

CS5540: PRINCIPLES OF BIG DATA MANAGEMENT

A system to store, analyze, and visualize Twitter's tweets

Project Phase 1 Document

Team # 7

Avni Mehta

Sneha Mishra

Arvind Tota

Step 1: Collect tweets using Twitter's Streaming API

Python Code for Tweets Collection

```
Collect_Tweets.py x
1  # -----
2  # * CS5540: Principles of Big Data Management
3  # * Project Phase 1
4  # * Team #7: Avni Mehta, Sneha Mishra, Arvind Tota
5  # -----
6
7  # Import necessary methods from tweepy library
8  from tweepy.streaming import StreamListener
9  from tweepy import OAuthHandler
10 from tweepy import Stream
11
12 # Import other libraries
13 import datetime
14
15
16 # This is a basic listener for received tweets to stdout.
17 class TwitterListener(StreamListener):
18
19     tweet_number = 0          # class variable
20
21     def __init__(self, max_tweets):
22         super().__init__()
23         self.max_tweets = max_tweets          # max number of tweets
24         self.last = datetime.datetime.now()   # last time for status
25
26     def on_data(self, data):
27
28         # Write to file
29         with open('twitter_data.txt', 'a') as my_file:
30             my_file.write(data)
31
32         # Count no. of tweets collected
33         self.tweet_number += 1
34
35
36     # Stop when the limit is reached
37     if self.tweet_number >= self.max_tweets:
38         print('\tStopping data collection : Limit of ' + str(self.max_tweets) + ' tweets reached.')
39         # Return False to the listener's on_data() of streaming.py API
40         return False
41
42     self.status(self.tweet_number)
43     return True
44
45 # Status method prints out tweet counts every ten minutes
46 def status(self, tweetCount):
47     now = datetime.datetime.now()
48     # Print status every 10 mins
49     if (now - self.last).total_seconds() > 600:
50         print('\t' + str(tweetCount) + ' tweets collected..')
51         self.last = now
52
53 # On error, print error status
54 def on_error(self, status):
55     """ Handles the response error status. """
56     print('Error status code : ' + str(status))
```

```
Collect_Tweets.py x
56
57
58 # Main Activity
59 if __name__ == '__main__':
60     print('\n-----')
61     print('Big Data Project :')
62     print('Collect 100K+ tweets, extract hashtags and urls and run wordcount using Hadoop & Spark')
63     print('-----')
64
65     # Variables that contains the user credentials to access Twitter API
66     access_token = "712180562-wnFa9ahIaiR7mFZrHyodaOmYepgl0cL2Rsr2bGfs"
67     access_token_secret = "GHYZH6CpQouyUS3EeRSbrGMDtxHPkBFfLM9r6dvX9MYHAW"
68     consumer_key = "bR6Laeo9MOxlvqCVNvBlwPuvn"
69     consumer_secret = "fmjeR54JsgF5yQouBBcmubWZCtGn602zNy8G2KEJFZk9G3Ahli"
70
71     n = 10000 # no of tweets to be collected
72
73     # Twitter authentication and the connection to Twitter Streaming API
74     listener = TwitterListener(n)
75     auth = OAuthHandler(consumer_key, consumer_secret)
76     auth.set_access_token(access_token, access_token_secret)
77     stream = Stream(auth, listener)
78
79     print('\nStep 1. Collecting tweets..')
80
81     # Filter twitter streams to capture data by the specified keywords and languages
82     stream.filter(track=['MachineLearning', 'BigData'], languages=['en'])
83
84
```

