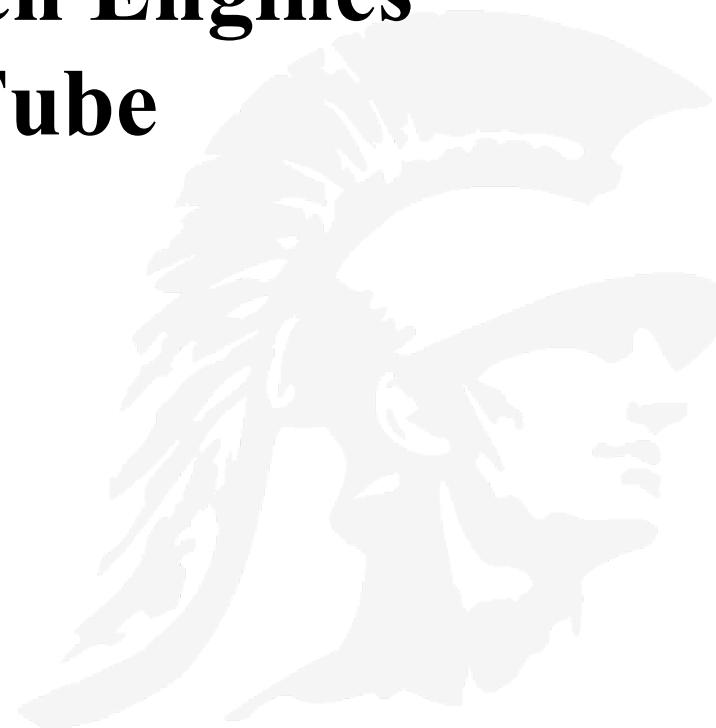


# Video Search Engines YouTube



# Video Search Engines

- A **video search engine** is a web-based search engine which crawls the web for video content.
  - Some video search engines use crawls to locate videos, while others allow content to be uploaded and hosted on their own servers, e.g. YouTube
- 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

# Some Video Search Engines

- ***blinkx*** (renamed as RhythmOne) was launched in 2004 and uses speech recognition and visual analysis to process spidered video rather than rely on metadata alone. *blinkx* claims to have the largest archive of video on the web and puts its collection at around 26,000,000 hours of content
  - <http://www.blinkxtv.com/> now goes to *360Daily.com*
- ***Picsearch Video Search*** has been licensed to search portals since 2006. Picsearch is a search technology provider who powers image, video and audio search for over 100 major search engines around the world
  - <https://www.picsearch.com/>
- ***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  - ***No longer active***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 ***no longer active***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  - ***No longer active***
- ***Check out <https://www.bing.com/videos/>***

## Some Video Hosting Sites

- 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
- Video search engines such as Google, Bing and others include videos from all these websites, but do not actually host the content
- Most of these websites which host video allow their videos to be embedded on other websites
- Oh yes, **youtube.com**, and yes, it is legal to crawl YouTube, see section 4.H of their Terms of Service, <https://www.youtube.com/static?template=terms>

# Video Streaming Services

- **Hulu** is an America subscription video on demand service jointly owned by Walt Disney, 21<sup>st</sup> Century Fox, Comcast, and Time Warner
  - In December 2017, Disney will acquire Fox's partial ownership, giving it a majority stake when the deal closes
  - It is primarily oriented towards instant streaming of television series', carrying current and past episodes of many series from its owners' respective television networks and other content partners
- **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

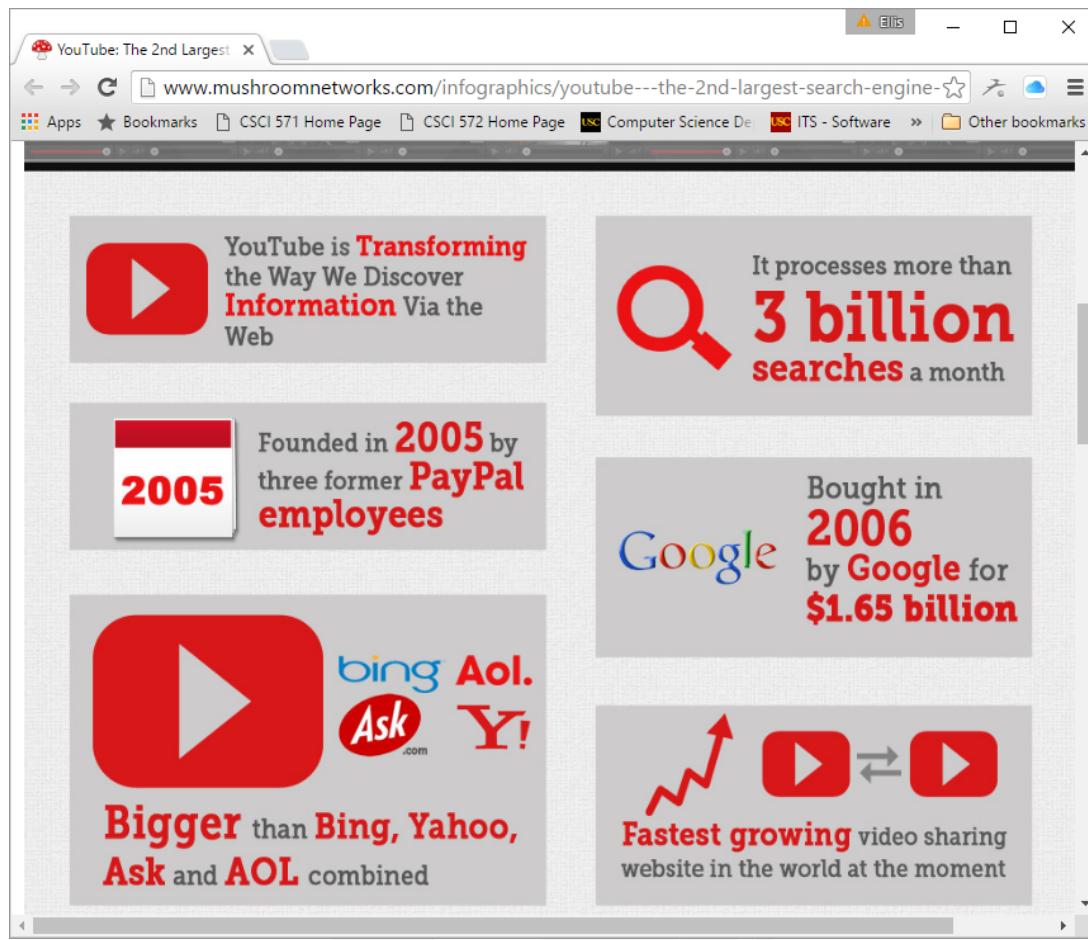
- **Entertainment**

# 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
- The site allows users to upload, view, rate, share, add to favorites, report and comment on videos
- YouTube makes use of WebM, H.264/MPEG-4, AVC and Adobe Flash to display videos
- Most of the content on YouTube has been uploaded by individuals, but media corporations including CBS, BBC, Vevo and Hulu offer some of their material via YouTube as part of the YouTube partnership program
- In December 2016, the website was ranked as the second most popular site by Alexa Internet, a web traffic analysis company.[\[1\]](#)

