

Exploring World's Interest in Paralympics through Twitter

Venkata Sravya Kalla, Thanaa Ghanem
Information and Computer Science Department
Metropolitan State University
St. Paul, MN, 55106
cu9426bs@metrostate.edu, thanaa.ghanem@metrostate.edu

Abstract

Social media services are considered critical factors in information dissemination and influence discovery. Among those, Twitter has become the de facto of microblogging services with millions of tweets posted every day. Combined with the widespread use of smart mobile devices, social media providers are currently able to enrich their data with geotags to represent location information. In this paper, we present a framework that can be used to analyze geotagged Twitter status updates and hence provide a not-so-expensive tool to extract useful insights about communities. As a use case, we present our results of using the framework to perform an analytical study of geotagged tweets to learn about people's interest in Paralympics that took place in 2016. The goal of this study is to determine how different places of the world express different interests in Paralympics games. The results of this study can be used by the sporting industry to appropriately plan their support of the various games at the different countries and/or states.

1 Introduction

Social media services are considered critical factors in information dissemination and influence discovery. Among those, Twitter has become the de facto of microblogging services with millions of tweets posted every day. Combined with the widespread use of smart mobile devices, social media providers are currently able to enrich their data with geotags to represent location information. This generated a plethora of spatial information in social media which enables rich spatial analytics tasks on social media data.

In this paper, we present a framework that can be used to analyze the geotagged Twitter status updates. The framework consists of the following three main components: 1) a *crawler* that uses Twitter API to collect tweets, 2) a *parser* that parses the collected tweets' JSON objects and extracts fields of interest, and 3) an *analyzer* that analyzes that extracted data. This framework can be used to perform a variety of analysis tasks on twitter data, which is a self-reported social media. The framework then can be used as a not-so-expensive tool to monitor communities and hence promoting various industries in these communities as appropriate.

As a use case, we used the proposed framework to study people's interest in Paralympics as one of the major international sport events that took place in 2016. The goal of this study is to determine how different places of the world express different interests in Paralympics games. Our study is mainly analyzing tweets to find information about: 1) the most discussed games, (2) the most commonly used hashtags related to Paralympics, (3) the diversity of people who participated in discussions, (4) the favorite game(s) of each country, and (5) the most followed sport on each day during the Paralympics event. The results of this study can be used by the sporting industry to appropriately plan their support of the various games at the different countries and/or states. We used various techniques to report the results of our analysis including word clouds, graphs, and maps. In addition, we implemented an interactive exploration tool for the spatial distribution of interests of the various games, which provides a low-effort tool for interested groups to explore our results.

The rest of this paper is organized as followed. Section 2 discusses the proposed framework with its three main components. Section 3 presents the details of our study on Paralympics. Finally, Section 4 concludes the paper.

2 A Framework to Analyze Geotagged Twitter Status Updates

In this section, we discuss the implementation details of a framework that can be used to analyze geotagged Twitter status updates. The framework is composed of the following three components: (1) a *crawler* to collect tweets using Twitter API, (2) a *parser* to parse the collected tweets and filter out tweets that are not of interest to the underlying analysis, and (3) an *analyzer* to perform analysis on the collected tweets, visualize the results, and highlight the findings. It is important to note that the proposed framework can be used to perform any other analytical study on Twitter data (e.g., find out who is talking about the *Oscars*). In order to perform a different analysis task, the three components need to be slightly adjusted to reflect the new subject of interest.