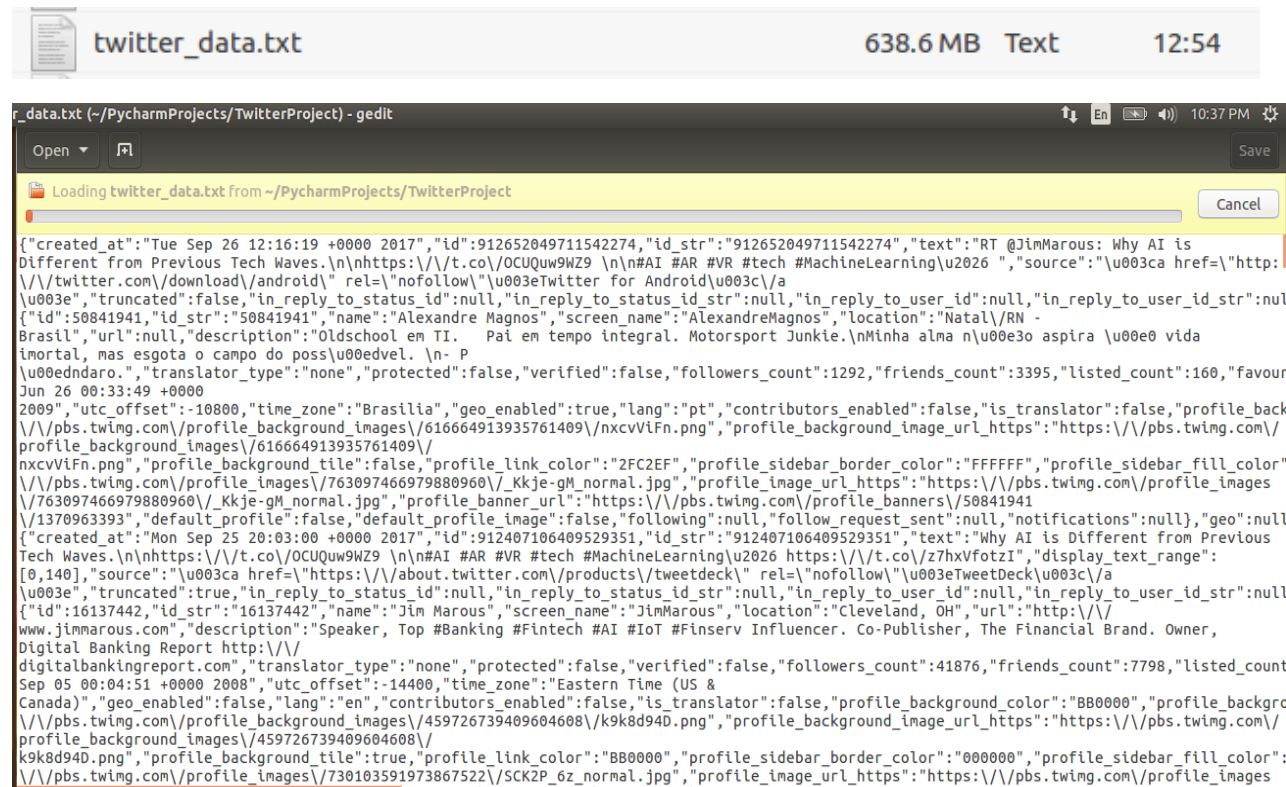
The above program collects tweets into a text file by using Twitter's streaming API Tweepy.

- i. In order to connect to the Twitter's Streaming API, we need to authenticate using credentials from our twitter developer account.
- ii. The program filters data by keywords ('MachineLearning', 'BigData'), and language ('en').
- iii. The data is appended to file 'twitter_data.txt'.
- iv. The program prints the total number of tweets collected every 10 mins on the console.

We have collected 100k+ tweets in batches.

Output File

twitter_data.txt (Not included in the output folder as size > 600MB)



