In [1]:

```
!aws s3 cp s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/ s3://jaybucket-demo1/videos/ --recursive
```

```
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Sect01.mp4 to s3://jaybucket-demo1/videos/Mod02
_Sect01.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Intro.mp4 to s3://jaybucket-demo1/videos/Mod02_
Intro.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Sect02.mp4 to s3://jaybucket-demo1/videos/Mod02
_Sect02.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_WrapUp.mp4 to s3://jaybucket-demo1/videos/Mod02
_WrapUp.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Intro.mp4 to s3://jaybucket-demo1/videos/Mod03_
Intro.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod01_Course Overview.mp4 to s3://jaybucket-demo1/vid
eos/Mod01_Course Overview.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Sect03.mp4 to s3://jaybucket-demo1/videos/Mod02
_Sect03.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect02_part2.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect02_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Sect05.mp4 to s3://jaybucket-demo1/videos/Mod02
_Sect05.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect03_part1.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect03_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect02_part1.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect02_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect03_part3.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect03_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect03_part2.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect03_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod02_Sect04.mp4 to s3://jaybucket-demo1/videos/Mod02
_Sect04.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect01.mp4 to s3://jaybucket-demo1/videos/Mod03
_Sect01.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect04_part2.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect04_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect04_part3.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect04_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect02_part3.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect02_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect07_part1.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect07_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect04_part1.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect04_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect05.mp4 to s3://jaybucket-demo1/videos/Mod03
```

```
_Sect05.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect06.mp4 to s3://jaybucket-demo1/videos/Mod03
_Sect06.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect07_part2.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect07_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect07_part3.mp4 to s3://jaybucket-demo1/video
s/Mod03_Sect07_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_Sect08.mp4 to s3://jaybucket-demo1/videos/Mod03
_Sect08.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod03_WrapUp.mp4 to s3://jaybucket-demo1/videos/Mod03
_WrapUp.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_Intro.mp4 to s3://jaybucket-demo1/videos/Mod04_
Intro.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_Sect01.mp4 to s3://jaybucket-demo1/videos/Mod04
_Sect01.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_Sect02_part1.mp4 to s3://jaybucket-demo1/video
s/Mod04_Sect02_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_Sect02_part2.mp4 to s3://jaybucket-demo1/video
s/Mod04_Sect02_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_WrapUp.mp4 to s3://jaybucket-demo1/videos/Mod04
_WrapUp.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect02_part2.mp4 to s3://jaybucket-demo1/video
s/Mod05_Sect02_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod04_Sect02_part3.mp4 to s3://jaybucket-demo1/video
s/Mod04_Sect02_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Intro.mp4 to s3://jaybucket-demo1/videos/Mod05_
Intro.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect02_part1_ver2.mp4 to s3://jaybucket-demo1/v
ideos/Mod05_Sect02_part1_ver2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect03_part1.mp4 to s3://jaybucket-demo1/video
s/Mod05_Sect03_part1.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_WrapUp_ver2.mp4 to s3://jaybucket-demo1/videos/
Mod05_WrapUp_ver2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod06_Intro.mp4 to s3://jaybucket-demo1/videos/Mod06_
Intro.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect01_ver2.mp4 to s3://jaybucket-demo1/videos/
Mod05_Sect01_ver2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect03_part3.mp4 to s3://jaybucket-demo1/video
s/Mod05_Sect03_part3.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect03_part4_ver2.mp4 to s3://jaybucket-demo1/v
ideos/Mod05_Sect03_part4_ver2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod05_Sect03_part2.mp4 to s3://jaybucket-demo1/video
```

```
s/Mod05_Sect03_part2.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod06_WrapUp.mp4 to s3://jaybucket-demo1/videos/Mod06
_WrapUp.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod07_Sect01.mp4 to s3://jaybucket-demo1/videos/Mod07
_Sect01.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod06_Sect02.mp4 to s3://jaybucket-demo1/videos/Mod06
_Sect02.mp4
copy: s3://aws-tc-largeobjects/CUR-TF-200-ACMNLP-1/video/Mod06_Sect01.mp4 to s3://jaybucket-demo1/videos/Mod06
_Sect01.mp4
```

