

# Web Scraper

```
In [1]: from bs4 import BeautifulSoup
import requests
import csv
import re
```

```
In [2]: ## The Indian Express
def indian_express_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text,"lxml")
    if article.find(id="pcl-full-content") == None:
        return None
    else:
        content = ""
        for para in article.find('div', id = "pcl-full-content").find_all("p"):
            content+=para.get_text()
        return content
```

```
In [3]: url = "https://indianexpress.com/article/india/diwali-pollution-air-quality-index-delhi-mumbai-kolkata-8228093/"
indian_express_get_content(url)
```

```
Out[3]: 'As Diwali was celebrated across the country Monday, the air quality index (AQI) in the National Capital Region (NCR) plummeted to 'very poor', while other major cities recorded 'moderate' or 'satisfactory' air quality. Delhi's air quality hit the "very poor" category with an AQI of 312. This is the first 'very poor' air day of the season for the national capital. Gurgaon and Noida also recorded air quality in the 'very poor' category, with respective AQIs of 322 and 305. An AQI between 301 and 400 is considered to be "very poor" and can cause respiratory illness on prolonged exposure, according to the Central Pollution Control Board (CPCB). The System of Air Quality and Weather Forecasting and Research (SAFAR), meanwhile, termed the air quality across Mumbai and Pune as 'moderate' and 'satisfactory', respectively. The CPCB recorded an AQI value of 136 yesterday, with the major pollutants being PM10 and PM2.5 in Mumbai. According to the CPCB, Bengaluru recorded an AQI index value of 100 with a 'satisfactory' air quality. A 'moderate' AQI value of 192 was recorded in Chennai. In line with a Supreme Court order, the Tamil Nadu government had specified that crackers should only be burst from 6 am to 7 am and 7 pm to 8 pm. The CPCB termed the air quality of the city of Kolkata as 'satisfactory' with an AQI value of 55. The major pollutants were identified as PM10 and NO2. Recently, the West Bengal Environment Minister Manas Bhuniya had said the new central control room at Paribesh Bhavan is monitoring sound and air pollution in real time on Diwali and Kali Puja.'
```

```
In [4]: ## The India Today
def india_today_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text,"lxml")
    title = article.find("h1").get_text()
    if (article.find('div',{ "class": "description" }) == None):
        return None, None, None
    else:
        content = ""
        for para in article.find(class_ = "description").find_all("p")[:-3]:
            content+=para.get_text()
        return content
```

```
In [5]: url = "https://www.indiatoday.in/india/story/coimbatore-blast-accused-call-recordings-isis-2292127-2022-11-01"
india_today_get_content(url)
```

```
Out[5]: "By Pramod Madhav: Junaidha Begum, the mother of one of the accused in the Coimbatore car blast case, has claimed that cops that her son only went to help Jamesha Mubeen shift his home. Mubeen was killed in the incident that occurred in Tamil Nadu's Coimbatore on October 23. Soon after the blast, the Coimbatore Police launched an investigation and found bomb-making chemicals at the house of the deceased. Six accused were arrested. They were identified as Mohammed Dalqa, Mohammed Azharuddin, Muhamed Riyaz, Feroz Ismail, Afzar, and Muhammad Nawaz Ismail. ALSO READ | 'My son being portrayed in bad light': Mother of accused who sold car used in Coimbatore blast According to the police, Mubeen was on his way to a target and was trying to avoid a checkpoint when the blast occurred, killing him as he tried to leave the car. Officials are not yet sure about what his target was. Footage from CCTV near Mubeen's residence showed the four men struggling to carry an item wrapped in a plastic sack. Suspicion arose whether this was the bomb that killed Jamesha. Officials are investigating the CCTV footage to find out what the group were carrying. Dalqa had purportedly sold the car to Mubeen, in which he was carrying the explosive. Hafsath Biwi, the mother of Dalqa, claimed that he had sold the car for a commission since he was dealing in second-hand cars and had no connection with the blast. Meanwhile, Riyaz's mother Junaidha Begum also told the police that Feroz and Riyaz had gone to Mubeen's house as he had called them to help him shift his home. ALSO READ | Coimbatore blast: TN BJP chief claims police had details on accused being 'ISIS sympathiser' India Today has now accessed the call recordings between Riyaz and his mother and Riyaz and Feroz on the day of the blast. Here are the audio transcripts of the calls. Audio 1 Junaidha Begum: Where are you Riyaz? \nRiyaz - I'm going to help him shift his house. Junaidha Begum: Where is the vehicle? \nRiyaz: I've given it to my brother-in-law. Junaidha Begum: I called you to see if we could get the auto. \nRiyaz: Is it? I have given the car to my brother-in-law. I don't have the auto with me either. Junaidha Begum: Where are you going? \nRiyaz: I'm out with Feroz. Junaidha Begum: Where are you? \nRiyaz: Near Kottaimedu (where Mubeen lived). Audio 2 Junaidha Begum: Where are you? \nRiyaz: I'm helping with shifting items at Mubeen's house. Junaidha Begum: Who all are there? \nRiyaz: Feroz, Nawaz, and I. Junaidha Begum: Return early. \nRiyaz: It's almost over. Audio 3 (Conversation between Feroz and Riyaz) Feroz: Hey, come here. Come quickly. \nRiyaz: Can't. It will take time. Feroz: How long? \nRiyaz: Maybe 15 minutes. Feroz: Okay, come fast. \nRiyaz: Why? What happened? Feroz: He (Mubeen) is calling to help him shift home. \nRiyaz: You go and help him. I can't go. Feroz: Why? \nRiyaz: There are people at home. Feroz: Okay, come here for a minute. \nRiyaz: I'm not going da."
```

```
In [6]: ## NDTV
```

```
def ndtv_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text, "lxml")
    if article.find('div',{'id':"ins_storybody"}) == None:
        return None, None
    else:
        content = ""
        for para in article.find('div',{'id':"ins_storybody"}).find_all("p"):
            content+=para.get_text()
        return content
```

