

# CS 751: Assignment 4

Rohit Lambi

Spring 2015

# Contents

1	Q1 . . . . .	2
	1.1 Solution . . . . .	2
	1.2 Code Listing . . . . .	3
2	Q2 . . . . .	7
	2.1 Solution . . . . .	7
	2.2 Code Listing . . . . .	8
3	Q3 . . . . .	11
	3.1 Solution . . . . .	11
	3.2 Code Listing . . . . .	16
4	Q4 . . . . .	19
	4.1 Solution . . . . .	19
	4.2 Code Listing . . . . .	29

# 1 Q1

1. Using the pages from A3 that boilerpipe successfully processed, download those representations again and reprocess them with boilerpipe.
  - 1.1 Document the time difference (e.g., Time(A4) - Time(A3)).
  - 1.2 Compute the Jaccard Distance x for each pair of pages (i.e., P(A3) and P(A4) for:
    - Unique terms (i.e., unigrams)
    - Bigrams
    - Trigrams
  - 1.3 See: [http://en.wikipedia.org/wiki/Jaccard\\_index](http://en.wikipedia.org/wiki/Jaccard_index) - WSDM 2010 paper: <http://www.l3s.de/~kohlschuetter/boilerplate/>
  - 1.4 For each of the 3 cases (i.e., 1-, 2-, 3-grams) build a Cumulative Distribution Function that shows the percentage change on the x-axis and the percentage of the population on the y-axis
  - 1.5 See: [http://en.wikipedia.org/wiki/Cumulative\\_distribution\\_function](http://en.wikipedia.org/wiki/Cumulative_distribution_function)

## 1.1 Solution

1. Time difference between the two sets of boilerpipe representations of the HTML documents is one month.
2. I already had one set of boilerpipe representations of these HTML documents from A3.
3. I used the same python program using 'boilerpipe' library from A3 to fetch just the text from the same 10,000 URIs of A3.
4. Next, I wrote a python program to find the unigram, bigrams and trigrams and find the Jaccard Distance x for each pair of pages (i.e., P(A3) and P(A4)). To find unigram, bigrams and trigrams I used 'nltk' library.
5. Following table lists few examples which shows the range of change measured:

Sno	Unigram JD	Bigram JD	Trigram JD
1	0.0	0.0	0.0
2	0.005	0.006	0.009
3	0.47	0.52	0.55
4	0.83	0.94	0.95
5	1.0	1.0	1.0

Figure 1: Examples

## 1.2 Code Listing

### extractTextWithBoilerpipe.py

```
1 import os
2 import thread
3 import threading
4 import csv
5 import datetime
6 from boilerpipe.extract import Extractor
7
8
9
10 iteration = '10'
11 INPUT_PATH = 'input-urls/' + iteration + '.txt'
12 SKIPPED_PATH = "skipped-urls/skipped-" + iteration + ".txt"
13
14 BOILERPIPE_TEXT = 'boilerpipe-text/' + iteration
15 if not os.path.exists(BOILERPIPE_TEXT):
16     print 'Creating folder - ' + BOILERPIPE_TEXT
17     os.makedirs(BOILERPIPE_TEXT)
18
19
20 def fetchWebPage(URL):
21
22     extractor = Extractor(url=URL)
23     extracted_text = extractor.getText()
24     return extracted_text
25
26
27 fieldNames = ['sno', 'seqNum', 'tcoUrl', 'url']
28
29 print datetime.datetime.now()
30
31 totalUrls = 0
32 skipCnt = 0
33 emptyContent = 0
34
35 skipped = open(SKIPPED_PATH, 'w')
36 with open(INPUT_PATH) as csvfile:
37     reader = csv.DictReader(csvfile, fieldnames=fieldNames, delimiter='\t')
38
39     for row in reader:
40
41         totalUrls += 1
42         try:
43             print 'Fetching ', str(row['sno']), ' ', row['url']
44             extracted_text = fetchWebPage(row['url'])
45
46             if not extracted_text:
47                 emptyContent += 1
48
49             fil = open(BOILERPIPE_TEXT + '/' + row['sno'], 'w')
50             fil.write(extracted_text.encode('UTF-8'))
51             fil.close()
52
53         except:
54             skipCnt += 1
55             skipped.write(row['sno'] + '\t' + row['url'] + '\n')
56
57 skipped.close()
58
59 print datetime.datetime.now()
60
61 summary = open("summary-" + iteration + ".txt", 'w')
62 summary.write("TotalUrls - " + str(totalUrls))
63 summary.write("\nSkipped - " + str(skipCnt))
64 summary.write("\nEmpty Content - " + str(emptyContent))
65 summary.close()
66
67
68
```

```
69 #print 'Waiting for all threads to complete'  
70 #while threading.activeCount() > 1:  
71 #    print str(threading.activeCount())  
72 #    pass  
73  
74 print 'Completed fetching all webpages'  
75 print datetime.datetime.now()
```

Listing 1: Python program to extract only text from HTML documents using boilerpipe

## findNgrams.py

```
1 import os
2 from nltk.util import ngrams
3 import re
4
5
6 def removePunctuation(paragraph):
7
8     paragraph = re.sub('[\",() ;?\\\[\\].{}#&_.*+=%!<>~0-9]', '', paragraph)
9     return paragraph
10
11
12 def normalizeString(paragraph):
13
14     words = paragraph.split()
15     normalizedWordList = []
16
17     for word in words:
18         word = word.lower()
19         word = word.strip()
20
21         #word = removePunctuation(word);
22         normalizedWordList.append(word)
23
24     return normalizedWordList
25
26 def getUniqueNgram(paragraph, n):
27     nGrams = ngrams(normalizeString(paragraph), n)
28     uniqueNGrams = set(nGrams)
29     return uniqueNGrams
30
31 def getNgrams(filePath):
32     f = open(filePath, 'r')
33     paragraph = f.read()
34     f.close()
35
36     uniGram = getUniqueNgram(paragraph, 1)
37     biGrams = getUniqueNgram(paragraph, 2)
38     triGrams = getUniqueNgram(paragraph, 3)
39
40     grams = {'uniGram': uniGram, 'biGrams': biGrams, 'triGrams': triGrams}
41
42     return grams
43
44
45 def writeNgrams(grams, oneGFile, twoGFile, threeGFile):
46     oneGFile.write('\t' + str(len(grams['uniGram'])))
47     twoGFile.write('\t' + str(len(grams['biGrams'])))
48     threeGFile.write('\t' + str(len(grams['triGrams'])))
49
50 def writeJaccardDistance_old(a3Grams, a4Grams, oneGFile, twoGFile, threeGFile):
51     oneGJD = str(calculateJaccardDistance(a3Grams['uniGram'], a4Grams['uniGram']))
52     twoGJD = str(calculateJaccardDistance(a3Grams['biGrams'], a4Grams['biGrams']))
53     threeGJD = str(calculateJaccardDistance(a3Grams['triGrams'], a4Grams['triGrams']))
54
55     oneGFile.write('\t' + oneGJD + '\t' + twoGJD + '\t' + threeGJD)
56
57 def writeJaccardDistance(a3Grams, a4Grams, fName, jdFile):
58     oneGJD = str(calculateJaccardDistance(a3Grams['uniGram'], a4Grams['uniGram']))
59     twoGJD = str(calculateJaccardDistance(a3Grams['biGrams'], a4Grams['biGrams']))
60     threeGJD = str(calculateJaccardDistance(a3Grams['triGrams'], a4Grams['triGrams']))
61     jdFile.write(str(fName) + '\t' + oneGJD + '\t' + twoGJD + '\t' + threeGJD + '\n')
62
63
64 def calculateJaccardDistance(a3Words, a4Words):
65     union = set(a3Words).union(a4Words)
66     intersect = set(a3Words).intersection(a4Words)
67     jacDist = (len(union) - len(intersect)) * 1.0 / len(union)
68     return jacDist
69
70
```

```

71 A3_INPUT_PATH = '../A3/boilerpipe_text/all'
72 A4_INPUT_PATH = 'boilerpipe_text/all'
73
74 #A3_INPUT_PATH = '../A3/sample-boilerpipe/'
75 #A4_INPUT_PATH = 'sample-boilerpipe/'
76
77 A3Files = os.listdir(A3_INPUT_PATH)
78 A4Files = os.listdir(A4_INPUT_PATH)
79
80 OUTPUT_DIR = 'grams'
81 if not os.path.exists(OUTPUT_DIR):
82     os.makedirs(OUTPUT_DIR)
83
84 oneGFile = open(OUTPUT_DIR + "/jaccard-distance.txt", 'w')
85
86 for fName in A3Files:
87     if fName in A4Files:
88         a3fPath = A3_INPUT_PATH + '/' + fName
89         a4fPath = A4_INPUT_PATH + '/' + fName
90
91         a3Grams = getNgrams(a3fPath)
92         a4Grams = getNgrams(a4fPath)
93
94         if len(a3Grams['triGrams']) != 0 and len(a4Grams['triGrams']) != 0:
95             writeJaccardDistance(a3Grams, a4Grams, fName, oneGFile)
96
97
98
99 oneGFile.close()

```

Listing 2: Python program to calculate jaccard distance for Unigram, Bigrams and Trigrams

## 2 Q2

1. 2.1 Using the pages from Q1 (A4), download all TimeMaps (including TimeMaps with 404 responses, i.e. empty or null TimeMaps)
  - Upload all the TimeMaps to github
2. 2.2 Build a CDF for number of mementos for each original URI (i.e., x-axis = number of mementos, y-axis = percentage of links)

