

← 1/40 → \* \* \* 9:08:20

# Videos!

••

University of Southern California

 USC **Viterbi**  
School of Engineering



# Video Search Engines YouTube et al

Copyright Ellis Horowitz 2011-2022

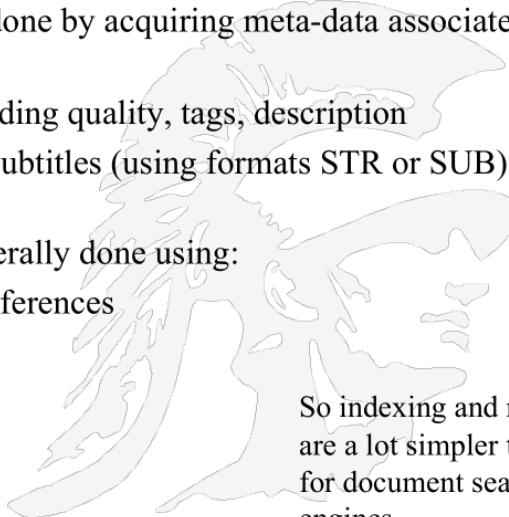
••

University of Southern California

USC **Viterbi**  
School of Engineering

## Video Search Engines – Quick Summary

- A **video search engine** is a web-based search engine which crawls the web primarily for video content.
  - YouTube is not strictly a video search engine as it does not crawl the web looking for video content
- The **indexing** of video content is normally done by acquiring meta-data associated with the video, e.g.
  - Author, title, creation date, duration, coding quality, tags, description
  - Other aspects of video recognition are subtitles (using formats STR or SUB) and transcription (using format TTXT)
- The **ranking** of videos under a query is generally done using:
  - Relevance: using metadata and user preferences
  - Ordered by date of upload
  - Ordered by number of views
  - Ordered by duration
  - Ordered by user rating



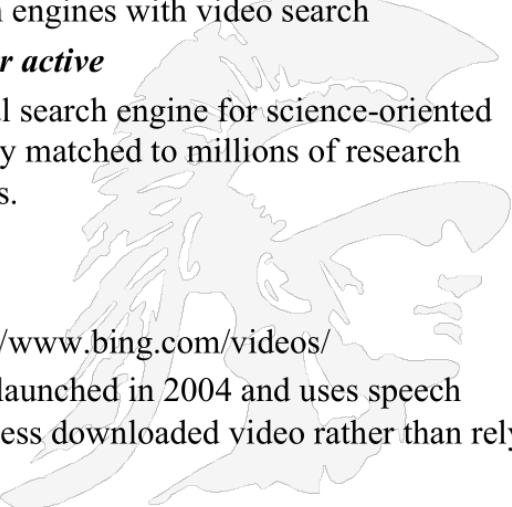
So indexing and ranking  
are a lot simpler than  
for document search  
engines

••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## Video Search Engines That *Crawl* for Content



- ***Those no longer existing***
  - ***CastTV*** was a Web-wide video search engine that was founded in 2006
    - ***No longer active***
  - ***Munax*** released their first version all-content search engine in 2005 and powers both nationwide and worldwide search engines with video search
    - <http://www.munax.com/> ***no longer active***
  - ***ScienceStage*** is an integrated universal search engine for science-oriented videos. All videos are also semantically matched to millions of research documents from open-access databases.
    - ***No longer active***
- ***A few remain***
  - ***Bing*** does crawl for videos, see <https://www.bing.com/videos/>
  - ***blinkx*** (renamed as RhythmOne) was launched in 2004 and uses speech recognition and visual analysis to process downloaded video rather than rely on metadata alone
    - <http://www.blinkxtv.com/> ***now redirects to 360Daily.com***

Copyright Ellis Horowitz, 2011-2022

3

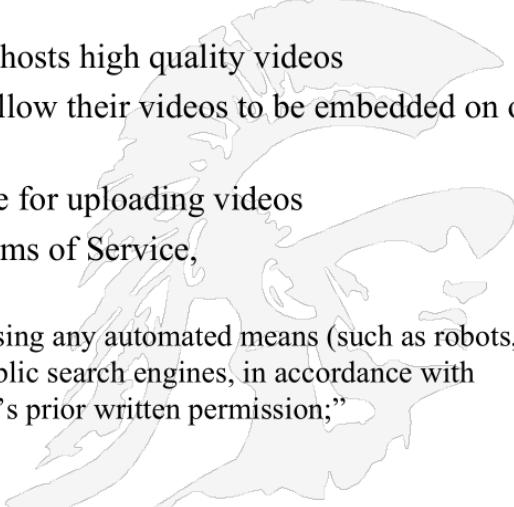
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## Video Search Engines That Host

- Largely because of the large file sizes involved, video hosting is highly concentrated on a fairly small number of websites
  - **vimeo.com**, first to support HD video, focuses on short, arty, films
  - **vevo.com**, a joint venture of Universal Music Group, Sony Music Entertainment and Warner Music Group
  - **dailymotion.com**, owned by Vivendi, hosts high quality videos
- Most of these websites which host video allow their videos to be embedded on other websites
- **YouTube.com** has become the defacto site for uploading videos
- It is legal to crawl YouTube, see their Terms of Service,  
[www.youtube.com/static?template=terms](http://www.youtube.com/static?template=terms)
- “3. You are not allowed to access the Service using any automated means (such as robots, botnets or scrapers) except (a) in the case of public search engines, in accordance with YouTube’s robots.txt file; or (b) with YouTube’s prior written permission;”



Copyright Ellis Horowitz, 2011-2022

4

••

University of Southern California

USC Viterbi School of Engineering

Video Search Engines That Stream Entertainment

- **Hulu** is an American subscription video on demand service jointly owned by Walt Disney, 21<sup>st</sup> Century Fox, Comcast, and Time Warner
  - In December 2017, Disney acquired Fox's partial ownership, giving it a majority stake; other owners include Comcast
- **Netflix** is an American subscription video on demand service, that originally delivered DVDs;
  - They develop their own content as well as offering content from major film distributors
- **Amazon Prime** is an American subscription video on demand service offering television and file shows for rent or purchase
- **Disney+** a recent entry
- There are many others: XtremeHD, Sling TV, Apple TV+, HBO Max, Acorn TV, etc

• Entertainment

Copyright Ellis Horowitz, 2011-2022

5

••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## Some Technologies Supporting Video Content

- **Subtitles:** there are two formats, one for subtitles and one for transcripts
  - There are three main types of video subtitling services:
    1. **open caption:** burned into the video
    2. **closed caption:** can be turned on/off, generally at the bottom of the screen
    3. **SDH (Subtitles for the Deaf and Hard of Hearing):** similar to closed caption, but includes words describing actions or moods
  - **SRT or SUB for subtitles**
    - SRT (. srt) stands for “SubRip Subtitle” file, and it's the most common subtitle/caption file format. It is a **text format**
    - **TTXT for transcripts**
- **Speech Recognition,** used to extract phrases from audio transcripts for better indexing
  - **Gaudi, Google Audio Indexing** uses voice recognition to locate the exact spot where words are spoken
  - <https://www.searchenginejournal.com/google-audio-search-will-it-ever-be-possible/397129/>
  - **Text Recognition:** uses OCR on video slides to detect words,
  - e.g. **TalkMiner System**, see [https://www.youtube.com/watch?v=7N6I\\_m9LywM](https://www.youtube.com/watch?v=7N6I_m9LywM)

Copyright Ellis Horowitz, 2011-2022

6

••

**University of Southern California**

**USC Viterbi**  
School of Engineering

# YouTube Background

- YouTube is an American video hosting website headquartered in San Bruno, California, created by three former PayPal employees: Chad Hurley, Steve Chen, Jawed Karim in February 2005.
- In November 2006, it was bought by Google for US\$1.65 billion
- In 2020 Google announced that YouTube generated revenue of \$19.8 billion
- The site allows users to upload, view, rate, share, add to favorites, report and comment on videos
- In January 2022, the website was ranked as the second most popular site by Alexa Internet, a web traffic analysis company (now owned by Amazon)
  - See also  
[https://en.wikipedia.org/wiki/List\\_of\\_most\\_popular\\_websites](https://en.wikipedia.org/wiki/List_of_most_popular_websites)

For details see Related Articles page, Mar 2020

Copyright Ellis Horowitz, 2011-2022

••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## YouTube as a Search Engine

