ARYAMAN MISHRA

19BCE1027

LAB 2 WEB MINING

Wikipedia link: https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship

```
from urllib.request import urlopen

html = urlopen('https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship')
print(html.read())
```

```
b'<!DOCTYPE html>\n<html class="client-nojs" lang="en" dir="ltr">\n<head>\n<meta charset="UTF-8"/>\n<title>2020 Formula One World Championsh
```

```
[2] from urllib.request import urlopen
    from bs4 import BeautifulSoup

    html = urlopen('https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship')
    bs = BeautifulSoup(html.read(), 'html.parser')
    print(bs.h1)
```

```
<h1 class="firstHeading" id="firstHeading">2020 Formula One World Championship</h1>
```

```
from urllib.request import urlopen
from urllib.error import HTTPError
from urllib.error import URLError

try:
    html = urlopen("https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship")
except HTTPError as e:
    print("The server returned an HTTP error")
except URLError as e:
    print("The server could not be found!")
else:
    print(html.read())
```

```
b'<!DOCTYPE html>\n<html class="client-nojs" lang="en" dir="ltr">\n<head>\n<meta charset="UTF-8"/>\n<title>2020 Formula One World Championship - Wikipedia<
```

```
from urllib.request import urlopen
from urllib.error import HTTPError
from bs4 import BeautifulSoup


def getTitle(url):
    try:
        html = urlopen(url)
    except HTTPError as e:
        return None
    try:
        bsObj = BeautifulSoup(html.read(), "lxml")
        title = bsObj.body.h1
    except AttributeError as e:
        return None
    return title


title = getTitle("https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship")
if title == None:
    print("Title could not be found")
else:
    print(title)
```

```
<h1 class="firstHeading" id="firstHeading">2020 Formula One World Championship</h1>
```

```
nameList = bs.findAll('h3')
for name in nameList:
    print(name.get_text())
```

```
Free practice drivers[edit]
Team changes[edit]
Driver changes[edit]
Changes from the 2019 calendar to the original 2020 calendar[edit]
Sporting regulations[edit]
Technical regulations[edit]
Initial response[edit]
Race postponements and cancellations[edit]
Rescheduled calendar[edit]
Regulatory changes[edit]
Solidarity campaign[edit]
Mercedes[edit]
Racing Point[edit]
Opening rounds[edit]
Mid-season rounds[edit]
Closing rounds[edit]
Grands Prix[edit]
Scoring system[edit]
World Drivers' Championship standings[edit]
World Constructors' Championship standings[edit]
Personal tools
```