See: <http://timetravel.mementoweb.org/guide/api/>

### 2.1 Solution

1. I wrote a python program to fetch the timemaps. The aggregator that I used is <http://labs.mementoweb.org/timemap/json/>
2. Following is the graph showing CDF for number of mementos for each original URI:

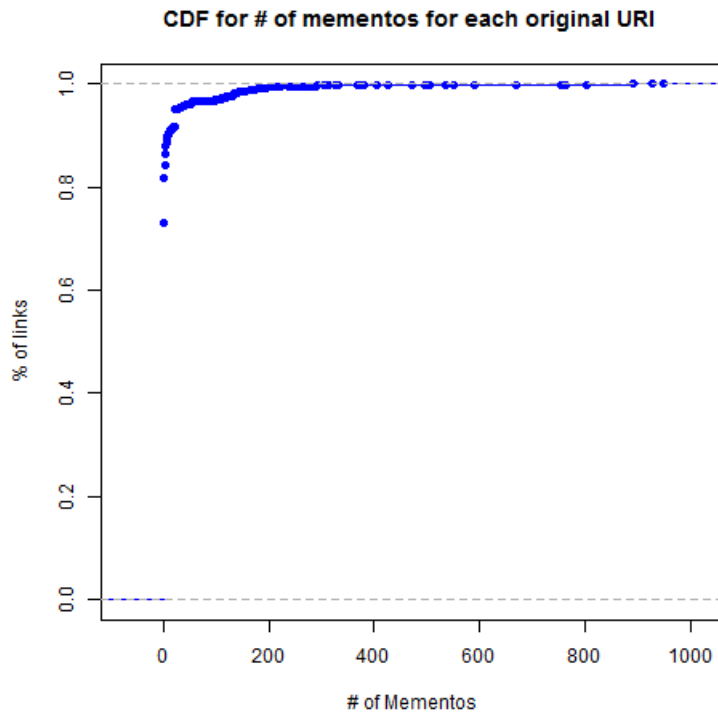


Figure 2: Examples



## 2.2 Code Listing

Python program to fetch the timemaps for the URIs from Q1 on which boilerpipe is successful

```
1  '''
2  Created on Feb 8, 2015
3
4  @author: rlambi
5  '''
6  import subprocess
7  import os
8  import thread
9  import threading
10 import csv
11 import datetime
12
13 def fetchTimemap(url, fileName):
14     url = "http://labs.mementoweb.org/timemap/json/" + url
15     print "\n", str(fileName), url
16     #subprocess.Popen(["wget","--output-document=" + fileName, url]) # + ".html"
17     command = "wget --output-document=" + fileName + " " + url
18     print command
19     os.system(command)
20
21 sites = 'timemaps'
22 if not os.path.exists(sites):
23     os.makedirs(sites)
24
25 os.chdir(sites)
26
27 fieldNames = [ 'sno', 'seqNum', 'tcoUrl', 'url' ]
28
29 print datetime.datetime.now()
30 with open('../input-urls/10.txt') as csvfile:
31     #with open('../all-urls.txt') as csvfile:
32     #with open('../sample-urls.txt') as csvfile:
33     reader = csv.DictReader(csvfile, fieldnames=fieldNames, delimiter='\t')
34     for row in reader:
35         fetchTimemap(row['url'], row['sno'])
36         #thread.start_new_thread(fetchWebPage, (row['url'], row['sno'], ))
37
38 print datetime.datetime.now()
39
40 #print 'Waiting for all threads to complete'
41 #while threading.activeCount() > 1:
42 #    print str(threading.activeCount())
43 #    pass
44
45 print 'Completed fetching all webpages'
46 print datetime.datetime.now()
```

## Python program to count the mementos for each Timemap

```
1 import json
2 import os
3
4 INPUT_PATH = '2.timemaps/'
5 OUTPUT_PATH = 'mementoCount/'
6 OUTPUT_FILE = 'mementoCount.txt'
7
8 if not os.path.exists(OUTPUT_PATH):
9     print 'Creating folder - ' + OUTPUT_PATH
10    os.makedirs(OUTPUT_PATH)
11
12 fileList = os.listdir(INPUT_PATH)
13 #fileList = ['1201']
14 oFile = open(OUTPUT_PATH + OUTPUT_FILE, 'w')
15 for fName in fileList:
16     f = open(INPUT_PATH + fName, 'r')
17     jsonContent = f.read()
18
19     if len(jsonContent) == 0:
20         mementoCount = 0
21     else:
22         fContent = json.loads(jsonContent)
23         #print fContent['mementos']['list'][0]['uri']
24         try:
25             mementoCount = len(fContent['mementos']['list'])
26         except KeyError:
27             mementoCount = 0
28
29     oFile.write(fName + '\t' + str(mementoCount) + '\n')
30     f.close()
31
32 oFile.close()
```

## R code to plot CDF

```
1 library(RColorBrewer)
2
3 data <- read.table("mementoCount.txt",header=FALSE)
4 finalData = rep(data[,1])
5 title <- "CDF for # of mementos for each original URI "
6 xlab <- "# of Mementos"
7 ylab <- "% of links"
8 P1 = ecdf(finalData)
9 png("mementoCount.png")
10
11 p1 <- plot(P1,col="blue", main=title , xlab=xlab , ylab=ylab)
```

### 3 Q3

1. 3.1 Using 20 links that have TimeMaps
  - With 20 or more mementos
  - Have existed 2 years or more (i.e., Memento-Datetime of “first memento” is April XX, 2013 or older)
  - Note: select from Q1/Q2 links, else choose them by hand
2. 3.2 For each link, create a graph that shows Jaccard Distance, relative to the first memento, through time
  - x-axis: continuous time, y-axis: Jaccard Distance relative to the first memento

#### 3.1 Solution

1. I hand picked 20 timemaps from Q2.
2. I wrote a python program to process the timemaps, get the memento URIs, download the boilerpipe representation.
3. Next, I included the code for calculation of jacard distance of all other mementos relative to the first memento.
4. Following are few graphs that shows Jaccard Distance relative to the first memento, through time:

