

pprint(videos)

```
파랑색 : 정현님 설명 부분,
#!/usr/bin/env pvthon
from future import print
import sys
                          빨강색: 용석, 성필님 설명 부분
import os
import argparse
from pytube import YouTube
                           (상세 내용은 별도 협의)
from pytube.utils import prin
from pytube.exceptions impor
from pprint import pprint
                          주황색 : 예찬님 설명 부분(정현님이
|def main(): 🗲
                          script.py정리한 keynote파일에 이미
   parser = argparse.Argumen
   parser.add argument("
                          어느정도 설명되어있습니다!
   parser.add argument(
   parser.add_argument("--resolution", "-r", dest="res", help=(
   parser.add_argument("--path", "-p", action=FullPaths, default=os.getcwd(),
   parser.add_argument("--filename", "-f", dest="filename", help=(
   args = parser.parse_args()
       vt = YouTube(args.url)
       videos = []
       for i, video in enumerate(yt.get_videos()):
          ext = video.extension
          res = video.resolution
          videos.append("{} {}".format(ext, res))
   except PvtubeError:
       sys.exit(1)
   if args.filename:
       yt.set_filename(args.filename)
   if args.ext or args.res:
       if not all([args.ext, args.res]):
          pprint(videos)
          svs.exit(1)
   if args.ext and args.res:
       vid = yt.get(args.ext, args.res)
       if not vid:
```

!!!발표 역할 분담!!!

nmand-line 인터페이스 형식 정의, 로드 할 수 있도록 정보를 전달하는 역할

```
if args.ext and args.res:
        vid = yt.get(args.ext, args.res)
        if not vid:
            pprint(videos)
            sys.exit(1)
   elif args.ext:
        videos = yt.filter(extension=args.ext)
        if not videos:
            svs.exit(1)
        vid = max(videos)
   elif args.res:
        videos = yt.filter(resolution=args.res)
        if not videos:
            sys.exit(1)
        vid = max(videos)
        vid = max(yt.get_videos())
        vid.download(path=args.path, on_progress=print_status)
   except KeyboardInterrupt:
        print("Download interrupted.")
        sys.exit(1)
if __name__ == '__main__':
 → main()
```

```
if __name__ == "__main__"
```

'이 파일이 command line interface에서 실행되는 경우' 의 의미

구현한 코드가 다른 파이썬 코드에 의해 module로 import 될 경우도 있고, command line interface에서 직접 실행될 경우도 있는데 위 코드는 command line interface에 의해 실행 될 경우에만 실행하고 싶은 코드 블록이 있을 경우 사 [참고] http://stackoverflow.com/questions/419163/what-does-if-name-main-do