Namespaces

Variants

Views

More

Search

Navigation

Contribute

Tools

Print/export

In other projects

Languages

```
[4] titles = bs.find_all(['h1', 'h2','h3','h4','h5','h6'])
    print([title for title in titles])
```

```
[<h1 class="firstHeading" id="firstHeading">2020 Formula One World Championship</h1>, <h2 id="mw-toc-heading">Contents</h2>, <h2><span class="mw-headline" id="Entries">Entries</span>
</h3>, <h3 class="vector-menu-heading" id="p-namespaces-label"><span>Namespaces</span>
</h3>, <h3 class="vector-menu-heading" id="p-variants-label"><span>Variants</span>
</h3>, <h3 class="vector-menu-heading" id="p-views-label"><span>Views</span>
</h3>, <h3 class="vector-menu-heading" id="p-cactions-label"><span>More</span>
</h3>, <h3>
<label for="searchInput">Search</label>
</h3>, <h3 class="vector-menu-heading" id="p-navigation-label"><span>Navigation</span>
</h3>, <h3 class="vector-menu-heading" id="p-interaction-label"><span>Contribute</span>
</h3>, <h3 class="vector-menu-heading" id="p-tb-label"><span>Tools</span>
</h3>, <h3 class="vector-menu-heading" id="p-coll-print_export-label"><span>Print/export</span>
</h3>, <h3 class="vector-menu-heading" id="p-wikibase-otherprojects-label"><span>In other projects</span>
</h3>, <h3 class="vector-menu-heading" id="p-lang-label"><span>Languages</span>
</h3>]
```

```
allText = bs.find_all('span')
print([text for text in allText])
```

```
[<span class="summary"><span class="dtstart">2020</span> <a href="/wiki/Formula_One" title="Formula One">FIA Formula One</a><br/>World Championship</span>, <span class="dtstart">2020<
<ul><li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/defa
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
</span>, <span class="cs1-format">(PDF)</span>, <span class="cs1-format">(PDF)</span>, <span class="reference-accessdate">. Retrieved <span class="nowrap">2 July</span> 2020</span>, <
```

```
nameList = bs.find_all(text='F1')
print(len(nameList))
```

```
68
```

```
title = bs.find_all(id='title', class_='text')
print([text for text in allText])
```

```
[<span class="summary"><span class="dtstart">2020</span> <a href="/wiki/Formula_One" title="Formula One">FIA Formula One</a><br/>World Championship</span>, <span class="dtstart">2020<
<ul><li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/defa
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
<li><link href="mw-data:TemplateStyles:r999302996" rel="mw-deduplicated-inline-style"/><cite class="citation web cs1"><a class="external text" href="https://www.fia.com/sites/default/
</span>, <span class="cs1-format">(PDF)</span>, <span class="cs1-format">(PDF)</span>, <span class="reference-accessdate">. Retrieved <span class="nowrap">2 July</span> 2020</span>, <
```

```
import requests
import urllib.request
import time
from bs4 import BeautifulSoup
import numpy as np
import pandas as pd
from urllib.request import urlopen
url = 'https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship'
html = urlopen(url)
soup = BeautifulSoup(html, 'html.parser')
```

```
tables = soup.find_all('table')
tables
```

```
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">19
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">17
</td>
<td style="background:  #EFCFFF; color:  ##000; text-align: auto;" title="">Ret
</td>
<td style="background:  #EFCFFF; color:  ##000; text-align: auto;" title="">Ret
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">17
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">13
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">17
</td>
<td style="background:  #EFCFFF; color:  ##000; text-align: auto;" title="">Ret
</td>
<td style="background:  #EFCFFF; color:  ##000; text-align: auto;" title="">Ret
</td>
```

```
<td align="left" rowspan="2"><span class="flagicon"><a href="/wiki/United_Kingdom" title="United Kingdom"><img alt="United Kingdom" class="thumbborder" data-file-height="600" data-
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">11
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">16
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">18
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">12
