TRUMP DAILY

Trump daily is a crawler that crawls the top 25 news contents that contains the top-featured man on the planet, US President Donald Trump. Currently, the crawler crawls from

- CNN
- Twitter

Installation

The crawler is built using python 3.6. So you would need to set up Python 3.6. The crawlers can be run as scheduled task (cron job) or can be run manually.

- Create a mysql/mariadb database named 'trumpdaily' in your localhost/hosted server
- Upload the SQL file 'trumpdaily.sql' to create the tables
- Install the required python3.6 libraries
- Run mainCrawler.py to start the crawler
- Connect your website or web application to query the feed from the database and display on the site

Required libraries

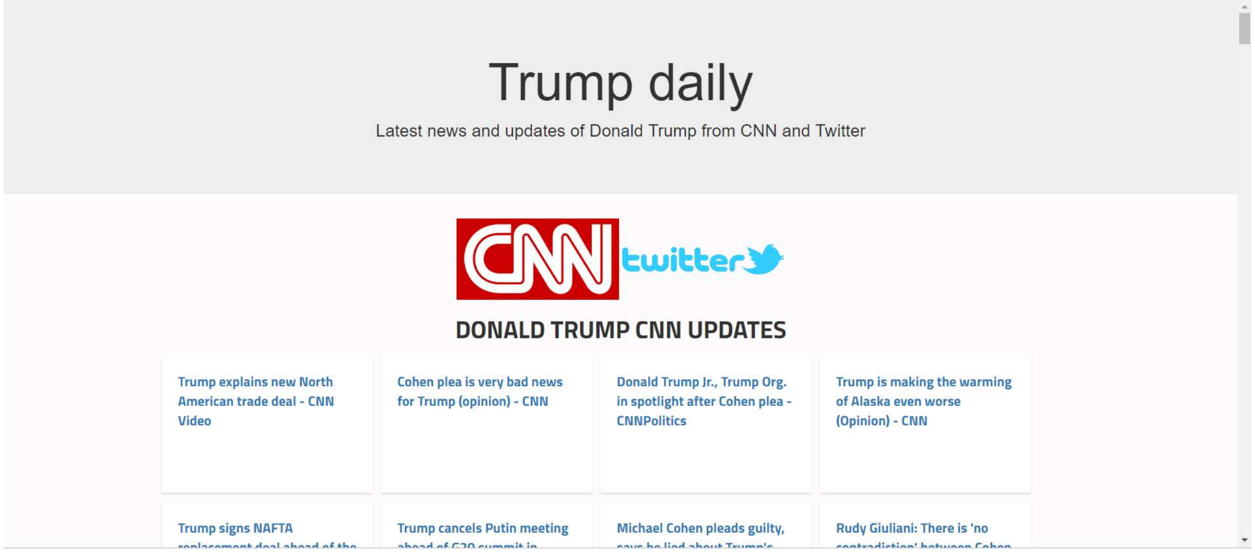
The python libraries and the additional tools that are required to set this project are:

Libraries/Tools	Purpose	Reference
MySQL	To save the collected feeds	https://www.mysql.com/
Github	Git repository	plugins/github/README.md
Selenium 3.141.0	To render dynamic web	https://pypi.org/project/selenium/

	pages of CNN	
Beautifulsoup4 4.6.3	To build effective crawler to parse through contents	https://pypi.org/project/beautifulsoup4/
Configparser 3.5.0	To parse configuration files	https://pypi.org/project/configparser/
time 1.0.0	Default python library to add delays into twitter scrapping	https://pypi.org/project/time/