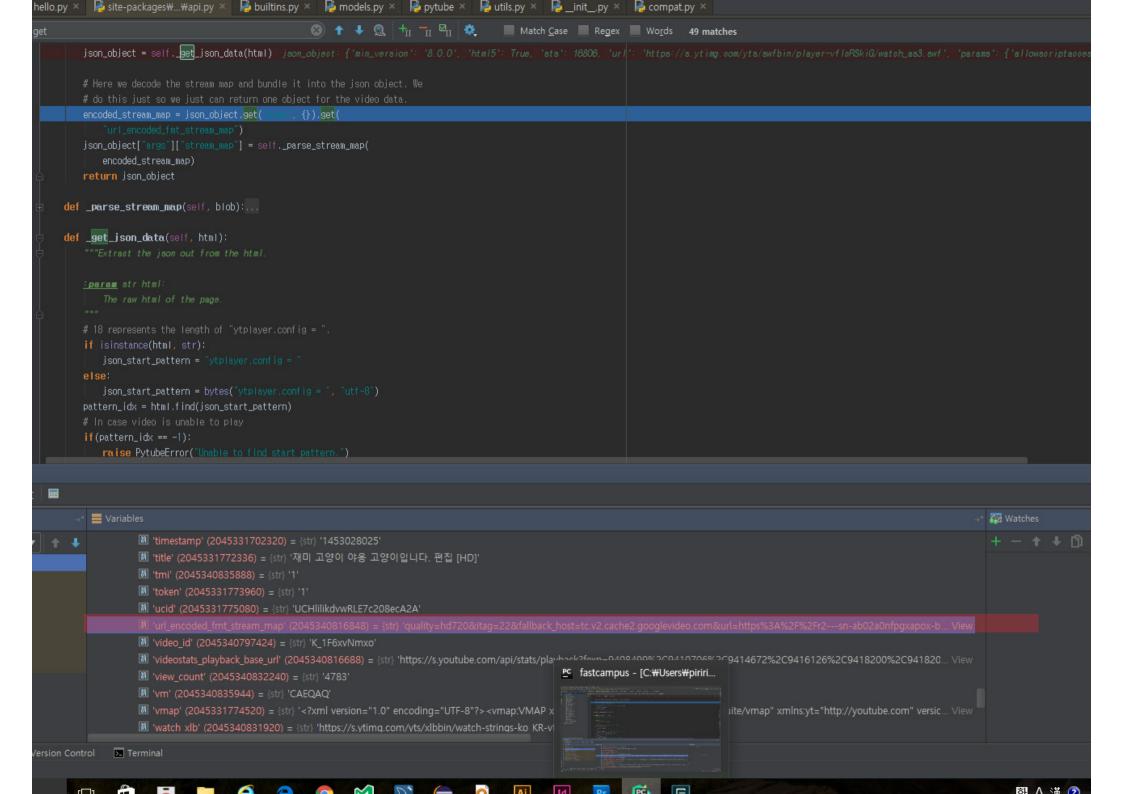
```
api.py API wrapper
                              class YouTube(object):
                                  """Class representation of a single instance of a YouTube session.
                                  def __init__(self, url=None):
                                      """Initializes YouTube API wrapper.
                                       :param str url:
                                           The url to the YouTube video.
                                      self._filename = None
                                      self._video_url = None
                                      self._js_cache = None
                                      self._videos = []
                                      if url:
                                          self.from_url(url)
                                                         def from url(self, url):
                                                             """Sets the url for the video.
                                                             :param str url:
args = parser.parse_args()
                                                                 The url to the YouTube video.
try:
   yt = YouTube(args.url)
                                                             self. video url = url
   videos = []
   for i, video in enumerate(yt.get_videos()):
                                                             # Reset the filename and videos list in case the same instance is
       ext = video.extension
                                                             # reused.
       res = video.resolution
                                                             self._filename = None
       videos.append("{} {}".format(ext, res))
                                                             self. videos = []
                                                             # Get the video details.
                                                             video data = self.get video data()
                                                             # Set the title from the title.
          def get_videos(self):
                                                             self.title = video_data.get("args", {}).get("title")
               """Gets all videos."""
                                                             # Rewrite and add the url to the javascript file, we'll need to fetch
               return self._videos
                                                             # this if YouTube doesn't provide us with the signature.
                                                             js_url = "http:" + video_data.get("assets", {}).get("js")
                                                             # Just make these easily accessible as variables.
                                                             stream_map = video_data.get("args", {}).get("stream_map")
                                                             video_urls = stream_map.get("url")
                                                             # of available videos.
                                                             for idx, url in enumerate(video_urls):
```

```
def from_url(self, url):
              class YouTube(object):
                 """Class representation of a single instance of a YouTube session.
                 def __init__(self, url=None):
                                                                             :param str url:
                     :param str url:
                       The url to the YouTube video.
                                                                           self. video url = url
                    self._filename = None
                    self._js_cache = None
                                                                             # reused.
                    self. videos = []
                    if url:
                                                                             self._filename = None
                        self.from url(url)
                                                                             self. videos = []
   get video data(self):
   """Gets the page and extracts out the video data."""
   # Reset the filename incase it was previously set.
   self.title = None
   response = urlopen(self.url)
   if not response:
       raise PvtubeError("Unable to open url: {0}".format(self.url))
   html = response.read()
   print(html)
   if isinstance(html, str):
       restriction_pattern = "og:restrictions:age"
   else:
       restriction_pattern = bytes("og:restrictions:age", "utf-8")
   if restriction_pattern in html:
       raise AgeRestricted("Age restricted video. Unable to download"
                                                                             # of available videos.