Step 2: Extracting Hashtags and URLs

Python Code for Extraction

```
Extract_Fields.py ×
1  # -----
2  # * CS5540: Principles of Big Data Management
3  # * Project Phase 1
4  # * Team #7: Avni Mehta, Sneha Mishra, Arvind Tota
5  # -----
6
7  # Import libraries
8  import json
9
10
11  # This functions takes .txt file as input
12  # and returns tweets as a list of json
13  def parse_tweets():
14
15      tweets_data_path = 'twitter_data.txt'
16      twt_data = []
17      tweets_file = open(tweets_data_path, "r")
18      for line in tweets_file:
19          try:
20              tweet = json.loads(line)
21              twt_data.append(tweet)
22          except:
23              continue
24
25      print('\tSuccessfully parsed data into json format')
26      return twt_data
27
28
29  # This function extracts hashtags and urls from tweets data
30  def extract_hashtags_urls(twt_data):
31
32      # Extract hashtags and urls into separate files
33      htfile = 'hashtags.txt'
34      urlfile = 'urls.txt'
35
36      htoutfile = open(htfile, 'w')
37      urloutfile = open(urlfile, 'w')
38
39      for i in range(len(twt_data)):
40
41          # Extracting hashtags
42          ht = twt_data[i].get('entities').get('hashtags')
43          for j in range(len(ht)):
44              htoutfile.write(ht[j].get('text'))
45              htoutfile.write(' ')
46
47          # Extracting urls (expanded urls only)
48          url = twt_data[i].get('entities').get('urls')
49          for k in range(len(url)):
50              urloutfile.write(url[k].get('expanded_url'))
51              urloutfile.write(' ')
52
53      # Print once extraction is complete
54      print('\tExtracted hashtags in file - ' + htfile)
55      print('\tExtracted urls in file - ' + urlfile)
56      print('\nExiting the application.')
57
58  # Main Application
```

```

58
59 # Main Activity
60 if __name__ == '__main__':
61
62     print('\n-----')
63     print('Big Data Project :')
64     print('Collect 100K+ tweets, extract hashtags and urls and run wordcount using Hadoop & Spark')
65     print('-----')
66
67     # Step 2: Parsing the data
68     print('\nStep 2. Parsing data..')
69     tweets_data = parse_tweets()
70     print(str(len(tweets_data)))
71
72     # Step 3: Extracting hashtags and urls
73     print('\nStep 3. Extracting hashtags and urls..')
74     #extract_hashtags_urls(tweets_data)

```

There are two methods in this program.

- i. The *parse_tweets* method takes the output from previous program as input, and parses the same as a list of json.
- ii. The *extract_hashtags_urls* method extracts hashtags and urls into two separate files.

Output files

hashtags.txt

```
hashtags - Notepad
File Edit Format View Help
AI AR VR tech MachineLearning GoodRead Technology Trends Wearables IoT Bigdata travel BigData Analytics IOT AI RT fintech
bigdata news DeepLearning AI MachineLearning BigData ML DL tech cybersecurity ecommerce ransomware Malware infosec AI
MachineLearning bigdata fintech MachineLearning gamedev BigData emergentech AI ML IoT Datascience BigData CyberSecurity
infosec VR CRISPR tech science BigData genomics BigData Analytics lifesciences ml machinelearning DeepLearning AI
MachineLearning BigData ML DL tech DataPlane Hortonworks datalakes AI Bigdata G7 Innovation7 BigData AI BIGDATA programming
developers softwaredevelopment ai java bigdata tech Algorithms DataMining BigData CyberSecurity BigData MachineLearning
DataScience AI MachineLearning abdsc iPhones AI BigData MachineLearning algorithms IoT InternetOfThings IoE IIoT AI BigData
Blockchain Fintech DifferentialPrivacy bigdata machinelearning BigData AI EmergingMarkets ArtificialIntelligence
MachineLearning DeepLearning AI MachineLearning BigData ML blockchain tech BigData PredictiveAnalytics Business Streaming
BigData Finance Machinelearning neuralnetworks javascript tensorflow MCubed banking fintech digital biometrics bigdata
BigData ArtificialIntelligence RT IoT InternetOfThings IoE IIoT AI BigData Blockchain Fintech BigData DeepLearning AI
MachineLearning BigData ML DL tech Blockchain CyberSecurity IoT infosec BigData smartcontracts DLT InternetOfThings
infographic Industry40 BigData SmartCity CyberSecurity IoT AI Marketing GrowthHacking BigData Mpgvip Defstar5 SEO
SocialMedia MakeYourOwnLane Digital Tech AI MachineLearning BigData Fintech Insurtech ML AI BigData 4IR Fintech defstar5
Mpgvip Tech Startup IoT IoE makeyourownlane Innovation GameChangers BigData NextGen Computing AI data IoT mobility FinTech
startuplife Machinelearning neuralnetworks javascript tensorflow MCubed DeepLearning AI MachineLearning BigData ML DL tech
Bitcoin Blockchain fintech bigdata infosec IoT AI GCPANEL timessquare DataScience Disruption BigData MarketExpansion
digital internet socialmedia search video ai iot bigdata IoT InternetOfThings IoE IIoT AI BigData Blockchain Fintech AI
BigData RT DataScience DataScientist BigData AI IoT opendata BigData analytics Agriculture AI Marketing GrowthHacking
BigData Mpgvip Defstar5 SEO SocialMedia MakeYourOwnLane machinelearning AI BigData RT IoT InternetOfThings IoE IIoT AI
BigData Blockchain Fintech programming developers softwaredevelopment ai java bigdata IoT InternetOfThings IoE IIoT AI
BigData Blockchain Fintech BigData DigitalMarketing CX Analytics CMO datascience INBOUND17 startup SmallBiz business tech
BigData LeanStartup health AI ml bigdata data science clinical venturecapital IoT InternetOfThings IoE IIoT AI BigData
Blockchain Fintech Internet SocialMediaDay Digital Bigdata mpgvip makeyourownlane Startup Marketing CRISPR tech science
BigData genomics CRISPR tech science BigData genomics IoT InternetOfThings IoE IIoT AI BigData Blockchain Fintech AI
```