# 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>

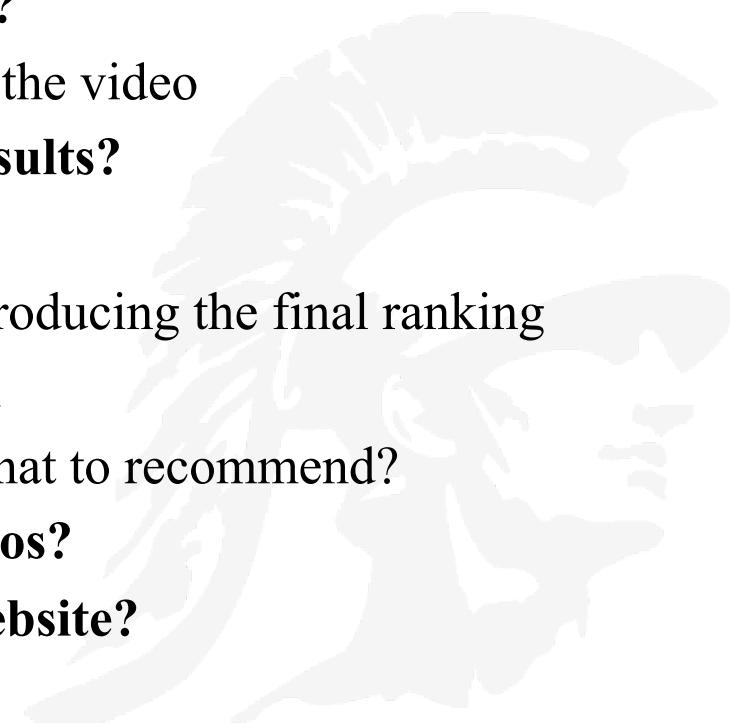


# YouTube Traffic - Some Facts

- **As of 2018:**
  - **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**

# YouTube Search Engine Issues to Consider

- **What video formats are acceptable?**
  - For Uploading
  - For Downloading (the YouTube players: desktop, iPhone, android)
- **How does YouTube identify videos?**
  - What does YouTube know about the video
- **How does YouTube rank search results?**
  - What factors are used
  - How are the factors weighed in producing the final ranking
- **YouTube Recommendation System**
  - How does YouTube determine what to recommend?
- **How Does YouTube distribute videos?**
- **How does YouTube monetize its website?**
  - YouTube's ContentID system



# YouTube is Not a Search Engine in the Conventional Sense

- **YouTube does not crawl the web for videos**
    - Its content comes to YouTube, how nice!
  - **YouTube only searches the content that has been uploaded**
  - **YouTube only returns videos, not web pages or pdf files**
  - **Perhaps YouTube is best thought of as a *content aggregator***
- 
- ***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 09/2018, there are 2249 YouTubers with over a million subscribers*

# YouTube Gathers Information When Videos are Uploaded

YouTube captures:

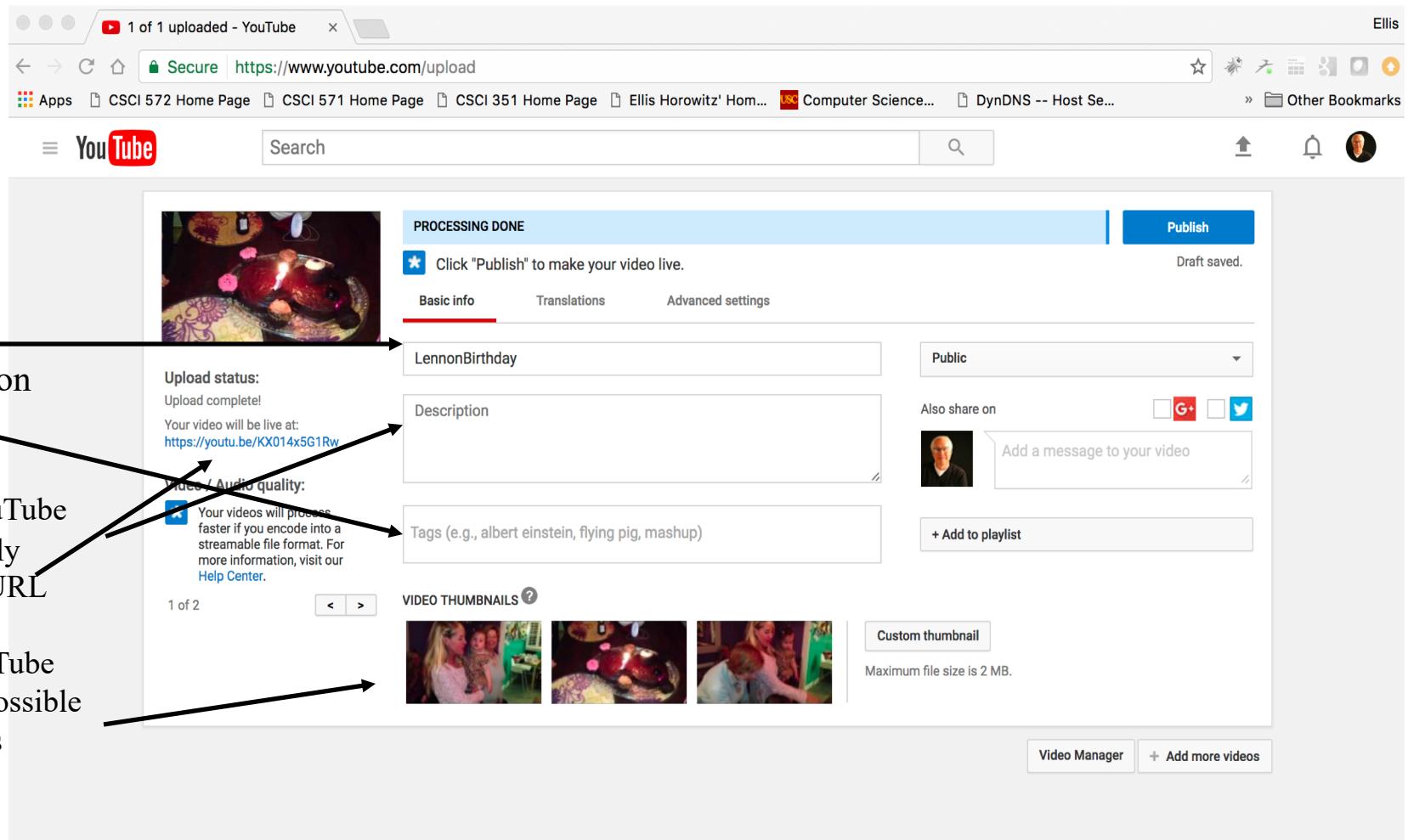
Name \_\_\_\_\_

Description \_\_\_\_\_

Tags \_\_\_\_\_

Note: YouTube immediately assigns a URL

Note: YouTube suggests possible thumbnails

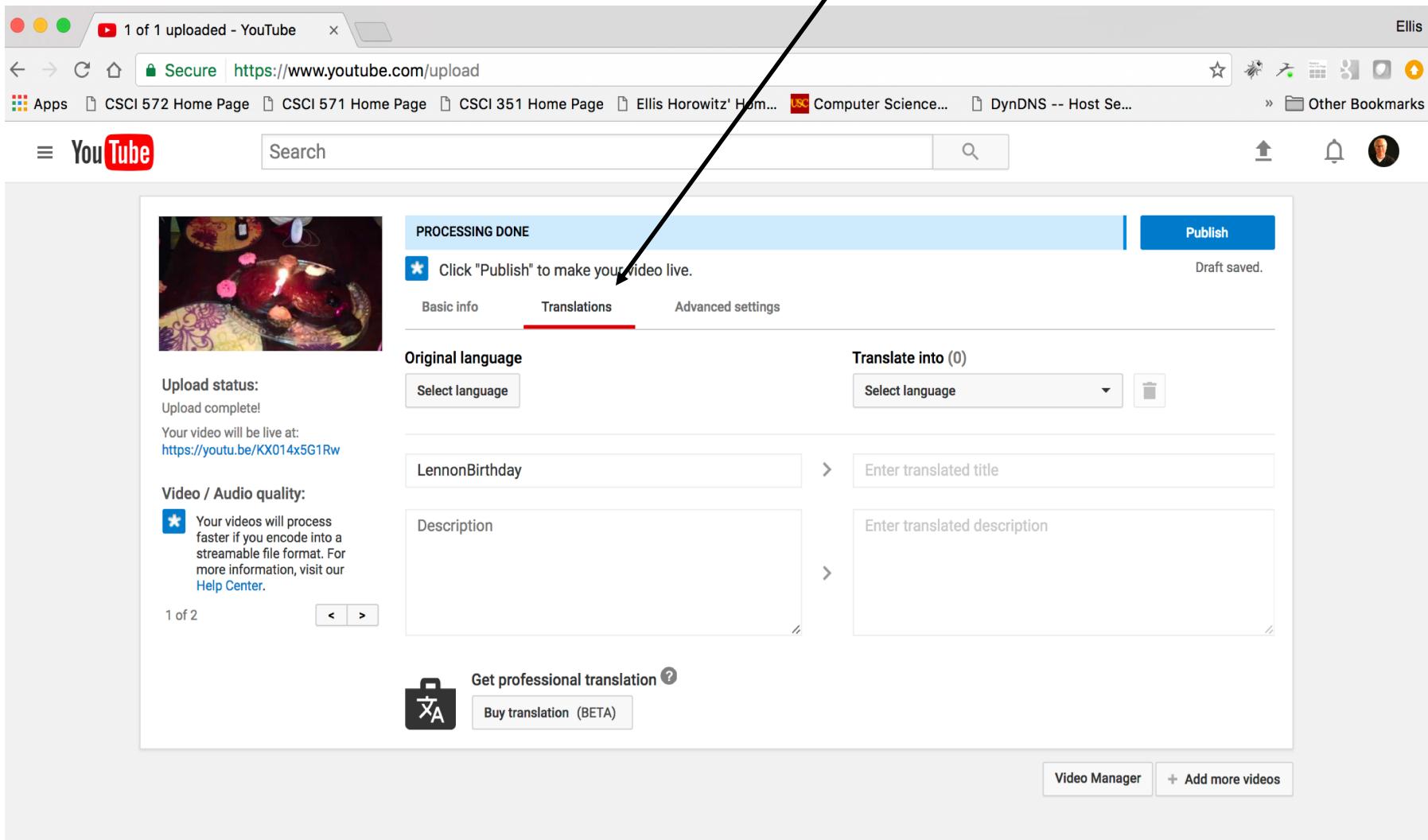


The screenshot shows the YouTube upload page after a video has been uploaded. The status bar at the top says "1 of 1 uploaded - YouTube". The main area displays the following information:

- PROCESSING DONE**: Click "Publish" to make your video live.
- Basic info** (selected): LennonBirthday
- Translations**
- Advanced settings**
- Upload status:** Upload complete!
- Your video will be live at:** <https://youtu.be/KX014x5G1Rw>
- Description**: (empty)
- Tags (e.g., albert einstein, flying pig, mashup)**: (empty)
- Visibility**: Public
- Also share on**: G+ (unchecked), Twitter (unchecked)
- Thumbnail**: Shows three suggested thumbnails from the video and a "Custom thumbnail" button.
- Video Manager** and **+ Add more videos** buttons.

Annotations with arrows point to the "Name" field, the "Description" field, the "Tags" field, the "Note: YouTube immediately assigns a URL" note, and the "Note: YouTube suggests possible thumbnails" note.