## Graphs

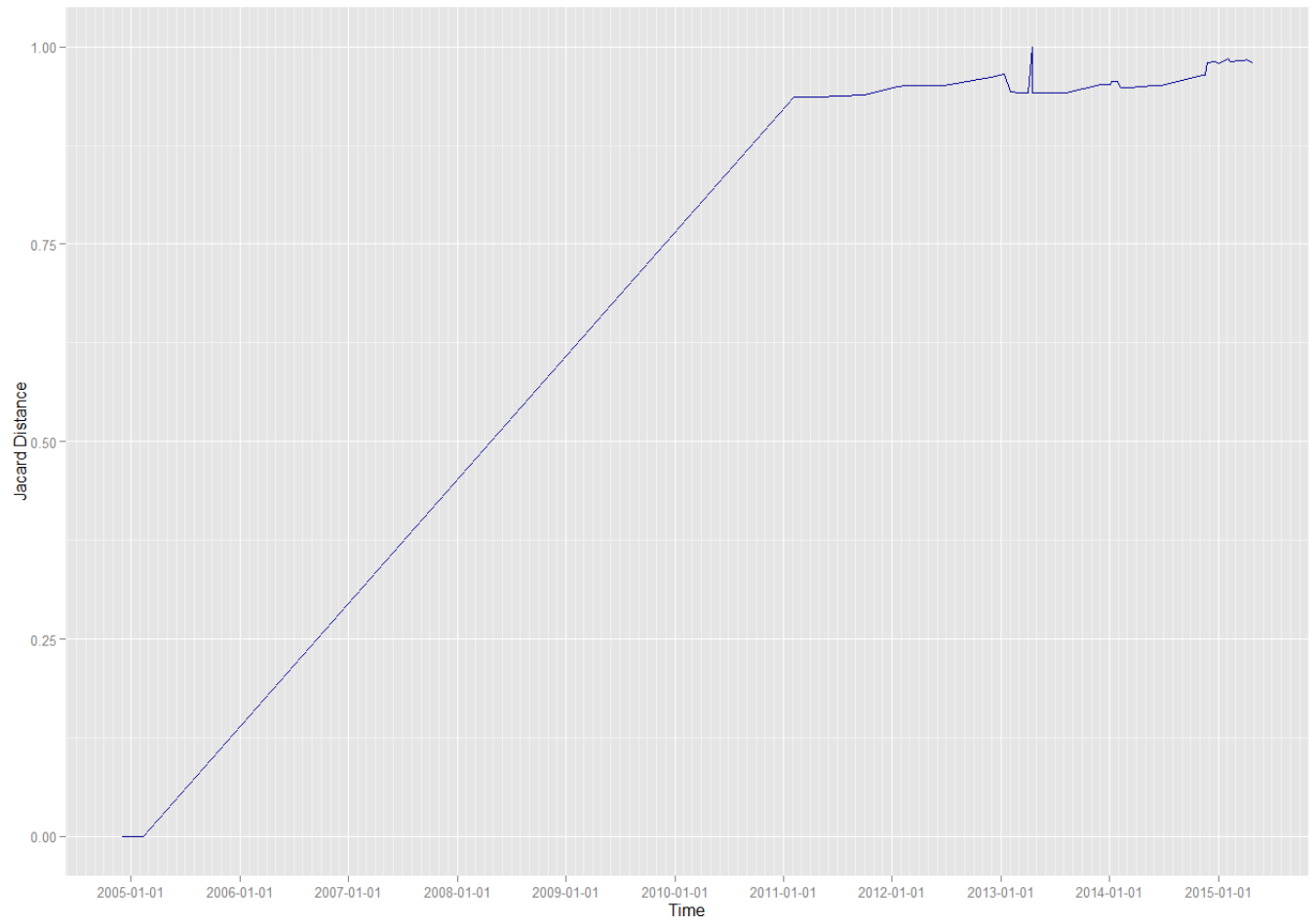


Figure 3: Memento 1

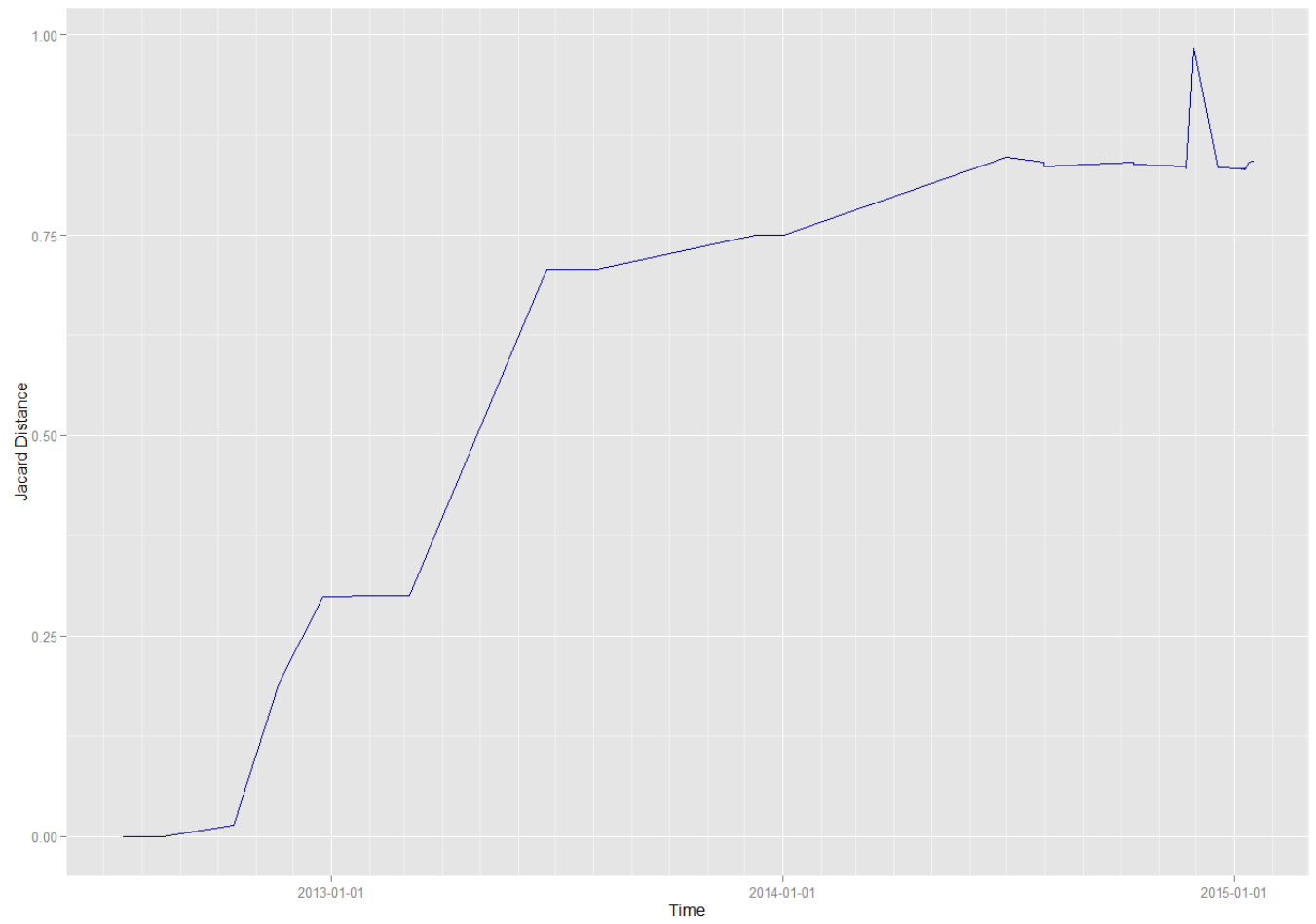


Figure 4: Memento 2

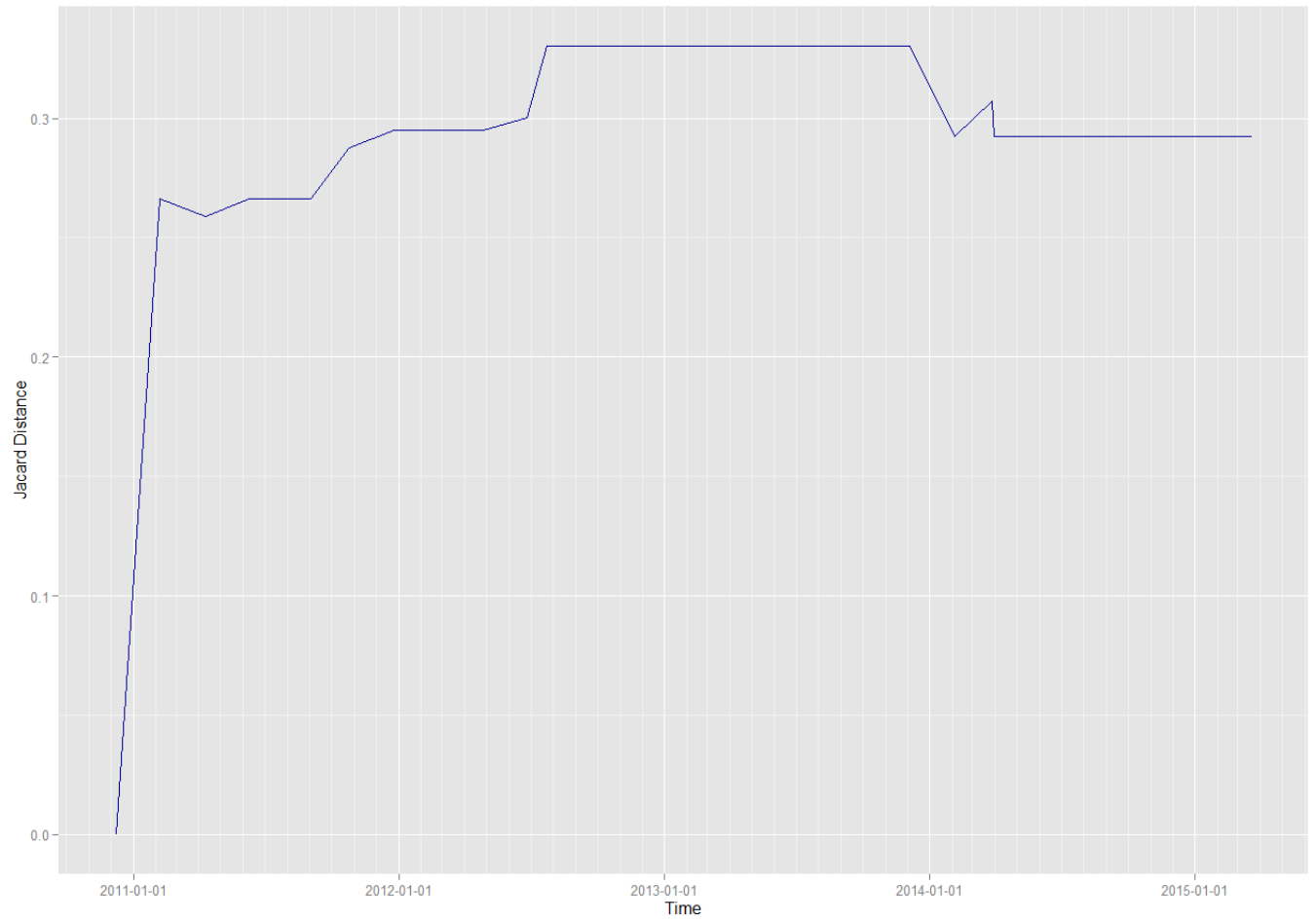


Figure 5: Memento 3

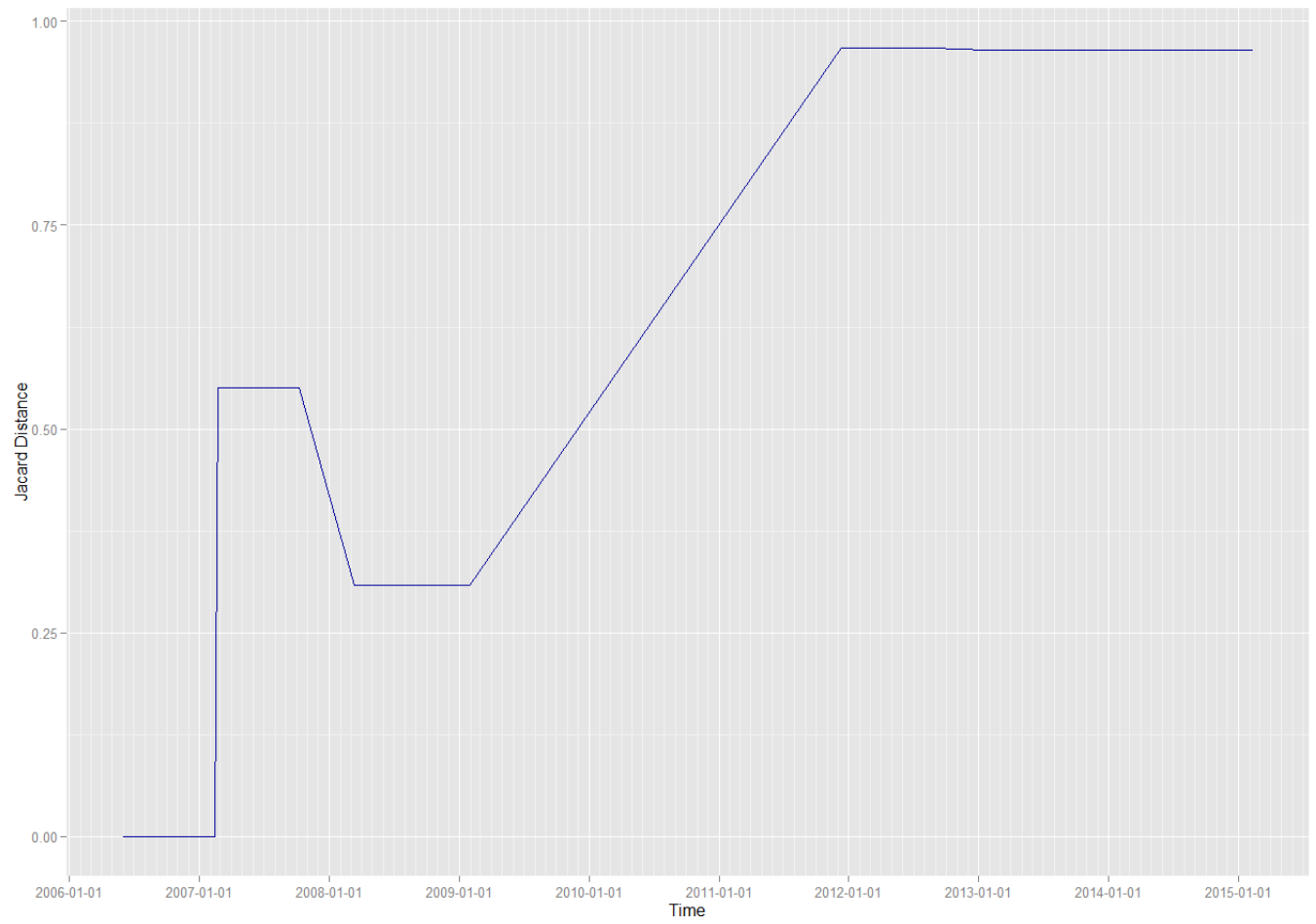


Figure 6: Memento 4



## 3.2 Code Listing

Python program to using boilerpipe library to extract just the text from HTML documents of the mementos for the selected 20 timemaps

```
1 import json
2 import os
3 from boilerpipe.extract import Extractor
4 from nltk.util import ngrams
5 import urllib2
6
7 INPUT_PATH = '3.selected-timemaps/'
8 OUTPUT_PATH = 'mementos/'
9 OUTPUT_FILE = 'mementoCount.txt'
10
11 def removePunctuation(paragraph):
12
13     paragraph = re.sub('[\",() :;?\\[\\].{}#&_.*+=%!<>~0-9]', '', paragraph)
14     return paragraph
15
16
17 def normalizeString(paragraph):
18
19     words = paragraph.split()
20     normalizedWordList = []
21
22     for word in words:
23         word = word.lower()
24         word = word.strip()
25
26         #word = removePunctuation(word);
27         normalizedWordList.append(word)
28
29     return normalizedWordList
30
31 def getUniqueNgram(paragraph, n):
32     nGrams = ngrams(normalizeString(paragraph), n)
33     uniqueNGrams = set(nGrams)
34     return uniqueNGrams
35
36 def calculateJaccardDistance(a3Words, a4Words):
37     union = set(a3Words).union(a4Words)
38     intersect = set(a3Words).intersection(a4Words)
39     jacDist = (len(union) - len(intersect)) * 1.0 / len(union)
40     return jacDist
41
42 def getNgrams(filePath):
43     f = open(filePath, 'r')
44     paragraph = f.read()
45     f.close()
46     uniGram = getUniqueNgram(paragraph, 1)
47     return uniGram
48
49 def saveBoilerpipeText(URL, mementoId, bPipeOpFilePath):
50     extractor = Extractor(url=URL)
51     extracted_text = extractor.getText()
52
53     fil = open(bPipeOpFilePath, 'w')
54     fil.write(extracted_text.encode('UTF-8'))
55     fil.close()
56
57
58
59 if not os.path.exists(OUTPUT_PATH):
60     print 'Creating folder - ' + OUTPUT_PATH
61     os.makedirs(OUTPUT_PATH)
62
63 fileList = os.listdir(INPUT_PATH)
64 #fileList = ['1201']
65
66
```

```

67 for fName in fileList:
68     print 'processing for ' + fName
69     iFile = open(INPUT_PATH + fName, 'r')
70     jsonContent = iFile.read()
71
72     if len(jsonContent) != 0:
73         fContent = json.loads(jsonContent)
74         #print fContent['mementos']['list'][0]['uri']
75         mementoURIs = fContent['mementos']['list']
76         BPIPE_OUTPUT_PATH = OUTPUT_PATH + fName + '/'
77
78         if not os.path.exists(BPIPE_OUTPUT_PATH):
79             os.makedirs(BPIPE_OUTPUT_PATH)
80
81         oFile = open(BPIPE_OUTPUT_PATH + 'jaccard-distance.txt', 'w')