In [7]: url = "https://www.ndtv.com/india-news/shallow-water-with-rocks-led-to-casualties-ndrf-chief-on-gujarat-cable-b  
ndtv\_get\_content(url)

Out[7]: '135 people died in the bridge collapse tragedy on SundayThe chief of India's disaster response force has thro  
wn some light on what may have caused the high casualties after a cable bridge snapped in Gujarat's Morbi city  
on Sunday, killing 135.The water towards both ends of the century-old cable bridge is shallow, about 10 feet or  
the length of a small car, and the people hit the rocky bottom after falling into the river with great force, k  
illing them, VVN Prasanna Kumar, the Commandant of the National Disaster Response Force (NDRF), which is part o  
f the search team in the River Machhu, told NDTV on Tuesday.The middle area of the river, whose water is stagna  
nt with almost no flow, is about 20 feet deep, Mr Kumar said, adding the navy divers who are searching for bodi  
es effortlessly reached the bottom of the river."But the water is muddy and visibility is low," Mr Kumar said.  
"We have been using an ultramodern remote underwater vehicle fitted with a camera and a sonar device to locate  
bodies," the NDRF Commandant told NDTV. "The floor of the river is full of rocks. This is why many people were  
injured and killed," he added.NDRF Commandant VVN Prasanna Kumar said they are using advanced tools to locate b  
odiesMost of the bodies were found under the broken bridge as the river did not have any flow and the water did  
not carry them away, Mr Kumar said.He said there could be only one or two bodies left to be found, based on dat  
a of missing people given by the civil administration.Prime Minister Narendra Modi on Tuesday\visited the ho  
spital in Morbi city where the injured people are being treated. The facility got a quick-fix facelift overnigh  
t ahead of his visit.The narrow cable bridge collapsed due to a huge rush of people, sources in the Gujarat for  
ensics laboratory have said. Some of the old cables that held the bridge were not changed during the seven-mont  
h renovation that started in March this year, sources have said.'

In [8]: ## Times of India  
def toi\_get\_content(url):  
 data = requests.get(url).content  
 soup = BeautifulSoup(data, "lxml")  
 div = soup.find("div", class\_ = "\_3YYSt clearfix ")  
 if div is None:  
 div = soup.find("div", class\_ = "\_3YYSt clearfix")  
 if div is not None:  
 return "".join(list(div.strings))  
 return None, None

In [9]: url = "https://timesofindia.indiatimes.com/city/ahmedabad/morbi-tragedy-why-safety-sops-for-amusement-rides-don  
toi\_get\_content(url)

