InnerTube

YOUTUBE CAPTION SEARCH

CS 410
Text Information Systems

University of Illinois Urbana-Champaign

The Team

- Arushi Garg (arushig6)
- Kyle Demeule (demeule2)
- Suraj Nayak (surajn3)

Problem

Searching the video user wants to view is a manual and laborious exercise

- Length of videos can be long.
- No simple way to visualize the content of a video.

Traditional approaches to searching:

- Guess based on title
- Manually scan video
- Read video comments (if available)

Solution

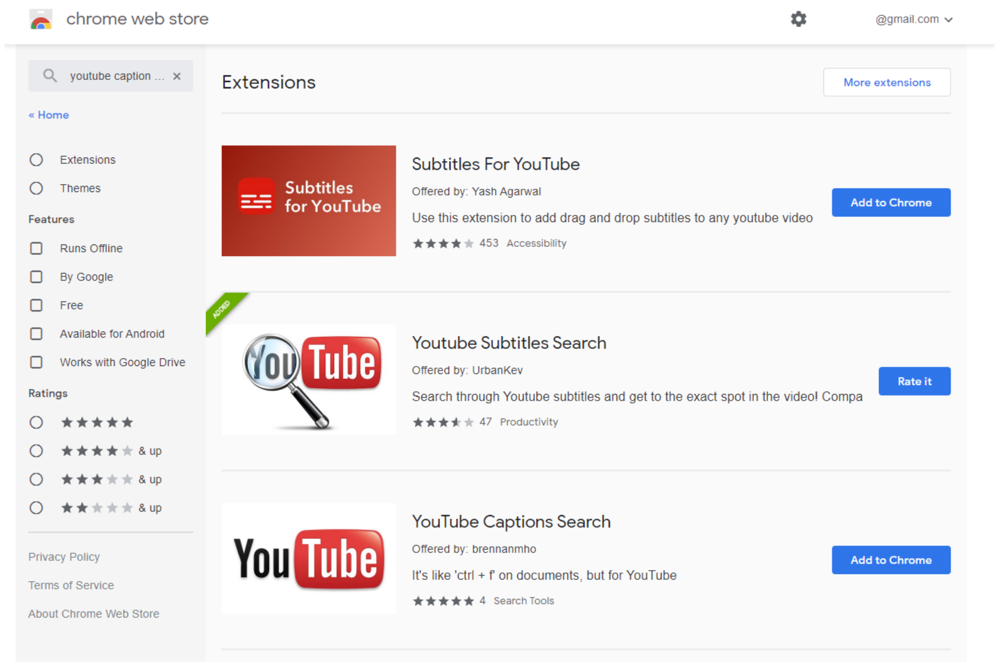
Utilize captions generated by YouTube to create video search engine

Implement as Chrome Extension to work seamlessly with YouTube

Existing Solutions

Several Plug-ins like this already existed

- However all of them are currently non-functioning.
  - Likely due to parsing YouTube DOM (which can change) or using non-official APIs.
- All are closed source.



Extensions

Subtitles For YouTube
Offered by: Yash Agarwal
Use this extension to add drag and drop subtitles to any youtube video
★★★★☆ 453 Accessibility