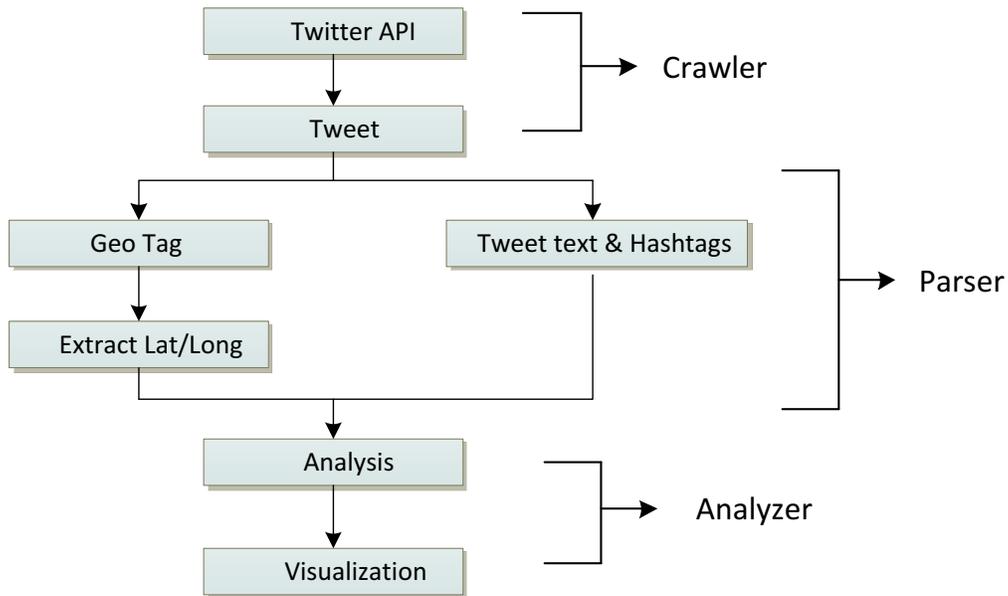


Figure 1: Framework to Analyze Tweets

In the following sections, we discuss implementation details for each component.

2.1 Crawling Twitter Data

On average, around 6,000 tweets are tweeted on twitter every second which means around 350,000 tweet per minute, 500 million tweets per day, or 200 billion per year. Twitter make a small percent (1%) of these tweets available to the public at no cost through Twitter APIs (Kumar, Morstatter, & Liu, 2014).

APIs to access Twitter data can be classified into the following two types based on their design and access method:

- *REST APIs* that use the pull strategy for data retrieval. In the pull strategy, tweets are not sent to the user until the user sends an explicit request.
- *Streaming APIs* that use the push strategy to provide a continuous stream of public tweets. In this strategy, once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user (Kumar, Morstatter, & Liu, 2014).

In this section, we explain the steps we followed to collect tweet using the Streaming API.

2.1.1. Getting Twitter API Authentication Keys:

Twitter APIs can only be accessed by authenticated requests. Twitter uses Open Authentication (OAuth) and each request must be signed with a valid user account. Figure 2 gives an overview of the steps involved in using OAuth to access Twitter API by applications. Mainly, the steps are as follows:

- Create an account from the following link: <https://twitter.com/signup>. Through this registration process, the user is issued a *consumer token* and a *secret* that are to be used by any application to authenticate itself to Twitter.
- The application then uses the provided *consumer key* and *secret* to create a unique Twitter link to which a user is directed for authentication. Twitter verifies the user's identity and issues an OAuth verifier or a PIN.
- The application then uses the PIN to request an *access token* and *access secret* that are unique to the user.
- The *access token* and *access secret* for a user do not change and can be cached by the application for future requests. Thus, this process only needs to be performed once.

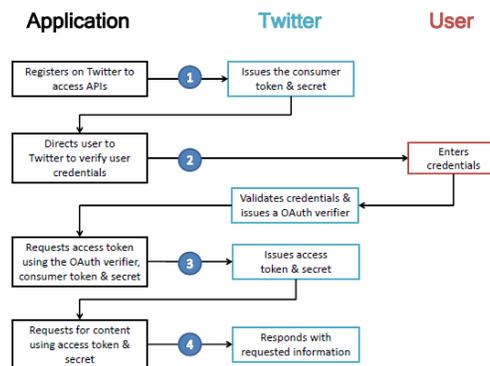


Figure 2: Open Authentication Workflow (Kumar, Morstatter, & Liu, 2014).

2.1.2 Implementing a Crawler in Java:

After creating a Twitter Developer's account and obtaining access credentials, we implemented a Java application, called crawler, to connect to Twitter and download tweets. The main Java library that is used is the Twitter4J library that is encapsulated in a file called `twitter4j-core-4.0.4.jar` that can be downloaded from the following link: <http://twitter4j.org/en/>. Once downloaded, the path to the library's jar file should be added to the crawler's application class path.

Figure 3 gives the Java code for the crawler application. The code consists of three main steps. In the first step, the application authenticates itself to the Twitter API framework by integrating the OAuth authorization credentials that were obtained as explained in the previous section. In the second step, an instance of type `TwitterStream` is created to keep the connection open, and an instance of class `StatusListener` is created to consume the tweets from Twitter. In the third step, we list the keywords of interests to our Paralympics analysis. These keywords are used to filter out tweets that are not needed for our purposes. It is important to note that, in order to use the crawler to