- **YouTube - The 2nd Largest Search Engine (cite:Infographic)**
- **YouTube processes more than 3 billion searches a month.**
- **It's bigger than Bing, Yahoo!, Ask and AOL combined!**
- <http://www.mushroomnetworks.com/infographics/youtube---the-2nd-largest-search-engine-infographic>



Copyright Ellis Horowitz, 2011-2022

8

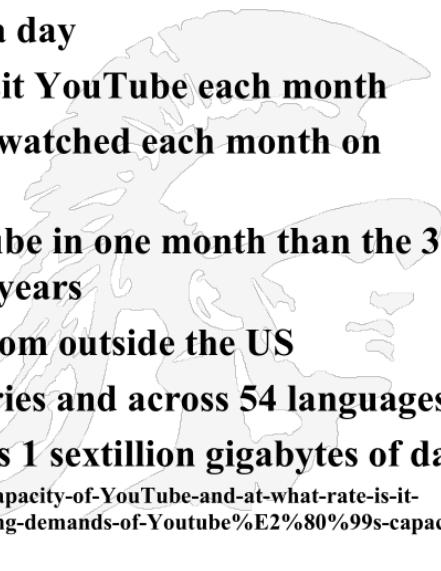
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## YouTube Traffic - Some Facts

- **As of 2021:**
  - **60 hours of video are uploaded every minute, or one hour of video is uploaded to YouTube every second.**
  - **Over 4 billion videos are viewed a day**
  - **Over 800 million unique users visit YouTube each month**
  - **Over 3 billion hours of video are watched each month on YouTube**
  - **More video is uploaded to YouTube in one month than the 3 major US networks created in 60 years**
  - **70% of YouTube traffic comes from outside the US**
  - **YouTube is localized in 39 countries and across 54 languages**
  - **It is estimated that YouTube holds 1 sextillion gigabytes of data**
  - <https://www.quora.com/What-is-the-total-size-storage-capacity-of-YouTube-and-at-what-rate-is-it-increasing-How-is-Google-keeping-up-with-the-increasing-demands-of-Youtube%E2%80%99s-capacity-given-that-thousands-of-videos-are-uploaded-every-day>



Copyright Ellis Horowitz, 2011-2022

9

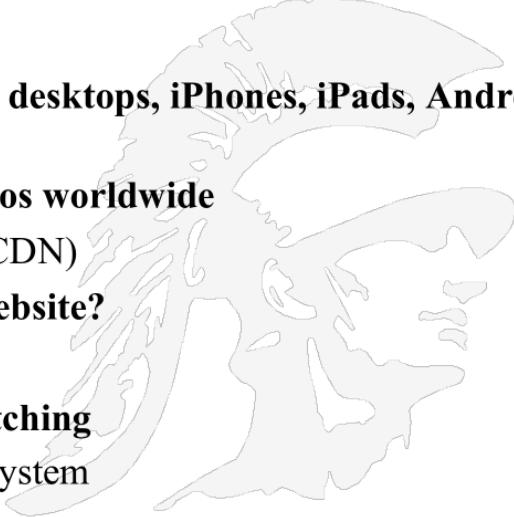
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## YouTube Search Engine Issues to Consider

- Since crawling, indexing and ranking are not big challenges for YouTube, what are the major hurdles
- 1. **What video formats are acceptable**
  - For uploading
  - For downloading
- 2. **How are videos to be displayed on: desktops, iPhones, iPads, Android devices, etc**
- 3. **How does YouTube distribute videos worldwide**
  - A content distribution network (CDN)
- 4. **How does YouTube monetize its website?**
  - YouTube's ContentID system
- 5. **How does YouTube keep users watching**
  - The YouTube Recommendation System



Copyright Ellis Horowitz, 2011-2022

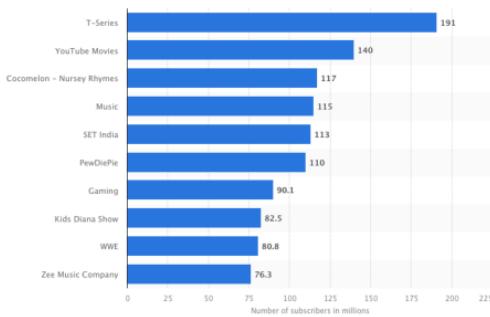
10

••

**University of Southern California**  **YouTube Channels**

**USC Viterbi**  
School of Engineering

- In order to upload a video you must be a registered user
- In addition YouTube offers a special type of account called a **channel**; channels include
  - thumbnails of videos you've uploaded,
  - members to whom you've subscribed,
  - videos from other members you've picked as favorites,
  - lists of members who are your friends,
  - your subscribers, and
- Biggest YouTube Channels as of 2021



Channel	Subscribers (in millions)
T-Series	191
YouTube Movies	140
Cocomelon – Nursery Rhymes	117
Music	115
SET India	113
PewDiePie	110
Gaming	90.1
Kids Diana Show	82.5
WWE	60.8
Zee Music Company	76.3

*With 1 million subscribers, a YouTuber will make between \$300,000 – \$2 million To be in the top 1000 YouTubers you must have ~1.8 million subscribers As of 09/2020, there are more than 2000 YouTubers with over a million subscribers*

<https://www.statista.com/statistics/277758/most-popular-youtube-channels-ranked-by-subscribers/>

© Copyright Ellis Horowitz, 2011-2022

11

••

**University of Southern California**  **USC**

## YouTube Gathers Information When Videos are Uploaded

YouTube captures:

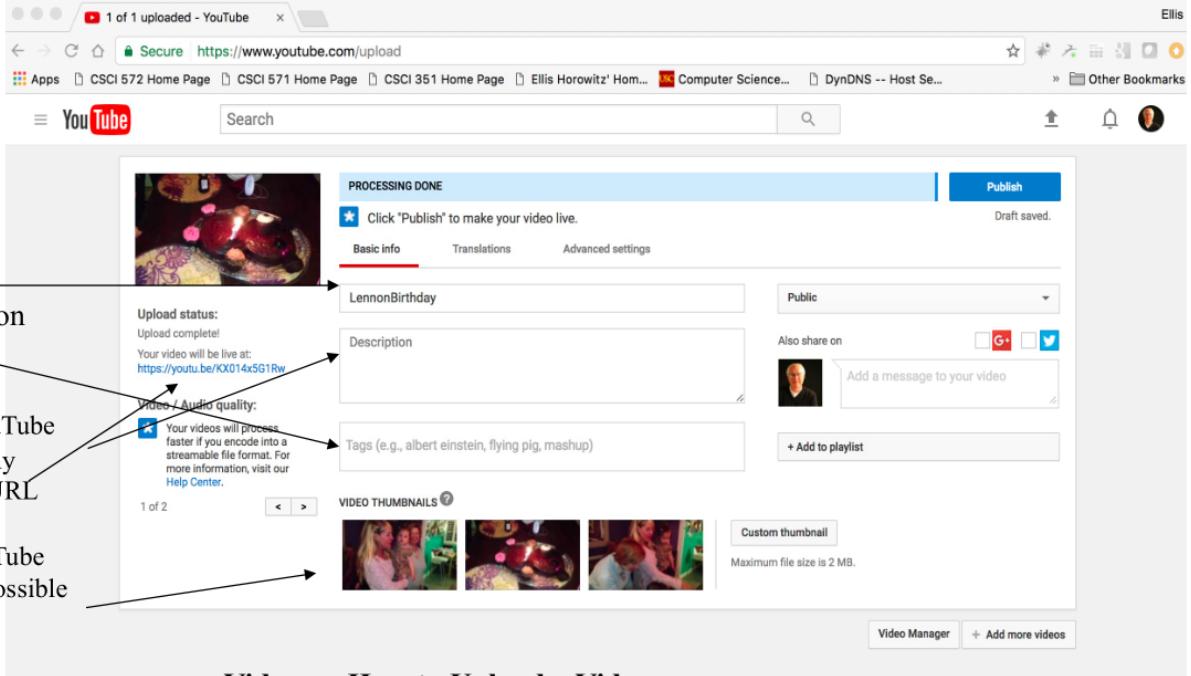
Name \_\_\_\_\_

Description \_\_\_\_\_

Tags \_\_\_\_\_

Note: YouTube immediately assigns a URL

Note: YouTube suggests possible thumbnails



**PROCESSING DONE**

Click "Publish" to make your video live.

**Basic info**    Translations    Advanced settings

**Upload status:**  
Upload complete!  
Your video will be live at:  
<https://youtu.be/KX014x5G1Rw>

**Description**

**Tags (e.g., albert einstein, flying pig, mashup)**

**VIDEO THUMBNAILS**

**Video & Audio quality:**  
Your video will process faster if you encode into a streamable file format. For more information, visit our Help Center.