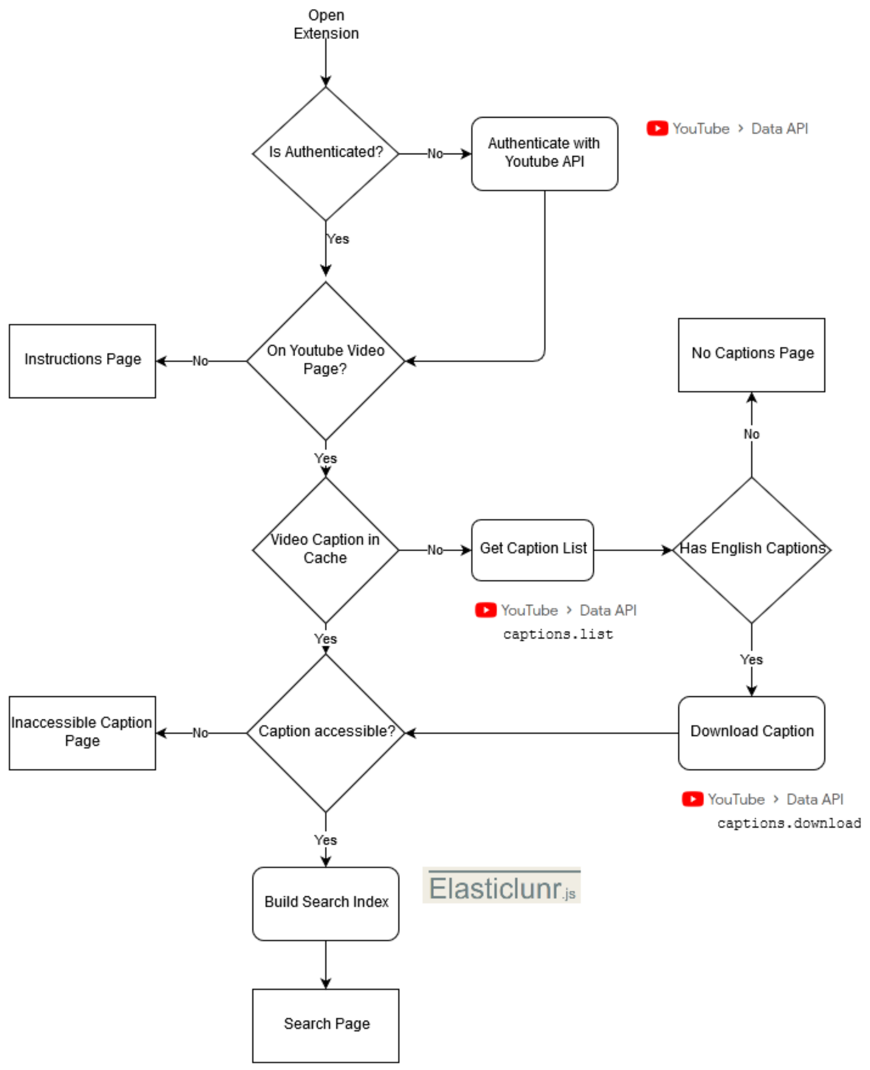
Youtube Subtitles Search
Offered by: UrbanKev
Search through Youtube subtitles and get to the exact spot in the video! Compa
★★★★☆ 47 Productivity

YouTube Captions Search
Offered by: brennanmho
It's like 'ctrl + f' on documents, but for YouTube
★★★★★ 4 Search Tools

Demo

Implementation

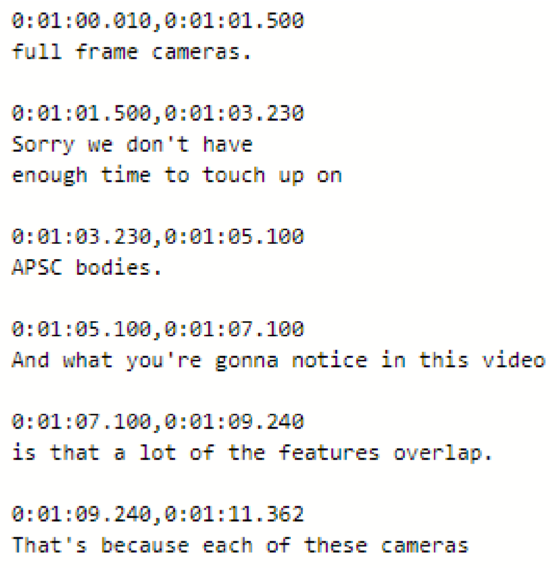
Flow on Open



Implementation: Index

How to go from Transcript -> Documents?

Transcript format:



```
0:01:00.010,0:01:01.500
full frame cameras.

0:01:01.500,0:01:03.230
Sorry we don't have
enough time to touch up on

0:01:03.230,0:01:05.100
APSC bodies.

0:01:05.100,0:01:07.100
And what you're gonna notice in this video

0:01:07.100,0:01:09.240
is that a lot of the features overlap.

0:01:09.240,0:01:11.362
That's because each of these cameras
```

Each segment is considered a document.

Implementation

Flow on Search

- Get search term
- Search index for search term
- Order results based on search score

Flow on Search Result Select

- Get selected search result
- Convert search result to seconds in video of result
- Inject code to use Youtube iFrame API to seek player to that time

Implementation: Technology Used

Chrome Extension
- Application Development

Youtube Data API
- Accessing captions

Youtube iFrame Player API
- Seeking to video section on search click

Elasticlunr.js
- Building search index

jQuery
- DOM manipulation, AJAX

Google API Client Library for JavaScript
- Authentication

Features

Speed (no need for an HTTP request for each search)

Dynamic (build index on the fly)

Stop words, stemming.

Cache downloaded captions

Limitations

This Extension uses the YouTube V3 API to access captions for videos.

Pros:

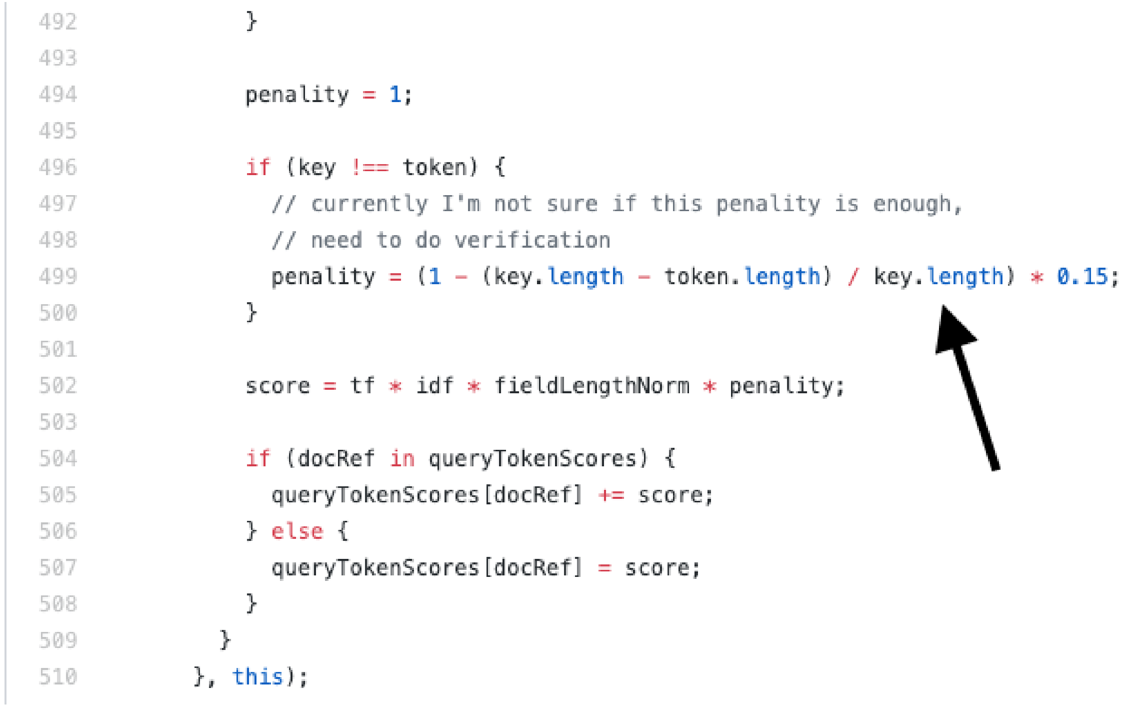
- Most inline approach with Youtube's TOS.
- Will likely have the most longevity.

Cons:

- YouTube requires oauth and heightened user permissions to access the `captions.download` endpoint.
- YouTube restricts third-party access to video captions, majority of videos on YouTube currently don't allow third party access.

Scoring the documents

The scoring function used in the application includes TF, IDF and Length Normalization



```
492       }
493
494       penality = 1;
495
496       if (key !== token) {
497         // currently I'm not sure if this penality is enough,
498         // need to do verification
499         penality = (1 - (key.length - token.length) / key.length) * 0.15;
500       }
501
502       score = tf * idf * fieldLengthNorm * penality;
503
504       if (docRef in queryTokenScores) {
505         queryTokenScores[docRef] += score;
506       } else {
507         queryTokenScores[docRef] = score;
508       }
509     }
510   }, this);
```

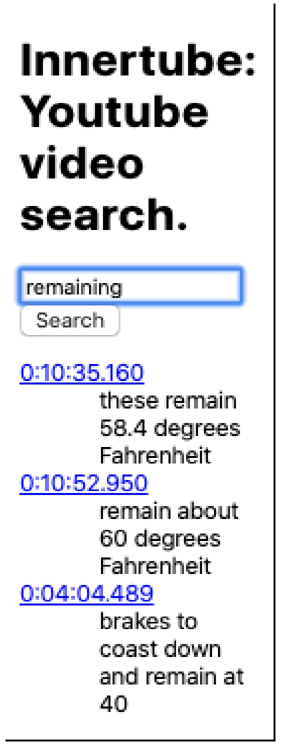
Source : https://github.com/weixsong/elasticlunr.js/blob/eb23a604db018d2a6a61487027be80ebe6da1dde/lib/index/legacy.js#L499

Stemming & Stop word removal

The plugin is also capable to perform search based on :

a. Root word

b. Removal of stop word from the search query



Innertube: Youtube video search.

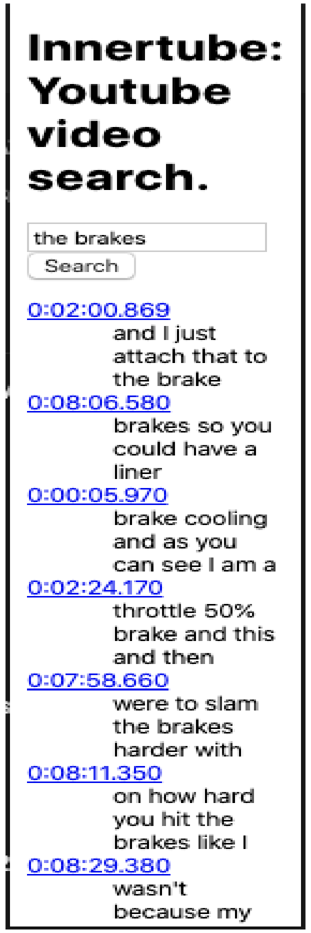
remaining

Search

0:10:35.160
these remain 58.4 degrees Fahrenheit

0:10:52.950
remain about 60 degrees Fahrenheit

0:04:04.489
brakes to coast down and remain at 40



Innertube: Youtube video search.

the brakes

Search

0:02:00.869
and I just attach that to the brake

0:08:06.580
brakes so you could have a liner

0:00:05.970
brake cooling and as you can see I am a

0:02:24.170
throttle 50% brake and this and then

0:07:58.660
were to slam the brakes harder with

0:08:11.350
on how hard you hit the brakes like I

0:08:29.380
wasn't because my

Running Locally

Runs entirely in the chrome extension; no extra servers required.

Steps:

- Download code (github repository).
- Load Extension locally (instruction source):
  - Navigate to `chrome://extensions`
  - Expand the Developer dropdown menu and click “Load Unpacked Extension”
  - Navigate to the local folder containing the extension’s code and click Ok
  - Assuming there are no errors, the extension should load into your browser
- Click plugin, follow authentication flow.
- Navigate to Youtube video (example).
- Click extension, enter search term.

Appendix

Not all videos allow caption downloads by third parties. Some samples that do:

Videos

- How Turbochargers Work
- Camera Review
- Differential Equations

Channels

- 3Blue1Brown
- Artifical Intelligence
- EngineeringExplained