collect tweets related to another topic of interest, only step 3 need to modified to replace the Paralympics keywords with the new words of interest.

```
public final class Crawler {
    public static void main(String[] args) {
        ConfigurationBuilder cb = new ConfigurationBuilder();
        cb.setDebugEnabled(true);

        //step 1:Using authenticated requests
        cb.setOAuthConsumerKey("XXX");
        cb.setOAuthConsumerSecret("XXX");
        cb.setOAuthAccessTokenSecret("XXX");
        final PrintWriter out = new PrintWriter(new
        FileWriter("tweetstxt"));

        //step 2: Instantiate Twitter Stream class.
        TwitterStream twitterStream = new
        TwitterStreamFactory(cb.build()).getInstance();
        StatusListener listener = new StatusListener() {
            public void onStatus(Status status) {
                System.out.println("@" +
                status.getUser().getScreenName() + " - " +
                status.getText());
                out.println(status);
            }
        };

        //step 3: Using Query class to filter tweets
        FilterQuery fq = new FilterQuery();
        String keywords[] = {"paralympics", "Rio2016"};
        fq.track(keywords);
        twitterStream.addListener(listener);
        twitterStream.filter(fq);
    }
}
```

Figure 3: A Java program to collect Tweets via Twitter’s Streaming API

We ran the twitter crawler from September 07, 2016 to September 14 ,2016 and collected around two million tweets that are related to Paralympics. The tweets that are returned by the crawler are in JSON format. Figure 4 gives an example JSON object for a tweet that was posted on Wednesday, September 7th, 2016 at 18:03:54 O’clock.

```
{
  "created_at": "Wed Sep 07 18:03:54 +0000 2016",
  "id": 773582583359832064,
  "id_str": "773582583359832064",
  "text": "Get ready, the #Paralympics #OpeningCeremony starts very soon!! #Rio2016 @OISphotos https://t.co/rs5lem6NY9",
  "source": "\u003ca href=\\"http://twitter.com/download/android\" rel=\\"nofollow\" \u003eTwitter for Android\u003c/a\u003e",
  "truncated": false,
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "id": 234399791,
    "id_str": "234399791",
    "name": "Paralympic Games",
    "screen_name": "Paralympics",
    "location": null,
    "url": "http://www.paralympic.org",
    "description": "Official Twitter page of the International Paralympic Committee (IPC), the global governing body of the Paralympic Movement .",
    "protected": false,
    "verified": true,
    "followers_count": 111033,
    "friends_count": 2717,
    "listed_count": 1358,
    "favourites_count": 1041,
    "statuses_count": 20990,
    "created_at": "Wed Jan 05 15:28:57 +0000 2011",
    "utc_offset": 7200,
    "time_zone": "Berlin",
    "geo_enabled": true,
    "lang": "en",
    "contributors_enabled": false,
    "is_translator": false,
    "profile_background_color": "131516",
    "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/37880000176428665/qljs-1Mo.jpeg",
    "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/37880000176428665/qljs-1Mo.jpeg",
    "profile_background_tile": false,
    "profile_link_color": "009999",
    "profile_sidebar_border_color": "FFFFFF",
    "profile_sidebar_fill_color": "E6E6E6",
    "profile_text_color": "333333",
    "profile_use_background_image": true,
    "profile_image_url": "http://pbs.twimg.com/profile_images/462133007562702848/NrOn7Amc_normal.jpeg",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/462133007562702848/NrOn7Amc_normal.jpeg",
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/234399791/1470389151",
    "default_profile": false,
    "default_profile_image": false,
    "following": null,
    "follow_request_sent": null,
    "notifications": null,
    "geo": null,
    "coordinates": null,
    "place": {
      "id": "6ab790448f3ae962",
      "url": "https://api.twitter.com/1.1/geo/id/6ab790448f3ae962.json",
      "place_type": "neighborhood",
      "name": "Maracan\u00e3",
      "full_name": "Maracan\u00e3, Rio de Janeiro",
      "country_code": "BR",
      "country": "Brazil",
      "bounding_box": {
        "type": "Polygon",
        "coordinates": [[[-43.239754, -22.918292], [-43.239754, -22.907972], [-43.212277, -22.907972], [-43.212277, -22.918292]]]
      },
      "attributes": {}
    },
    "contributors": null,
    "is_quote_status": false,
    "retweet_count": 25,
    "favorite_count": 35,
    "entities": {
      "hashtags": [
        {
          "text": "Paralympics",
          "indices": [15, 27]
        },
        {
          "text": "OpeningCeremony",
          "indices": [28, 44]
        },
        {
          "text": "Rio2016",
          "indices": [64, 72]
        }
      ],
      "urls": [],
      "user_mentions": [
        {
          "screen_name": "Paralympics",
          "name": "Paralympic Games",
          "id": 234399791,
          "id_str": "234399791",
          "indices": [3, 15]
        },
        {
          "screen_name": "OISphotos",
          "name": "OIS Photos",
          "id": 2590317558,
          "id_str": "2590317558",
          "indices": [90, 100]
        }
      ],
      "media": [
        {
          "id": 773582556537257984,
          "id_str": "773582556537257984",
          "indices": [101, 124],
          "media_url": "http://pbs.twimg.com/media/CrxRPI7XgAAK7IY.jpg",
          "media_url_https": "https://pbs.twimg.com/media/CrxRPI7XgAAK7IY.jpg",
          "url": "https://t.co/rs5lem6NY9",
          "display_url": "pic.twitter.com/rs5lem6NY9",
          "expanded_url": "https://twitter.com/Paralympics/status/773582583359832064/photo/1",
          "type": "photo",
          "sizes": {
            "medium": {
              "w": 1200,
              "h": 800,
              "resize": "fit"
            },
            "small": {
              "w": 680,
              "h": 453,
              "resize": "fit"
            },
            "thumb": {
              "w": 150,
              "h": 150,
              "resize": "crop"
            },
            "large": {
              "w": 2048,
              "h": 1365,
              "resize": "fit"
            }
          },
          "source_status_id": 773582583359832064,
          "source_status_id_str": "773582583359832064",
          "source_user_id": 234399791,
          "source_user_id_str": "234399791"
        }
      ],
      "favorited": false,
      "retweeted": false,
      "possibly_sensitive": false,
      "filter_level": "low",
      "lang": "en",
      "timestamp_ms": "1473272240922"
    }
  }
}
```