**Custom thumbnail**  
Maximum file size is 2 MB.

**Video Manager**    + Add more videos

**Video on How to Upload a Video**  
<https://support.google.com/youtube/answer/57407>

Copyright Ellis Horowitz 2011-2022

••

University of Southern California  USC

## Uploading to YouTube Second Input Screen

1 of 1 uploaded - YouTube

Secure https://www.youtube.com/upload

Ellis

Apps CSCl 572 Home Page CSCl 571 Home Page CSCl 351 Home Page Ellis Horowitz' Hom... Computer Science... DynDNS -- Host Se... Other Bookmarks

You Tube Search

PROCESSING DONE

Click "Publish" to make your video live.

Basic info Translations Advanced settings

Draft saved.

Original language Select language

Translate into (0) Select language

LennonBirthday Enter translated title

Description Enter translated description

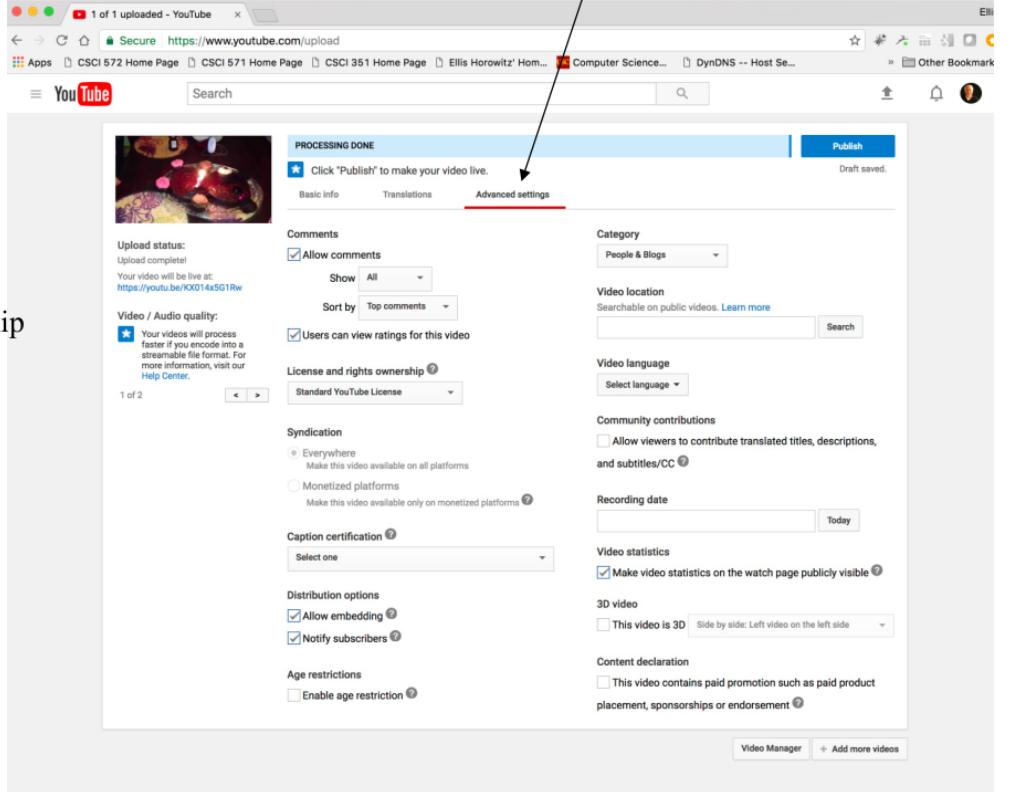
Get professional translation Buy translation (BETA)

Video Manager Add more videos

••

**University of Southern California**  

## Uploading to YouTube Third Input Screen



The screenshot shows the YouTube upload interface. At the top, it says "1 of 1 uploaded - YouTube". Below that is a navigation bar with links like "Secure https://www.youtube.com/upload", "CSCI 572 Home Page", "CSCI 571 Home Page", "CSCI 351 Home Page", "Ellis Horowitz' Home Page", "Computer Science...", "DynDNS -- Host Se...", and "Other Bookmark". The main area is titled "Uploading to YouTube Third Input Screen". It shows a thumbnail of a video and the message "PROCESSING DONE Click 'Publish' to make your video live." There are tabs for "Basic info", "Translations", and "Advanced settings", with "Advanced settings" being the active tab. On the left, there are sections for "Comments" (with "Allow comments" checked), "Video / Audio quality" (with a note about encoding), "License and rights ownership" (set to "Standard YouTube License"), "Syndication" (with "Everywhere" selected), "Caption certification" (with "Select one"), "Distribution options" (with "Allow embedding" and "Notify subscribers" checked), and "Age restrictions" (with "Enable age restriction" unchecked). On the right, there are sections for "Category" (set to "People & Blogs"), "Video location" (with a search bar and "Search" button), "Video language" (with a "Select language" dropdown), "Community contributions" (with "Allow viewers to contribute translated titles, descriptions, and subtitles/CC" unchecked), "Recording date" (with a date picker set to "Today"), "Video statistics" (with "Make video statistics on the watch page publicly visible" checked), "3D video" (with "This video is 3D" and a dropdown for "Side by side: Left video on the left side"), "Content declaration" (with "This video contains paid promotion such as paid product placement, sponsorships or endorsement" unchecked), and "Video Manager" (+ Add more videos).

YouTube allows the creator to specify:

- License and ownership
- Syndication
- Caption
- Embedding
- Age restrictions
- Categories

...

••

University of Southern California  USC

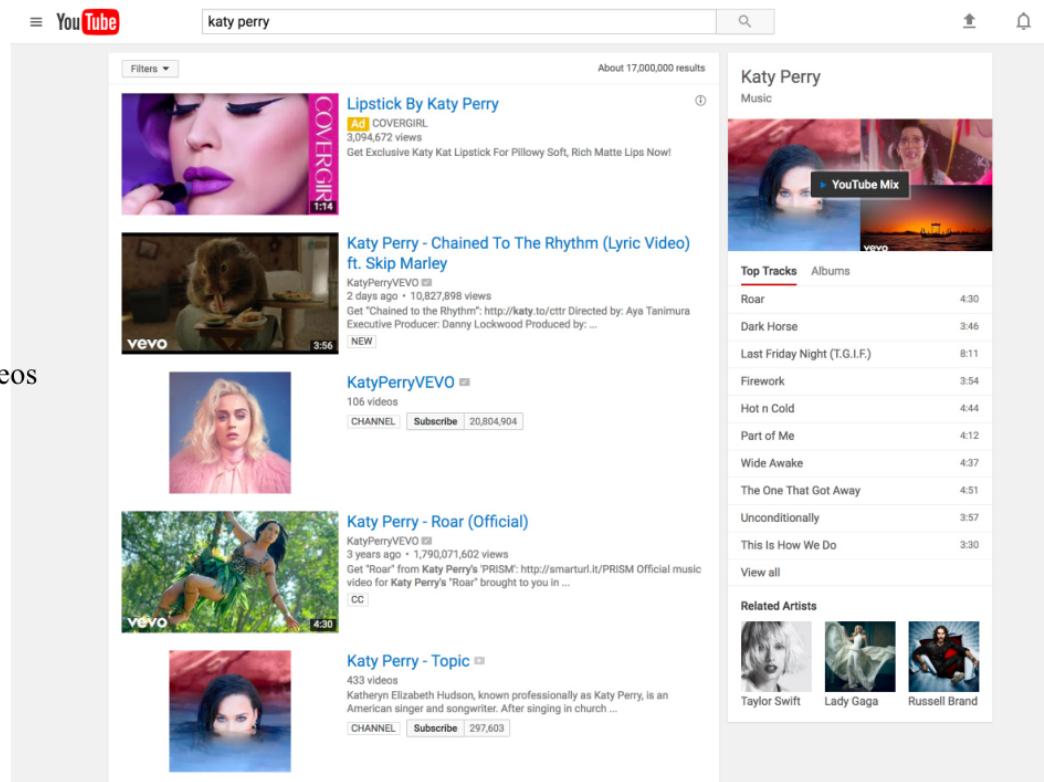
## Business Model: Ads, Ads, Ads Sample YouTube Search Results for Katy Perry

First result is an Ad

2<sup>nd</sup> and 4<sup>th</sup> results are stored at Vevo

3<sup>rd</sup> and 5<sup>th</sup> results are links to a Katy Perry channel with 106 videos

To the right is a mix of Katy Perry songs and some “related” artists



**Lipstick By Katy Perry**  
Ad COVERGIRL  
3,094,672 views  
Get Exclusive Katy Kat Lipstick For Pillow Soft, Rich Matte Lips Now!

