

reddit

Home Popular Answers BETA Explore All Start a community

GAMES ON REDDIT

Syllo Merge syllables fast 167k monthly players

Honk Farm Merge Valley Jump Cat Discover More Games

CUSTOM FEEDS

Create Custom Feed

RECENT

/r/chileIT

/r/vibecoding

/r/chile

/r/calculsheet

/r/GYM

COMMUNITIES

Manage Communities

/r/aweeanagang

/r/chile

/r/CuartetoDeNos

/r/FortNiteBR

/r/Gta5Modding

/r/gtaglitches

/r/gtaonline

/r/TheVirginZone

/r/Highdeology

/r/ValorantComp...

/r/wallstreetbets

/r/argentina

/r/calculsheet

/r/ChatGPTProm...

/r/ChileIT

/r/chileIT

/r/compsci

/r/CryptoCurrency

/r/Cyberpunka...

/r/Daytrading

/r/EdgingTalk

/r/FortniteSavet...

/r/GTA

/r/GYM

/r/hentai

/r/pcmasterrace

/r/RepubliCadeC...

/r/Steam

/r/SteamArtwork...

/r/WallStreetbet...

/r/yoy_ctm

RESOURCES

← r/AWSCertifications • 2y ago

hml1366

...

How to Download pdf/videos from AWS Academy

Hi all,

I am taking a Big Data course at college in which we have been g access to AWS Academy for pdf and video materials.

The access will be until the end of the course only, but I'd like to download the pdf and video materials into my pc to for future reference.

Any idea how I can download materials from AWS Academy portal? I tried Inspect Element -> Network method but the link is from emergingtalent.contentcontroller.com which prohibits seeing the material.

Is there any way at all to download material from AWS Academy?

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

Sort by: Best

Search Comments

asplayer12 • 2y ago • Edited 2y ago

In the network tab of dev tool right click copy Curl like this

```
Linux x86_64; rv:109.0) Gecko/20100101 Firefox/115.0" -H "Accept: video/webm,video/ogg,video/*;q=0.5" -H "User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:121.0) Gecko/20101001 Firefox/121.0" -H "Accept: */*" -H "Accept-Language: en-US,en;q=0.5" -H "Accept-Encoding: gzip, deflate, br" -H "Referer: https://emergingtalent.contentcontroller.com/ScormEngineInterface/defaultui/player/cms-a-wj/Default/cms-media/la.html" -H "Connection: keep-alive" -H "Cookie: CloudFront-Policy=XXXXX; CloudFront-Key-Pair-ID=XXXXXXX" -H "Sec-Fetch-Dest: empty" -H "Sec-Fetch-Mode: cors" -H "Sec-Fetch-Site: same-origin" -H "TE: trailers"
```

paste this in to your terminal and add "-o filename.mp4"

The point of this is to preserve the header for the request

hmd1366 OP • 2y ago

I cannot find the option for "Copy Curl"

<https://bb.co/716v8B>

asplayer12 • 2y ago • Edited 2y ago

sorry i meant cURL

BTW i also wrote a small python script that semi-automatically download the video and pdf. Yeah, the code is dirty and i could probably automate it even further but i couldn't be bothered.

Just paste your request header to the header variable and the normal url (Not the cURL) + the name of the video to the dictionary.

Im not sure how you would copy just the request header in chrome but in firefox just click a request and toggle "raw" <https://imgur.com/ehTAWbE> and make sure to remove the first line

```
import requests
import os
from urllib.parse import urlparse

header = ""paste header here""

# ("url of .vtt or .mp4 or .pdf ","filename with no extentions")
jobs = {
    'https://emergingtalent.contentcontroller.com/vault/3b92a5a-1f5a-41c5-8ce8-d11d5f',
    'https://emergingtalent.contentcontroller.com/vault/3b92a5a-1f5a-41c5-8ce8-d11d5f',
    'https://emergingtalent.contentcontroller.com/vault/7b5a7c1-d4a0-4909-8a88-d83001'
}

buf = header.splitlines()

header_dict = {} # formatting the header to a dict
for i in buf:
    i = i.split(":", 1)
    i[0] = i[0].replace(":", "")
    i[1] = i[1].replace(" ", "")
    header_dict[i[0]] = i[1]

# print(header_dict["User-Agent"])

for url, filename in jobs.items():

    r = requests.get(url=url, headers=header_dict)

    a = urlparse(url)
    a = os.path.basename(a.path)

    asff = file_extension = os.path.splitext(a)
    filename = filename.replace(".", "")
    filename = filename.replace(" ", "")
    filename = filename.replace("/", "")

    filename = f"/path/to/your/folder/module_08/{filename}{file_extension}"

    print(f'Downloaded {filename}')
    open(filename, 'wb').write(r.content)
```

red_sweater_bandit • 1y ago

I'm replying to this comment to say your python script worked. I was able to download the student guide PDFs by copying the raw request header (without the first line as you stated) and updating the 'jobs' dictionary with my own course url from the pdf get request. Hopefully this will help someone else just like it helped me.