Figure 4: Example Tweet in JSON Format

Each tweet's JSON object contains around 40 attributes. In the following section we discuss our implementation of a parser that can be used to parse a the tweet's JSON object to extract only some attributes that are required by the analyzer.

2.2 Parsing Tweet's JSON:

We implemented a parser to extract required attributes from the collected tweets' JSON objects. Figure 5 gives the java code for the parser application. We used the Google gson library to parse the JSON objects. This library can be downloaded from the following link: <http://www.java2s.com/Code/Jar/g/Downloadgson222jar.htm>. The *JsonParser* class parse a JSON string into a tree of Java objects (Jenkov, 2016). Different attributes from the JSON object can be extracted using `get(<attribute name>)` method as illustrated in Figure 5. We used the parser to extract the following attributes that are of interest to our analysis:

- *created_at*: time when this Tweet was created.
- *text*: the actual text of the status update.
- *geo*: Represents the geographic location of this Tweet.

- *coordinates*: The longitude and latitude of the Tweet's location, as an array in the form of [latitude, longitude].
- *hashtags*: hashtags that are mentioned in the tweet's text.

```
import com.google.gson.*;

public class JsonToTweet {
    public static Tweet parseTweetJSON(String tweetJSON) {

        Tweet newTweet = new Tweet();
        JsonObject object =
        JsonParser().parse(tweetJSON).getAsJsonObject();
        //retrieving the tweet's text from JSON object
        JsonElement text = object.get("text");
        if((text!= null)&&(!text.toString().equals(" "))) {
            newTweet.setTweetText(text.getAsString());
            //retrieving tweet's time from JSON object
            element = object.get("created_at");
            newTweet.setCreatedAt(element.getAsString());
            //other attributes can be retrieved similarly
        }
    }
}
```

Figure 5: Tweet's JSON Parser

2.3 Analyzer:

We used the parser to extract the information from JSON object and store it in a spread sheet for further analysis. Then, we can generate summaries of the data using grouping and counting operations. For example, to find out what are the most frequently used hashtags, we group on the hashtag attribute and count how many rows are there in each group. The summaries are visualized for easy interpretation. We used various visualization tools to represent our findings, including bar graphs, tables, word cloud, and maps. In the following section, we discuss in details our summaries along with our findings.

3 Paralympics Analysis

The 2016 Summer Paralympics is the 15th Summer Paralympic event that was governed by the International Paralympics Committee, and were held in Rio de Janeiro, Brazil, from September 7, 2016 to September 18, 2016. This event is the world's largest multi-sport event for Para athletes. Amongst the sweeping vistas of Rio de Janeiro, over 4,350 athletes from 159 countries competed in 22 sports, across 526 medal events ("Rio 2016 Summer Paralympics", 2016). The main aim to

3.2 Most involved countries in Paralympics discussions:

Athletes from 159 National Paralympic countries competed in Rio 2016 Paralympics. We are using twitter data to identify, which part of the world participated in discussions on Paralympics. This analysis can be helpful to identify nations which are not involved in discussions and help the Sporting committee to take required steps to support the sports in those regions.