**Katy Perry - Chained To The Rhythm (Lyric Video)  
ft. Skip Marley**  
KatyPerryVEVO 2 days ago • 10,827,898 views  
Get "Chained to the Rhythm": <http://katy.to/ctr> Directed by: Aya Tanimura Executive Producer: Danny Lockwood Produced by: ... NEW

**KatyPerryVEVO**  
106 videos  
CHANNEL | [Subscribe](#) 20,804,904

**Katy Perry - Roar (Official)**  
KatyPerryVEVO 3 years ago • 1,790,071,602 views  
Get "Roar" from Katy Perry's PRISM: <http://smarturl.it/PRISM> Official music video for Katy Perry's "Roar" brought to you in ... CC

**Katy Perry - Topic**  
433 videos  
Katheryn Elizabeth Hudson, known professionally as Katy Perry, is an American singer and songwriter. After singing in church ... CHANNEL | [Subscribe](#) 297,603

**Katy Perry**  
Music  
  
Top Tracks Albums  
Roar 4:30  
Dark Horse 3:46  
Last Friday Night (T.G.I.F.) 8:11  
Firework 3:54  
Hot n Cold 4:44  
Part of Me 4:12  
Wide Awake 4:37  
The One That Got Away 4:51  
Unconditionally 3:57  
This Is How We Do 3:30  
View all

Related Artists  
  
Taylor Swift  
Lady Gaga  
Russell Brand

••

**University of Southern California**

**USC Viterbi**  
School of Engineering

**Ranking: Ads, Views, Age  
YouTube Search Results**

Begins with an ad

The next 4 results are ordered by the number of views: 420,004, 369,979, 228,004

Subsequent listings are a mixture of highly viewed videos, but older, e.g. Lec 1 MIT has 3 million+ views but is 7 years old

It is not obvious how the ranking was determined

**computer science**

About 18,600,000 results

**Technology For Students**  
About Best Buy  
41,606 views  
Check Out Best Buy's Student Device Management Programs For Schools!

**Lecture 0 - Introduction to Computer Science I**  
Asim Ali  
2 years ago • 420,004 views  
This is first lecture from the series of course "Introduction to Computer Science I", Harvard OpenCourseWare with Instructor David ...

**Computer Science a good major?**  
ENGINEERED TRUTH 88  
3 years ago • 369,979 views  
You should ask a lot of people for advice. In my opinion, most people in the world should get their bachelors in CS before working ...

**Computer science is for everyone | Hadi Partovi | TEDxRainier**  
TEDx Talks  
2 years ago • 228,044 views  
This talk was given at a local TEDx event, produced independently of the TED Conferences. This persuasive talk shows how ...

**Computer Science vs Self-Taught vs Coding Bootcamp (ft. Quincy Larson)**  
ENGINEERED TRUTH 88  
4 months ago • 155,464 views  
Quincy Larson is the creator of FreeCodeCamp.com, the #1 way to learn code for free. FreeCodeCamp is also the most starred ...

**Computer science education: why does it suck so much and what if it didn't? | Ashley Gavin |...**  
TEDx Talks  
1 year ago • 220,105 views  
Ashley's talk shines a light on the major problem that is American Computer Science education. In 2020, 1.4 million new jobs will ...

**Computer science education: why does it suck so much and what if it didn't? | Ashley Gavin |...**  
TEDx Talks  
1 year ago • 220,105 views  
Ashley's talk shines a light on the major problem that is American Computer Science education. In 2020, 1.4 million new jobs will ...

**Lec 1 | MIT 6.00 Introduction to Computer Science and Programming, Fall 2008**  
MIT OpenCourseWare  
7 years ago • 3,423,564 views  
Lecture 1: Goals of the course; what is computation; introduction to data types, operators, and variables Instructors: Prof. Prof. ...

**Question: How Important is Math in a Computer Science Degree?**  
Ell the Computer Guy Live  
1 year ago • 119,331 views  
I would like to know how hard it is the mathematics part in the computer science undergraduate course. I love computers and ...

**Computer Science Explained in less than 3 minutes**  
shawn demlane  
2 years ago • 257,833 views  
Computer Programming is an amazing field of compilation, amazement, difficulty, but above all, fun. Computer Programming ...

**Computer Science Tutor**  
77 videos  
CHANNEL | Subscribe | 6,009

**Vlog: What to expect in a Computer Science course**  
lcc0512  
1 year ago • 25,738 views  
Being pretty near graduation now, I decide that, by reflecting upon my own experience, answer some of the most commonly asked ...

••

**University of Southern California**

**USC Viterbi**  
School of Engineering

# YouTube Advanced Search Ranking Filters

- During a search YouTube provides filters for users to refine their search:
  - UPLOAD DATE
  - TYPE
  - DURATION
  - FEATURES
  - SORT BY

Copyright Ellis Horowitz, 2011-2022

17

••

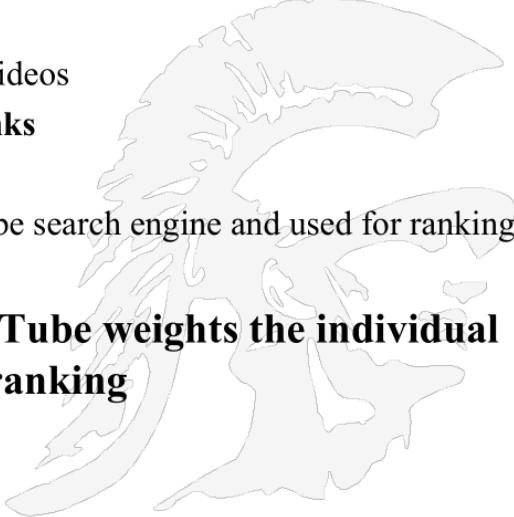
University of Southern California 

 USC **Viterbi**  
School of Engineering

# YouTube Ranking Factors

- YouTube uses the following metrics for ranking search results:
- 1. **Meta Data**
  - video titles, descriptions and tags are core ranking factors
  - include links to a website and social profiles
- 2. **Video Quality**
  - HD ranks higher than low quality videos
- 3. **Number of views, likes, shares and links**
- 4. **Subtitles and Closed Captions**
  - captions are crawled by the YouTube search engine and used for ranking

- **What is not known is how YouTube weights the individual factors to make up their final ranking**



Copyright Ellis Horowitz, 2011-2022

18

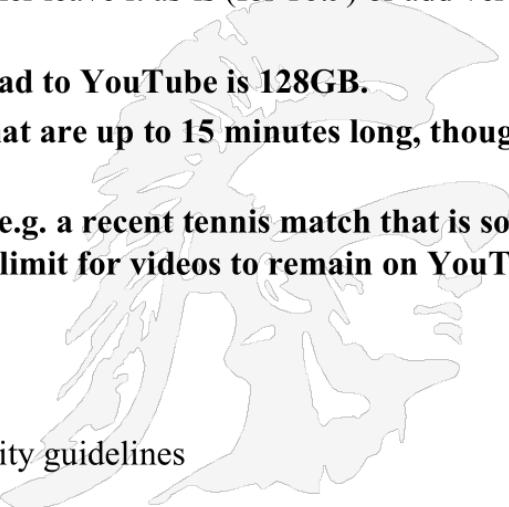
••

University of Southern California  USC

 USC **Viterbi**  
School of Engineering

## YouTube Upload Characteristics

- **YouTube Upload Characteristics**
  - YouTube supports 8 video formats for uploading: MOV, MP4 (MPEG4), AVI, WMV, FLV, 3GP, MPEGPS, WebM
  - **Aspect Ratio:** the standard aspect ratios are: 4:3 or 16:9. When the video is uploaded to the site, YouTube will either leave it as-is (for 16:9) or add vertical black bars (for 4:3)
  - **The maximum file size you can upload to YouTube is 128GB.**
  - **By default, you can upload videos that are up to 15 minutes long, though that can be extended**
  - **Many videos have a short life cycle, e.g. a recent tennis match that is soon forgotten, however, there is no time limit for videos to remain on YouTube, unless**
    - You delete the video.
    - You delete your account.
    - You violate copyright or community guidelines