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">18
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">17
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">16
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">11
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">11
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">16
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">14
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">14
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">11
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">16
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">12
</td>
<td style="background:  #CFCFFF; color:  ##000; text-align: auto;" title="">16
</td>
<th style="vertical-align:middle">Constructor
</th>
<th><a href="/wiki/2020_Austrian_Grand_Prix" title="2020 Austrian Grand Prix">AUT</a><br/><span class="flagicon"><a href="/wiki/Austria" title="Austria"><img alt="Austria" class="th
</th>
<th><a href="/wiki/2020_Styrian_Grand_Prix" title="2020 Styrian Grand Prix">STY</a><br/><span class="flagicon"><a href="/wiki/Austria" title="Austria"><img alt="Austria" class="thum
</th>
<th><a href="/wiki/2020_Hungarian_Grand_Prix" title="2020 Hungarian Grand Prix">HUN</a><br/><span class="flagicon"><a href="/wiki/Hungary" title="Hungary"><img alt="Hungary" class='
</th>
<th><a href="/wiki/2020_British_Grand_Prix" title="2020 British Grand Prix">GBR</a><br/><span class="flagicon"><a href="/wiki/United_Kingdom" title="United Kingdom"><img alt="Unitec
</th>
<th><a href="/wiki/70th_Anniversary_Grand_Prix" title="70th Anniversary Grand Prix">70A</a><br/><span class="flagicon"><a href="/wiki/United_Kingdom" title="United Kingdom"><img alt
</th>
<th><a href="/wiki/2020_Spanish_Grand_Prix" title="2020 Spanish Grand Prix">ESP</a><br/><span class="flagicon"><a href="/wiki/Spain" title="Spain"><img alt="Spain" class="thumbborde
</th>
<th><a href="/wiki/2020_Belgian_Grand_Prix" title="2020 Belgian Grand Prix">BEL</a><br/><span class="flagicon"><a href="/wiki/Belgium" title="Belgium"><img alt="Belgium" class="thum
</th>
<th><a href="/wiki/2020_Italian_Grand_Prix" title="2020 Italian Grand Prix">ITA</a><br/><span class="flagicon"><a href="/wiki/Italy" title="Italy"><img alt="Italy" class="thumbborde
</th>
<th><a href="/wiki/2020_Tuscan_Grand_Prix" title="2020 Tuscan Grand Prix">TUS</a><br/><span class="flagicon"><a href="/wiki/Italy" title="Italy"><img alt="Italy" class="thumbborder'
</th>
<th><a href="/wiki/2020_Russian_Grand_Prix" title="2020 Russian Grand Prix">RUS</a><br/><span class="flagicon"><a href="/wiki/Russia" title="Russia"><img alt="Russia" class="thumbbc
</th>
<th><a href="/wiki/2020_Eifel_Grand_Prix" title="2020 Eifel Grand Prix">EIF</a><br/><span class="flagicon"><a href="/wiki/Germany" title="Germany"><img alt="Germany" class="thumbbor
</th>
<th><a href="/wiki/2020_Portuguese_Grand_Prix" title="2020 Portuguese Grand Prix">POR</a><br/><span class="flagicon"><a href="/wiki/Portugal" title="Portugal"><img alt="Portugal" cl
</th>
<th><a href="/wiki/2020_Emilia_Romagna_Grand_Prix" title="2020 Emilia Romagna Grand Prix">EMI</a><br/><span class="flagicon"><a href="/wiki/Italy" title="Italy"><img alt="Italy" cla
</th>
<th><a href="/wiki/2020_Turkish_Grand_Prix" title="2020 Turkish Grand Prix">TUR</a><br/><span class="flagicon"><a href="/wiki/Turkey" title="Turkey"><img alt="Turkey" class="thumbbc
</th>
<th><a href="/wiki/2020_Bahrain_Grand_Prix" title="2020 Bahrain Grand Prix">BHR</a><br/><span class="flagicon"><a href="/wiki/Bahrain" title="Bahrain"><img alt="Bahrain" class="thum
</th>
<th><a href="/wiki/2020_Sakhir_Grand_Prix" title="2020 Sakhir Grand Prix">SKH</a><br/><span class="flagicon"><a href="/wiki/Bahrain" title="Bahrain"><img alt="Bahrain" class="thumbb
</th>
<th><a href="/wiki/2020_Abu_Dhabi_Grand_Prix" title="2020 Abu Dhabi Grand Prix">ABU</a><br/><span class="flagicon"><a href="/wiki/United_Arab_Emirates" title="United Arab Emirates">
```

```
[27] import re
     def process_num(num):
         return float(re.sub(r'[^\w\s.]','',num))
     num1 = float(re.sub(r'[^\w\s.]','','1,156.30'))
     num1
```

```
1156.3
```

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import re

html = urlopen('https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship')
bs = BeautifulSoup(html, 'html.parser')
images = bs.find_all('img', {'src':re.compile('\.\.\/img\/gifts/img.*\.jpg')})
for image in images:
    print(image['src'])
```



```
bs.find_all(lambda tag: len(tag.attrs) == 2)
```

```
<a href="/wiki/Stavelot" title="Stavelot">Stavelot</a>,
<a href="/wiki/Italian_Grand_Prix" title="Italian Grand Prix">Italian Grand Prix</a>,
<a href="/wiki/Italy" title="Italy"><img alt="Italy" class="thumbborder" data-file-height="1000" data-file-width="1500" decoding="async" height="15" src="//upload.wikimedia.org/w
<a href="/wiki/Monza" title="Monza">Monza</a>,
<a href="/wiki/Italy" title="Italy"><img alt="Italy" class="thumbborder" data-file-height="1000" data-file-width="1500" decoding="async" height="15" src="//upload.wikimedia.org/w
<a href="/wiki/Mugello_Circuit" title="Mugello Circuit">Autodromo Internazionale del Mugello</a>,
<a href="/wiki/Scarperia_e_San_Piero" title="Scarperia e San Piero">Scarperia e San Piero</a>,
<a href="/wiki/Russian_Grand_Prix" title="Russian Grand Prix">Russian Grand Prix</a>,
<a href="/wiki/Russia" title="Russia"><img alt="Russia" class="thumbborder" data-file-height="600" data-file-width="900" decoding="async" height="15" src="//upload.wikimedia.org/
<a href="/wiki/Sochi_Autodrom" title="Sochi Autodrom">Sochi Autodrom</a>,
<a href="/wiki/Sochi" title="Sochi">Sochi</a>,
<a href="/wiki/Germany" title="Germany"><img alt="Germany" class="thumbborder" data-file-height="600" data-file-width="1000" decoding="async" height="14" src="//upload.wikimedia.
<a href="/wiki/N%C3%BCrburgring" title="Nürburgring">Nürburgring</a>,
<a href="/wiki/N%C3%BCrburg" title="Nürburg">Nürburg</a>,
<a href="/wiki/Portuguese_Grand_Prix" title="Portuguese Grand Prix">Portuguese Grand Prix</a>,
<a href="/wiki/Portugal" title="Portugal"><img alt="Portugal" class="thumbborder" data-file-height="400" data-file-width="600" decoding="async" height="15" src="//upload.wikimedi
<a href="/wiki/Algarve_International_Circuit" title="Algarve International Circuit">Autódromo Internacional do Algarve</a>,
<a href="/wiki/Portim%C3%A3o" title="Portimão">Portimão</a>,
<a href="/wiki/Emilia_Romagna_Grand_Prix" title="Emilia Romagna Grand Prix">Emilia Romagna Grand Prix</a>,
<a href="/wiki/Italy" title="Italy"><img alt="Italy" class="thumbborder" data-file-height="1000" data-file-width="1500" decoding="async" height="15" src="//upload.wikimedia.org/w
<a href="/wiki/Imola" title="Imola">Imola</a>,
<a href="/wiki/Turkish_Grand_Prix" title="Turkish Grand Prix">Turkish Grand Prix</a>,
<a href="/wiki/Turkey" title="Turkey"><img alt="Turkey" class="thumbborder" data-file-height="800" data-file-width="1200" decoding="async" height="15" src="//upload.wikimedia.org
<a href="/wiki/Istanbul_Park" title="Istanbul Park">Istanbul Park</a>,
<a href="/wiki/Tuzla,_Istanbul" title="Tuzla, Istanbul">Tuzla</a>,
<a href="/wiki/Bahrain_Grand_Prix" title="Bahrain Grand Prix">Bahrain Grand Prix</a>,
<a href="/wiki/Bahrain" title="Bahrain"><img alt="Bahrain" class="thumbborder" data-file-height="900" data-file-width="1500" decoding="async" height="14" src="//upload.wikimedia.
<a href="/wiki/Bahrain_International_Circuit" title="Bahrain International Circuit">Bahrain International Circuit</a>,
<a href="/wiki/Sakhir" title="Sakhir">Sakhir</a>,
```

```python
bs.find_all('', text='Lewis Hamilton')
```

```
['Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
 'Lewis Hamilton',
```

```python
[32] bs.find_all(lambda tag: tag.get_text() == 'Ferrari')
```

```
[<a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>,
 <a href="/wiki/Scuderia_Ferrari" title="Scuderia Ferrari">Ferrari</a>]
```

```
from urllib.request import urlopen
from bs4 import BeautifulSoup

html = urlopen('https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship')
bs = BeautifulSoup(html, 'html.parser')
for link in bs.find_all('a'):
    if 'href' in link.attrs:
        print(link.attrs['href'])
```

