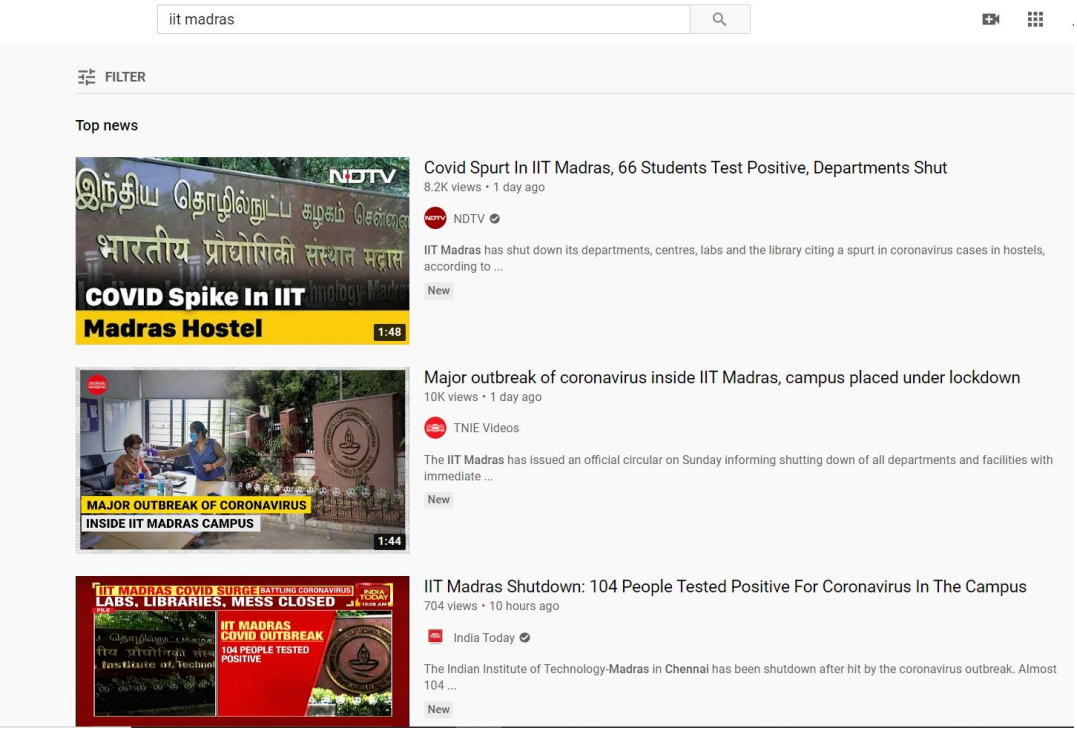
Problem 8(iii) Framework for Youtube search

- Since about 500 hours of content is uploaded to YouTube every minute, so storing the videos in a proper data structure is very important, for searching.

- First of all, for videos to be searched easily they need to be sorted in some manner, but sorting in alphabetical order is not feasible since so much content is uploaded every min, so we can't just keep on sorting them alphabetically.

- Neither sorting alphabetically is useful because e.g. if we search for IIT Madras, clearly not all the results start with the word IIT Madras, rather the video name contains IIT Madras somewhere in the middle (as shown in the image below), so sorting alphabetically doesn't make sense.