   # Extract out the json data from the html response body.
   json_object = self._get_json_data(html)
                                                                                 trv:
   # Here we decode the stream map and bundle it into the ison object. We
   # do this just so we just can return one object for the video data.
   encoded_stream_map = json_object.get("args", {}).get(
                                                                                          continue
   json_object["args"]["stream_map"] = self._parse_stream_map(
       encoded_stream_map)
   return json_object
                                                                                     continue
> self 객체에 있는 url 경로를 이용하여 urlopen(), response.read()
메소드로 html 정보를 읽어온다
                                                                                 # cipher from the js.
> html 내의 연령 제한 여부를 확인(html이 str 형태인지 여부에 따라 별
도 처리)
                                                                                                "cipher.")
> html 내의 json_data를 추출 후 json_object로 리턴
> json_object내의 stream map을 3단계에 걸쳐서 이쁘게 parsing
                                                                             self._js_cache = None
```

```
"""Sets the url for the video.
    The url to the YouTube video.
# Reset the filename and videos list in case the same instance is
# Get the video details.
video data = self.get video data()
# Set the title from the title.
self.title = video_data.get("args", {}).get("title")
# Rewrite and add the url to the javascript file, we'll need to fetch
# this if YouTube doesn't provide us with the signature.
js url = "http:" + video data.get("assets", {}).get("js")
# Just make these easily accessible as variables.
stream_map = video_data.get("args", {}).get("stream_map")
video urls = stream map.get("url")
# For each video url, identify the quality profile and add it to list
for idx, url in enumerate(video_urls):
    log.debug("attempting to get quality profile from url: %s", url)
        itag, quality profile = self. get quality profile from url(url)
        if not quality_profile:
           log.warn("unable to identify profile for itag=%s", itag)
   except (TypeError, KeyError) as e:
        log.exception("passing on exception %s", e)
   # Check if we have the signature, otherwise we'll need to get the
   if "signature=" not in url:
        log.debug("signature not in url, attempting to resolve the "
        signature = self._get_cipher(stream_map["s"][idx], js_url)
       url = "{0}&signature={1}".format(url, signature)
   self._add_video(url, self.filename, **quality_profile)
# Clear the cached is. Make sure to keep this at the end of
# `from_url()` so we can mock inject the js in unit tests.
```

```
Match Case Regex Words 49 matches
> ison object내의 stream map을 찾는 과정!
       # Here we decode the stream map and bundle it into the ison object. We
       # do this just so we just can return one object for the video data.
       encoded_stream_map = json_object.get(
       json_object["args"]["stream_map"] = self._parse_stream_map(
           encoded_stream_map)
       return ison_object
   def _parse_stream_map(self, blob):...
   def _get_json_data(self, html):
        :param str html:
           json_start_pattern = "ytplayer.config = "
           json_start_pattern = bytes("ytplayer.config = ", "utf-8")
       pattern_idx = html.find(json_start_pattern)
       if (pattern_idx == -1):
           raise PytubeError("Unable to find start pattern.")
Variables
                                                                                                                                                                                               Ca W
           thml = {bytes} b'<!DOCTYPE html><html lang="ko" data-cast-api-enabled="true"><head><style name="www-roboto">@font-face{font-family:\"Roboto\";font-style:normal;font-weight:4...
           son object = {dict} {'min version': '8.0.0', 'html5': True, 'sts': 16806, 'url': 'https://s.ytimg.com/yts/swfbin/player-yflcRSkig/watch as3.swf', 'params': {'allowscriptaccess': 'always', 'bgcolor'
                   🔞 'account_playback_token' (2045331736448) = {str} 'QUFFLUhqbGNwSkl4Q1ZCYWhUMU1Za0dyZ0o1TE5OZmU2QXxBQ3Jtc0traTVhT1VfeTNQLVhuZWpodjl6d0RkcU43R3FqQTFaeDV... View
                   'ad3_module' (2045331702960) = {str} '1'
                   'ad_device' (2045340796464) = {str} '1'
                   'ad_flags' (2045340798064) = {str} '1'
                   3 'ad_logging_flag' (2045331702128) = {str} '1'
                   8 'ad_slots' (2045340798256) = {str} '0'
                   👪 'ad tag' (2045340836616) = {str} 'https://ad.doubleclick.net/N4061/pfadx/com.ytpwatch.comedy;av=1;kpeid=HlilikdvwRLE7c208ecA2A;kpid=10853451;kvid=K 1F6xyNmxo;mpvid:... View
```