```
https://commons.wikimedia.org/wiki/Category:2020_in_Formula_One
https://af.wikipedia.org/wiki/2020_Formule_Een-seisoen
https://ar.wikipedia.org/wiki/%D8%A8%D8%B7%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%B9%D8%A7%D9%84%D9%85_%D9%84%D8%B3%D8%A8%D8%A7%D9%82%D8%A7%D8%AA_%D9%81%D9%88%D8%B1%D9%85%D9%88%D9%
https://az.wikipedia.org/wiki/2020_Formula_1_m%C3%B6vs%C3%BCm%C3%BC
https://bs.wikipedia.org/wiki/Formula_1_%E2%80%93_sezona_2020.
https://br.wikipedia.org/wiki/Kevezadeg_bed_ar_Formulenn_1_2020
https://ca.wikipedia.org/wiki/Temporada_2020_de_F%C3%B3rmula_1
https://cs.wikipedia.org/wiki/Formule_1_v_roce_2020
https://de.wikipedia.org/wiki/Formel-1-Weltmeisterschaft_2020
https://et.wikipedia.org/wiki/2020._aasta_Vormel_1_hooaeg
https://el.wikipedia.org/wiki/%CE%A0%CE%B1%CE%B3%CE%BA%CF%8C%CF%83%CE%BC%CE%B9%CE%BF_%CF%80%CF%81%CF%89%CF%84%CE%AC%CE%B8%CE%BB%CE%B7%CE%BC%CE%B1_%CE%A6%CF%8C%CF%81%CE%BC%CE%BF
https://es.wikipedia.org/wiki/Temporada_2020_de_F%C3%B3rmula_1
https://fa.wikipedia.org/wiki/%D9%85%D8%B3%D8%A7%D8%A8%D9%82%D8%A7%D8%AA_%D9%81%D8%B1%D9%85%D9%88%D9%84_%DB%8C%DA%A9_%D9%81%D8%B5%D9%84_%DB%B2%DB%B0%DB%B2%DB%B0
https://fr.wikipedia.org/wiki/Championnat_du_monde_de_Formule_1_2020
https://gd.wikipedia.org/wiki/Farpais_Foirmle_a_h-Aon_na_Cruinne_2020
https://gl.wikipedia.org/wiki/Campionato_Mundial_de_F%C3%B3rmula_1_de_2020
https://ko.wikipedia.org/wiki/2020%EB%85%84_%ED%8F%AC%EB%AE%AC%EB%9F%AC_%EC%9B%90_%EC%8B%9C%EC%A6%8C
https://hr.wikipedia.org/wiki/Formula_1_-_sezona_2020.
https://id.wikipedia.org/wiki/Formula_Satu_musim_2020
https://it.wikipedia.org/wiki/Campionato_mondiale_di_Formula_1_2020
https://he.wikipedia.org/wiki/%D7%A4%D7%95%D7%A8%D7%9E%D7%95%D7%9C%D7%94_1_%D7%A2%D7%95%D7%A0%D7%AA_2020
https://lv.wikipedia.org/wiki/2020._gada_F1_sezona
https://lt.wikipedia.org/wiki/2020_m._Formul%C4%97s_1_sezonas
```

```
from urllib.request import urlopen
from bs4 import BeautifulSoup
import re

html = urlopen('https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship')
bs = BeautifulSoup(html, 'html.parser')
for link in bs.find('div', {'id':'bodyContent'}).find_all(
    'a', href=re.compile('^(/wiki/)((?!:).)*$')):
    if 'href' in link.attrs:
        print(link.attrs['href'])
```

```
/wiki/1983_Formula_One_World_Championship
/wiki/1984_Formula_One_World_Championship
/wiki/1985_Formula_One_World_Championship
/wiki/1986_Formula_One_World_Championship
/wiki/1987_Formula_One_World_Championship
/wiki/1988_Formula_One_World_Championship
/wiki/1989_Formula_One_World_Championship
/wiki/1990_Formula_One_World_Championship
/wiki/1991_Formula_One_World_Championship
/wiki/1992_Formula_One_World_Championship
/wiki/1993_Formula_One_World_Championship
/wiki/1994_Formula_One_World_Championship
/wiki/1995_Formula_One_World_Championship
/wiki/1996_Formula_One_World_Championship
/wiki/1997_Formula_One_World_Championship
/wiki/1998_Formula_One_World_Championship
/wiki/1999_Formula_One_World_Championship
/wiki/2000_Formula_One_World_Championship
/wiki/2001_Formula_One_World_Championship
/wiki/2002_Formula_One_World_Championship
/wiki/2003_Formula_One_World_Championship
/wiki/2004_Formula_One_World_Championship
/wiki/2005_Formula_One_World_Championship
```

```python
from urllib.request import urlopen
from urllib.error import HTTPError
from bs4 import BeautifulSoup
import sys


def getTitle(url):
    try:
        html = urlopen(url)
    except HTTPError as e:
        print(e)
        return None
    try:
        bsObj = BeautifulSoup(html, "html.parser")
        title = bsObj.body.h1
    except AttributeError as e:
        return None
    return title

title =
getTitle("https://en.wikipedia.org/wiki/2020_Formula_One_World_Championship")
if title == None:
    print("Title could not be found")
else:
    print(title)
```

```
D:\Anaconda\python.exe "C:/Users/aryam/Desktop/Fall Sem 2021/Web Mining Lab/LAB 2 12-8-21/3-exceptionHandling.py"
<h1 class="firstHeading" id="firstHeading">2020 Formula One World Championship</h1>

Process finished with exit code 0
```