THE WEBSITE

The url to the website is trumpdaily.deciphertechglobal.com



Trump daily

Latest news and updates of Donald Trump from CNN and Twitter

DONALD TRUMP CNN UPDATES

Trump explains new North American trade deal - CNN Video

Cohen plea is very bad news for Trump (opinion) - CNN

Donald Trump Jr., Trump Org. in spotlight after Cohen plea - CNNPolitics

Trump is making the warming of Alaska even worse (Opinion) - CNN

Trump signs NAFTA replacement deal ahead of the

Trump cancels Putin meeting ahead of G20 summit in

Michael Cohen pleads guilty, says he lied about Trump's

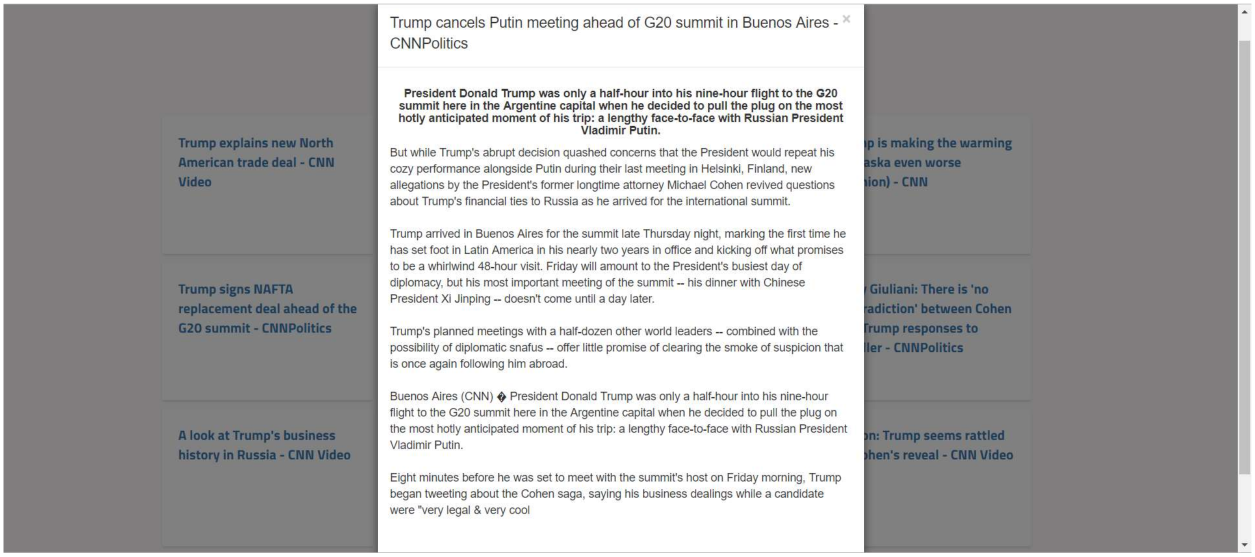
Rudy Giuliani: There is 'no contradiction' between Cohen

How to use the website:

- The website is divided into two sections, CNN and Twitter
- If you click CNN logo, it will take you to the CNN section
- If you click the Twitter logo, it will take you to the Twitter section

CNN Section:

- Designed in a modal form
- If you click on any news title it will pop up the news and its details like this


Trump explains new North American trade deal - CNN Video
Trump signs NAFTA replacement deal ahead of the G20 summit - CNNPolitics
A look at Trump's business history in Russia - CNN Video
p is making the warming aska even worse ion) - CNN
Giuliani: There is 'no radiction' between Cohen Trump responses to ller - CNNPolitics
n: Trump seems rattled hen's reveal - CNN Video
×

Trump cancels Putin meeting ahead of G20 summit in Buenos Aires - CNNPolitics

President Donald Trump was only a half-hour into his nine-hour flight to the G20 summit here in the Argentine capital when he decided to pull the plug on the most hotly anticipated moment of his trip: a lengthy face-to-face with Russian President Vladimir Putin.

But while Trump's abrupt decision quashed concerns that the President would repeat his cozy performance alongside Putin during their last meeting in Helsinki, Finland, new allegations by the President's former longtime attorney Michael Cohen revived questions about Trump's financial ties to Russia as he arrived for the international summit.

Trump arrived in Buenos Aires for the summit late Thursday night, marking the first time he has set foot in Latin America in his nearly two years in office and kicking off what promises to be a whirlwind 48-hour visit. Friday will amount to the President's busiest day of diplomacy, but his most important meeting of the summit -- his dinner with Chinese President Xi Jinping -- doesn't come until a day later.

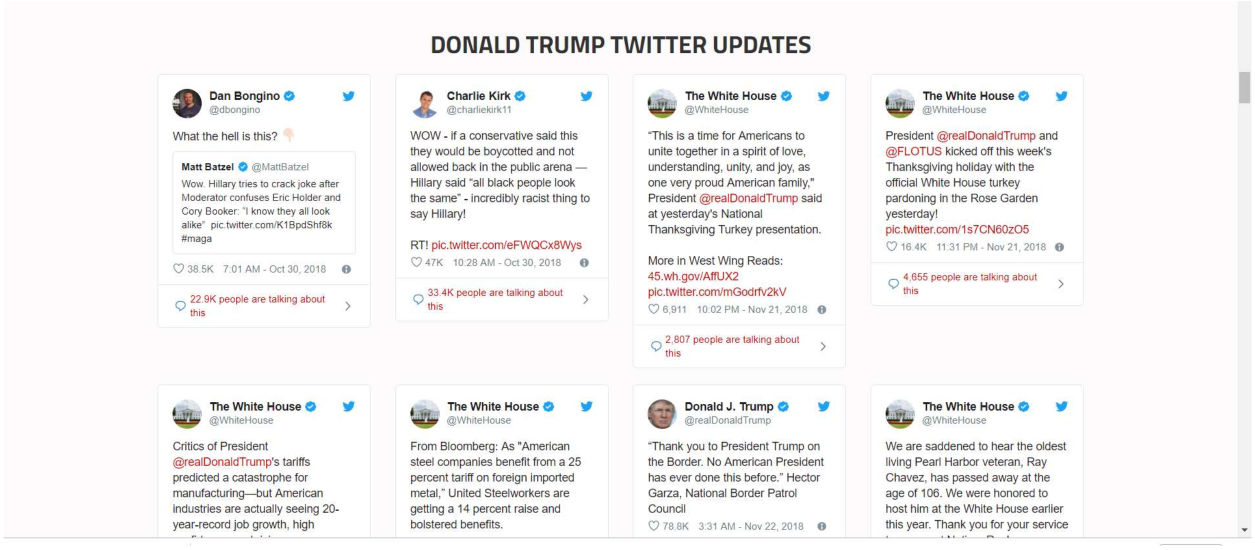
Trump's planned meetings with a half-dozen other world leaders -- combined with the possibility of diplomatic snafus -- offer little promise of clearing the smoke of suspicion that is once again following him abroad.