urls.txt

```
urls - Notepad
File Edit Format View Help
http://bit.ly/2yhET2x http://paper.li/suchetkaushik/1320063382?edition_id=7f005270-a2b4-11e7-8d59-0cc47a0d15fd
http://bit.ly/2hwmUxX https://buff.ly/2hp7ejk http://ipfconline.fr/blog/2017/05/22/fine-list-of-50-top-world-big-data-
experts-to-follow-in-2017-with-moz-social-score/ https://www.irishtimes.com/life-and-style/health-family/taking-a-
scissors-to-our-dna-the-gene-editing-technology-set-to-revolutionise-the-fight-against-disease-1.3222642
https://buff.ly/2fP8c59 http://dld.bz/fsmxQ http://analyticsindiamag.com/top-6-regression-algorithms-used-data-mining-
applications-industry/ http://contexti.com/blog/cyber-security-strengthened-big-data-analytics-machine-learning/
https://buff.ly/2xzczzR https://www.computerworld.com/article/3227826/mobile-wireless/the-latest-iphones-show-why-ai-is-
the-new-electricity.html http://internetofthingsrecruiting.com/medical-iot-actually-secure
http://www.popsoci.com/hurricane-harm-help-ecosystem https://buff.ly/2y27ufj
https://twitter.com/i/web/status/912652225255821314 http://rismedia.com/2017/09/24/big-data-big-opportunities-how-
predictive-analytics-changing-real-estate/ http://bit.ly/2wmGUNu http://bit.ly/2jV7YIG
http://internetofthingsrecruiting.com/medical-iot-actually-secure http://internetofthingsrecruiting.com/medical-iot-
actually-secure https://www.forbes.com/sites/gregoryferenstein/2017/08/10/thoughts-on-how-online-data-science-courses-
stack-up-to-a-masters-degree/#5a2661252c88 https://en.m.wikipedia.org/wiki/Jack_Wild
https://twitter.com/i/web/status/912652393250246656 http://ow.ly/IVvr30fp2uK
http://internetofthingsrecruiting.com/medical-iot-actually-secure http://internetofthingsrecruiting.com/medical-iot-
actually-secure https://goo.gl/vOHZ79 http://internetofthingsrecruiting.com/medical-iot-actually-secure
https://www.irishtimes.com/life-and-style/health-family/taking-a-scissors-to-our-dna-the-gene-editing-technology-set-to-
revolutionise-the-fight-against-disease-1.3222642 https://www.irishtimes.com/life-and-style/health-family/taking-a-
scissors-to-our-dna-the-gene-editing-technology-set-to-revolutionise-the-fight-against-disease-1.3222642
http://internetofthingsrecruiting.com/medical-iot-actually-secure http://ow.ly/IVvr30fp2uK
http://internetofthingsrecruiting.com/medical-iot-actually-secure https://twitter.com/i/web/status/912652492810399744
https://blogs.wsj.com/cio/2017/09/08/what-your-ceo-is-reading-agricultures-silicon-valley-boards-and-digitization/
http://internetofthingsrecruiting.com/medical-iot-actually-secure https://lnkd.in/f8ep3FP http://nyti.ms/2wh8eIX
https://buff.ly/2xd6svo https://twitter.com/i/web/status/91265259995838464 http://www.nasdaq.com/article/how-facebook-is-
```