# Uploading to YouTube 2nd of 3 Input Screens

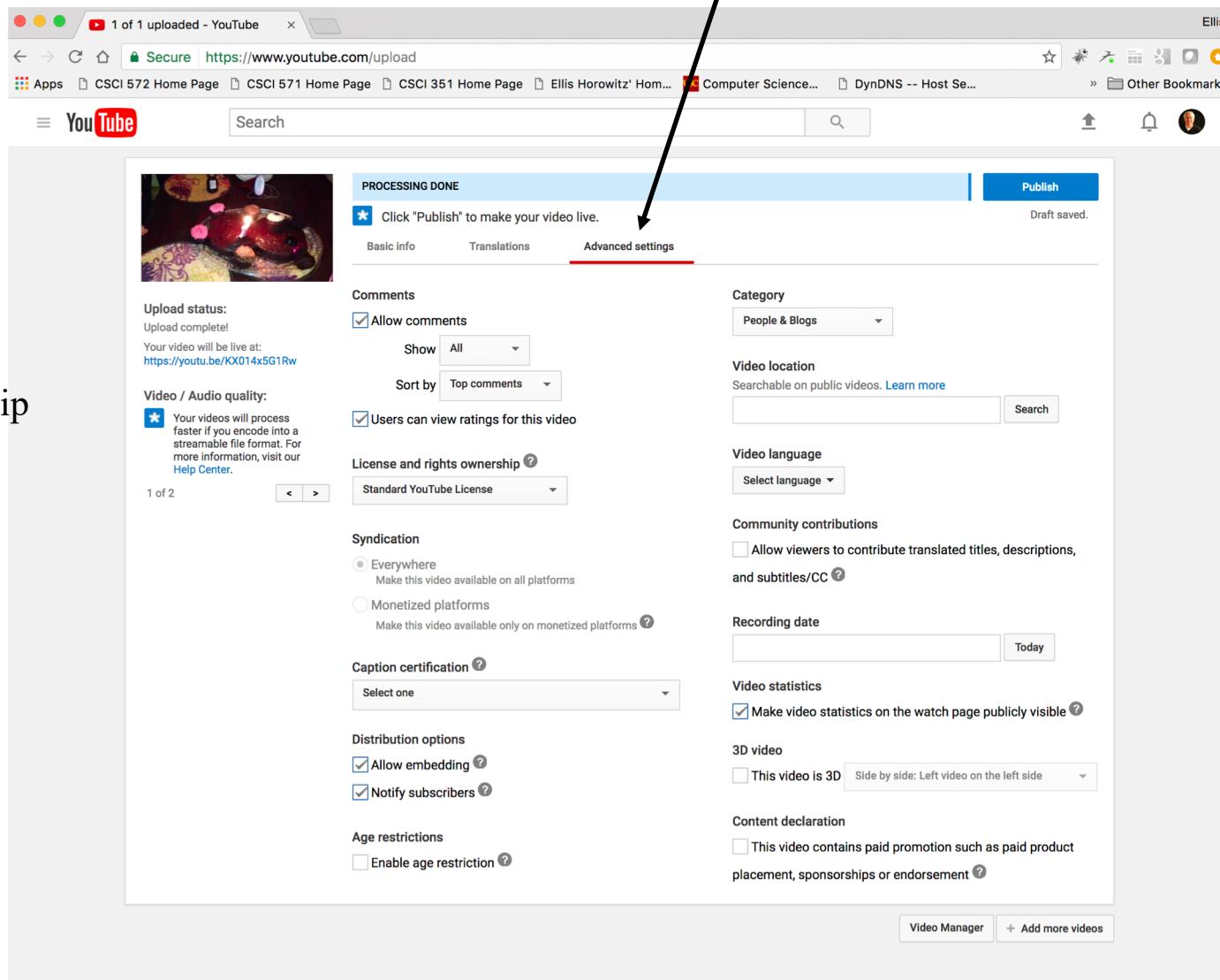


The screenshot shows the YouTube upload interface. At the top, a banner indicates "1 of 1 uploaded - YouTube". The URL in the address bar is <https://www.youtube.com/upload>. The main content area shows a thumbnail of a video featuring a lit candle on a decorated plate. A message says "PROCESSING DONE" and "Click 'Publish' to make your video live." Below this, tabs for "Basic info", "Translations", and "Advanced settings" are visible, with "Translations" being the active tab. On the left, "Upload status" shows "Upload complete!" and the video will be live at <https://youtu.be/KX014x5G1Rw>. Under "Video / Audio quality", a note says videos will process faster if encoded into a streamable file format. The "Translations" section allows selecting languages for title and description. At the bottom, there's a "Get professional translation" button and a "Buy translation (BETA)" link.

# Uploading to YouTube Third Input Screen

YouTube allows  
the creator to  
specify:

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



The screenshot shows the YouTube upload interface after processing is done. A large arrow points from the top right towards the "Advanced settings" tab, which is currently selected. The interface includes sections for Comments, License and rights ownership, Syndication, Caption certification, Distribution options, Age restrictions, and various optional settings like 3D video and content declaration.

**PROCESSING DONE**  
Click "Publish" to make your video live.

**Comments**  
 Allow comments  
Show All  
Sort by Top comments  
 Users can view ratings for this video

**Category**  
People & Blogs

**Video location**  
Searchable on public videos. [Learn more](#)

**Video language**  
Select language

**Community contributions**  
 Allow viewers to contribute translated titles, descriptions, and subtitles/CC

**Recording date**  
Today

**Video statistics**  
 Make video statistics on the watch page publicly visible

**3D video**  
 This video is 3D Side by side: Left video on the left side

**Content declaration**  
 This video contains paid promotion such as paid product placement, sponsorships or endorsement

**Basic info**   **Translations**   **Advanced settings**

1 of 2

Allow embedding  
Notify subscribers

Enable age restriction

Video Manager   + Add more videos

# Ads, Ads, Ads

## Sample YouTube Search Results for Katy Perry

First result is an Ad

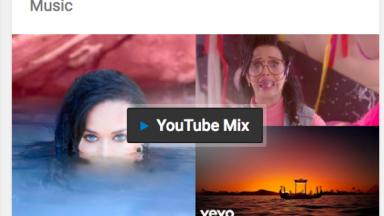
2<sup>nd</sup> and 4<sup>th</sup> results are stored at Vevo

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

YouTube katy perry

About 17,000,000 results

	<b>Lipstick By Katy Perry</b> <b>Ad</b> COVERGIRL 3,094,672 views Get Exclusive Katy Kat Lipstick For Pillow Soft, Rich Matte Lips Now!	
	<b>Katy Perry - Chained To The Rhythm (Lyric Video) ft. Skip Marley</b> KatyPerryVEVO 2 days ago • 10,827,898 views Get "Chained to the Rhythm": <a href="http://katy.to/cttr">http://katy.to/cttr</a> Directed by: Aya Tanimura Executive Producer: Danny Lockwood Produced by: ... NEW	<b>Katy Perry</b> Music 
	<b>KatyPerryVEVO</b> 106 videos CHANNEL Subscribe 20,804,904	<b>Top Tracks</b> 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
	<b>Katy Perry - Roar (Official)</b> KatyPerryVEVO 3 years ago • 1,790,071,602 views Get "Roar" from Katy Perry's 'PRISM': <a href="http://smarturl.it/PRISM">http://smarturl.it/PRISM</a> Official music video for Katy Perry's "Roar" brought to you in ... CC	<b>Related Artists</b>  Taylor Swift  Lady Gaga  Russell Brand
	<b>Katy Perry - Topic</b> 433 videos Katheryn Elizabeth Hudson, known professionally as Katy Perry, is an American singer and songwriter. After singing in church ... CHANNEL Subscribe 297,603	

# 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

YouTube computer science

About 18,600,000 results

**Technology For Students**  
Ad 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 ...  
CC

**Computer Science a good major?**  
ENGINEERED TRUTH  
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  
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. CC

**Question: How Important is Math in a Computer Science Degree?**  
Eli 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 ...