To perform this analysis, we parsed the tweets to extract geocode from where the tweet is originated. Then, we used Maps to visualize the data with Geocodes and Each dot on the map refers to the tweet originated location.



Figure 7: Map indicating tweet originated place

We can see from the figure 7 that some countries (e.g., Russia and UAE) were not participating in the discussions. Due to Russia's doping scandal, International Paralympic committee banned athletes from competing at the Rio 2016 Paralympics ("Rio Paralympics 2016", 2016). This might be the reason behind Russia' silence. This finding proves that twitter can be opted as research data to obtain accurate results.

3.3 Most discussed game per each date:

In this experiment, we identified the most discussed games on each day during Paralympics event. We parsed the tweets to extract the date on which the tweet was posted and also extracted the sport(s) that are mentioned in the tweet. Then, we performed grouping by sport and date and counted the rows related to each group.

Table 1 gives the most discussed sport on each day of the event. It is important to note that this finding matches the Paralympics schedule. For example, the opening ceremony of Paralympics took place on September, 7th, 2016 and the Paralympic cauldron was lit by legendary Brazilian para - swimmer Clodoaldo Silva. Our results show that many tweets on that day were about

Paralympic swimming which were related to Para swimmer Silva and dance related to performances in the ceremony ("Summer Paralympic Games",2016).

Date	Sport
7 Sep, 2016	Paralympic Swimming, Dance
8 Sep, 2016	Paralympic Athletics, cycling
9 Sep, 2016	Paralympic Athletics, Cycling, Paralympic Swimming
10 Sep, 2016	Paralympic Athletics, Wheelchair Tennis, Para Triathlon
11 Sep, 2016	Paralympic Athletics, Paralympic Swimming, Para Triathlon, Wheelchair Tennis,
12 Sep, 2016	Paralympic Athletics, Wheelchair Tennis
13 Sep, 2016	Paralympic Athletics, Wheelchair Tennis
14 Sep ,2016	Cycling, Para Triathlon,

Table 1: Sport followed on each date during Paralympics Event.

4 Conclusions

The main goal of our project is to prove that social media gives a not-so-expensive tool for information dissemination. We first presented a framework that can be used to analyze geotagged tweets. We then used the proposed framework to explore world’s interest in Paralympics 2016. We performed our analysis in three main steps 1) collected tweets from twitter which are related to Paralympics 2016, 2) implemented a parser to extract relevant data from tweets, and then 3) analyzed the collected data and presented the results using reports like word cloud, bar graphs and maps.

According to our analysis, Paralympic athletics, Paralympic swimming and wheel chair tennis are the most discussed games in Summer 2016 Paralympics. Team Netherlands emerges to be the most followed team in Paralympics followed by Team Great Britain. Word cloud created from our analysis gives a good picture on sports, teams and countries which were more mentioned in discussions about Paralympics. Also, we found that countries like Russia was not involved in discussions. This finding was consistent with the fact the Russia was banned from participating in the event. This proves that twitter can be used as tool to extract accurate results.

References:

- Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter Data Analytics*. New York, NY: Springer New York.
- Yamamoto, Y. (2007). *Twitter4J - A Java library for the Twitter API*. Retrieved March 24, 2017, from <http://twitter4j.org/en/index.html>.
- Raamkumar, A. S. (2014). *Whats in a Country Name - Twitter Hashtag Analysis of #Singapore*. Retrieved March 24, 2017, from <https://www.semanticscholar.org/paper/Whats-in-a-Country-Name-Twitter-Hashtag-Analysis/Raamkumar/619cb5422bfde18c015e20bc116aa7508464e96d>.
- Jenkov, J. (2016). *GSON - JsonParser*. Retrieved March 24, 2017, from <http://tutorials.jenkov.com/java-json/gson-jsonparser.html>.
- Taylor, A. (2016, September 12). *The 2016 Summer Paralympics*. Retrieved March 24, 2017, from <https://www.theatlantic.com/photo/2016/09/the-2016-summer-paralympics/499622/>
- *Rio 2016 Summer Paralympics*. (n.d.). Retrieved March 24, 2017, from <http://www.paralympics.org.nz/Events/Rio-2016-Summer-Paralympics>
- *Summer Paralympic Games*. (2017, March 22). Retrieved March 24, 2017, from https://en.wikipedia.org/wiki/Summer_Paralympic_Games
- *Rio Paralympics 2016: Russian athletes banned after doping scandal - BBC Sport*. (2016). Retrieved March 24, 2017, from <http://www.bbc.com/sport/disability-sport/37002582>