82         for i, URI in enumerate(mementoURIs):
83             mementoId = i + 1
84             bPipeOpFilePath = BPIPE_OUTPUT_PATH + str(mementoId)
85
86             try:
87                 print str(mementoId), '\t', URI['uri']
88                 saveBoilerpipeText(URI['uri'], mementoId, bPipeOpFilePath)
89                 currentMementoWords = getNgrams(bPipeOpFilePath)
90
91                 if mementoId == 1:
92                     firstMementoWords = currentMementoWords
93
94                 if currentMementoWords != None:
95                     jacDist = calculateJaccardDistance(firstMementoWords,
96                                                         currentMementoWords)
97                     oFile.write('1\t' + str(mementoId) + '\t' + str(jacDist) + '\t'
98                                + URI['datetime'] + '\n')
99             except urllib2.HTTPError:
100                 print 'HTTP Error'
101                 continue
102             except:
103                 print 'Error'
104                 continue
105
106         oFile.close()
107     iFile.close()

```

## R code to plot CDF

```
1 j4 <- read.csv(file="j7.txt",header=TRUE,sep="\t")
2
3 j4$Time <- as.Date(strptime(j4$Time, "%Y-%m-%dT%H:%M:%S"))
4
5 ggplot(j4, aes(Time, Distance)) + geom_line(colour="#000099") +
6 scale_x_date(breaks = "1 year", minor_breaks = "1 month") + xlab("Time") + ylab("Jacard  
Distance")
```

## 4 Q4

1. 3.1 Choose a news-related event
2. 3.2 Use twarc.py to collect 1000 tweets, every day for 5 different days
  - See: <https://github.com/edsu/twarc>
3. 3.3 For each day:
  - Create a wall
  - Build a tag/word cloud for each day
  - Create a map using GeoJSON and Github
    - See: <https://help.github.com/articles/mapping-geojson-files-on-github/>
4. 3.4 Discuss in detail lessons learned, experiences, etc.

### 4.1 Solution

1. I wrote a python program using 'Twarc' library to collect daily 1000 tweets.
2. The news-related event that I chose was "Nepal Earthquake".
3. Once I was done collecting the tweets, I used different tools in the Twarc package to get the wall, word cloud and geojson from these saved tweets.
4. Wall is a HTML document displaying the tweets with user handle, Tweet text, displays URLs if any, date time of the Tweet and displays the number of retweets.
5. Word Cloud is a HTML document which displays the cloud of most frequently used words in the list of tweets and displays these words as an image beautifully.
6. GeoJSON tool gives the locations from where the Tweet originated. It can give the location only if the Tweet has location information. Using this location information, github.com plots these on map.
7. Following are the screenshots of walls, word clouds and geojsons for 5 different days.

