from youtube\_transcript\_api import YouTubeTranscriptApi

from youtube\_transcript\_api.formatters import TextFormatter

import pandas as pd

df = pd.read\_excel(r'C:\Users\Aletta Meinsma\surfdrive2\Shared\SCS\projects\Quantum in TEDx talks\Paper\All\_data\_1.xlsx', sheet\_name = 'List\_data')

videoid\_list = df['VideoID'].tolist()

title\_list = df['Title'].tolist()

speaker\_list = df['Speaker'].tolist()

place\_list = df['Place'].tolist()

Year\_list = df['Year'].tolist()

Nr\_list = df['Number'].tolist()

for nr, videoId, title, speaker, place, year in zip(Nr\_list, videoid\_list,title\_list, speaker\_list, place\_list, Year\_list):

try:

transcript = YouTubeTranscriptApi.get\_transcript(videoId)

formatter = TextFormatter()

text\_formatted = formatter.format\_transcript(transcript)

path = r'C:\Users\Aletta Meinsma\Dropbox\PhD\Research\Project 1\Try\_Transcripts'

file\_name = str(videoId)

with open(path+'/'+str(nr) + '. ' + file\_name+'.txt', 'w', encoding='utf-8') as txt\_file:

txt\_file.write(title + '\n' + speaker + ', ' + place + ', ' + str(year) + '\n')

txt\_file.writelines(text\_formatted)

except:

path = r'C:\Users\Aletta Meinsma\Dropbox\PhD\Research\Project 1\Try\_Transcripts'

file\_name = str(videoId)

with open(path+'/'+str(nr)+ '. ' + file\_name+'.txt', 'w', encoding='utf-8') as txt\_file:

txt\_file.write(title + '\n' + speaker + ', ' + place + ', ' + str(year) + '\n')

txt\_file.writelines("TranscriptsDisabled")