Thanks u/asplayer12 lol

MediocrePlatform6870 • 1y ago

hey i didnt understand can u plz tell me how to do {

benjtom • 1mo ago

Same here, I was helped very much from this lol

xhaarz-admin • 2y ago

Could you please explain how to get the pdf url? I'm taking an AWS Academy course and the system used is Canvas LMS. Each page displayed on student guide is loaded with the cms javascript library and it loads every page as a mediafile.

asplayer12 • 2y ago

1. login to awsacademy
2. open the devtool of your browser (I'm using Firefox ctrl+shift+c) and go to the network tab
3. go to any student guide
4. on the devtool search for "pdf"
5. You should be able to find the pdf url https://bb.co/9_Y2m3aR8 (the one in the bottom)

Unless somehow the content management system is different for your modules, if that's the case then I have no idea.

xhaarz-admin • 2y ago

Thanks, that worked. Btw I tried with the cURL command but after running it shows the no authorized message *Content can only be accessed by the launch process.* Please launch your course again.
Any idea how to download the pdf?

Chauru10 • 2y ago

u/asplayer12 I'm getting "You are not authenticated to access this content. Reason: Access GUID is unregistered. Please relaunch the course." what headers are you using? I'm using the one of the request that came in the ge to of the pdf

asplayer12 • 2y ago • Edited 2y ago

```
curl https://emergingtalent.contentcontroller.com/vault/2XXXX-XXXX-XXXX-XXXX-XXXXXXX/XXXX-XXXX-XXXX-XXXX-XXXXXXX/XXXX-XXXX-XXXX-XXXX-XXXXXXX.pdf --compressed -H 'User-Agent: Mozilla/5.0 (X11; Linux x86_64; rv:121.0) Gecko/20101001 Firefox/121.0' -H 'Accept: */*' -H 'Accept-Language: en-US,en;q=0.5' -H 'Accept-Encoding: gzip, deflate, br' -H 'Referer: https://emergingtalent.contentcontroller.com/ScormEngineInterface/defaultui/player/cms-a-wj/Default/cms-media/la.html' -H 'Connection: keep-alive' -H 'Cookie: CloudFront-Policy=XXXXX; CloudFront-Key-Pair-ID=XXXXXXX' -H 'Sec-Fetch-Dest: empty' -H 'Sec-Fetch-Mode: cors' -H 'Sec-Fetch-Site: same-origin' -H 'TE: trailers'
```

Make sure your header has a referer and the cookies. Also, try doing right click and resend the request and see if you get status code 200

Nani_The_Fock • 1y ago

I tried this, CMD is telling me that "-compressed" isn't compatible with my libcurl version so i got rid of that tag. The download works if i add "-o file.pdf" to the end of the cURL string, but the pdf itself cannot be opened.

Any suggestions? How do I go about updating my libcurl version? Is the "-compressed" tag actually required?

Funny enough, copying as cURL doesn't seem to work on Chromium (I'm using Brave) for some reason? Works on Firefox just fine though.

esreavib • 1y ago • Edited 1y ago

i could probably automate it even further but i couldn't be bothered

I've just spent hours trying to automate this by scraping the modules pages to obtain the video, subs and pdf files, but i could never get selenium to extract the required URLs from the nested iframes where they reside (requests did even earlier).

Do you have any pointers on how to accomplish this? My goal would be to make an offline copy of the course content so that i could keep studying the vids and PDFs without connection. Having to go through each page with the DevTools Network tab open and copy manually every single URL is mind numbing :)

Alone_Location 6184 • 1y ago

Thank, but i get:
Traceback (most recent call last):
File "script_download_aws.py", line 38, in <module>
header_dict[0] = i[1]
IndexError: list index out of range

AccomplishedDetail42 • 7mo ago

THIS WORKS!!

benjtom • 1mo ago

You saved me so much hassle, it worked like a charm. Thank you so much for this.

Inoob • 2y ago

I have followed the same way but only 155 kb of file is being downloaded. What might be the reason for it?

Nani_The_Fock • 1y ago

Doesn't work. Copied as cURL (Windows), pasted into CMD, lastly added "-o filename.mp4". Gave me an unopenable .mp4 file in my User directory.

Maybe something changed on AWS side?

(deleted) • 2y ago

Fit Relationship7077 • 5d ago

thank you so much for this. it works asf

Zarhor • 1y ago

to download the pdf files, just search to <https://emergingtalent.contentcontroller.com/ScormEngineInterface/dispatch/> and go to [location link](#), then use network tab on dev tools to take the pdf link without block. To download the videos i use fetchw extension

(deleted) • 1mo ago

Can confirm this is still working. Thank you!!

spingus • 9mo ago

Thank you for posting --I don't understand how to do it though!

I loaded the module with the PDF while dev tools is open. I found some occurrences of <https://emergingtalent.contentcontroller.com/ScormEngineInterface/dispatch/> and a couple had a location link (that was a lot longer than the one in the image you provided) and how I use that to find the pdf link without block on the network tab?

I really appreciate your help! I do not understand html :(

Zarhor • 9mo ago

I've checked here and they've changed it a bit, making a lot of things appear instead of the correct ones, but it's still easy to get it. Open the emergingtalent page with Inspect/Dev tools open, go to the Network tab and search for: <https://emergingtalent.contentcontroller.com/vault/>, the result will be the PDF.

spingus • 9mo ago

Thank you for the response! I got to exactly what you suggested...and then it gave me the message: 'Content can only be accessed by the launch process. Please launch your course again.'

the urls are pdfs, just not accessible:-