## Walls

## Title Here

created on the command line with [twarc](#)

















 <b>Andrew Dawson</b> awjdaws Perseverance and Grace as Nepal Is Turned Upside Down <a href="http://t.co/d0Xy1FtOia">http://t.co/d0Xy1FtOia</a> 0 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>We need u zayn</b> DaizZii_1D RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>Ari Sidharta</b> arsdhr RT @therealbanksy: Praying for strength and courage to all those affected in #Nepal #India #Earthquake #NepalQuake <a href="http://t.co/ty6gcyg9GG">http://t.co/ty6gcyg9GG</a> 2867 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>Joanna.</b> palegrungehes RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:27 +0000 2015
 <b>anodyne</b> DiChristine RT @USEmbassyNepal: Great primer from @USAID on how to help and stay informed on #NepalQuake: <a href="http://t.co/P0r8lXRdk1">http://t.co/P0r8lXRdk1</a> 7 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>MM_Feels</b> RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>zaynie</b> SpicyChichi RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:27 +0000 2015	 <b>DiadaTerra</b> ruisaldanha RT @geographile: Survivors and Dead Evacuated from Everest Base Camp <a href="http://t.co/kSE4pynvL7">http://t.co/kSE4pynvL7</a> Thorough summation of conditions on/near Everest... 1 Retweets Mon Apr 27 01:01:27 +0000 2015
 <b>Palmy P.</b> p_ppalmy RT @UNICEF_Thailand: Earthquake in Nepal. UNICEF is assessing immediate needs of children there, readying for rapid humanitarian response. ... 975 Retweets Mon Apr 27 01:01:26 +0000 2015	 <b>▶↑&amp;HARRY&amp;↓◀</b> HxranDxrableHaz RT @Harry_Styles: Thinking about everyone involved in the Earthquake in Nepal.. Such awful news. 125317 Retweets Mon Apr 27 01:01:26 +0000 2015	 <b>Itzel Malik Horan</b> itzel0808 RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:26 +0000 2015	 <b>DARREN</b> Dazzler2711 RT @RedCrossAU: URGENT: We have launched an appeal to assist those affected by the #NepalEarthquake: <a href="http://t.co/2Lu0YJA6CY">http://t.co/2Lu0YJA6CY</a> 296 Retweets Mon Apr 27 01:01:26 +0000 2015
 <b>Vibhu Mehta</b> vibhubebe RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:25 +0000 2015	 <b>Cintia</b> Cintia2k9 RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:25 +0000 2015	 <b>PLEASE JUSTIN</b> OMGmyFellings RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 85883 Retweets Mon Apr 27 01:01:25 +0000 2015	 <b>Faika Hoggan</b> hydowahodos RT @stockzeus: Let's all take a moment to say a prayer for those affected by the Earthquake in Nepal. May God bless you and be with you. 220 Retweets Mon Apr 27 01:01:25 +0000 2015

Figure 7: Wall 1

## Title Here

created on the command line with [twarc](#)

















 <b>horan</b> lubnasf RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNi9m">http://t.co/NXBJftNi9m</a> or text: DONATE5 ... 39787 Retweets Tue Apr 28 00:42:10 +0000 2015	 <b>Ara</b> nxall_h RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNi9m">http://t.co/NXBJftNi9m</a> or text: DONATE5 ... 39787 Retweets Tue Apr 28 00:42:10 +0000 2015	 <b>Armine</b> arminkaaa RT @KimKardashian: Praying for victims of the earthquake in Nepal. Pls. help @prabalgurung raise relief money by donating here <a href="http://t.co/2gn7raSsl9">http://t.co/2gn7raSsl9</a> 1990 Retweets Tue Apr 28 00:42:10 +0000 2015	 <b>キリエ</b> KiryNet RT @TelegraphNews: Nepal earthquake: Everest cleared, but dozens of Britons still missing <a href="http://t.co/TzeDMbxc3r">http://t.co/TzeDMbxc3r</a> <a href="http://t.co/jNkdAQunTb">http://t.co/jNkdAQunTb</a> 39 Retweets Tue Apr 28 00:42:10 +0000 2015
 <b>Laurel</b> hasselhopf RT @patagonia: Our hearts go out to all impacted by the earthquake in Nepal. Be a part of the relief with @wavesforwater: <a href="http://t.co/trwOR74VoP">http://t.co/trwOR74VoP</a> 47 Retweets Tue Apr 28 00:42:10 +0000 2015	 <b>fatima jamil</b> fatjimjam RT @prophiphop: Murder, Protests, riots, police states. And Nepal crumbled under an earthquake and drag each other out of rubble. We have a... 188 Retweets Tue Apr 28 00:42:10 +0000 2015	 <b>Wini Amiwati</b> winiamia RT @NGIndonesia: Gempa di Nepal kali ini memang mematikan, apa penyebababnya? <a href="http://t.co/ZJuixCWbtQ">http://t.co/ZJuixCWbtQ</a> <a href="http://t.co/etfb1mA5Dk">#prayfornepal</a> eart... 2 Retweets Tue Apr 28 00:42:09 +0000 2015	 <b>Apollonia .</b> babygirlx3 RT @ICEYTREY: Bruce Jenner still front page of Yahoo when 1500 people died in an earthquake in Nepal this morning 1600 Retweets Tue Apr 28 00:42:09 +0000 2015
 <b>lexinerus</b> lexinerus ReTw acrossUpdate: Facebook, Google and other tech firms aid in Nepal earthquake relief - San Jose Mercury News: I... <a href="http://t.co/5vyJrkSUC">http://t.co/5vyJrkSUC</a> 0 Retweets Tue Apr 28 00:42:09 +0000 2015	 <b>sophia</b> tousledhes RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNi9m">http://t.co/NXBJftNi9m</a> or text: DONATE5 ... 39787 Retweets Tue Apr 28 00:42:09 +0000 2015	 <b>Please Louis</b> Niallfaan RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre... 124565 Retweets Tue Apr 28 00:42:08 +0000 2015	 <b>Michael Soussan</b> michaelsoussan RT @CNN: Be careful of scams when donating. Here are the organizations on the ground providing help: <a href="http://t.co/SLCU28r6j8">http://t.co/SLCU28r6j8</a> <a href="http://t.co/Sx6...">http://t.co/Sx6...</a> 824 Retweets Tue Apr 28 00:42:08 +0000 2015
 <b>Rene Horton</b> rene_horton RT @RNBkfast: According to @UNICEF, almost 1 million children are now in need of urgent humanitarian assistance: <a href="http://t.co/eADBncozHL">http://t.co/eADBncozHL</a> ... 2 Retweets Tue Apr 28 00:42:07 +0000 2015	 <b>WI Libertarian</b> WI_Libertarian Israel Sends 260 Emergency Responders To Help Victims Of Nepal Earthquake, Human Rights Watch Director Slams Them... <a href="https://t.co/4qlPxTS3ev">https://t.co/4qlPxTS3ev</a> 0 Retweets Tue Apr 28 00:42:07 +0000 2015	 <b>It is what it is</b> CuteQuirks RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNi9m">http://t.co/NXBJftNi9m</a> or text: DONATE5 ... 39787 Retweets Tue Apr 28 00:42:05 +0000 2015	 <b>Nabin Gurung</b> Skipshot RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNi9m">http://t.co/NXBJftNi9m</a> or text: DONATE5 ... 39787 Retweets Tue Apr 28 00:42:05 +0000 2015

Figure 8: Wall 2

# Title Here

created on the command line with [twarc](#)

















 <p><b>Just Financial News</b> justfinancenews</p> <p>Nepal Earthquake: India promises help till normalcy returns <a href="http://t.co/ieeHjfnEAI">http://t.co/ieeHjfnEAI</a></p> <p>0 Retweets Wed Apr 29 00:13:33 +0000 2015</p>	 <p><b>★..C&amp;ç¥..★</b> ivceci</p> <p>RT @lGeociencias: El Himalaya, tierra de colisión de continentes <a href="http://t.co/ZyE7K9xshw">http://t.co/ZyE7K9xshw</a> <a href="http://t.co/a4fRopzOXS">http://t.co/a4fRopzOXS</a></p> <p>32 Retweets Wed Apr 29 00:13:33 +0000 2015</p>	 <p><b>Lucy Hale is life</b> sheyla_sheyla1</p> <p>RT @selenagomez: Help children affected by the #NepalEarthquake - support UNICEF's relief efforts <a href="http://t.co/Vhiury5w2n">http://t.co/Vhiury5w2n</a> via @UNICEFUSA #Ne...</p> <p>6999 Retweets Wed Apr 29 00:13:33 +0000 2015</p>	 <p><b>Ignacia Calderon</b> IgnaciaaMolina</p> <p>RT @Harry_Styles: To help those affected by the devastating earthquake in Nepal, text: DONATE5 to 70008 and give £5. Thank you @savechildre...</p> <p>135501 Retweets Wed Apr 29 00:13:33 +0000 2015</p>
 <p><b>Mariana Blas</b> MinaBlas</p> <p>RT @UncleBlazer: weird putting things in perspective when the news goes from basically race riots in the US to almost 5000 dead in Nepal fr...</p> <p>929 Retweets Wed Apr 29 00:13:32 +0000 2015</p>	 <p><b>Safe बिपु 100coolacharya</b></p> <p>"@FCBarcelona: A minute's silence is observed in memory of the victims of the earthquake in Nepal #FCBGetafe #FCBlive" Barca #respect</p> <p>0 Retweets Wed Apr 29 00:13:31 +0000 2015</p>	 <p><b>beccy owen</b> beccyowen</p> <p>@WestYorkshireUK NEW Beccy Owen SINGLE JUST RELEASED for #NepalQuakeRelief <a href="https://t.co/oDJAdJVUQk">https://t.co/oDJAdJVUQk</a> ... PLEASE RT</p> <p>0 Retweets Wed Apr 29 00:13:30 +0000 2015</p>	 <p><b>Daniela</b> Dani_Berenice</p> <p>RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNI9m">http://t.co/NXBJftNI9m</a> or text: DONATE5 ...</p> <p>72635 Retweets Wed Apr 29 00:13:30 +0000 2015</p>
 <p><b>Sar..</b> SarWilliams1</p> <p>Nepal has declared three days of mourning for the victims of Saturday's earthquake in which more than 5,000... <a href="http://t.co/5Z0mJEQB1S">http://t.co/5Z0mJEQB1S</a></p> <p>0 Retweets Wed Apr 29 00:13:29 +0000 2015</p>	 <p><b>Madison Pavan</b> madspavan</p> <p>RT @epagez: 5,000 people have died in the Nepal earthquake but the news thinks heroin smugglers deserve more attention</p> <p>3 Retweets Wed Apr 29 00:13:28 +0000 2015</p>	 <p><b>MARCUS</b> mdxvis</p> <p>RT @CNN: Nepal earthquake: Death toll passes 4,600 as rescuers face challenges. <a href="http://t.co/3BJcEbiC0E">http://t.co/3BJcEbiC0E</a> <a href="http://t.co/AE94goyxji">http://t.co/AE94goyxji</a></p> <p>1055 Retweets Wed Apr 29 00:13:27 +0000 2015</p>	 <p><b>Catalina Jaime</b> CatalinaJaime2</p> <p>RT @vfung: Going to respond to the Nepal earthquake? Make sure to check-in with @HumanitarianID. #nepalquake #NepalQuakeRelief <a href="http://t.co/...">http://t.co/...</a></p> <p>6 Retweets Wed Apr 29 00:13:27 +0000 2015</p>
 <p><b>bestworldnews</b> bestworldnewsss</p> <p>Nepal earthquake: UK aid donations reach £15m - BBC News <a href="http://t.co/gWWCK33fEW">http://t.co/gWWCK33fEW</a> #worldnews</p> <p>0 Retweets Wed Apr 29 00:13:27 +0000 2015</p>	 <p><b>nikki micky :) mehNikki</b></p> <p>RT @nprnews: "I am going wherever you go, Mum." A 10-year-old's view of the Nepal quake <a href="http://t.co/LsofvkFcb3">http://t.co/LsofvkFcb3</a> @tangledjourneys <a href="http://t.co/...">http://t.co/...</a></p> <p>48 Retweets Wed Apr 29 00:13:26 +0000 2015</p>	 <p><b>Anubhab Ray</b> anubhab</p> <p>RT @DelhiTweeter: Imagine Kejriwal was PM during Nepal #earthquake, His statement would have been "Hum kuch nahin kar sakte Ji, Army Presid..."</p> <p>795 Retweets Wed Apr 29 00:13:25 +0000 2015</p>	 <p><b>. isucksmalik</b></p> <p>RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBJftNI9m">http://t.co/NXBJftNI9m</a> or text: DONATE5 ...</p> <p>72635 Retweets Wed Apr 29 00:13:22 +0000 2015</p>