Copyright Ellis Horowitz, 2011-2022

19

••

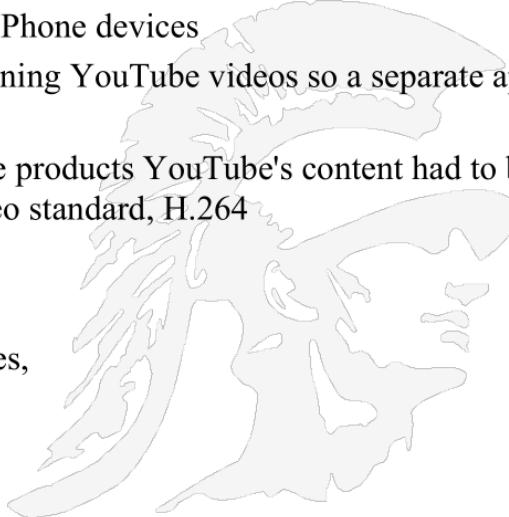
University of Southern California



USC **Viterbi**  
School of Engineering

## YouTube Videos Run On Multiple Platforms

- **Desktops/laptops**
  - Videos are played in your browser assuming it supports HTML5
  - This avoided the need to use Adobe Flash Player
- **Smartphones**
  - YouTube apps exist for Android and iPhone devices
    - There is no native support for running YouTube videos so a separate app is required
  - For YouTube's videos to run on Apple products YouTube's content had to be transcoded into Apple's preferred video standard, H.264
- **Other Devices**
  - Apple TV, Fire TV, iPod Touch,
  - TiVo, PlayStation, Wii Game consoles,
  - Xbox Live, Roku Players
  - Google Chromecast



Copyright Ellis Horowitz, 2011-2022

20

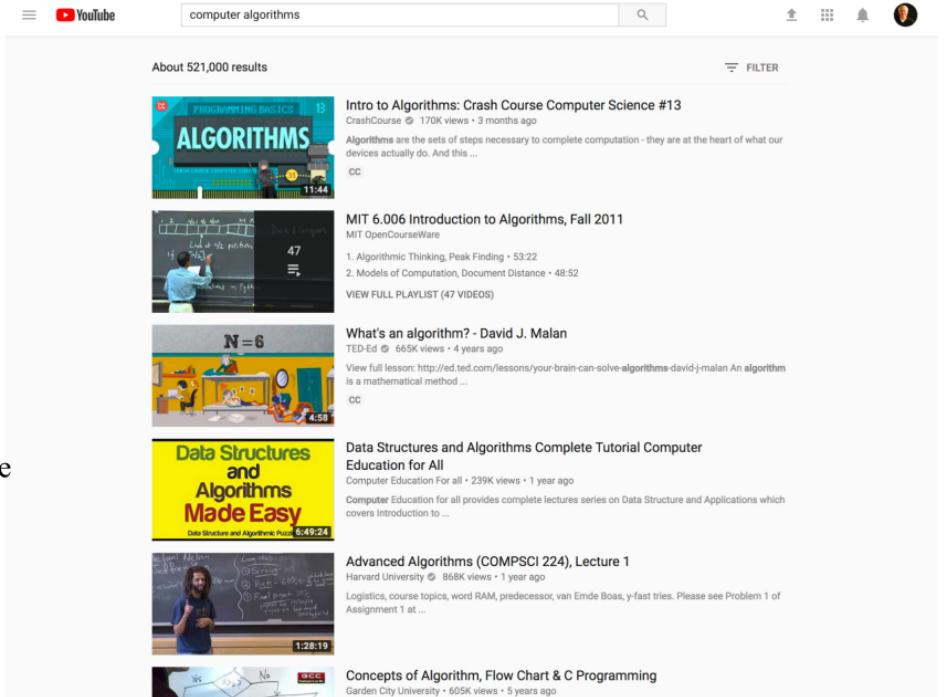
••

**University of Southern California**  

## YouTube Makes Recommendations to Retain Viewers

- **YouTube Search Results Example for query “computer algorithms”**
- **Assume we choose the first result**

Recommendations are made to maximize watch time



<https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n867596>

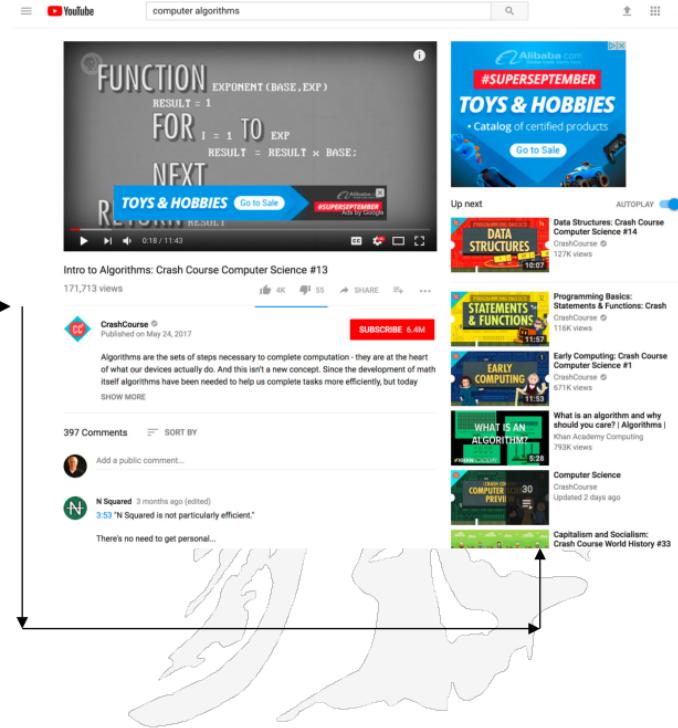
Copyright Ellis Horowitz, 2011-2022

21

••

**University of Southern California**   

## YouTube Recommendation Algorithm



- Given the query “computer algorithms” followed by a selection, YouTube makes recommendations for subsequent videos
- Recommendations account for 60% of all video clicks

Copyright Ellis Horowitz, 2011-2022

22

••

University of Southern California 

 USC **Viterbi** School of Engineering

## YouTube Recommendation System Uses Graph Properties

- Association Rule Mining
  - For each pair of videos  $v_i, v_j$  compute co-visitation counts, i.e. they count how often they were co-watched; if  $c_{i,j}$  is the co-visitation count, then relatedness is defined as

$$r(v_i, v_j) = \frac{c_{ij}}{f(v_i, v_j)}$$

where  $c_i$  and  $c_j$  are the total occurrence counts across all sessions for videos  $v_i$  and  $v_j$ .  $f(v_i, v_j)$  is a normalization function that takes the global popularity of both the seed video and the candidate video into account; e.g.  $f(v_i, v_j) = c_i * c_j$

The set of related videos,  $R_i$  for a given seed video  $v_i$  is determined by taking the top N candidate videos ranked by their scores  $r(v_i, v_j)$

Related videos induce a directed graph over the set of videos, namely:  
 For each pair of videos  $(v_i, v_j)$ , there is an edge  $e_{ij}$  from  $v_i$  to  $v_j$  iff  $v_j$  is in  $R_i$

For details see: *The YouTube Recommendation System*  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.9301&rep=rep1&type=pdf>

Copyright Ellis Horowitz, 2011-2022

23

Here is the paper referenced above.