In [23]:
```
{
    "Effect": "Allow",
    "Action": "transcribe:StartTranscriptionJob",
    "Resource": "arn:aws:transcribe:<region>:<account-id>:transcription-job/*"
}
```

Out[23]:
```
{'Effect': 'Allow',
 'Action': 'transcribe:StartTranscriptionJob',
 'Resource': 'arn:aws:transcribe:<region>:<account-id>:transcription-job/*'}
```

In [26]:
```
import json
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
with open('./combined_output.json') as f:
    alldata = json.load(f)
print(alldata.keys())
```

```
dict_keys(['demo1-job_name_1.1_1', 'demo1-job_name_1.1_2', 'demo1-job_name_1.1_3', 'demo1-job_name_1.1_4', 'de
mo1-job_name_1.1_5', 'demo1-job_name_1.1_6', 'demo1-job_name_1.1_7', 'demo1-job_name_1.1_8', 'demo1-job_name_
1.1_9', 'demo1-job_name_1.1_10', 'demo1-job_name_1.1_11', 'demo1-job_name_1.1_12', 'demo1-job_name_1.1_13', 'd
emo1-job_name_1.1_14', 'demo1-job_name_1.1_15', 'demo1-job_name_1.1_16', 'demo1-job_name_1.1_17', 'demo1-job_n
ame_1.1_18', 'demo1-job_name_1.1_19', 'demo1-job_name_1.1_20', 'demo1-job_name_1.1_21', 'demo1-job_name_1.1_2
2', 'demo1-job_name_1.1_23', 'demo1-job_name_1.1_24', 'demo1-job_name_1.1_25', 'demo1-job_name_1.1_26', 'demo1
-job_name_1.1_27', 'demo1-job_name_1.1_28', 'demo1-job_name_1.1_29', 'demo1-job_name_1.1_30', 'demo1-job_name_
1.1_31', 'demo1-job_name_1.1_32', 'demo1-job_name_1.1_33', 'demo1-job_name_1.1_34', 'demo1-job_name_1.1_35',
'demo1-job_name_1.1_36', 'demo1-job_name_1.1_37', 'demo1-job_name_1.1_38', 'demo1-job_name_1.1_39', 'demo1-job
_name_1.1_40', 'demo1-job_name_1.1_41', 'demo1-job_name_1.1_42', 'demo1-job_name_1.1_43', 'demo1-job_name_1.1_
44', 'demo1-job_name_1.1_45', 'demo1-job_name_1.1_46'])
```

In [32]:
```
nltk.download('punkt')
nltk.download('wordnet')
lemmatizer = WordNetLemmatizer()
for k, v in alldata.items():
    tokens = word_tokenize(v)
    print(tokens)
    normalized_text = " ".join([lemmatizer.lemmatize(token) for token in tokens])
    print(normalized_text)
    alldata[k] = normalized_text
```

```
[nltk_data] Downloading package punkt to /home/ec2-user/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package wordnet to /home/ec2-user/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!

['Hi', 'and', 'welcome', 'to', 'Amazon', 'Academy', 'of', 'Machine', 'Learning', 'Foundations', 'in', 'this',
'module', ',', 'you', "'ll", 'learn', 'about', 'the', 'course', 'objectives', ',', 'various', 'job', 'roles',
'in', 'the', 'machine', 'learning', 'domain', 'and', 'where', 'you', 'can', 'go', 'to', 'learn', 'more', 'abou
t', 'machine', 'learning', '.', 'After', 'completing', 'this', 'module', ',', 'you', 'should', 'be', 'able',
'to', 'identify', 'course', 'prerequisites', 'and', 'objectives', 'indicate', 'the', 'role', 'of', 'the', 'dat
a', 'scientist', 'in', 'business', 'and', 'identify', 'resources', 'for', 'further', 'learning', '.', 'We',
"'re", 'now', 'going', 'to', 'look', 'at', 'the', 'prerequisites', 'for', 'taking', 'this', 'course', '.', 'Be
fore', 'you', 'take', 'this', 'course', ',', 'we', 'recommend', 'that', 'you', 'first', 'complete', 'Aws', 'Ac
ademy', 'Cloud', 'Foundations', '.', 'You', 'should', 'also', 'have', 'some', 'general', 'technical', 'knowled
ge', 'of', 'it', 'including', 'foundational', 'computer', 'literacy', 'skills', 'like', 'basic', 'computer',
'concepts', ',', 'email', 'file', 'management', 'and', 'a', 'good', 'understanding', 'of', 'the', 'internet',
'.', 'We', 'also', 'recommend', 'that', 'you', 'have', 'intermediate', 'skills', 'with', 'Python', 'programmin
g', 'and', 'a', 'general', 'knowledge', 'of', 'applied', 'statistics', '.', 'Finally', ',', 'general', 'busine
ss', 'knowledge', 'is', 'important', 'for', 'this', 'course', '.', 'This', 'includes', 'insight', 'into', 'ho
w', 'information', 'technology', 'is', 'used', 'in', 'business', '.', 'It', "'s", 'also', 'important', 'to',
```

