Reference

- This example is from [https://towardsdatascience.com/easily-scrape-and-summarize-news-articles-using-python-dfc7667d9e74 (https://towardsdatascience.com/easily-scrape-and-summarize-news-articles-using-python-dfc7667d9e74)](https://towardsdatascience.com/easily-scrape-and-summarize-news-articles-using-python-dfc7667d9e74)
- Added some notes and modified a little bit

Scraping and Summarizing News Articles

- This notebook gives a short demonstration of code to scrape and summarize news articles.

In [2]: ▶|
```python
1  # Imports
2  import requests
3  from bs4 import BeautifulSoup
4  from gensim.summarization import summarize
```

We'll scrape and summarize the following article:

- [https://www.npr.org/2019/07/10/740387601/university-of-texas-austin-promises-free-tuition-for-low-income-students-in-2020 (https://www.npr.org/2019/07/10/740387601/university-of-texas-austin-promises-free-tuition-for-low-income-students-in-2020)](https://www.npr.org/2019/07/10/740387601/university-of-texas-austin-promises-free-tuition-for-low-income-students-in-2020)

In [3]: ▶|
```python
1  # Retrieve page text
2  url = 'https://www.npr.org/2019/07/10/740387601/university-of-texas-aust
3  page = requests.get(url).text
```

In [4]: ▶|
```python
1  # Turn page into BeautifulSoup object to access HTML tags
2  soup = BeautifulSoup(page)
```

- We can use Chrome's Inspect tool to find HTML tags by right-clicking on the page and choosing "Inspect."
- Then, click on the little button to find HTML tags for a specific part of the page. That will look like this:

In [5]: ▶|
```python
1  # Get headline
2  headline = soup.find('h1').get_text()
3  print(headline)
```

```
University of Texas-Austin Promises Free Tuition For Low-Income Students In
2020
```

- The main text of the article is surrounded by the "p" tag.
- This time we'll have to find all of the "p" tags contained on the page since the paragraphs of the article are each contained in a "p" tag.

```python
# Get text from all <p> tags.
p_tags = soup.find_all('p')
# Get the text from each of the "p" tags and strip surrounding whitespac
p_tags_text = [tag.get_text().strip() for tag in p_tags]
p_tags_text
```

Out[6]: ['Vanessa Romo',
 'Claire McInerny',
 'From',
 'The University of Texas-Austin announced Tuesday it is offering full tuit
ion scholarships to in-state undergraduates whose families make $65,000 or
less a year.\n   \n   \n   \n
Jon Herskovitz/Reuters\n   \n   \nhide captio
n',
 "Four year colleges and universities have difficulty recruiting talented s
tudents from the lower end of the economic spectrum who can't afford to att
end such institutions without taking on massive debt. To remedy that — at l
east in part — the University of Texas-Austin announced it is offering full
tuition scholarships to in-state undergraduates whose families make $65,000
or less per year.",
 "The University of Texas System Board of Regents voted unanimously on Tues
day to establish a $160 million endowment, drawing from the state's Permane
nt University Fund to begin the program in the fall of 2020.",
 '"Recognizing both the need for improved access to higher education and th
e high value of a UT Austin degree, we are dedicating a distribution from t
he Permanent University Fund to establish an endowment that will directly b
enefit students and make their degrees more affordable," Chairman of the Bo
ard of Regents Kevin Eltife said after the vote.',
 '"This will benefit students of our great state for years to come," he add
ed.',
 'The endowment — which includes money from oil and gas royalties earned on
state-owned land in West Texas — more than doubles an existing program offe
ring free tuition to students whose families make less than $30,000. It als
o expands financial assistance to middle class students whose families earn
up to $125,000 a year, compared to the current $100,000.',
 "UT-Austin is among the Texas system's more affordable universities; tuiti
on and fees cost about $10,500. The endowment will foot the bill for those
charges but not additional living expenses, including room and board, which
can add another $17,000 in annual fees.",
 'Officials expect about 8,600 students — nearly a quarter of undergraduate
s, including those already enrolled — will benefit from the full tuition pr
ogram and another 5,700 will gain access to additional financial support, t
he release said. Transfer students will also be eligible, but graduate stud
ents will not.',
 '"There is no greater engine of social and economic mobility than a colleg
e degree, and this initiative ensures that more Texans will benefit from a
high-quality UT Austin education," University Chancellor James Milliken sai
d in a statement.',
 'Cindy Muñoz, an incoming freshman from Houston, was on campus Wednesday f
or orientation. She was thrilled to learn of the endowment. Her mother is a
single parent who has been working as a housekeeper and doing other odd job
s to make ends meet.',
 'Until Tuesday Muñoz planned to pay for most of her college with loans. "I
knew I was going to go into student debt," she said, adding that she was wi
lling to take on the loans to go to her dream school.',
 '"I had that already in my head because my sister went to college and she
```

had to take out loans, and she\'s still dealing with loans today and she\'s 38 years old," Muñoz said.',
 "Once it goes into effect, the new financial aid program will apply to Muñoz and cover all of her tuition. She said she'll only need loans to pay for some of her living expenses.",
 '"It kind of makes me feel like I chose the right school," Munoz said. "Because a school that is willing to see past your socio-economic status, look at you as a student, a person who just wants to advance their education and pursue a career, I think they look past monetary issues and they\'re like, \'You want to do this? Then we\'re going to help you out.\' "',
 'The school noted it landed on the $65,000 cap because the average median household income in Texas in 2017 was $59,206.',
 'As tuition rates continue to climb and financial aid has continued to lag behind, a growing number of public and private four-year institutions are offering free tuition to students below a certain income.',
 "In 2008, Texas A&M began offering free tuition to students whose families' income was under $60,000. Michigan established the Kalamazoo Promise, which covers four years of tuition at any public college in the state as well as 15 private colleges. Louisiana offers a more restrictive program that comes with certain caveats, including a minimum aptitude test score.",
 'Last year, New York launched The Excelsior Scholarship, allowing middle-class families and individuals making up to $125,000 a year to qualify for tuition-free enrollment at all CUNY and SUNY two and four-year colleges.',
 'But some states are finding it difficult to sustain such programs. Oregon launched the Oregon Promise in 2016 but the program there had to be scaled back after a budget shortage, Pew reported.',
 "Editor's note: NPR member station KUT's broadcast license is owned by the University of Texas. Its newsroom operates independently.",
 'NPR thanks our sponsors',
 'Become an NPR sponsor']

In [7]: ▶|

```
1  # Filter out sentences that contain newline characters '\n' or don't con
2  sentence_list = [sentence for sentence in p_tags_text if not '\n' in sen
3  sentence_list = [sentence for sentence in sentence_list if '.' in senten
4  sentence_list
```

Out[7]: ["Four year colleges and universities have difficulty recruiting talented students from the lower end of the economic spectrum who can't afford to attend such institutions without taking on massive debt. To remedy that — at least in part — the University of Texas-Austin announced it is offering full tuition scholarships to in-state undergraduates whose families make $65,000 or less per year.",
 "The University of Texas System Board of Regents voted unanimously on Tuesday to establish a $160 million endowment, drawing from the state's Permanent University Fund to begin the program in the fall of 2020.",
 '"Recognizing both the need for improved access to higher education and the high value of a UT Austin degree, we are dedicating a distribution from the Permanent University Fund to establish an endowment that will directly benefit students and make their degrees more affordable," Chairman of the Board of Regents Kevin Eltife said after the vote.',
 '"This will benefit students of our great state for years to come," he added.',
 'The endowment — which includes money from oil and gas royalties earned on state-owned land in West Texas — more than doubles an existing program offering free tuition to students whose families make less than $30,000. It also expands financial assistance to middle class students whose families earn up to $125,000 a year, compared to the current $100,000.',
 "UT-Austin is among the Texas system's more affordable universities; tuition and fees cost about $10,500. The endowment will foot the bill for those charges but not additional living expenses, including room and board, which can add another $17,000 in annual fees.",
 'Officials expect about 8,600 students — nearly a quarter of undergraduates, including those already enrolled — will benefit from the full tuition program and another 5,700 will gain access to additional financial support, the release said. Transfer students will also be eligible, but graduate students will not.',
 '"There is no greater engine of social and economic mobility than a college degree, and this initiative ensures that more Texans will benefit from a high-quality UT Austin education," University Chancellor James Milliken said in a statement.',
 'Cindy Muñoz, an incoming freshman from Houston, was on campus Wednesday for orientation. She was thrilled to learn of the endowment. Her mother is a single parent who has been working as a housekeeper and doing other odd jobs to make ends meet.',
 'Until Tuesday Muñoz planned to pay for most of her college with loans. "I knew I was going to go into student debt," she said, adding that she was willing to take on the loans to go to her dream school.',
 '"I had that already in my head because my sister went to college and she had to take out loans, and she\'s still dealing with loans today and she\'s 38 years old," Muñoz said.',
 '"Once it goes into effect, the new financial aid program will apply to Muñoz and cover all of her tuition. She said she\'ll only need loans to pay for some of her living expenses.",
 '"It kind of makes me feel like I chose the right school," Munoz said. "Because a school that is willing to see past your socio-economic status, look at you as a student, a person who just wants to advance their education and pursue a career, I think they look past monetary issues and they

```
\'re like, \'You want to do this? Then we\'re going to help you out.\'
"',
 'The school noted it landed on the $65,000 cap because the average media
n household income in Texas in 2017 was $59,206.',
 'As tuition rates continue to climb and financial aid has continued to l
ag behind, a growing number of public and private four-year institutions
are offering free tuition to students below a certain income.',
 "In 2008, Texas A&M began offering free tuition to students whose famili
es' income was under $60,000. Michigan established the Kalamazoo Promise,
which covers four years of tuition at any public college in the state as
well as 15 private colleges. Louisiana offers a more restrictive program
that comes with certain caveats, including a minimum aptitude test scor
e.",
 'Last year, New York launched The Excelsior Scholarship, allowing middle
-class families and individuals making up to $125,000 a year to qualify f
or tuition-free enrollment at all CUNY and SUNY two and four-year college
s.',
 'But some states are finding it difficult to sustain such programs. Oreg
on launched the Oregon Promise in 2016 but the program there had to be sc
aled back after a budget shortage, Pew reported.',
 "Editor's note: NPR member station KUT's broadcast license is owned by t
he University of Texas. Its newsroom operates independently."]
```

In [8]:
```python
# Combine list items into string.
article = ' '.join(sentence_list)
```

In [9]:
```python
summary = summarize(article, ratio=0.3)
```

```
In [10]:    ▶|    1  print(f'Length of original article: {len(article)}')
                  2  print(f'Length of summary: {len(summary)} \n')
                  3  print(f'Headline: {headline} \n')
                  4  print(f'Article Summary:\n{summary}')
```

Length of original article: 4594
Length of summary: 1859

Headline: University of Texas-Austin Promises Free Tuition For Low-Income Students In 2020

Article Summary:
Four year colleges and universities have difficulty recruiting talented students from the lower end of the economic spectrum who can't afford to attend such institutions without taking on massive debt.
To remedy that — at least in part — the University of Texas-Austin announced it is offering full tuition scholarships to in-state undergraduates whose families make $65,000 or less per year.
The University of Texas System Board of Regents voted unanimously on Tuesday to establish a $160 million endowment, drawing from the state's Permanent University Fund to begin the program in the fall of 2020.
"Recognizing both the need for improved access to higher education and the high value of a UT Austin degree, we are dedicating a distribution from the Permanent University Fund to establish an endowment that will directly benefit students and make their degrees more affordable," Chairman of the Board of Regents Kevin Eltife said after the vote.
The endowment — which includes money from oil and gas royalties earned on state-owned land in West Texas — more than doubles an existing program offering free tuition to students whose families make less than $30,000.
Officials expect about 8,600 students — nearly a quarter of undergraduates, including those already enrolled — will benefit from the full tuition program and another 5,700 will gain access to additional financial support, the release said.
As tuition rates continue to climb and financial aid has continued to lag behind, a growing number of public and private four-year institutions are offering free tuition to students below a certain income.
In 2008, Texas A&M began offering free tuition to students whose families' income was under $60,000.
Michigan established the Kalamazoo Promise, which covers four years of tuition at any public college in the state as well as 15 private colleges.

```
In [ ]:    ▶|    1
```