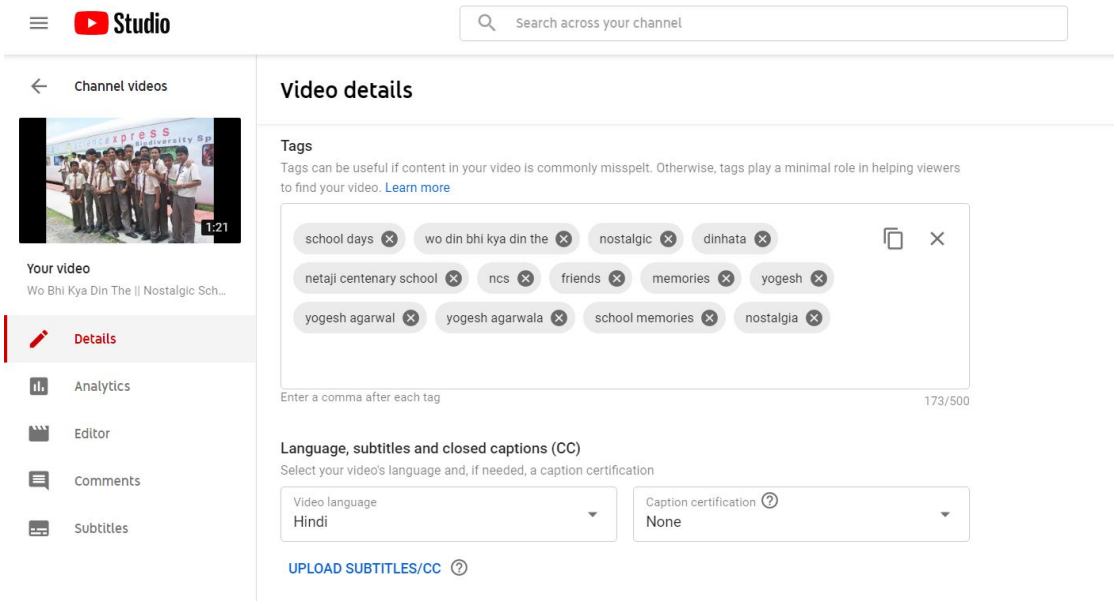


- If not alphabetically, then how should we sort the videos? One way is to sort the videos based on relevance (likes, views,comments) but the problem with this is that we can't compare e.g. a **music video** with a **sports video** so how do we decide relevance?

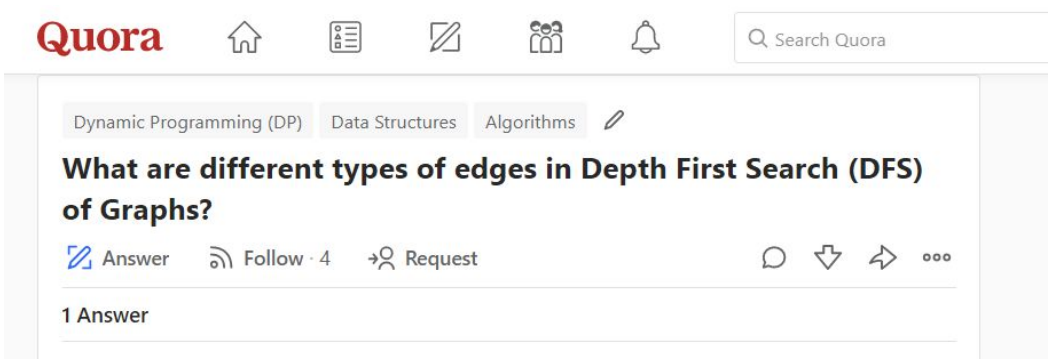
My suggestion:

- We should create different categories of topics e.g. when a video named "***Covid Spurt In IIT Madras, 66 Students Test Positive, Departments Shut***" is uploaded then while uploading based on the keywords used in the video name and the description we can put the video in several categories e.g. **this video can be put in categories like Covid, IIT Madras, IIT etc.**

- These keywords can be **auto-generated** (this technique is used by **quora** as well) based on the keywords used in the video name and description. Also the uploader can be given a option to use Tags **manually** (as shown in image below), which will help youtube to put the video in those categories and make it easily searchable, but there must be a limit to the number of keywords that can be used so that the uploader uses only relevant keywords.



- Similarly, quora generate tags for each question to make searching the questions easier. (**in below question we can see three tags** Dynamic Programming , Data Structures, Algorithms)



Benefits of categories/tags:

- We can use hashtables to store videos in each category, as then we can store the video name as the **key** and likes, views etc. as its **value**

```
Category1={
        'name1' :{
                Likes: 47k
                Likes (vs. dislikes): 23
                Watchtime: 12hrs
                Comments: 428
        },
        ‘name2’ :{
                Likes: 4k
                Likes (vs. dislikes): 53
                Watchtime: 2hrs
                Comments: 351
        }
        .
        .
        .
}
Category2={
        .
        .
        .
}
```

- So once we put the videos in a category, we can sort the videos in each category based on number of views, likes, comments, click through rates and few other factors.

- When a search is made we can look for the category and from the category we can suggest the best video, as each category is presorted based on relevance.

- This is model is much better than storing all videos in a stack, because the amount of content uploaded in each minute is so huge that we won’t be able to look for the best video, without using the concept of keywords.