Out[9]: 'AHMEDABAD: In 2019, when an amusement park ride in Kankaria crashed, killing two and injuring 28, the state go  
vernment tasked a seven-member committee with preparing standard operating procedures (SOPs) to enforce safety  
protocols for all recreational rides. However, the SOPs were not made applicable to suspension bridges where si  
ghtseers are allowed ticketed entry because the activity is not classified as 'recreational'. Though separate  
rules exist for bridges, these are mostly commuter bridges. Had Morbi's suspension bridge been put under the '  
'recreational ride' category, regular safety checks and accountability could have further prevented the traged  
y. Read Also135 killed in Morbi bridge collapse, 170 rescued: Gujarat minister Rajendra TrivediAs many as 135 p  
eople have been killed in the Morbi bridge collapse tragedy and 170 others rescued, Gujarat minister Rajendra T  
rivedi said on Tuesday."All the districts have technical committees in place that inspect recreational rides an  
d issue safety certificates. However, as suspension footbridges like the one in Morbi are not classified as 'r  
ecreational', despite being ticketed, the rules for amusement rides don't apply to bridges," said an official  
in the state roads and building department who did not wish to be named. In Ahmedabad alone, between 2012 and 2  
020, there have been seven major accidents at melas or temporary fairs. These include deaths caused due to elec  
trocution, fall, and injuries. Read AlsoWhy govt has a lot to answer in Morbi tragedyAs the death toll in the b  
ridge collapse mounted to 135, local authorities have been quick to blame the private contractor for the lapses  
that led to the tragedy. But there are several questions they still need to answerThe rules that were framed by  
the seven-member committee after the Kankaria disaster came into effect in February 2020 in Ahmedabad district.  
The SOPs were the first of their kind; they ensured safety checks and inspection rules for recreational rides.  
Later, in 2021, the state government, through a notification, asked all districts to set up similar technical c  
ommittees to implement these rules. Read AlsoGujarat Morbi Footbridge: 'Renovation firm only painted, polished  
cables'"The contractor seems to have just painted and polished all the cables ahead of the bridge being reopen  
ed to the public on October 26. We haven't found anything so far to confirm if any of those worn-out cables we  
re changed," said a police officer privy to the probe. "We are now looking into thisThe seven-member committee  
set up for Ahmedabad district under the chairmanship of executive engineer of roads and building department inc  
luded the R&B department's deputy executive engineer, two teachers from LD Engineering College and Ahmedabad P  
olytechnic, and officials of the industrial health and safety department and municipal corporation's engineeri  
ng department. So far, the committee has inspected and issued safety certificates to 16 amusement rides in the  
city. The suspension bridges were excluded from the purview of these committees, said roads and buildings depar  
tment sources. The state government could have set up a separate system to ensure safety inspections of such st  
ructures, they said. Read AlsoGujarat: Morbi tragedy exposes rot in civic systemThe Morbi tragedy has exposed t  
he rot in civic institutions and their failure to ensure safe and reliable infrastructure. The contractor, Ore  
v a group, was to approach the district collector for a structural safety certificate before beginning operations  
on October 26.Sources added that for the inspection of bridges, the operator is required to apply to the collec  
tor's office and then the R&B department issues structural stability certificates after inspection. The govern  
ment is yet to set up SOPs for these bridges, the sources added. '

In [10]: ## The Hindu  
def get\_content(url):  
 html = requests.get(url)  
 article = BeautifulSoup(html.text, "lxml")  
 content = ""

```

paras = article.find_all("p")[1:-5]
for para in paras:
    content+=para.get_text()
return content

```

```

In [11]: url = "https://www.thehindu.com/news/national/other-states/four-militants-killed-in-anti-militancy-operations-in-south-kashmir"
get_content(url)

```

```

Out[11]: 'CONNECT WITH US November 01, 2022 10:14 pm | Updated November 02, 2022 12:37 pm IST - SRINAGAR \nCOMMENTS\n SH
ARE\n\nREAD LATER\n\nSecurity personnel stand guard near encounter site at Samthan Bijbehara area of Anantnag i
n South Kashmir on November 1, 2022.\n] Photo Credit: Nissar Ahmad\nFour militants were killed in two separate
anti-militancy operations in south Kashmir's Anantnag and Pulwama districts on Tuesday, and a likely " fidayeen
" (suicide) attack in the Valley, the J&K Police said."Lashkar-e-Taiba (LeT) commander Mukhtar Bhat is among th
e three killed terrorists [in Pulwama's Awantipora area]. As per source, he, along with a foreign terrorist, wa
s going for ' fidayeen' attack on security forces' camp. The\xa0Awantipora police and the Army averted a major
terror incident," Additional Director General of Police Vijay Kumar said.\xa0One AK-47 rifle, one AK-56 rifle a
nd one pistol were recovered from the encounter site at Pulwama's Awantipora.Mr. Kumar said Bhat was involved i
n several terror crimes, including the killing of one Assistant Sub-Inspector of the Central Reserve Police For
ce and two Railway Protection Force personnel. "It's a big success for us," Mr. Kumar said.The third deceased p
erson, believed to be a local, could not be identified immediately.The militant killed in another anti-militanc
y operation in Anantnag's Bijbehara has been identified as Shakir Ahmad, a resident of Ladermud. "One terrorist
has been killed in the Bijbehara encounter. The operation is going on," the police said, adding, "The killed te
rrorist was linked with Hizbul Mujahideen and was involved in several terror crime cases, including attacks on
police and security forces."
Meanwhile, the J&K Police arrested three militants in Srinagar. "10 kilogrammes
bucket Improvised Explosive Device (IED) and two hand grenades were recovered from these militants. The IED is
being destroyed in-situ in the Rangreth area by the Bomb Disposal Squad. A case has been registered under Secti
ons of the Unlawful Activities Prevention Act, Arms Act, and Explosives Act," the police said.\xa0\nCOMMENTS\n'

```

```

In [12]: def news_extract(heading, url):

def india_today_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text,"lxml")
    if (article.find('div',{'class':"description"}) == None):
        return None
    else:
        content = ""
        for para in article.find(class_ = "description").find_all("p")[:3]:
            content+=para.get_text()
        return content

def indian_express_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text,"lxml")
    if article.find(id="pcl-full-content") == None:
        return None
    else:
        content = ""
        for para in article.find('div', id = "pcl-full-content").find_all("p"):
            content+=para.get_text()
        return content

def hindu_get_content(url):
    html = requests.get(url)
    article = BeautifulSoup(html.text,"lxml")
    content = ""
    paras = article.find_all("p")[1:-5]
    for para in paras:
        content+=para.get_text()
    return content

def ndtv_get_content(url):
    ar = requests.get(url)
    article = BeautifulSoup(ar.text,"lxml")
    if article.find('div',{'id':"ins_storybody"}) == None:
        return None, None
    else:
        content = ""
        for para in article.find('div',{'id':"ins_storybody"}).find_all("p"):
            content+=para.get_text()
        return content

def toi_get_content(url):
    data = requests.get(url).content
    soup = BeautifulSoup(data,"lxml")
    div = soup.find("div", class_ = "_3YYSt clearfix ")
    if div is None:
        div = soup.find("div", class_ = "_3YYSt clearfix")
    if div is not None:
        time = soup.find("div", class_="yYIU- byline").find("span").text
        time = re.sub("Updated:|IST|AM|PM", "", time, flags=re.IGNORECASE).strip()
        # time = datetime.strptime(time, "%b %d, %Y, %H:%M")
        return "".join(list(div.strings))
    return None

if (heading == 'India Today'): return india_today_get_content(url)
if (heading == 'The Indian Express'): return indian_express_get_content(url)
if (heading == 'The_Hindu'): return hindu_get_content(url)