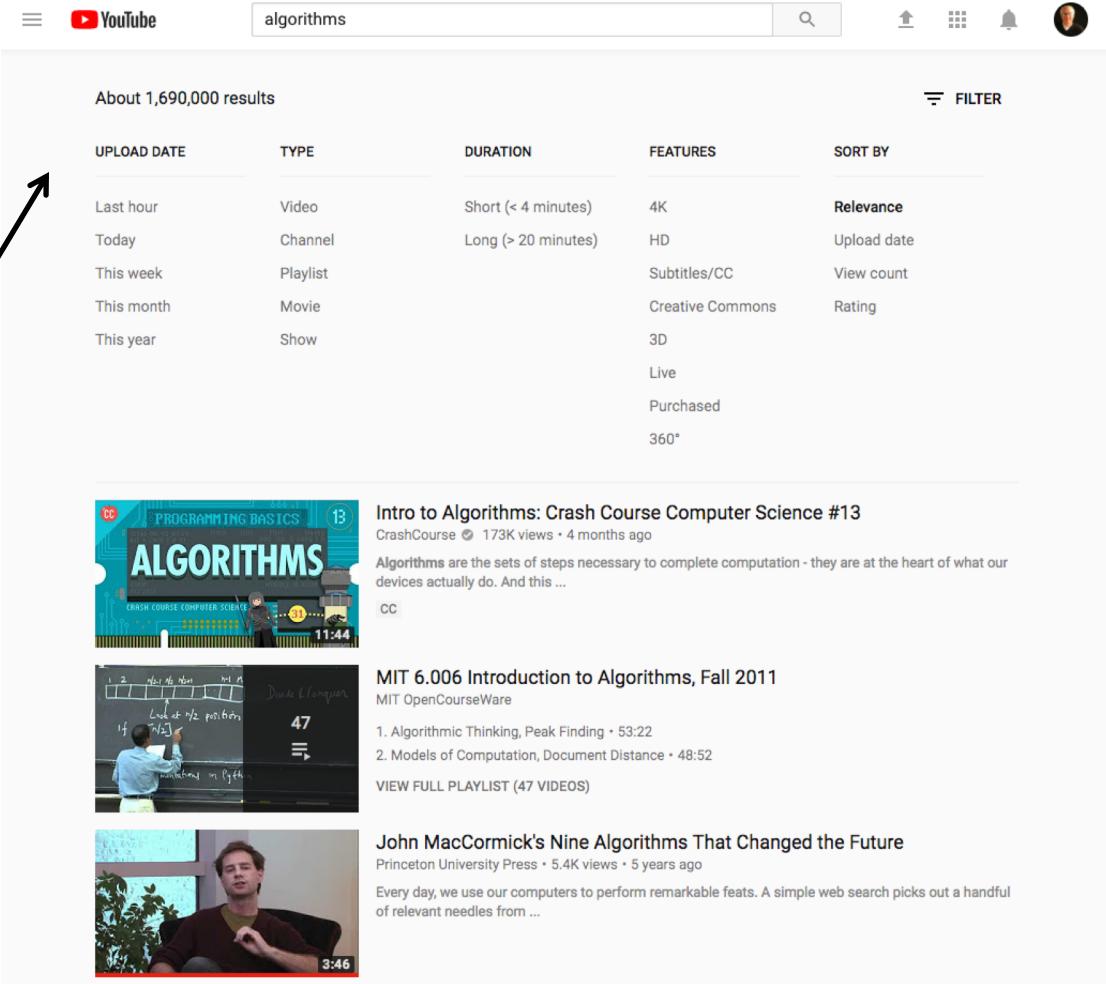
**Computer Science Explained in less than 3 minutes**  
shaun diem-lane  
2 years ago • 257,833 views  
Computer Programming is an amazing field of complication, amazement, difficulty, but above all, fun. Computer Programming ...

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

**Vlog: What to expect in a Computer Science course**  
lcc0612  
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 ...

# YouTube Channels and Filters

- YouTube lets users create a channel, which serves as the home page for the user's account.
- During a search YouTube provides filters for users to refine their search:
  - UPLOAD DATE
  - TYPE
  - DURATION
  - FEATURES
  - SORT BY



The screenshot shows a YouTube search results page for the query "algorithms". At the top, there is a search bar with the text "algorithms" and a magnifying glass icon. To the right of the search bar are several icons: a camera, a filter, a search icon, a refresh icon, a bell, and a profile picture. Below the search bar, the text "About 1,690,000 results" is displayed, followed by a "FILTER" button. A large black arrow points from the word "SORT BY" in the list above to the "FILTER" button on the search results page.

UPLOAD DATE	TYPE	DURATION	FEATURES	SORT BY
Last hour	Video	Short (< 4 minutes)	4K	<b>Relevance</b>
Today	Channel	Long (> 20 minutes)	HD	Upload date
This week	Playlist		Subtitles/CC	View count
This month	Movie		Creative Commons	Rating
This year	Show		3D	
			Live	
			Purchased	
			360°	

**Intro to Algorithms: Crash Course Computer Science #13**  
 CrashCourse • 173K views • 4 months ago  
 Algorithms are the sets of steps necessary to complete computation - they are at the heart of what our devices actually do. And this ...  
 CC

**MIT 6.006 Introduction to Algorithms, Fall 2011**  
 MIT OpenCourseWare  
 1. Algorithmic Thinking, Peak Finding • 53:22  
 2. Models of Computation, Document Distance • 48:52  
[VIEW FULL PLAYLIST \(47 VIDEOS\)](#)

**John MacCormick's Nine Algorithms That Changed the Future**  
 Princeton University Press • 5.4K views • 5 years ago  
 Every day, we use our computers to perform remarkable feats. A simple web search picks out a handful of relevant needles from ...

# 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 can be 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

# Caveats and Facts

- **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 by visiting <https://www.youtube.com/verify>
- **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

# YouTube Videos Run On Multiple Platforms

- In July 2010, the mobile version of the site was relaunched based on HTML5 avoiding the need to use Adobe Flash Player
- Most modern smartphones are capable of accessing YouTube videos
  - YouTube apps exist for Android and iPhone devices
  - For YouTube's videos to run on Apple products YouTube's content had to be transcoded into Apple's preferred video standard, H.264
- YouTube videos also can be viewed on devices including
  - Apple TV, iPod Touch,
  - TiVo, PlayStation, Wii Game consoles,
  - Xbox Live, and other devices

# YouTube Makes Recommendations to Retain Viewers

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

computer algorithms

About 521,000 results

**Intro to Algorithms: Crash Course Computer Science #13**  
CrashCourse 170K views • 3 months ago  
Algorithms are the sets of steps necessary to complete computation - they are at the heart of what our devices actually do. And this ...  
CC

**MIT 6.006 Introduction to Algorithms, Fall 2011**  
MIT OpenCourseWare  
1. Algorithmic Thinking, Peak Finding • 53:22  
2. Models of Computation, Document Distance • 48:52  
VIEW FULL PLAYLIST (47 VIDEOS)

**What's an algorithm? - David J. Malan**  
TED-Ed 665K views • 4 years ago  
View full lesson: <http://ed.ted.com/lessons/your-brain-can-solve-algorithms-david-j-malan> An algorithm is a mathematical method ...  
CC

**Data Structures and Algorithms Made Easy**  
Data Structure and Algorithmic Puzzles 6:49:24  
Computer Education for All 239K views • 1 year ago  
Computer Education for all provides complete lectures series on Data Structure and Applications which covers Introduction to ...

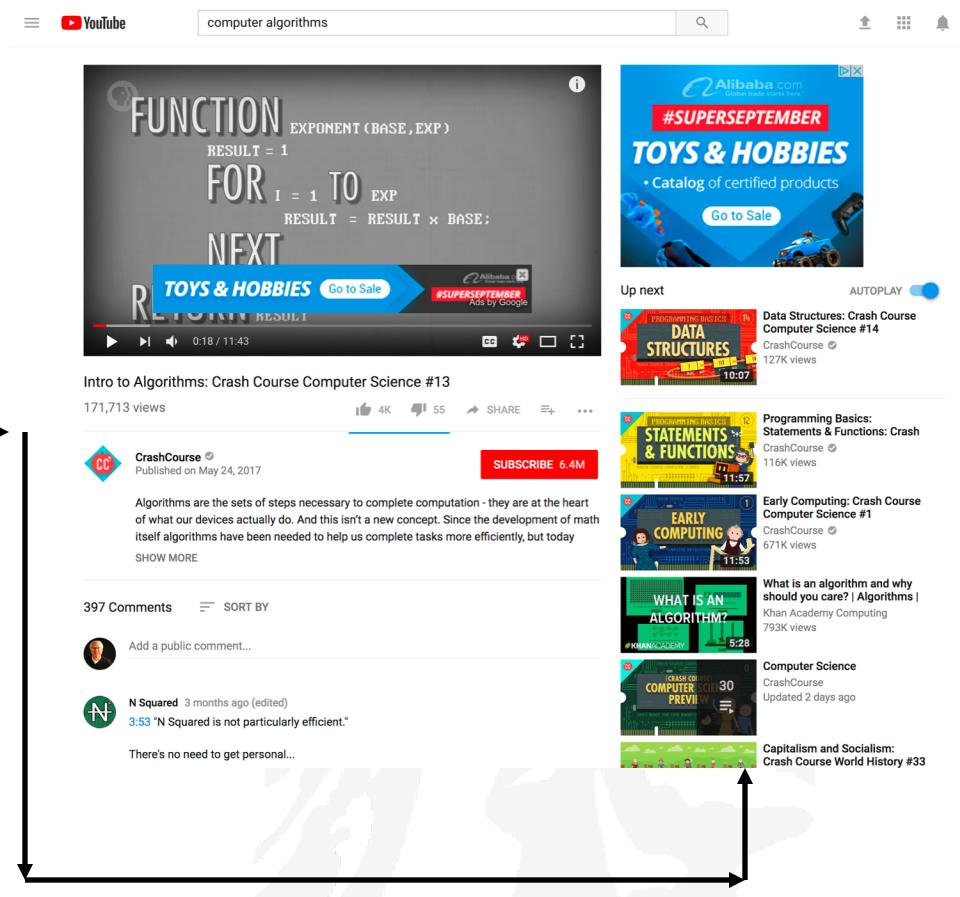
**Advanced Algorithms (COMPSCI 224), Lecture 1**  
Harvard University 868K views • 1 year ago  
Logistics, course topics, word RAM, predecessor, van Emde Boas, y-fast tries. Please see Problem 1 of Assignment 1 at ...

**Concepts of Algorithm, Flow Chart & C Programming**  
Garden City University 605K views • 5 years ago

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

# 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



# YouTube Recommendation System

- Since videos are constantly being updated and videos may quickly go viral, the recommendation system must constantly be updated
- Elements that are considered are the following:
  - First, a user's personal activity is initially determined (watched, favorited, liked, people he subscribes to, videos he watched to the end)
  - Next, an expanding set of videos is produced by traversing the co-visitation graph (other videos people watched after watching a given video) – see the next slide

# YouTube Recommendation System

- **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>

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

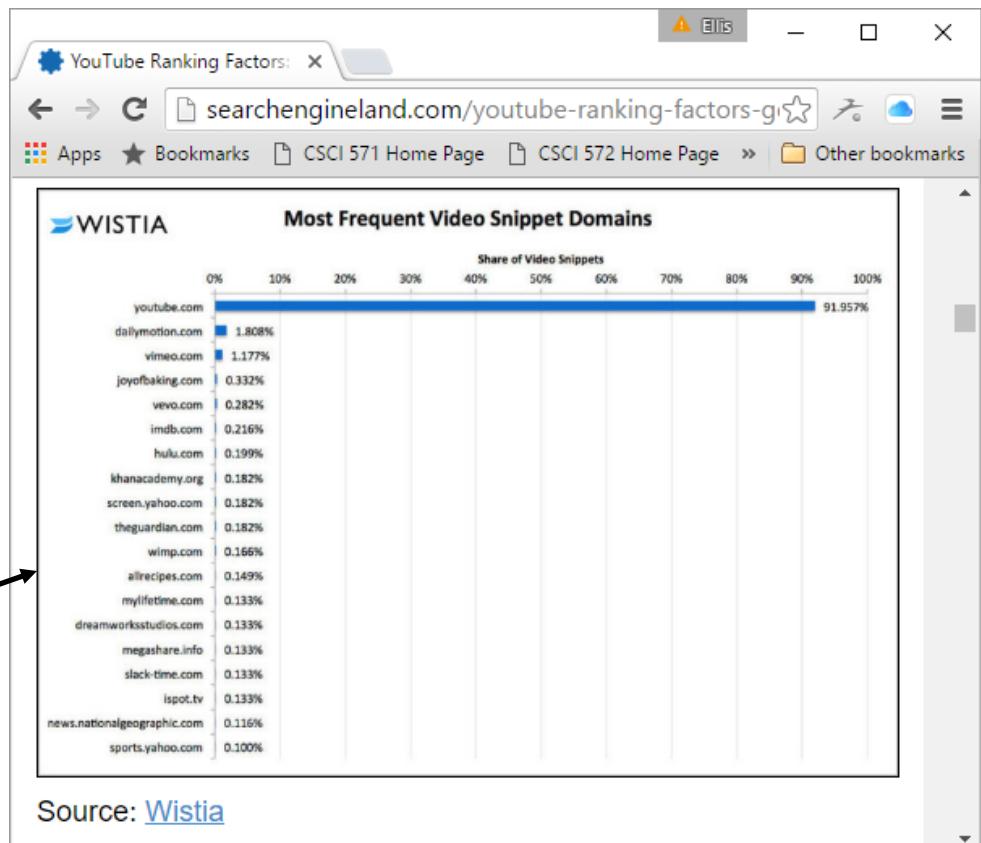
Google weeded out the video competition in Web search by predominantly displaying **only video-rich snippets** for YouTube videos back in 2014.