In [34]:
```
!pip install yake
```

```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://pip.repos.neuron.amazonaws.com
(https://pip.repos.neuron.amazonaws.com)
Collecting yake
  Downloading yake-0.4.8-py2.py3-none-any.whl (60 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 60.2/60.2 kB 4.9 MB/s eta 0:00:00
Collecting segtok
  Downloading segtok-1.5.11-py3-none-any.whl (24 kB)
Requirement already satisfied: networkx in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages
(from yake) (3.0)
Requirement already satisfied: numpy in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (fr
om yake) (1.22.3)
Requirement already satisfied: click>=6.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-package
s (from yake) (8.1.3)
Collecting tabulate
  Downloading tabulate-0.9.0-py3-none-any.whl (35 kB)
Requirement already satisfied: jellyfish in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages
(from yake) (0.9.0)
Requirement already satisfied: regex in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (fr
om segtok->yake) (2022.10.31)
Installing collected packages: tabulate, segtok, yake
Successfully installed segtok-1.5.11 tabulate-0.9.0 yake-0.4.8
```

In [38]:
```
import yake
with open('./combined_output.json') as f:
    data = json.load(f)
language = "en"
kw_extractor = yake.KeywordExtractor(lan=language )
for key, value in data.items():
    key_words = kw_extractor.extract_keywords(value)
    print(f" showing all the keywords from the videos {key}:")
    for k in key_words:
        print(k[0])
    print()
```

```
 showing all the keywords from the videos demo1-job_name_1.1_1:
Machine Learning
Machine
Learning
Amazon machine learning
Amazon
Section
machine learning pipeline
Amazon machine
data
machine learning problems
Machine Learning Foundations
managed Amazon machine
certified machine learning
machine learning services
Academy Cloud Foundations
module
Amazon Sagemaker
learn
```

In [39]:
```
pip install spacy
```

```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://pip.repos.neuron.amazonaws.com
(https://pip.repos.neuron.amazonaws.com)
Requirement already satisfied: spacy in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (3.
5.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy) (8.1.9)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy) (2.4.6)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy) (2.0.7)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy) (2.0.8)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /home/ec2-user/anaconda3/envs/python3/lib/python
3.10/site-packages (from spacy) (1.0.4)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy) (3.3.0)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/s
ite-packages (from spacy) (3.0.8)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/si
te-packages (from spacy) (1.1.1)
Requirement already satisfied: setuptools in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-package
s (from spacy) (65.6.3)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /home/ec2-user/anaconda3/envs/python3/lib/python
3.10/site-packages (from spacy) (3.0.12)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy) (2.28.1)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in /home/ec2-user/anaconda3/envs/python3/
lib/python3.10/site-packages (from spacy) (1.10.7)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy) (6.3.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.
10/site-packages (from spacy) (1.0.9)
Requirement already satisfied: packaging>=20.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pa
ckages (from spacy) (21.3)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy) (4.64.1)
Requirement already satisfied: pathy>=0.10.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pack
ages (from spacy) (0.10.1)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy) (0.7.0)
Requirement already satisfied: numpy>=1.15.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pack
```

```
ages (from spacy) (1.22.3)
Requirement already satisfied: jinja2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (f
rom spacy) (3.1.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from packaging>=20.0->spacy) (3.0.9)
Requirement already satisfied: typing-extensions>=4.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4->spacy) (4.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site
-packages (from requests<3.0.0,>=2.13.0->spacy) (2022.12.7)
Requirement already satisfied: charset-normalizer<3,>=2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.1.1)
Requirement already satisfied: idna<4,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packa
ges (from requests<3.0.0,>=2.13.0->spacy) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/s
ite-packages (from requests<3.0.0,>=2.13.0->spacy) (1.26.8)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.0.4)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site
-packages (from thinc<8.2.0,>=8.1.8->spacy) (0.7.9)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from typer<0.8.0,>=0.3.0->spacy) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pa
ckages (from jinja2->spacy) (2.1.1)
Note: you may need to restart the kernel to use updated packages.
```