> Terminal



> 힘들게 찾아온 stream map을 3단계에 걸쳐서 이쁘게 parsing!

```
type=video%2Fmp4%3B+codecs%3D%22avc1.64001F%2C+mp4a.40.2%22&quality=hd720&fallback host=tc.v21.cache7.googlevideo.com&itag=22&url=https%3A%2F%2Fr8---sn-3u-b
['type=video%2Fmp4%3B+codecs%3D%22avc1.64001F%2C+mp4a.40.2%22&quality=hd720&fallback host=tc.v21.cache7.qooqlevideo.com&itaq=22&url=https%3A%2F%2Fr8---sn-3u-
 type=video%2Fwebm%3B+codecs%3D%22vp8.0%2C+vorbis%22&qualitv=medium&fallback host=tc.v13.cache1.qooglevideo.com&itag=43&url=https%3A%2F%2Fr8---sn-3u-bh2es.q
 'type=video%2Fmp4%3B+codecs%3D%22avc1.42001E%2C+mp4a.40.2%22&guality=medium&fallback host=tc.v3.cache4.googlevideo.com&itag=18&url=https%3A%2F%2Fr8---sn-3u-
 'type=video%2Fx-flv&quality=small&fallback_host=tc.v24.cache8.googlevideo.com&itag=5&url=https%3A%2F%2Fr8—-sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fup
 'type=video%2F3gpp%3B+codecs%3D%22mp4v.20.3%2C+mp4a.40.2%22&guality=small&fallback host=tc.v13.cache8.googlevideo.com&itag=36&url=https%3A%2F%2Fr8---sn-3u-b
 type=video%2F3gpp%3B+codecs%3D%22mp4v.20.3%2C+mp4a.40.2%22&guality=small&fallback_host=tc.v12.cache8.googlevideo.com&itag=17&url=https%3A%2F%2Fr8---sn-3u-b
[['type=video%2Fmp4%3B+codecs%3D%22avc1.64001F%2C+mp4a.40.2%22',
  'quality=hd720',
 'fallback_host=tc.v21.cache7.googlevideo.com',
 'itag=22',
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fratebypass%3Dyes%26mime%3Dvideo%252Fmp4%26key%3Dyt6%26itag%3D22%26pcm2cms%3Dyes%26upn
['tvpe=video%2Fwebm%3B+codecs%3D%22vp8.0%2C+vorbis%22',
 'quality=medium',
 'fallback host=tc.v13.cache1.googlevideo.com',
 'itag=43',
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fratebypass%3Dyes%26mime%3Dvideo%252Fwebm%26key%3Dyt6%26itag%3D43%26pcm2cms%3Dyes%26up
['type=video%2Fmp4%3B+codecs%3D%22avc1.42001E%2C+mp4a.40.2%22',
 'quality=medium',
 'fallback_host=tc.v3.cache4.googlevideo.com',
 'itag=18',
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fratebypass%3Dyes%26mime%3Dvideo%252Fmp4%26key%3Dyt6%26itag%3D18%26pcm2cms%3Dyes%26upn
['type=video%2Fx-flv',
 'quality=small',
 'fallback_host=tc.v24.cache8.googlevideo.com',
 'itag=5',
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fupn%3DZLfGzxg7JME%26mime%3Dvideo%252Fx-flv%26key%3Dyt6%26itag%3D5%26pcm2cms%3Dyes%26s
['type=video%2F3qpp%3B+codecs%3D%22mp4v.20.3%2C+mp4a.40.2%22',
 'quality=small',
 'fallback_host=tc.v13.cache8.googlevideo.com',
 'itag=36'.
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fupn%3DZLfGzxg7JME%26mime%3Dvideo%252F3gpp%26key%3Dyt6%26itag%3D36%26pcm2cms%3Dyes%26s
['type=video%2F3qpp%3B+codecs%3D%22mp4v.20.3%2C+mp4a.40.2%22',
 'quality=small',
 'fallback_host=tc.v12.cache8.googlevideo.com',
 'url=https%3A%2F%2Fr8---sn-3u-bh2es.googlevideo.com%2Fvideoplayback%3Fupn%3DZLfGzxg7JME%26mime%3Dvideo%252F3gpp%26kev%3Dvt6%26itag%3D17%26pcm2cms%3Dves%26s
```

```
def get guality profile from url(self, video url):
   """Gets the quality profile given a video url. Normally we would just
    use `urlparse` since itags are represented as a get parameter, but
    YouTube doesn't pass a properly encoded url.
    :param str video url:
        The malformed url-encoded video url.
    reg exp = re.compile('itag=(\d+)')
    itag = reg exp.findall(video url)
   if itag and len(itag) == 1:
        itag = int(itag[0])
       # Given an itag, refer to the YouTube quality profiles to get the
       # properties (media type, resolution, etc.) of the video.
       quality profile = QUALITY PROFILES.get(itag)
       if not quality_profile:
            return itag. None
       # Here we just combine the quality profile keys to the
       # corresponding quality profile, referenced by the itag.
       return itag, dict(list(zip(QUALITY PROFILE KEYS, quality profile)))
   if not itag:
        raise PytubeError("Unable to get encoding profile, no itag found.")
   elif len(itag) > 1:
        log.warn("Multiple itags found: %s", itag)
        raise PytubeError("Unable to get encoding profile, multiple itags "
    return False
```

```
QUALITY PROFILES = {
     # flash
     17: ("3gp", "144p", "MPEG-4 Visual", "Simple", "0,05", "AAC", "24"),
     # webm
     100: ("webm", "360p", "VP8", "3D", "N/A", "Vorbis", "128"),
     # mpeq4
    22: ("mp4", "720p", "H.264", "High", "2-2.9", "AAC", "192"),
82: ("mp4", "360p", "H.264", "3D", "0.5", "AAC", "96"),
83: ("mp4", "240p", "H.264", "3D", "0.5", "AAC", "96"),
    84: ("mp4", "720p", "H.264", "3D", "2-2.9", "AAC", "152"),
# The keys corresponding to the quality/codec map above.
QUALITY PROFILE KEYS = (
```

```
def _add_video(self, url, filename, **kwargs): ←
   """Adds new video object to videos.
   :param str url:
       The signed url to the video.
   :param str filename:
       The filename for the video.
   :param kwarqs:
       Additional properties to set for the video object.
   video = Video(url, filename, **kwargs)
   self._videos.append(video)
   self._videos.sort()
   return True
  > 마침내 self. video에 Video 클래스의 객체 할당
  try:
     yt = YouTube(args.url)
     videos = []
     for i, video in enumerate(yt.get_videos()):
         ext = video.extension
         res = video.resolution
         videos.append("{} {}".format(ext, res))
```

```
def from url(self, url):
    """Sets the url for the video.
    :param str url:
        The url to the YouTube video.
   self. video url = url
    # Reset the filename and videos list in case the same instance is
    # reused.
    self._filename = None
    self. videos = []
    # Get the video details.
    video_data = self.get_video_data()
    # Set the title from the title.
    self.title = video_data.get("args", {}).get("title")
    # Rewrite and add the url to the javascript file, we'll need to fetch
    # this if YouTube doesn't provide us with the signature.
    js url = "http:" + video data.get("assets", {}).get("js")
    # Just make these easily accessible as variables.
    stream_map = video_data.get("args", {}).get("stream_map")
    video urls = stream map.get("url")
    # For each video url, identify the quality profile and add it to list
    # of available videos.
    for idx, url in enumerate(video urls):
        log.debug("attempting to get quality profile from url: %s", url)
       try:
            itag, quality profile = self. get quality profile from url(url)
            if not quality_profile:
               log.warn("unable to identify profile for itag=%s", itag)
               continue
       except (TypeError, KeyError) as e:
            log.exception("passing on exception %s", e)
            continue
       # Check if we have the signature, otherwise we'll need to get the
       # cipher from the js.
       if "signature=" not in url:
            log.debug("signature not in url, attempting to resolve the "
                     "cipher.")
            signature = self._get_cipher(stream_map["s"][idx], js_url)
           url = "{0}&signature={1}".format(url, signature)
       self._add_video(url, self.filename, **quality_profile)
    # Clear the cached is. Make sure to keep this at the end of
    # `from_url()` so we can mock inject the js in unit tests.
    self._js_cache = None
```