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>

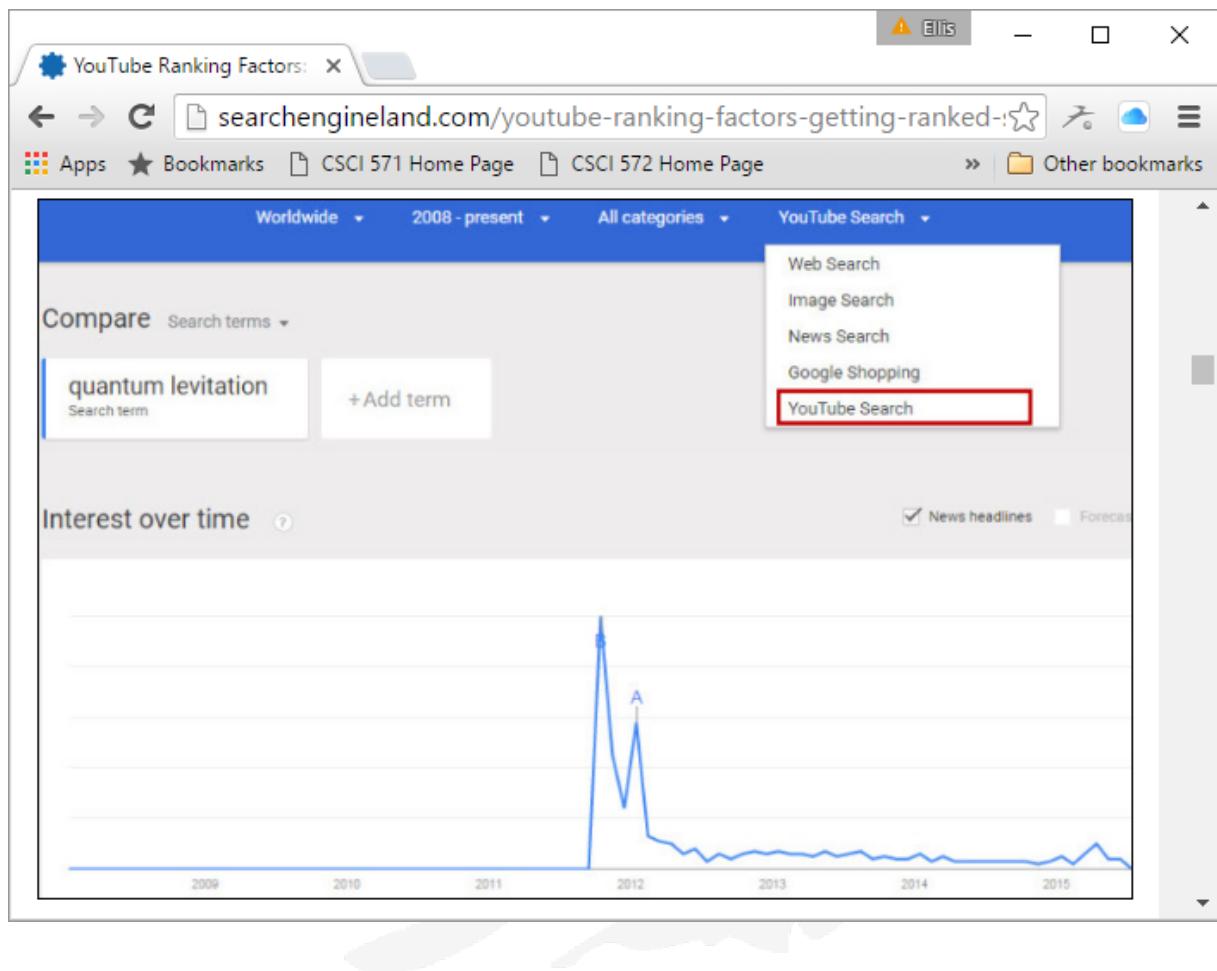
## Google Search is Biased Towards YouTube Videos



# Google is Biased Towards YouTube Videos

- Google also made an update to Google trends recently by including YouTube trending topics in the tool.

This shows that YouTube search traffic is significant enough that Google needed to incorporate it to paint a better picture of trending topics and stories across the Web.



# 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 companies that offer CDN services such as Akamai, Limelight and Level 3
- See the Akamai video for 10 minutes (Tom Leighton),
- [https://www.youtube.com/watch?v=Ni\\_60cbMydg](https://www.youtube.com/watch?v=Ni_60cbMydg)

# 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:**
- **YouTube assigns a fixed-length, 11 character string, unique identifier to each video, see**
- [http://curvetube.com/Will\\_YouTube\\_Ever\\_Run\\_Out\\_Of\\_Video\\_IDs\\_/gocwRvLhDf8.video](http://curvetube.com/Will_YouTube_Ever_Run_Out_Of_Video_IDs_/gocwRvLhDf8.video)
- **A Content Distribution Network**
  - <https://www.youtube.com/watch?v=6yrijdhvAtI>

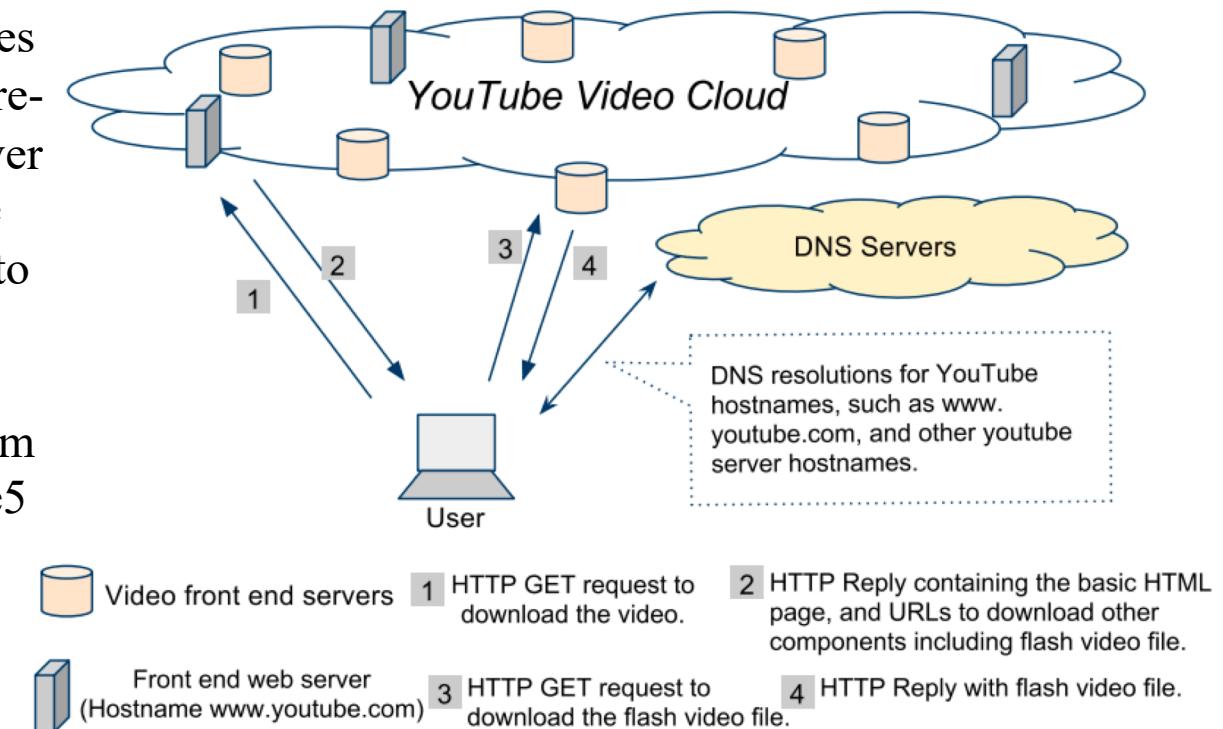
# YouTube's Content Distribution Network Basics

A local DNS server resolves [www.youtube.com](http://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>