Step 3: Loading file in HDFS


```
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$ SHADOOP_HOME/bin/hdfs dfs -copyFromLocal hashtags.txt /twitterproject
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$ SHADOOP_HOME/bin/hdfs dfs -copyFromLocal urls.txt /twitterproject
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$ SHADOOP_HOME/bin/hdfs dfs -ls /twitterproject
Found 2 items
-rw-r--r--  1 avni supergroup    4119450 2017-09-29 21:01 /twitterproject/hashtags.txt
-rw-r--r--  1 avni supergroup    3104002 2017-09-29 21:01 /twitterproject/urls.txt
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$
```

Both output files viz. hashtags.txt and urls.txt are loaded to HDFS

Step 4: Running word count on hashtags and URLs

Word count for Hashtags

```

 version 2.2.0

Using Scala version 2.11.8 (OpenJDK 64-Bit Server VM, Java 1.8.0_131)
Type in expressions to have them evaluated.
Type :help for more information.

scala> val htFile = sc.textFile("hdfs://localhost:9000/twitterproject/hashtags.txt")
htFile: org.apache.spark.rdd.RDD[String] = hdfs://localhost:9000/twitterproject/hashtags.txt MapPartitionsRDD[1] at textFile at <console>:24

scala> val htWords = htFile.flatMap(line => line.split(" "))
htWords: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[2] at flatMap at <console>:26

scala> val htCounts = htWords.map(word => (word.toLowerCase, 1)).reduceByKey(_ + _)
htCounts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[4] at reduceByKey at <console>:28

scala> val htSortedCount = htCounts.sortBy(_._2)
htSortedCount: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[9] at sortBy at <console>:30

scala> htSortedCount.saveAsTextFile("hdfs://localhost:9000/twitterproject/htoutput")
```

- i. A RDD is created from the input file hashtags.txt.
- ii. Next, each line in the RDD is split into words and stored in htWords.
- iii. Then a Key-Value pair RDD is created with key as lowered-case word and value 1. By using the reduce operation, all pairs with key are reduced and values are added. This is stored in htCounts.
- iv. htCounts is sorted by values in descending order and stored in htSortedCount.
- v. It is then saved as a text file at the given hdfs location

Word count for URLs

```

scala> val urlFile = sc.textFile("hdfs://localhost:9000/twitterproject/urls.txt")
urlFile: org.apache.spark.rdd.RDD[String] = hdfs://localhost:9000/twitterproject/urls.txt MapPartitionsRDD[12] at textFile at <console>:24

scala> val urlWords = urlFile.flatMap(line => line.split(" "))
urlWords: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[13] at flatMap at <console>:26

scala> val urlCounts = urlWords.map(word => (word, 1)).reduceByKey(_ + _)
urlCounts: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[15] at reduceByKey at <console>:28

scala> val urlSortedCount = urlCounts.sortBy(_._2)
urlSortedCount: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[20] at sortBy at <console>:30

scala> urlSortedCount.saveAsTextFile("hdfs://localhost:9000/twitterproject/urloutput")
```

The word count for URLs is gathered in similar manner as hashtags with the exception of lower-casing of word.

Spark Jobs UI

10.0.2.15:4040/jobs/

Search

☆

📄

↓

🏠

🔒

☰

SPARK

Spark

2.2.0

Jobs

Stages

Storage

Environment

Executors

SQL

Spark shell application UI

Spark Jobs (?)

User: avni

Total Uptime: 15 min

Scheduling Mode: FIFO

Completed Jobs: 4

▶ Event Timeline

Completed Jobs (4)

Job Id ▾	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
3	saveAsTextFile at <console>:33	2017/09/29 21:18:31	2 s	2/2 (1 skipped)	4/4 (2 skipped)
2	sortBy at <console>:30	2017/09/29 21:18:18	2 s	2/2	4/4
1	saveAsTextFile at <console>:33	2017/09/29 21:15:11	3 s	2/2 (1 skipped)	4/4 (2 skipped)
0	sortBy at <console>:30	2017/09/29 21:14:50	7 s	2/2	4/4

Step 5: Copying Output from HDFS to local

Output folders created in HDFS

```
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$ $HADOOP_HOME/bin/hdfs dfs -ls /twitterproject
Found 4 items
-rw-r--r-- 1 avni supergroup 4119450 2017-09-29 21:01 /twitterproject/hashtags.txt
drwxr-xr-x - avni supergroup 0 2017-09-29 21:15 /twitterproject/htoutput
drwxr-xr-x - avni supergroup 0 2017-09-29 21:18 /twitterproject/urloutput
-rw-r--r-- 1 avni supergroup 3104002 2017-09-29 21:01 /twitterproject/urls.txt
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject$ $HADOOP_HOME/bin/hdfs dfs -copyToLocal /twitterproject
```

Output folders copied to local

```
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject/twitterproject$ ls -l
total 7064
-rw-r--r-- 1 avni avni 4119450 Sep 29 21:20 hashtags.txt
drwxrwxr-x 2 avni avni 4096 Sep 29 21:20 htoutput
drwxrwxr-x 2 avni avni 4096 Sep 29 21:20 urloutput
-rw-r--r-- 1 avni avni 3104002 Sep 29 21:20 urls.txt
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject/twitterproject$
```

```
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject/twitterproject/htoutput$ ls -l
total 84
-rw-r--r-- 1 avni avni 53198 Sep 29 21:20 part-00000
-rw-r--r-- 1 avni avni 32005 Sep 29 21:20 part-00001
-rw-r--r-- 1 avni avni 0 Sep 29 21:20 _SUCCESS
```

```
avni@avni-VirtualBox:~/PycharmProjects/TwitterProject/twitterproject/urloutput$ ls -l
total 1112
-rw-r--r-- 1 avni avni 1137832 Sep 29 21:20 part-00000
-rw-r--r-- 1 avni avni 0 Sep 29 21:20 part-00001
-rw-r--r-- 1 avni avni 0 Sep 29 21:20 _SUCCESS
```

Hashtags Wordcount Output

part-00000	
1	(bigdata,61730)
2	(ai,45466)
3	(machinelearning,36453)
4	(iot,27962)
5	(datascience,16016)
6	(ml,15359)
7	(fintech,14944)
8	(tech,12270)
9	(analytics,9994)
10	(blockchain,9869)
11	(iiot,9032)
12	(deeplearning,7286)
13	(cybersecurity,5987)
14	(artificialintelligence,5908)
15	(defstar5,5243)
16	(datascientist,4710)
17	(internetofthings,4487)
18	(cloud,4248)
19	(di,3966)
20	(makeyourownlane,3882)
21	(iioe,3675)
22	(innovation,3442)
23	(insurtech,3181)
24	(marketing,3129)
25	(industry40,3126)
26	(mpgvip,3125)
27	(seo,3022)
28	(startups,2798)
29	(startup,2724)
30	(data,2644)
31	(banking,2551)
32	(infosec,2450)
33	(smm,2390)
34	(digitaltransformation,2355)
35	(digital,2166)
36	(socialmedia,2054)
37	(growthhacking,2018)
38	(m2m,1907)
Normal text file	
length: 53198 lines: 3726 Ln: 1 Col: 1 Sel: 0 0	
UNIX UTF-8 INS	