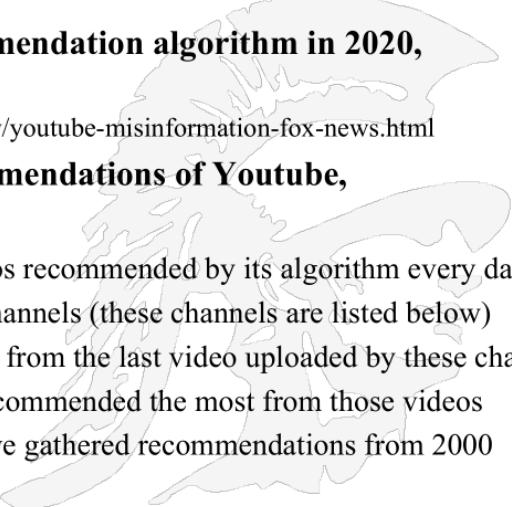
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## Media Sites (including YouTube) Move Away from False Information

- **YouTube's recommendation algorithm used to send people to misinformation, e.g. see**
  - <https://www.youtube.com/watch?v=FI8tFmBIPak> (3 min)
  - <https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478>
- **As a result YouTube changed its recommendation algorithm in 2020, eliminating so-called “fringe” sites**
  - <https://www.nytimes.com/2020/11/03/technology/youtube-misinformation-fox-news.html>
- **There is a website that tracks the recommendations of Youtube,**
- **<https://algotransparency.org/>**
- “We used a multi-step program to analyze videos recommended by its algorithm every day”
  - Step 1: We start from a list of 1000+ US channels (these channels are listed below)
  - Step 2: We gather all recommended videos from the last video uploaded by these channels
  - Step 3: We compute which channel was recommended the most from those videos
  - Step 4: We repeat step 2 and 3 until we have gathered recommendations from 2000 channels
  - Step 5: For each video that was observed, we count and display from how many channels it was recommended



Copyright Ellis Horowitz, 2011-2022

24

••

**University of Southern California**  

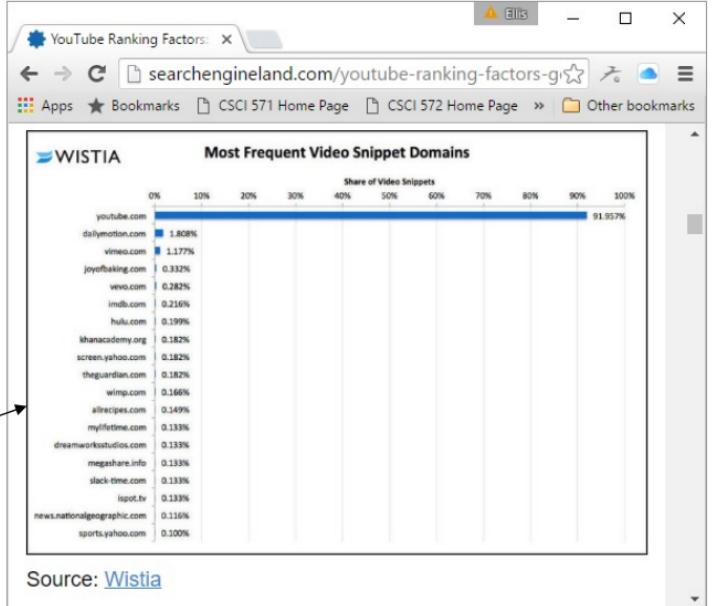
## Google Search is Biased Towards YouTube Videos

A **video rich snippet** means that when someone searches for something on Google, you can have a small tiny **video** show up next to your result to let the user know that particular result (yours) has a **video** to help

Google weeded out the video competition in Web search by predominantly displaying **only video-rich snippets** for YouTube videos back in 2014.

Here is a graph outlining the percentage share of video-rich snippets in Google; 91% are from YouTube

see  
<https://wistia.com/blog/where-did-my-video-snippets-go>



Domain	Share of Video Snippets
youtube.com	91.95%
dailymotion.com	1.86%
vimeo.com	1.17%
joyoffaking.com	0.33%
vevo.com	0.28%
imdb.com	0.21%
hulu.com	0.19%
khanacademy.org	0.18%
screen.yahoo.com	0.18%
theguardian.com	0.18%
wimp.com	0.16%
allrecipes.com	0.14%
mylifetime.com	0.13%
dreamworksstudios.com	0.13%
megashare.info	0.13%
slack-time.com	0.13%
ispot.tv	0.13%
news.nationalgeographic.com	0.11%
sports.yahoo.com	0.10%

Source: [Wistia](#)

e.g. try “tutorial on bitcoin”

Copyright Ellis Horowitz 2011-2022

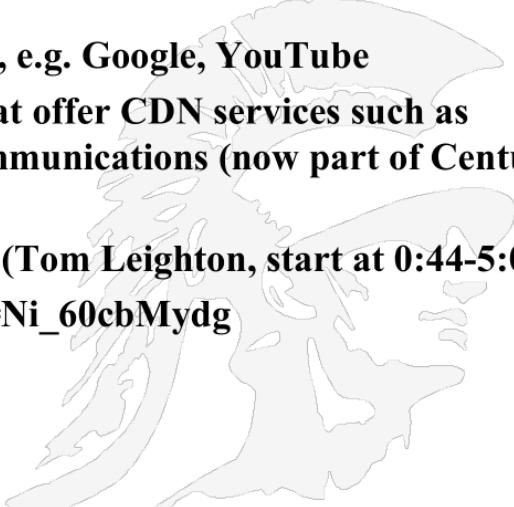
••

University of Southern California  USC

 USC **Viterbi**  
School of Engineering

## Content Delivery Networks

- A content distribution network (CDN) consists of a large set of content servers and a means for dynamically selecting servers based on knowledge of the location of the user and possibly the content being requested
- Some sights operate their own CDN, e.g. Google, YouTube
- There are third party companies that offer CDN services such as Akamai, Limelight and Level 3 Communications (now part of Century Link)
- See the Akamai video for 5 minutes (Tom Leighton, start at 0:44-5:00),
- [https://www.youtube.com/watch?v=Ni\\_60cbMydg](https://www.youtube.com/watch?v=Ni_60cbMydg)



Copyright Ellis Horowitz, 2011-2022

26

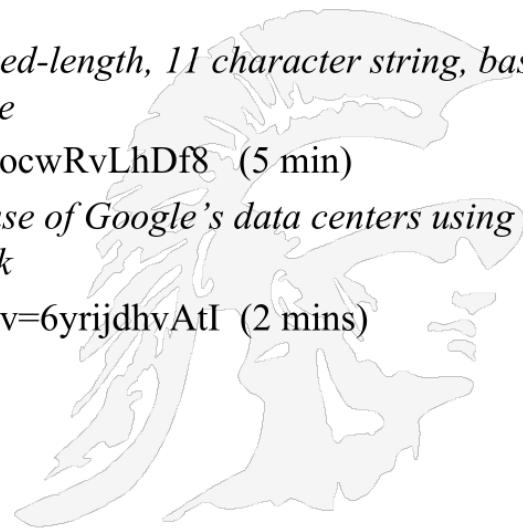
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## YouTube Video Delivery System

- **Two Critical Technology Challenges for YouTube:**
  - *how to identify billions of videos*
  - *How to efficiently deliver the video to the desktop/mobile device*
- **The Solutions:**
- **Identification:** *YouTube assigns a fixed-length, 11 character string, base 64, unique identifier to each video, see*
  - <https://www.youtube.com/watch?v=gocwRvLhDf8> (5 min)
- **Efficient Delivery:** *YouTube makes use of Google's data centers using them as a content distribution network*
  - <https://www.youtube.com/watch?v=6yrijdhvAtI> (2 mins)



Copyright Ellis Horowitz, 2011-2022

27

••



## YouTube (Google's) Content Delivery Datacenters

- A map of Google's data centers, see
- <https://www.google.com/about/datacenters/inside/locations/index.html>



**Figure 4: Geographical distribution of YouTube Video Cache Locations.**

Copyright Ellis Horowitz, 2011-2022

28

••

University of Southern California

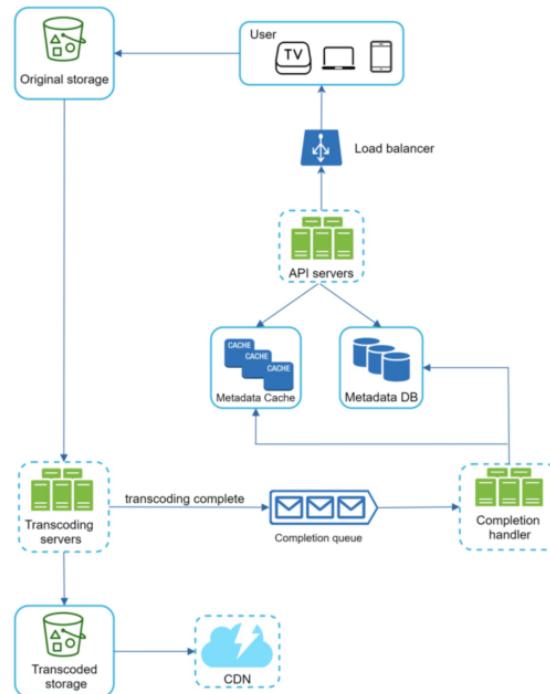


**USC Viterbi**  
School of Engineering

## Uploading a YouTube Video

1. videos are uploaded from a desktop to a central Data Center
2. the video is then transcoded into multiple formats
3. transcoded copies are sent to the Content Distribution Network

Video transcoding is a technique of converting a video into multiple different formats and resolutions to make it playable across different devices and bandwidths. The technique is also known as *video encoding*. This enables YouTube to stream videos in different resolutions such as *144p, 240p, 360p, 480p, 720p, 1080p & 4K*.