```

```
if (heading == 'NDTV'): return ndtv_get_content(url)
if (heading == 'Times_of_India'): return toi_get_content(url)
```

```
In [13]: url = "https://www.indiatoday.in/india/story/coimbatore-blast-accused-call-recordings-isis-2292127-202
news_company = "India Today"
news_extract(news_company, url)
```

```
Out[13]: "By Pramod Madhav: Junaidha Begum, the mother of one of the accused in the Coimbatore car blast case, has claim
ed that cops that her son only went to help Jamesha Mubeen shift his home. Mubeen was killed in the incident th
at occurred in Tamil Nadu's Coimbatore on October 23. Soon after the blast, the Coimbatore Police launched an in
vestigation and found bomb-making chemicals at the house of the deceased. Six accused were arrested. They were
identified as Mohammed Dalqa, Mohammed Azharuddin, Muhamed Riyaz, Feroz Ismail, Afzar, and Muhammad Nawaz Ismai
l. ALSO READ | 'My son being portrayed in bad light': Mother of accused who sold car used in Coimbatore blast Ac
cording to the police, Mubeen was on his way to a target and was trying to avoid a checkpoint when the blast oc
curred, killing him as he tried to leave the car. Officials are not yet sure about what his target was. Footage
from CCTV near Mubeen's residence showed the four men struggling to carry an item wrapped in a plastic sack. Su
spicion arose whether this was the bomb that killed Jamesha. Officials are investigating the CCTV footage to fi
nd out what the group were carrying. Dalqa had purportedly sold the car to Mubeen, in which he was carrying the
explosive. Hafsath Biwi, the mother of Dalqa, claimed that he had sold the car for a commission since he was de
aling in second-hand cars and had no connection with the blast. Meanwhile, Riyaz's mother Junaidha Begum also to
ld the police that Feroz and Riyaz had gone to Mubeen's house as he had called them to help him shift his home.
ALSO READ | Coimbatore blast: TN BJP chief claims police had details on accused being 'ISIS sympathiser' India T
oday has now accessed the call recordings between Riyaz and his mother and Riyaz and Feroz on the day of the bl
ast. Here are the audio transcripts of the calls. Audio 1 Junaidha Begum: Where are you Riyaz? \nRiyaz - I'm going
to help him shift his house. Junaidha Begum: Where is the vehicle? \nRiyaz: I've given it to my brother-in-law. Ju
naidha Begum: I called you to see if we could get the auto. \nRiyaz: Is it? I have given the car to my brother-in
-law. I don't have the auto with me either. Junaidha Begum: Where are you going? \nRiyaz: I'm out with Feroz. Juna
idha Begum: Where are you? \nRiyaz: Near Kottaimedu (where Mubeen lived). Audio 2 Junaidha Begum: Where are you? \n
Riyaz: I'm helping with shifting items at Mubeen's house. Junaidha Begum: Who all are there? \nRiyaz: Feroz, Nawa
z, and I. Junaidha Begum: Return early. \nRiyaz: It's almost over. Audio 3 (Conversation between Feroz and Riyaz) Fe
roz: Hey, come here. Come quickly. \nRiyaz: Can't. It will take time. Feroz: How long? \nRiyaz: Maybe 15 minutes. F
eroz: Okay, come fast. \nRiyaz: Why? What happened? Feroz: He (Mubeen) is calling to help him shift home. \nRiyaz:
You go and help him. I can't go. Feroz: Why? \nRiyaz: There are people at home. Feroz: Okay, come here for a minut
e. \nRiyaz: I'm not going da."
```

```
In [14]: url = "https://www.thehindu.com/news/national/other-states/four-militants-killed-in-anti-militancy-operations-i
news_company = "The_Hindu"
news_extract(news_company, url)
```

```
Out[14]: 'CONNECT WITH US November 01, 2022 10:14 pm | Updated November 02, 2022 12:37 pm IST - SRINAGAR \nCOMMENTS\n SH
ARE\n\nREAD LATER\n\nSecurity personnel stand guard near encounter site at Samthan Bijbehara area of Anantnag i
n South Kashmir on November 1, 2022.\n| Photo Credit: Nissar Ahmad\nFour militants were killed in two separate
anti-militancy operations in south Kashmir's Anantnag and Pulwama districts on Tuesday, and a likely " fidayeen
" (suicide) attack in the Valley, the J&K Police said. "Lashkar-e-Taiba (LeT) commander Mukhtar Bhat is among th
e three killed terrorists [in Pulwama's Awantipora area]. As per source, he, along with a foreign terrorist, wa
s going for ' fidayeen' attack on security forces' camp. The\xa0Awantipora police and the Army averted a major
terror incident," Additional Director General of Police Vijay Kumar said.\xa0One AK-47 rifle, one AK-56 rifle a
nd one pistol were recovered from the encounter site at Pulwama's Awantipora. Mr. Kumar said Bhat was involved i
n several terror crimes, including the killing of one Assistant Sub-Inspector of the Central Reserve Police For
ce and two Railway Protection Force personnel. "It's a big success for us," Mr. Kumar said. The third deceased p
erson, believed to be a local, could not be identified immediately. The militant killed in another anti-militanc
y operation in Anantnag's Bijbehara has been identified as Shakir Ahmad, a resident of Ladermud. "One terrorist
has been killed in the Bijbehara encounter. The operation is going on," the police said, adding, "The killed te
rrorist was linked with Hizbul Mujahideen and was involved in several terror crime cases, including attacks on
police and security forces." \xa0Meanwhile, the J&K Police arrested three militants in Srinagar. "10 kilogrammes
bucket Improvised Explosive Device (IED) and two hand grenades were recovered from these militants. The IED is
being destroyed in-situ in the Rangreth area by the Bomb Disposal Squad. A case has been registered under Secti
ons of the Unlawful Activities Prevention Act, Arms Act, and Explosives Act," the police said.\xa0\nCOMMENTS\n'
```