URLs Wordcount Output

part-00000	
1	(http://bit.ly/NM00D_PB ,389)
2	(http://afkinside.com/140800/south-africa-establishes-first-data-science-academy-3-8m-investment ,348)
3	(http://maria-1ohnsen.com/multilingualSEO-blog/the-surprising-truth-about-self-aware-ai-robots/ ,331)
4	(http://bit.ly/2xCHXsG ,317)
5	(https://www.forbes.com/sites/benkerschberg/2017/06/20/how-data-science-helps-us-ask-the-right-questions-and-why-ibm-never-became-the-king-of-photocopies ,317)
6	(https://www.forbes.com/sites/bernardmarr/2017/06/06/the-9-best-free-online-big-data-and-data-science-courses ,286)
7	(https://bit.ly/2xvGNP9 ,288)
8	(http://www.rleweb.com/ ,277)
9	(https://www.nextbigfuture.com/2017/09/quantum-machine-learning.html ,267)
10	(http://bit.ly/2xvB7ana ,258)
11	(http://bit.ly/2xvCHH1 ,240)
12	(https://www.dmagazine.com/business-economy/2017/09/mark-cuban-invest-in-ai-or-get-left-behind/ ,222)
13	(https://www.cio.com/article/3222879/certifications/7-data-science-certifications-that-will-pay-off.html ,204)
14	(https://goo.gl/3vvvzf ,199)
15	(http://www.livemint.com/Opinion/2cmc8NnUz13UMihH1Ux4v/Digital-bank-of-the-future.html ,197)
16	(https://www.forbes.com/sites/forbesagencycouncil/2017/07/28/influencer-marketing-and-the-power-of-data-science ,193)
17	(https://singularityhub.com/2017/05/21/abundance-is-possible-but-only-if-everyones-along-for-the-ride/ ,191)
18	(https://www.wired.com/story/ikea-place-ar-kit-augmented-reality/?mbid=nl_092117_gadgetlab_intro ,188)
19	(
20	(http://www.techrepublic.com/article/report-average-enterprise-data-breach-cost-rises-to-1-3m/?hbid=27537035630225091086340873861452&ftag=TRF-03-10aaa6b&ftag=TRF-03-10aaa6b ,173)
21	(http://bit.ly/2xvB7ol ,170)
22	(https://www.forbes.com/sites/quora/2017/02/06/is-data-science-too-easy ,168)
23	(http://internetofthingsrecruiting.com/how-secure-are-home-iot-devices ,168)
24	(https://www.r-bloggers.com/why-i-use-r-for-data-science-an-ode-to-r ,166)
25	(http://internetofthingsrecruiting.com/stop-losing-candidates-counter-offers ,164)
26	(http://internetofthingsrecruiting.com/internet-things-timeline ,162)
27	(http://internetofthingsrecruiting.com/div-applications-expand-reach-mindshare-iot ,160)
28	(http://www.bigcloud.io/data-scientists-how-to-perfect-your-cv ,159)
29	(http://internetofthingsrecruiting.com/why-should-companies-care-about-iot ,157)
30	(https://internetofthingsrecruiting.com/ddos-attacks-using-iot-devices ,157)
31	(http://www.itweb.co.za/index.php?ad=164835&options=com_content&view=article&catid=69#WcV3XpFFark_twitter ,157)
32	(http://internetofthingsrecruiting.com/unique-new-users-iot ,156)
33	(http://www.dataeverity.net/data-science-use-cases ,154)
34	(https://www.cognitomediam.com/news/2017/09/26/the-future-of-financial-markets-1 ,154)
35	(https://www.forbes.com/sites/kalevlestaru/2017/07/30/from-subject-to-soundbite-results-to-hype-communicating-data-science/#58c6318e6c35 ,150)
36	(http://internetofthingsrecruiting.com/byod-it-security-and-the-internet-of-things ,149)
37	(http://internetofthingsrecruiting.com/sigma-z-wave-enters-public-domain ,149)
Normal text file	
length: 1137832 lines: 21855 Ln: 1 Col: 1 Sel: 0 0	
UNIX UTF-8 INS	