Copyright Ellis Horowitz 2011-2022

••

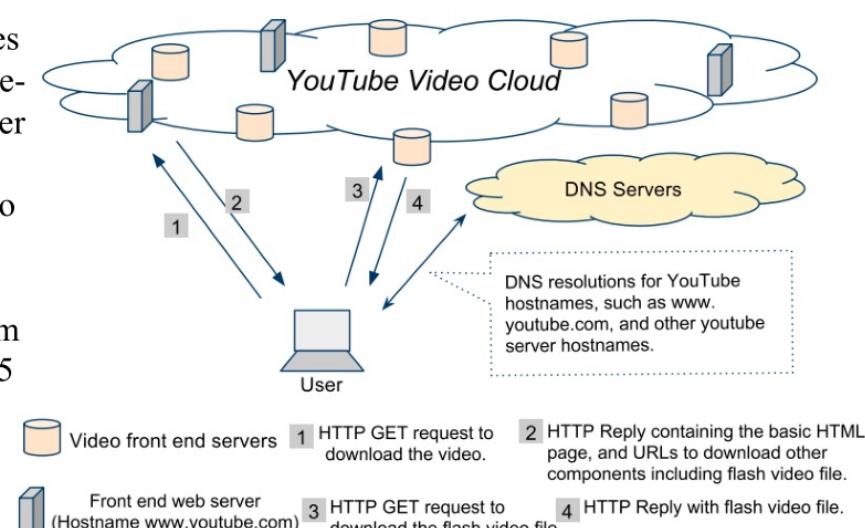
University of Southern California  USC 

## USC Viterbi YouTube's Content Distribution Network Downloading a YouTube Video

A local DNS server resolves www.youtube.com and is redirected to a YouTube server which downloads the page information and a pointer to a YouTube server that can deliver the video, e.g. v23.lscache5.c.youtube.com

The request to v23.lscache5

..  
may be further resolved



4 steps describing the delivery of a YouTube video

<http://www-users.cs.umn.edu/~zhzhang/Papers/youtube-tech-report.pdf>

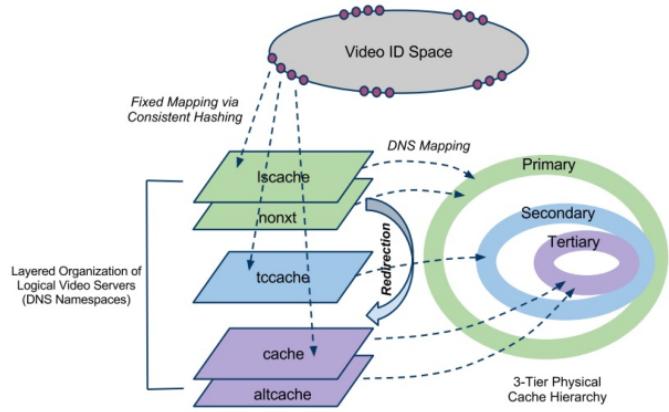
Copyright Ellis Horowitz 2011-2022

••

**University of Southern California**  

# YouTube Delivery System

- The design of the YouTube video delivery system consists of three components:
  - a “flat” video id space,
  - a multi-layered logical server organization consisting of five anycast namespaces (and two unicast namespaces), and
  - a 3-tiered physical cache hierarchy with (at least) 38 primary locations, 8 secondary and 5 tertiary locations.



The diagram illustrates the YouTube Architectural Design. At the top is the **Video ID Space**, represented by a grey oval containing numerous small dots. Below it is the **Layered Organization of Logical Video Servers (DNS Namespaces)**, which consists of four layers: **iscache** (green), **honxt** (light green), **tccache** (blue), and **cache** (purple). Arrows labeled **Fixed Mapping via Consistent Hashing** point from the Video ID Space to each layer. To the right is the **3-Tier Physical Cache Hierarchy**, shown as three concentric circles labeled **Primary**, **Secondary**, and **Tertiary**. A curved arrow labeled **DNS Mapping** points from the logical servers to the Primary tier. A curved arrow labeled **Redirection** points from the Primary tier to the Secondary tier.

**Figure 3: YouTube Architectural Design.**

<https://www-users.cse.umn.edu/~zhang089/Papers/youtube-tech-report.pdf>

••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## References to YouTube's CDN

- There are four research papers that investigated and discussed the YouTube CDN, they are:
  1. *Vivisecting YouTube: An Active Measurement Study*, 2012, cited by Jefay
  2. *Dissecting Video Server Selection Strategies in the YouTube CDN*, 2011, cited by Jefay
  3. *YouTube Traffic Dynamics and Its InterPlay with a Tier-1 ISP*, 2010
  4. <https://www-users.cse.umn.edu/~zhang089/Papers/youtube-tech-report.pdf>
- All of the papers describe a complicated re-direction scheme to find the nearest data center to serve the video; they attempt to minimize Round Trip Time or RTT
- For rarely-called-for videos the “*Dissecting*” paper did a study requesting in California a rare video and observed that the first request came from the Netherlands, but future requests were served from California
- - Conclusion: videos are constantly being moved around to be closer to the place that is requesting them

Copyright Ellis Horowitz, 2011-2022

32

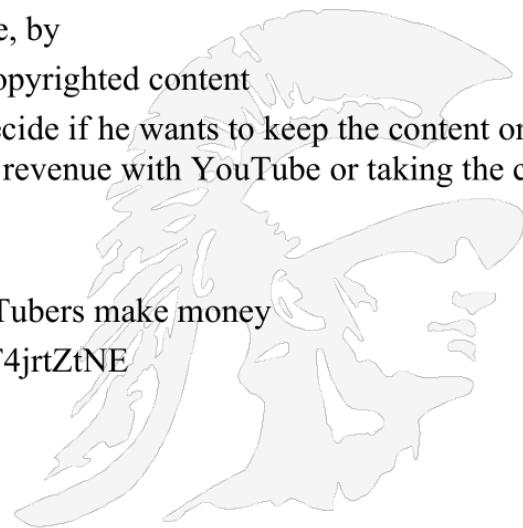
••

University of Southern California  USC

 USC Viterbi  
School of Engineering

## Monetizing YouTube

- **YouTube challenges in the early days**
  - YouTube had no way of making money and its infrastructure is very expensive
  - YouTube was being sued by content creators as many of YouTube's videos were uploaded illegally
  - YouTube **solved both problems** at once, by
    - Developing a system for spotting copyrighted content
    - Allowing the copyright owner to decide if he wants to keep the content on the site and let ads appear, splitting the revenue with YouTube or taking the content down
  - Here is a video that describes how YouTubers make money
  - <https://www.youtube.com/watch?v=v8F4jrtZtNE>
  - (8 min)



Copyright Ellis Horowitz, 2011-2022

33

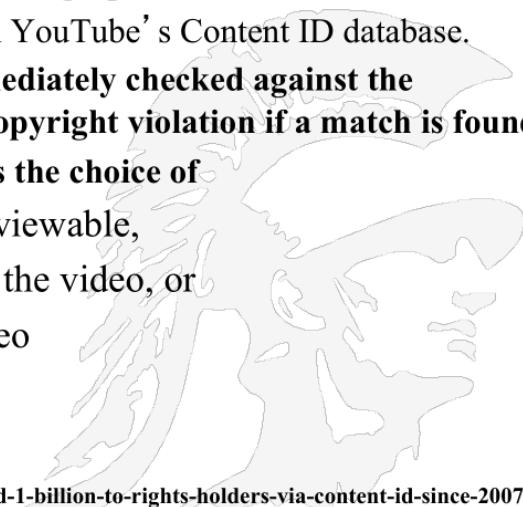
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

## ContentID

- YouTube's solution was to create a fingerprint database of copyrighted content, called Content ID
- YouTube solicited cooperation from content owners asking them to submit copies of their content so YouTube could fingerprint them
  - There are millions of reference files in YouTube's Content ID database.
- When a new video is uploaded, it is immediately checked against the database, and the video is flagged as a copyright violation if a match is found.
- When this occurs, the content owner has the choice of
  1. blocking the video to make it unviewable,
  2. tracking the viewing statistics of the video, or
  3. adding advertisements to the video



<https://arstechnica.com/tech-policy-policy/2014/10/youtube-has-paid-1-billion-to-rights-holders-via-content-id-since-2007/>