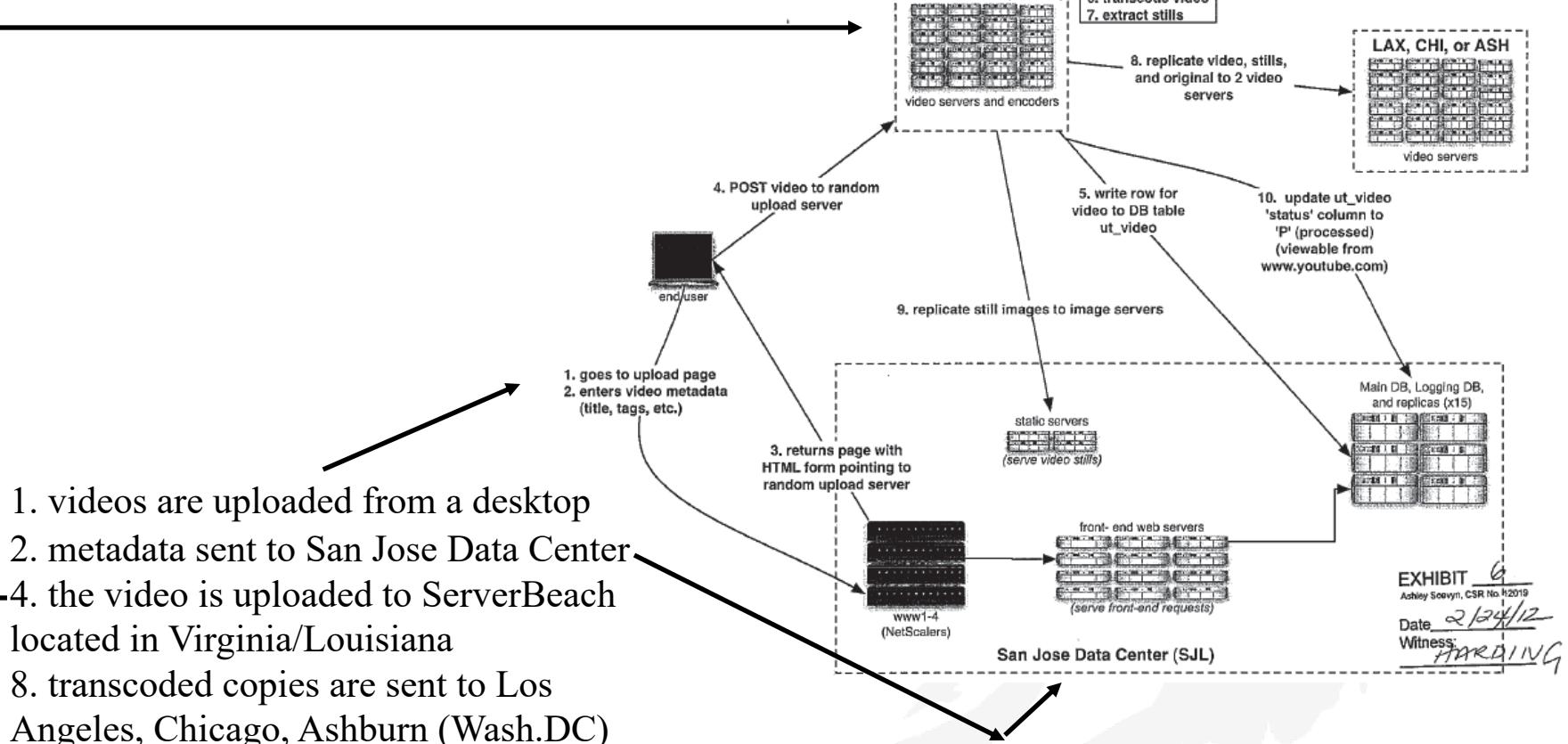
# YouTube Content Delivery Datacenters

- Adhikari et al report that YouTube uses a 3-tiered hierarchical caching scheme with 38 primary cache locations, 8 secondary locations, and 5 tertiary locations, distributed over five continents
- These roughly correspond to Google's data centers, see
- <https://www.google.com/about/datacenters/inside/locations/index.html>



Figure 4: Geographical distribution of YouTube Video Cache Locations.

# YouTube Upload Architecture



## References to YouTube's CDN

- There are three 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
- Both papers try to determine how a video server is selected, and in all cases there is 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 *Dissecting* 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 moved where they are wanted

# Monetizing YouTube

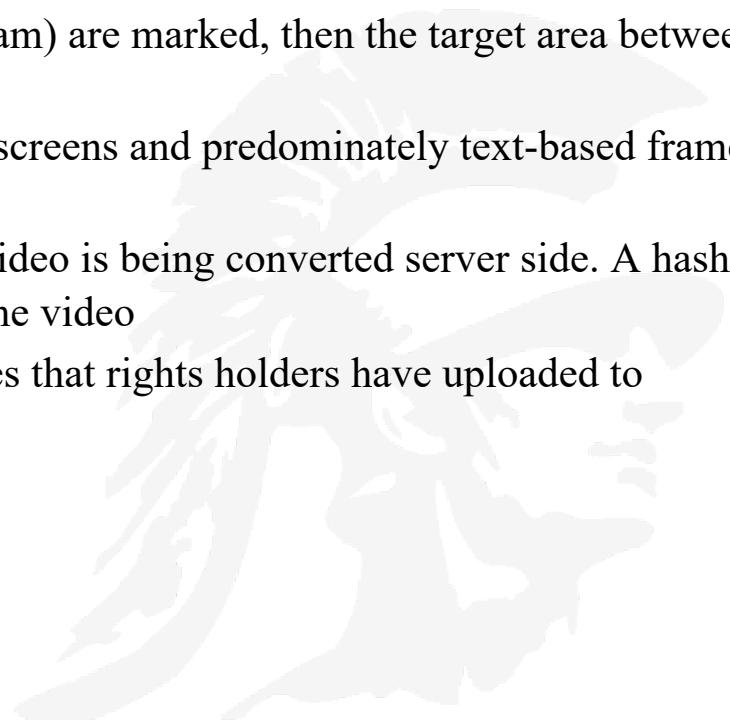
- **YouTube challenges in the early days**
  - YouTube had no way of making money, and its infrastructure is very expensive
  - YouTube was being sued by content creators, as many of YouTube's videos were uploaded illegally
  - YouTube **solved both problems** at once, by
    - Developing a system for 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

# ContentID

- YouTube's solution was to create a fingerprint database of copyrighted content, called Content ID
- ContentID technology creates an ID Fingerprint File for copyrighted audio and video material, and stores it in a database.
- 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 video is uploaded, it is checked against the database, and flags the video as a copyright violation if a match is found.
- When this occurs, the content owner has the choice of blocking the video to make it unviewable, tracking the viewing statistics of the video, or 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/](https://arstechnica.com/tech-policy-policy/2014/10/youtube-has-paid-1-billion-to-rights-holders-via-content-id-since-2007/)

# More Details on ContentID

- 
1. User uploads a video.
  2. YouTube then queues up the video to be processed server-side. It is encoded and quite a few different versions are created (HTML5 version, H.264, WebM VP8, HD, non-HD, etc).
  3. 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
  4. Frames from a video (most likely minus blank screens and predominately text-based frames) a sample of a section of the video is taken.
    - The frames are being analyzed when the video is being converted server side. A hash is created from those sampled frames from the video
  5. Content ID is based off audio and video samples that rights holders have uploaded to YouTube.