In [40]:
```
!python -m spacy download en_core_web_sm
```

```
Looking in indexes: https://pypi.org/simple, (https://pypi.org/simple,) https://pip.repos.neuron.amazonaws.com
(https://pip.repos.neuron.amazonaws.com)
Collecting en-core-web-sm==3.5.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.5.0/en_core_web_sm-
3.5.0-py3-none-any.whl (https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.5.0/en_co
re_web_sm-3.5.0-py3-none-any.whl) (12.8 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.8/12.8 MB 39.7 MB/s eta 0:00:0000:0100:01
Requirement already satisfied: spacy<3.6.0,>=3.5.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from en-core-web-sm==3.5.0) (3.5.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (8.1.9)
Requirement already satisfied: pathy>=0.10.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pack
ages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (0.10.1)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /home/ec2-user/anaconda3/envs/python3/lib/python
3.10/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.0.4)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (4.64.1)
Requirement already satisfied: numpy>=1.15.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pack
ages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.22.3)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.28.1)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.4.6)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/si
te-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.1.1)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4 in /home/ec2-user/anaconda3/envs/python3/
lib/python3.10/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.10.7)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.0.7)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (6.3.0)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/s
ite-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.0.8)
Requirement already satisfied: setuptools in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-package
s (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (65.6.3)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.3.0)
Requirement already satisfied: typer<0.8.0,>=0.3.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (0.7.0)
```

```
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /home/ec2-user/anaconda3/envs/python3/lib/python
3.10/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.0.12)
Requirement already satisfied: jinja2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packages (f
rom spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.1.2)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.
10/site-packages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.0.9)
Requirement already satisfied: packaging>=20.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pa
ckages (from spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from packaging>=20.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.0.9)
Requirement already satisfied: typing-extensions>=4.2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from pydantic!=1.8,!=1.8.1,<1.11.0,>=1.7.4->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (4.4.
0)
Requirement already satisfied: idna<4,>=2.5 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-packa
ges (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/s
ite-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (1.26.8)
Requirement already satisfied: certifi>=2017.4.17 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site
-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2022.12.7)
Requirement already satisfied: charset-normalizer<3,>=2 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from requests<3.0.0,>=2.13.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.1.1)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.1
0/site-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (0.0.4)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site
-packages (from thinc<8.2.0,>=8.1.8->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (0.7.9)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/sit
e-packages (from typer<0.8.0,>=0.3.0->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (8.1.3)
Requirement already satisfied: MarkupSafe>=2.0 in /home/ec2-user/anaconda3/envs/python3/lib/python3.10/site-pa
ckages (from jinja2->spacy<3.6.0,>=3.5.0->en-core-web-sm==3.5.0) (2.1.1)
✔ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')
```