Buenos Aires (CNN) � President Donald Trump was only a half-hour into his nine-hour flight to the G20 summit here in the Argentine capital when he decided to pull the plug on the most hotly anticipated moment of his trip: a lengthy face-to-face with Russian President Vladimir Putin.

Eight minutes before he was set to meet with the summit's host on Friday morning, Trump began tweeting about the Cohen saga, saying his business dealings while a candidate were "very legal & very cool

Twitter section:

- Feed collected from the Twitter tweet IDs
- The IDs are then used to generate Twitter widget using `https://platform.twitter.com/widgets.js`



DONALD TRUMP TWITTER UPDATES

Dan Bongino @dbongino

What the hell is this? 👇

Matt Batzel @MattBatzel
Wow. Hillary tries to crack joke after Moderator confuses Eric Holder and Cory Booker: "I know they all look alike" pic.twitter.com/K1BpdShf8k #maga

♡ 38.5K 7:01 AM - Oct 30, 2018

22.9K people are talking about this

Charlie Kirk @charliekirk11

WOW - if a conservative said this they would be boycotted and not allowed back in the public arena — Hillary said "all black people look the same" - incredibly racist thing to say Hillary!

RT! pic.twitter.com/eFWQCx8Wys

♡ 47K 10:28 AM - Oct 30, 2018

33.4K people are talking about this

The White House @WhiteHouse

"This is a time for Americans to unite together in a spirit of love, understanding, unity, and joy, as one very proud American family," President @realDonaldTrump said at yesterday's National Thanksgiving Turkey presentation.

More in West Wing Reads:
45.wh.gov/AffUX2
pic.twitter.com/mGodrfv2kV

♡ 6,911 10:02 PM - Nov 21, 2018

2,807 people are talking about this

The White House @WhiteHouse

President @realDonaldTrump and @FLOTUS kicked off this week's Thanksgiving holiday with the official White House turkey pardoning in the Rose Garden yesterday!
pic.twitter.com/1s7CN60zO5

♡ 16.4K 11:31 PM - Nov 21, 2018

4,655 people are talking about this

The White House @WhiteHouse

Critics of President @realDonaldTrump's tariffs predicted a catastrophe for manufacturing—but American industries are actually seeing 20-year-record job growth, high

The White House @WhiteHouse

From Bloomberg: As "American steel companies benefit from a 25 percent tariff on foreign imported metal," United Steelworkers are getting a 14 percent raise and bolstered benefits.

Donald J. Trump @realDonaldTrump

"Thank you to President Trump on the Border. No American President has ever done this before." Hector Garza, National Border Patrol Council

♡ 78.8K 3:31 AM - Nov 22, 2018

The White House @WhiteHouse

We are saddened to hear the oldest living Pearl Harbor veteran, Ray Chavez, has passed away at the age of 106. We were honored to host him at the White House earlier this year. Thank you for your service

CNN CRAWLING

1. To crawl CNN and find news about Trump, the starting point was chosen as the home page (www.cnn.com)
2. Using search engine analytics similarweb (www.similarweb.com), two primary category sections of CNN was chosen which has the highest probability to contain top news about Trump, they are "politics" and "us".
3. Different crawlers were made for each of the pages to crawl the urls of the news finding and scraping from the element named "cd__headline"
4. After getting the URLs, the URLs were passed to a class named "singleCNNNews", which crawls each URL and extract title (element parsed= title), short summary (element parsed=meta description) and description (multiple methods used for different templates used by CNN) (if available)
5. Accumulated data is then returned in the form of JSON and then stored in the database. Error handling has been added to handle any error in between this process

TWITTER CRAWLING

The ideology behind twitter crawling to get Trump tweets is to get the Tweet IDs from the two official accounts of Donald Trump: @ POTUS and @ realDonaldTrump.

1. At first the page is rendered and scrolled down automatedly 5 times to make larger collection of tweets in the page source
2. Then the data-stream-id, which contains the Tweet ID, is collected from all the 'js-stream-tweet' elements of the page
3. The IDs are then sent to the database