> 모든 준비를 마친 self._video는 다운로드를 하고자 할 때, Video 클래스의 객체를 생성합니다.

```
def download(self, path, chunk_size=8 * 1024, on_progress=None,
             on_finish=None, force_overwrite=False):
    """Downloads the video."""
    path = os.path.normpath(path)
    if os.path.isdir(path):
        filename = "{0}.{1}".format(self.filename, self.extension)
        path = os.path.join(path, filename)
   # TODO: If it's not a path, this should raise an ``OSError``.
    # TODO: Move this into cli, this kind of logic probably shouldn't be
   # handled by the library.
   if os.path.isfile(path) and not force overwrite:
        raise OSError("Conflicting filename:'{0}'".format(self.filename))
    # TODO: Split up the downloading and OS jazz into separate functions.
    response = urlopen(self.url)
    meta_data = dict(response.info().items())
    file_size = int(meta_data.get("Content-Length") or
                    meta_data.get("content-length"))
    self. bytes received = 0
    start = clock()
    # TODO: Let's get rid of this whole try/except block, let ``OSErrors``
    # fail loudly.
    try:
       with open(path, 'wb') as dst_file:
            while True:
                self. buffer = response.read(chunk size)
                # Check if the buffer is empty (aka no bytes remaining).
                if not self._buffer:
                    if on_finish:
                        # TODO: We possibly want to flush the
                        # `_bytes_recieved`` buffer before we call
                        # ``on finish()``.
                        on_finish(path)
                    break
                self. bytes received += len(self. buffer)
                dst_file.write(self._buffer)
                if on_progress:
                    on_progress(self._bytes_received, file_size, start)
    except KeyboardInterrupt:
        # TODO: Move this into the cli, ``KeyboardInterrupt`` handling
        # should be taken care of by the client. Also you should be allowed
        # to disable this.
        os.remove(path)
        raise KeyboardInterrupt(
```