Figure 9: Wall 3

## Title Here

created on the command line with [twarc](#)

















 <b>Aracely Arce</b> aracelyarceb RT @CNN: CNN obtained powerful images of a baby rescued at least 22 hrs after the Nepal quake hit: <a href="http://t.co/EzS0P0bnzm">http://t.co/EzS0P0bnzm</a> <a href="http://t.co/vJwXA...">http://t.co/vJwXA...</a> 1576 Retweets Thu Apr 30 03:16:45 +0000 2015	 <b>Alec Couros</b> courosa "Baby boy rescued from Nepal earthquake rubble" <a href="http://t.co/8d1LLYkV3n">http://t.co/8d1LLYkV3n</a> w/ photos 0 Retweets Thu Apr 30 03:16:43 +0000 2015	 <b>Mrinal Thakur</b> chickmt123 #letsFootball #atk TIME : Meet the U.S. first responders on the ground in Nepal <a href="http://t.co/8PQ5FPaREV">http://t.co/8PQ5FPaREV</a> ( <a href="http://t.co/8PQ5FPaREV">http://t.co/8PQ5FPaREV</a> ) 0 Retweets Thu Apr 30 03:16:41 +0000 2015	 <b>Simplifications</b> RT Survival stories from the Nepal earthquake (+video) - Christian Science Monitor 0 Retweets Thu Apr 30 03:16:41 +0000 2015
 <b>Tishtriya</b> npracheeti Shame on you, Pakistan! Sending beef to Nepal earthquake-affected, knowing it hurts Hindu sentiments <a href="https://t.co/MmFJ3979V0">https://t.co/MmFJ3979V0</a> 0 Retweets Thu Apr 30 03:16:40 +0000 2015	 <b>Kelly Montgomery</b> kellymontgomery RT @REI: Please consider supporting the relief effort now underway in Nepal: <a href="http://t.co/ud84X2QB8N">http://t.co/ud84X2QB8N</a> (3/3) <a href="http://t.co/RTqZME6gNy">http://t.co/RTqZME6gNy</a> 39 Retweets Thu Apr 30 03:16:40 +0000 2015	 <b>Rahul Tripathi</b> rahultri1977 RT @HDL_Global: Katua Pakistan serves beef to Nepal earthquake survivors #NepalQuakeRelief <a href="http://t.co/1mwFgBJ1xc">http://t.co/1mwFgBJ1xc</a> <a href="http://t.co/UPodc4nbo2">http://t.co/UPodc4nbo2</a> 3 Retweets Thu Apr 30 03:16:40 +0000 2015	 <b>Yash Agarwal</b> Yash_Agarwal97 RT @AskAnshul: Breaking: Pakistan serves Beef to Hindu Nation Nepal earthquake survivors. <a href="http://t.co/IGXiXCRj4">http://t.co/IGXiXCRj4</a> 36 Retweets Thu Apr 30 03:16:40 +0000 2015
 <b>Resourceful Aust</b> ResourcefulAust @CAREAustralia #Nepal earthquake: No food, clothes or shelter, say localsat epicentre of quake <a href="http://t.co/xZQixq78vL">http://t.co/xZQixq78vL</a> #auspol #wapol #ACT 0 Retweets Thu Apr 30 03:16:40 +0000 2015	 <b>Arjun's Rajeshree</b> ♡ Rajeshreebanerj RT @OMGFunnies_: Help number for the earthquake in Nepal. RT! <a href="http://t.co/Li1TGp1rhB">http://t.co/Li1TGp1rhB</a> 242 Retweets Thu Apr 30 03:16:38 +0000 2015	 <b>Chicago Press</b> Chicago_CP Chicago-area man trapped on mountain in Nepal after quake: The death toll in Nepal after this week's earthquake... <a href="http://t.co/61FopEXDvp">http://t.co/61FopEXDvp</a> 0 Retweets Thu Apr 30 03:16:35 +0000 2015	 <b>Raouf zainatieh</b> THE68JORDANIAN UN launches Nepal earthquake appeal. No coverage , no help ! I guess Baltimore for more drama ! <a href="http://t.co/em6cWexWTz">http://t.co/em6cWexWTz</a> 0 Retweets Thu Apr 30 03:16:34 +0000 2015
 <b>Pushkar N Kulkarni</b> pushkar_nk RT @IndiaToday: Pakistan serves beef to Nepal earthquake survivors #NepalQuakeRelief <a href="http://t.co/ynYuVvYZ5N">http://t.co/ynYuVvYZ5N</a> <a href="http://t.co/AalqRXLex7">http://t.co/AalqRXLex7</a> 37 Retweets Thu Apr 30 03:16:33 +0000 2015	 <b>Mohammed Azeem</b> SureShotWinner RT @ndtv: 'Help!' Stranded trekkers send out SOS after Nepal earthquake <a href="http://t.co/j1DuERjI4M">http://t.co/j1DuERjI4M</a> <a href="http://t.co/E6bVPIFINM">http://t.co/E6bVPIFINM</a> 16 Retweets Thu Apr 30 03:16:32 +0000 2015	 <b>WaywardThings</b> waywardthings #LiveScience Nepal Earthquake Lifted Kathmandu, but Shrank Everest <a href="http://t.co/fuUJO7t5mo">http://t.co/fuUJO7t5mo</a> 0 Retweets Thu Apr 30 03:16:31 +0000 2015	 <b>Arm Urb</b> arm_urbina RT @NatGeo: Did Nepal earthquake change Mount Everest's height? <a href="http://t.co/NwwYE1A0jD">http://t.co/NwwYE1A0jD</a> 344 Retweets Thu Apr 30 03:16:31 +0000 2015

Figure 10: Wall 4



## Title Here

created on the command line with [twarc](#)

















 <p><b>chloepaddington</b> chloepaddington</p> <p>RT @prabalgurung: Pls. help my fundraiser: Nepal Earthquake Fund. Donate &amp; spread the word. Every dollar counts-PG. <a href="http://t.co/y2rJRYCbGu">http://t.co/y2rJRYCbGu</a> ... 183 Retweets Fri May 01 05:07:59 +0000 2015</p>	 <p><b>madan giri</b> shadiemaniac</p> <p>RT @HuffingtonPost: Disastrous politics and poverty left Nepal unprepared for the earthquake <a href="http://t.co/2jAJeuWE3G">http://t.co/2jAJeuWE3G</a> <a href="http://t.co/nVawFNH5al">http://t.co/nVawFNH5al</a> 52 Retweets Fri May 01 05:07:58 +0000 2015</p>	 <p><b>KevilKhadka</b> KevilKhadka</p> <p>RT @DDNewsLive: Foreign Secretary S. Jaishankar and National Security Adviser Ajit Doval reach #Kathmandu to review relief work in earthquake... 17 Retweets Fri May 01 05:07:58 +0000 2015</p>	 <p><b>Merebear</b> merebear113</p> <p>RT @Louis_Tomlinson: Just donated @savechildrenuk #NepalEarthquake appeal. If you can too, go here <a href="http://t.co/NXBjftNI9m">http://t.co/NXBjftNI9m</a> or text: DONATE5 ... 87010 Retweets Fri May 01 05:07:57 +0000 2015</p>
 <p><b>Annamaria Fernandez</b> squareone111</p> <p>RT @FoxNews: After a magnitude-7.8 earthquake hit #Nepal, rescuers found a 4-month-old baby boy alive &amp; uninjured in the rubble! <a href="https://t.co/https://t.co/WughoWq108">https://t.co/https://t.co/WughoWq108</a> 736 Retweets Fri May 01 05:07:57 +0000 2015</p>	 <p><b>पोखरेली शुशिल</b> subeedi55</p> <p>RT @cnni: Remember the miracle baby who has become a symbol of hope for Nepal? He's safe now in his sister's arms. <a href="http://t.co/mvMZCYaKg">http://t.co/mvMZCYaKg</a> 75 Retweets Fri May 01 05:07:56 +0000 2015</p>	 <p><b>Carinity</b> carinity_org</p> <p>Carinity has a number of Nepalese staff across many of our facilities and would like to call on everyone for... <a href="http://t.co/kC6ekDycVf">http://t.co/kC6ekDycVf</a> 0 Retweets Fri May 01 05:07:55 +0000 2015</p>	 <p><b>Rohit Kumar Agarwal</b> rohitagarwal86</p> <p>RT @rahulkanwal: (Video) My ground report on the efforts being made by @RSSorg to provide relief to earthquake ravaged Nepal. <a href="http://t.co/nRJwNMzIYt">http://t.co/nRJwNMzIYt</a> 28 Retweets Fri May 01 05:07:55 +0000 2015</p>
 <p><b>KevilKhadka</b> KevilKhadka</p> <p>RT @NRaule: Nepalese people loss everything but still they don't loss their smile &amp; hope. #nepalquake #earthquake #Nepal 4 Retweets Fri May 01 05:07:54 +0000 2015</p>	 <p><b>NewsOnFire</b> AshikOnFire</p> <p>BSGN Nepal quake toll rises to 6,166: The toll in the massive earthquake that devastated Nepal has risen to 6,... <a href="http://t.co/ikdZehCBXn">http://t.co/ikdZehCBXn</a> 0 Retweets Fri May 01 05:07:54 +0000 2015</p>	 <p><b>Anil Nayak</b> anayak8</p> <p>RT @rahulkanwal: (Video) My ground report on the efforts being made by @RSSorg to provide relief to earthquake ravaged Nepal. <a href="http://t.co/nRJwNMzIYt">http://t.co/nRJwNMzIYt</a> 28 Retweets Fri May 01 05:07:53 +0000 2015</p>	 <p><b>Nev+are TEAM PAC</b> AmedBorn</p> <p>RT @WSJ: Five remarkable rescues bring hope after almost a week of devastation in Nepal. <a href="http://t.co/37BzLTsblG">http://t.co/37BzLTsblG</a> <a href="http://t.co/ratDgZ1MF1">http://t.co/ratDgZ1MF1</a> 267 Retweets Fri May 01 05:07:53 +0000 2015</p>
 <p><b>Vibhore Jaiswal</b> JaiswalVibhore</p> <p>Earthquake in Nepal. Golden Temple to send 1 lac and Delhi Gurdwaras to send 25k meals daily. @aashishgogna92 @AashaBisht 0 Retweets Fri May 01 05:07:52 +0000 2015</p>	 <p><b>Genea Garry</b> jgg2214</p> <p>RT @cnnbrk: Nepal earthquake death toll rises to more than 6,100. <a href="http://t.co/96CjNxThpr">http://t.co/96CjNxThpr</a> 991 Retweets Fri May 01 05:07:52 +0000 2015</p>	 <p><b>राजेश शुक्ला</b> rajshukla20</p> <p>RT @rahulkanwal: (Video) My ground report on the efforts being made by @RSSorg to provide relief to earthquake ravaged Nepal. <a href="http://t.co/nRJwNMzIYt">http://t.co/nRJwNMzIYt</a> 28 Retweets Fri May 01 05:07:48 +0000 2015</p>	 <p><b>Fromal Photography</b> FromalPhoto</p> <p>RT @CMcGranahan: Talented Nepal-based photographer Tom Kelly + son Liam document the earthquake <a href="http://t.co/w6FuY82JqP">http://t.co/w6FuY82JqP</a> a photographic labor... 7 Retweets Fri May 01 05:07:47 +0000 2015</p>