Copyright Ellis Horowitz, 2011-2022

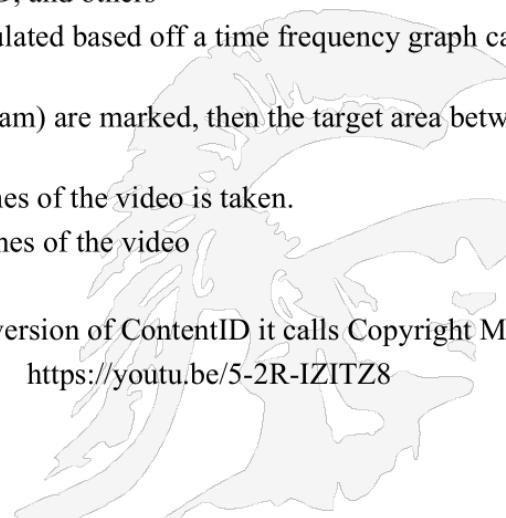
34

••



## More Details on ContentID

1. Content ID is based off audio and video samples that rights holders have uploaded to YouTube
  2. User uploads a video.
  3. YouTube then queues up the video to be processed i.e. it is transcoded into multiple formats including:
    - HTML5, H.264, WebM VP8, HD, non-HD, and others
  4. *If the video contains audio*, a hash is then calculated based off a time frequency graph called a spectrogram.
    - Target zones (peak points in the spectrogram) are marked, then the target area between them is also taken and hashed
  5. *For the video portion*, a sample section of frames of the video is taken.
    - A hash is created from those sampled frames of the video
- Note recently YouTube has introduced a new version of ContentID it calls Copyright Match
  - See the following videos for details, (2 min). <https://youtu.be/5-2R-IZITZ8>

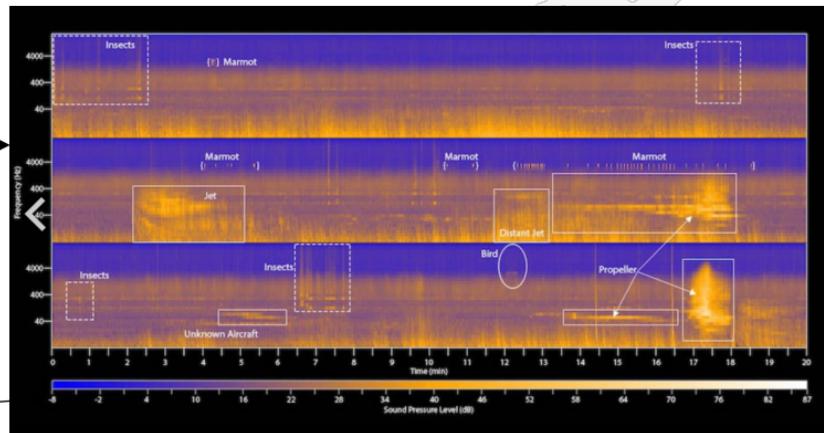


••

University of Southern California  USC

## Creating an Acoustic Fingerprint

- The audio signal is digitized and converted to a spectrogram – a time-frequency graph**
  - The graph below plots three dimensions of audio: frequency versus amplitude versus time
  - A common format is a graph with two dimensions: one axis represents time, and the other axis represents frequency; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or color of each point in the image.



frequency →

Time axis →

36

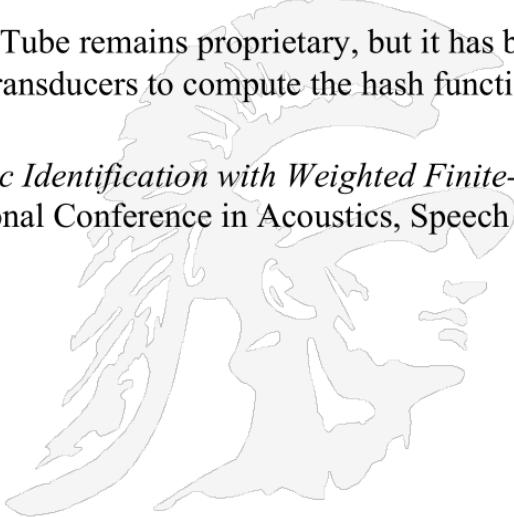
••

University of Southern California 

 USC **Viterbi**  
School of Engineering

# How Good is Content ID

- According to stats released by YouTube **99.5 percent** of all copyright issues specifically related to sound recordings are automatically resolved by Content ID
- In addition to music, Content ID also identifies 98% of copyright claims, including those tied to film, TV, gaming
- The actual hashing algorithm used by YouTube remains proprietary, but it has been suggested that YouTube uses finite-state transducers to compute the hash function, e.g. see
- Eugene Weinstein, Pedro J. Moreno; *Music Identification with Weighted Finite-State Transducers*, Proceedings of the International Conference in Acoustics, Speech and Signal Processing (ICASSP), 2007



Copyright Ellis Horowitz, 2011-2022

37

••

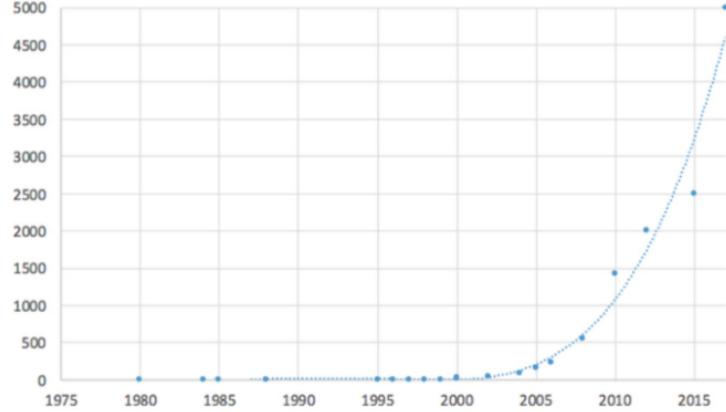
University of Southern California 

 USC **Viterbi**  
School of Engineering

## Will YouTube Ever Run Out of Storage

- The storage you can buy with \$100 has grown exponentially — or equivalently, the cost of storing 1GB of videos has decreased exponentially

Gigabytes of storage you can buy with \$100



Year	Gigabytes of storage (\$100)
1980	~10
1985	~10
1990	~10
1995	~10
2000	~10
2005	~100
2010	~1500
2012	~2000
2014	~2500
2015	~3000
2016	~4500
2017	~5000

Kryder's Law  
[https://en.wikipedia.org/wiki/Mark\\_Kryder#Kryder%27s\\_law\\_projection](https://en.wikipedia.org/wiki/Mark_Kryder#Kryder%27s_law_projection)



Copyright Ellis Horowitz, 2011-2022

38

••

**University of Southern California**  **USC**

**USC Viterbi**  
School of Engineering <https://www.quora.com/What-is-the-total-capacity-of-YouTube-storage>

**Total Capacity of YouTube Storage**

 1280x720	mp4	 download - 59.01 MB	
 640x360	mp4	 download - 15.34 MB	
 640x360	webm	 download - 19.07 MB	
 400x240	flv	 download - 8.51 MB	
 320x240	3gp	 download - 5.94 MB	
 176x144	3gp	 download - 2.12 MB	
 4k (no audio)	mp4	 download - 297.69 MB	

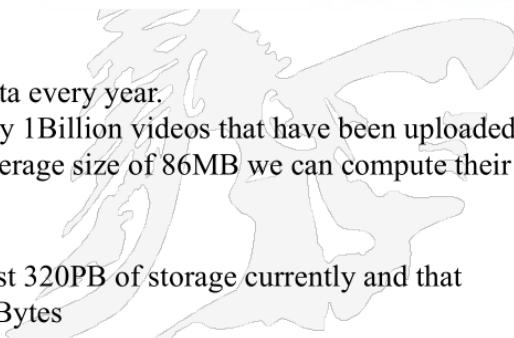
**24TB \* 4x (for profiles) \* 365 days = 35PB/year**

So YouTube needs to store roughly 35PB of new data every year.

From multiple sources we know that there is roughly 1Billion videos that have been uploaded to YouTube to date. Assuming each video has an average size of 86MB we can compute their total storage needs as:

**86MB \* 4 (for profiles) \* 1,000,000,000 = 320PB**

So it is estimated that YouTube needs to have at least 320PB of storage currently and that the storage needs are growing each year by 35 PetaBytes



Copyright Ellis Horowitz, 2011-2022

39