In [62]:
```
import spacy
import json
nlp = spacy.load('en_core_web_sm')
all_keyphrases = {}
with open('./combined_output.json') as f:
    data = json.load(f)
for video_id, video_text in data.items():
    doc = nlp(video_text)
    k_phrases = [chunk.text for chunk in doc.noun_chunks]
    print('Video:', video_id)
    print('All the text from the video:', video_text)
    print('All the key phrases from the video:', k_phrases)
    all_keyphrases[video_id] = k_phrases
with open('all_keyphrases.json', 'w') as f:
    json.dump(all_keyphrases, f)
print('All the key phrases saved to all_keyphrases.json')
```

```
Video: demo1-job_name_1.1_1
All the text from the video: Hi and welcome to Amazon Academy of Machine Learning Foundations in this module,
you'll learn about the course objectives, various job roles in the machine learning domain and where you can g
o to learn more about machine learning. After completing this module, you should be able to identify course pr
erequisites and objectives indicate the role of the data scientist in business and identify resources for furt
her learning. We're now going to look at the prerequisites for taking this course. Before you take this cours
e, we recommend that you first complete Aws Academy Cloud Foundations. You should also have some general techn
ical knowledge of it including foundational computer literacy skills like basic computer concepts, email file
management and a good understanding of the internet. We also recommend that you have intermediate skills with
Python programming and a general knowledge of applied statistics. Finally, general business knowledge is impor
tant for this course. This includes insight into how information technology is used in business. It's also imp
ortant to have business related skill sets such as communication skills, leadership skills, and an orientation
towards customer service. In this course, you'll be introduced to the key concepts of machine learning, its to
ols and its uses you'll also be introduced to and work with some of the AWS services for machine learning. Yo
u'll learn how to recognize how machine learning and deep learning are part of artificial intelligence. Descri
be artificial intelligence and machine learning terminology. Identify how machine learning can be used to solv
e a business problem. Describe the machine learning process. List the tools available to data scientists and i
dentify when to use machine learning instead of traditional software development methods. As part of this cour
se, you'll also learn how to implement a machine learning pipeline. This includes how to formulate a problem f
```

In [97]:
```
combine_output_dict = {}
with open('./combined_output.json', 'r') as f:
    for line in f:
        line_dict = json.loads(line)
        combine_output_dict.update(line_dict)
with open('combine_output_1.json', 'w') as f:
    json.dump(combine_output_dict, f)
print(f"combine output saved to combine_output_1.json")
```

```
combine output saved to combine_output_1.json
```

In [98]:
```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /home/ec2-user/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[98]:
```
True
```

In [94]:
```
import spacy
import string
import pandas as pd
from ipywidgets import widgets
from nltk.corpus import stopwords
from IPython.display import clear_output
def text_normalize(t):
    t = t.lower()
    t = t.translate(str.maketrans("", "", string.punctuation))
    stop_words = set(stopwords.words('english'))
    tokens = t.split()
    filtered_tokens = [token for token in tokens if token not in stop_words]
    t = " ".join(filtered_tokens)
    return t
def keywords_match(query_words, transcript_words):
    match_words = query_words.intersection(transcript_words)
    s = len(match_words) / len(query_words)
    return s
def seek_videos(button):
    clear_output(wait=True)
    query_Str = inputbox.value
    l = l_dropdown.value.lower()
    minimum_score = score_slider.value
    query = text_normalize(query_Str)
    querywords = set(extract_key_phrases(query))
    print("Query keywords are : {}".format(querywords))
    if len(querywords) == 0:
        with output_box:
            print("sorry!there are no any videos: {}".format(query_Str))
    else :
        df = pd.read_csv("Intermediate_res.csv")
        df['transcription_key_tags'] = df['transcription_key_tags'].apply(lambda x: set(x.split()))
        df['matching_score'] = df['transcription_key_tags'].apply(lambda x: keywords_match(querywords, x))
        df = df.sort_values(by='matching_score', ascending=False)
        videos = df[df['matching_score'] >= minimum_score].head(10)
        with output_box:
            display(videos[['path']])

inputbox = widgets.Text(
    placeholder='Enter your search query',
    description='Search:',
```

```
    layout=widgets.Layout(width='40%')
)
inputbox.style.background = '#F8F8F8'
l_dropdown = widgets.Dropdown(
    options=['English', 'Spanish', 'French','Gujarati','Hindi'],
    value='English',
    description='Language:',
    layout=widgets.Layout(width='25%')
)
l_dropdown.style.background = '#F8F8F8'
score_slider = widgets.FloatSlider(
    value=0.5,
    min=0,
    max=1,
    step=0.1,
    description='Min Score:',
    layout=widgets.Layout(width='50%')
)
score_slider.style.background = '#F8F8F8'
search_button = widgets.Button(description='Search')
search_button.style.button_color = 'green'
output_box = widgets.Output()
search_button.on_click(seek_videos)
display(widgets.VBox([inputbox, l_dropdown, score_slider, search_button, output_box]))
```

```
VBox(children=(Text(value='', description='Search:', layout=Layout(width='40%'), placeholder='Enter your sear
c…
```

In [ ]: