. *EDULEARN14 Proceedings* (2014), 5825–5834.

[11] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, Jan Silovsky, Georg Stemmer, and Karel Vesely. 2011. The Kaldi Speech Recognition Toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society. IEEE Catalog No.: CFP11SRW-USB.