Figure 11: Wall 5

## Word Clouds

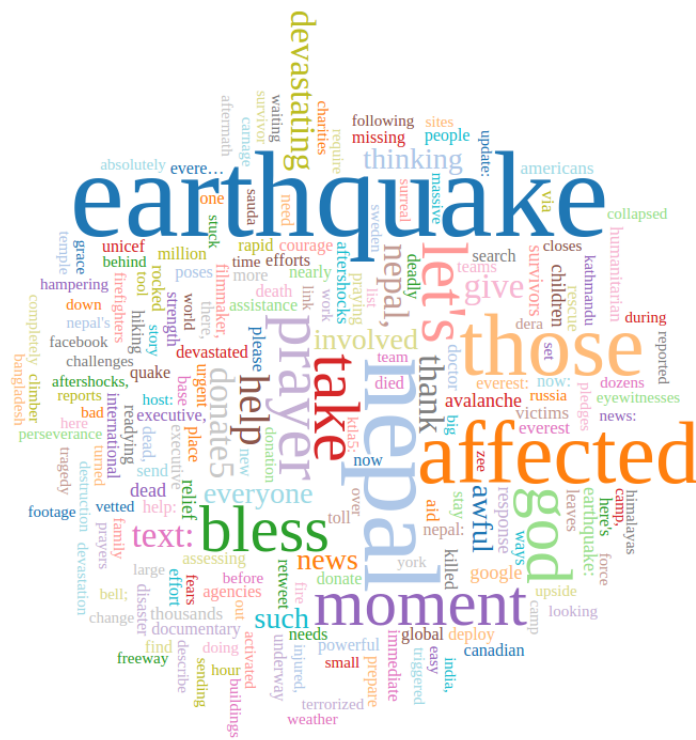


Figure 12: Word Cloud 1









## GeoJSONs

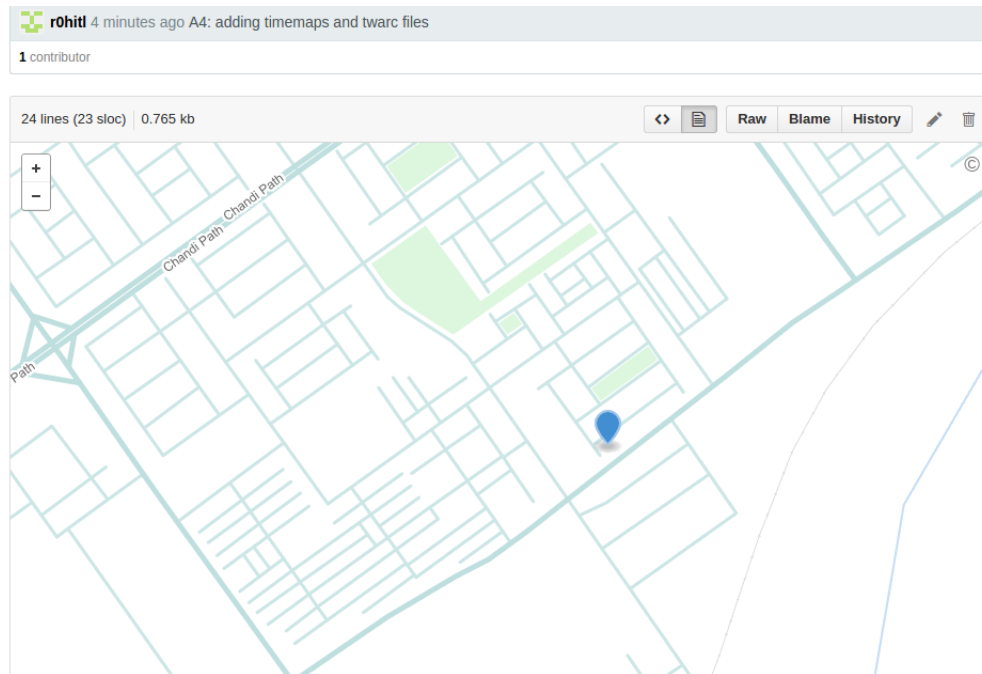


Figure 17: GeoJSON 1

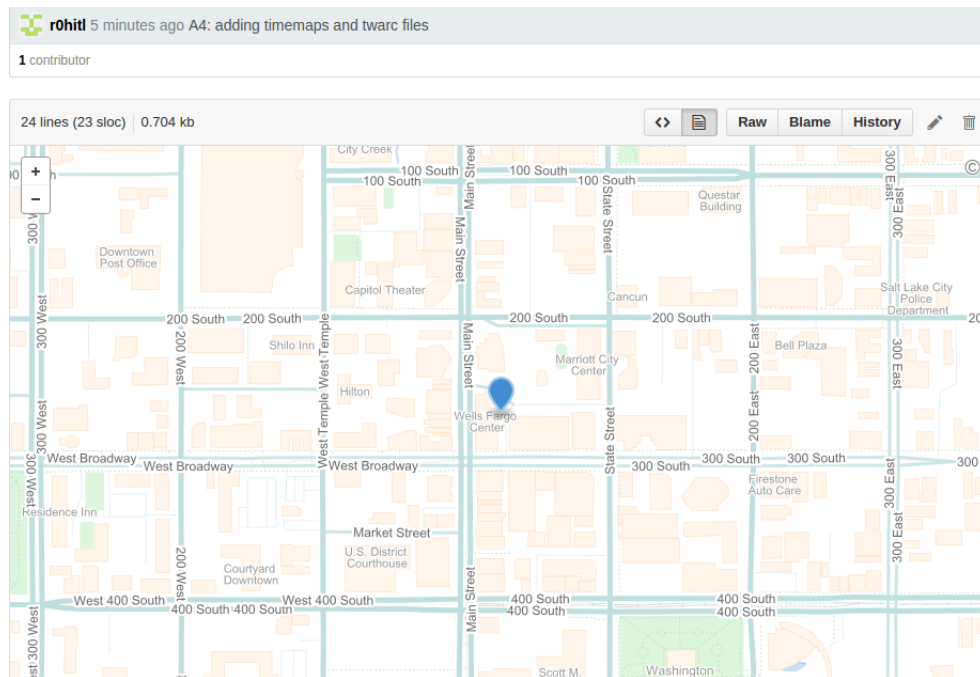


Figure 18: GeoJSON 2



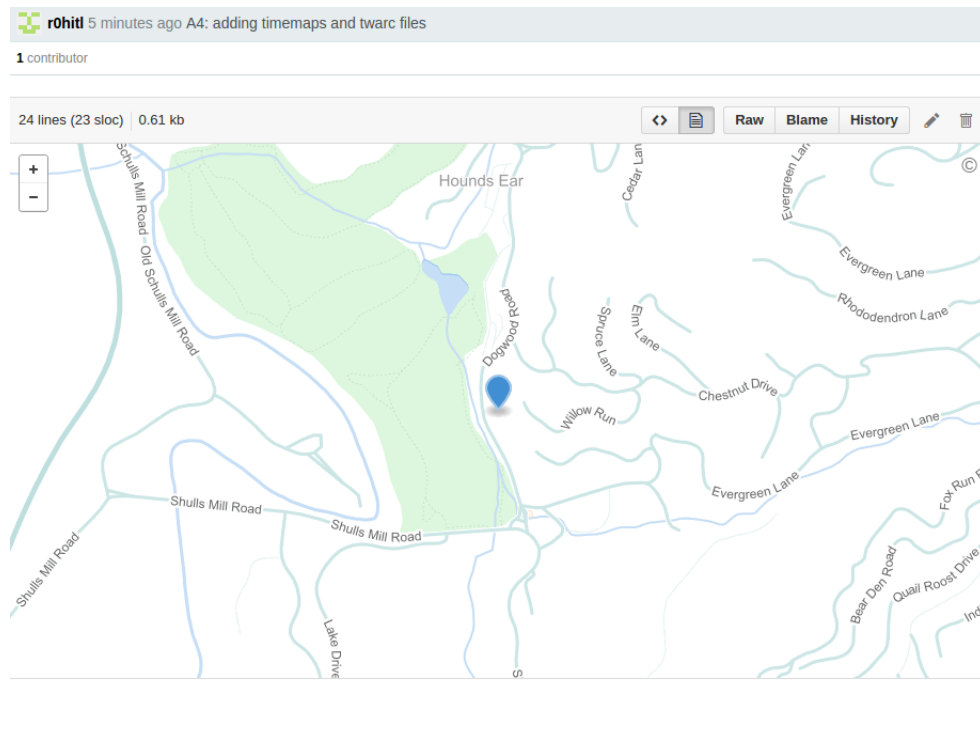


Figure 19: GeoJSON 3

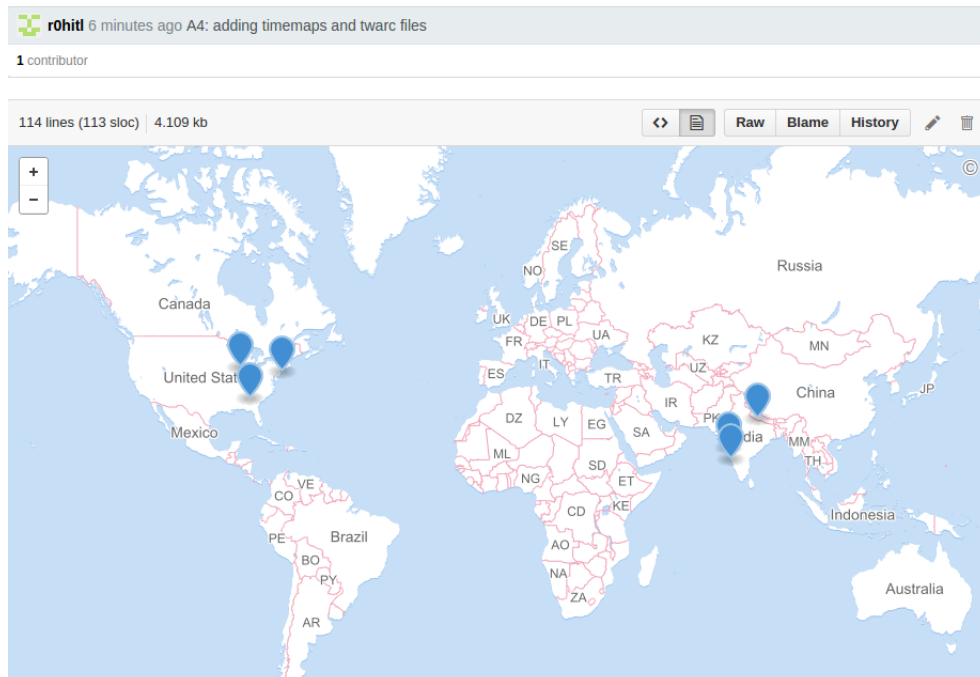


Figure 20: GeoJSON 4

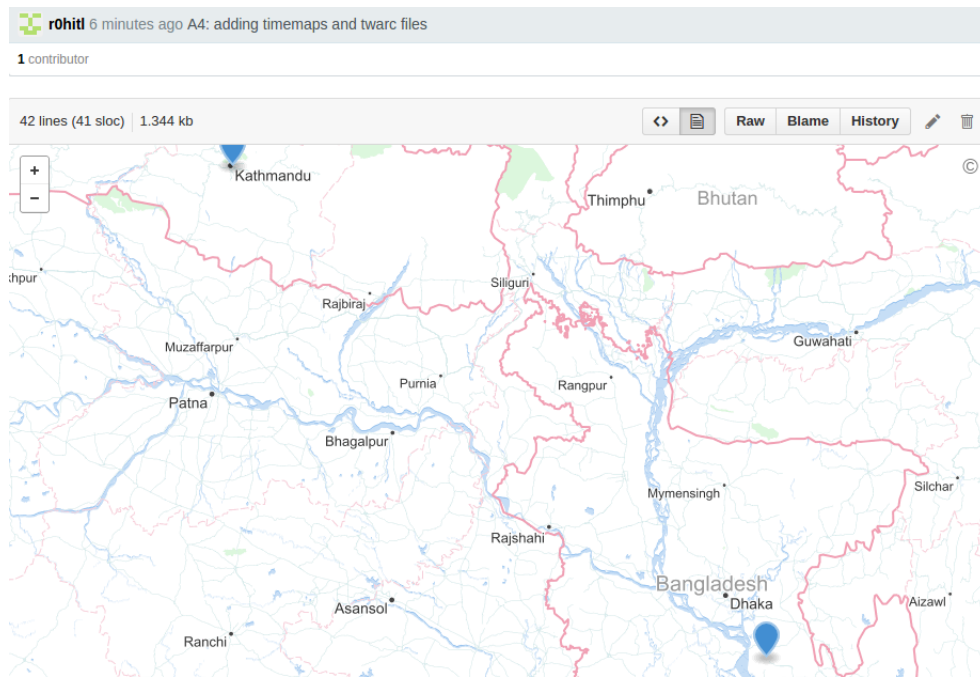


Figure 21: GeoJSON 5

## 4.2 Code Listing

Python program to fetch to fetch tweets using 'twarc' library

```
1 '''
2 @author: rlambi
3 '''
4 import os
5 import json
6 from twarc import Twarc
7
8 CONSUMER_KEY = "iuKUndPfIF5aWnNl0Ayq9Ztgt"
9 CONSUMER_SECRET = "QuNsF4gL2LssmbcdtKpyLZGiQctz98T4hXWcAKrBYGh72ZTFC8"
10
11 OAUTH_TOKEN = "549294315-P89swbZzgiP2n9bq6fW2T2jm5etru6Wr6TN08Lg3"
12 OAUTH_TOKEN_SECRET = "NMzDaS5doFtHxXxebE68AunmRHTsFfLxwAkk3LsDN75JH"
13
14 '''
15 os.system('python twarc/twarc.py --consumer_key iuKUndPfIF5aWnNl0Ayq9Ztgt --
16         consumer_secret QuNsF4gL2LssmbcdtKpyLZGiQctz98T4hXWcAKrBYGh72ZTFC8 --access_token
17         549294315-P89swbZzgiP2n9bq6fW2T2jm5etru6Wr6TN08Lg3 --access_token_secret
18         NMzDaS5doFtHxXxebE68AunmRHTsFfLxwAkk3LsDN75JH --search "statue of liberty bomb" >
19         tweets1.json')
20
21 '''
22 t = Twarc(CONSUMER_KEY, CONSUMER_SECRET, OAUTH_TOKEN, OAUTH_TOKEN_SECRET)
23
24 fil = open('tweets_nepal_5.txt', 'w')
25 counter = 0
26 for tweet in t.search("nepal earthquake"):
27     counter += 1
28     json_tweet = json.dumps(tweet)
29     print(json_tweet + '\n')
30     fil.write(json_tweet + '\n')
31
32     if counter == 1000:
33         break
34
35 fil.close()
36
37 print 'End'
```