



Trabalho Banco de Dados



Extração de Tweets do Twitter



Motivação

Com objetivo de adicionar mais informações ao primeiro trabalho de banco de dados (extração de anúncios do OLX <https://github.com/pedrohcosta/trab01>), realizamos a extração de tweets com termos referentes ao anuncios para uma possivel combinação dos termos juntamente com analise das informações.

Autenticação

A API REST do twitter permite a autenticação através de dois mecanismos baseados no padrão OAuth:

Requisitos básicos

Sistema operacional

1. Linux (Ubuntu)

Linguagem de programação

1. Python 3.x

Bibliotecas utilizadas

1. <http://tweepy.readthedocs.io/en/v3.5.0/>
2. <http://initd.org/psycpg/>

Instalação das bibliotecas

Tweepy

1. `sudo pip3 install tweepy`

Psycopg2

1. `sudo apt-get install python3-psycopg2`

Autenticação application-only

Autenticação application-only: a autenticação não fica vinculada a um usuário específico, mas sim a uma aplicação previamente registrada. Quando autenticado com esse mecanismo, sua aplicação não poderá realizar algumas operações típicas de um usuário, como postar tweets, por exemplo. Esse tipo de autenticação é mais indicado para aplicações que não terão um usuário interagindo com a rede social. Um exemplo seria um app que vai extrair dados do twitter para realizar análises.

Autenticação de usuário

A autenticação se dá diretamente por um usuário, de forma que a aplicação possa realizar operações comuns a usuários. Esse tipo de mecanismo é mais indicado para o caso de aplicativos que vão interagir com a rede social pelo usuário, como um app que posta as músicas mais ouvidas pelo usuário na semana.

Autenticação Application-only

Antes de mais nada, precisamos conhecer alguns conceitos básicos de autenticação usando o padrão OAuth. Para que possamos nos identificar como usuários de um serviço que utiliza um mecanismo de autenticação baseado em OAuth, é preciso que tenhamos os seguintes dados:

API Key: uma chave utilizada para que o nosso app se identifique perante o Twitter.

API Secret: um segredo usado pelo nosso app para provar que é o dono da API Key.

Access Token: depois de identificado junto ao serviço, nosso app precisa enviar o access token para que o serviço possa verificar qual é o nível de acesso que o app possui.

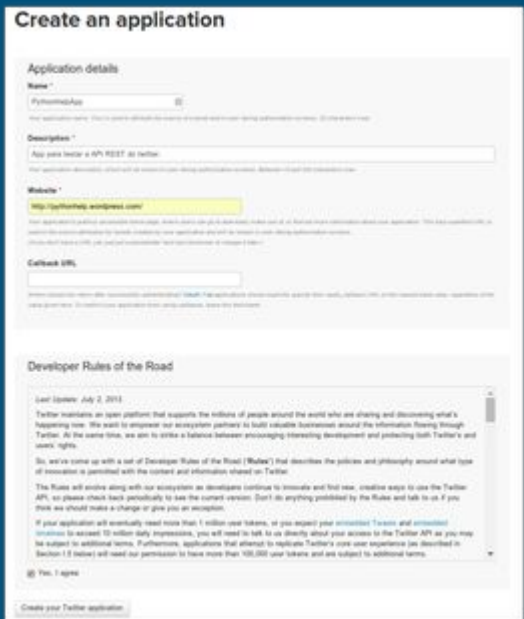
Access Token Secret: segredo usado pelo nosso app para provar que é o dono do access token.

Como Obter Essas Chaves com o Twitter

Para obter as chaves de acesso ao serviço é preciso primeiramente registrar um aplicativo junto ao Twitter.

Acesse <https://apps.twitter.com/>

Clique no botão “Create new app” para registrar um aplicativo para acesso à API. Primeiramente, você terá que preencher alguns dados básicos, como mostrado na imagem abaixo:



Create an application

Application details

Name *

Your application name. Once you submit this form, it will be visible to other developers using the Twitter API. It can be changed later.

Description *

Your application description. It will be visible to other developers using the Twitter API. It can be changed later.

Website *

Your website URL. It will be visible to other developers using the Twitter API. It can be changed later.

Callback URL

Where you want to receive the OAuth tokens. It will be visible to other developers using the Twitter API. It can be changed later.

Developer Rules of the Road

Last Update: July 2, 2013

Twitter maintains an open platform that supports the millions of people around the world who are sharing and discovering what's happening now. We want to empower our ecosystem partners to build valuable businesses around the information flowing through Twitter. At the same time, we aim to strike a balance between encouraging interesting development and protecting both Twitter's and users' rights.

So, we've come up with a set of Developer Rules of the Road (Rules) that describes the policies and philosophy around what type of innovation is permitted with the content and information shared on Twitter.

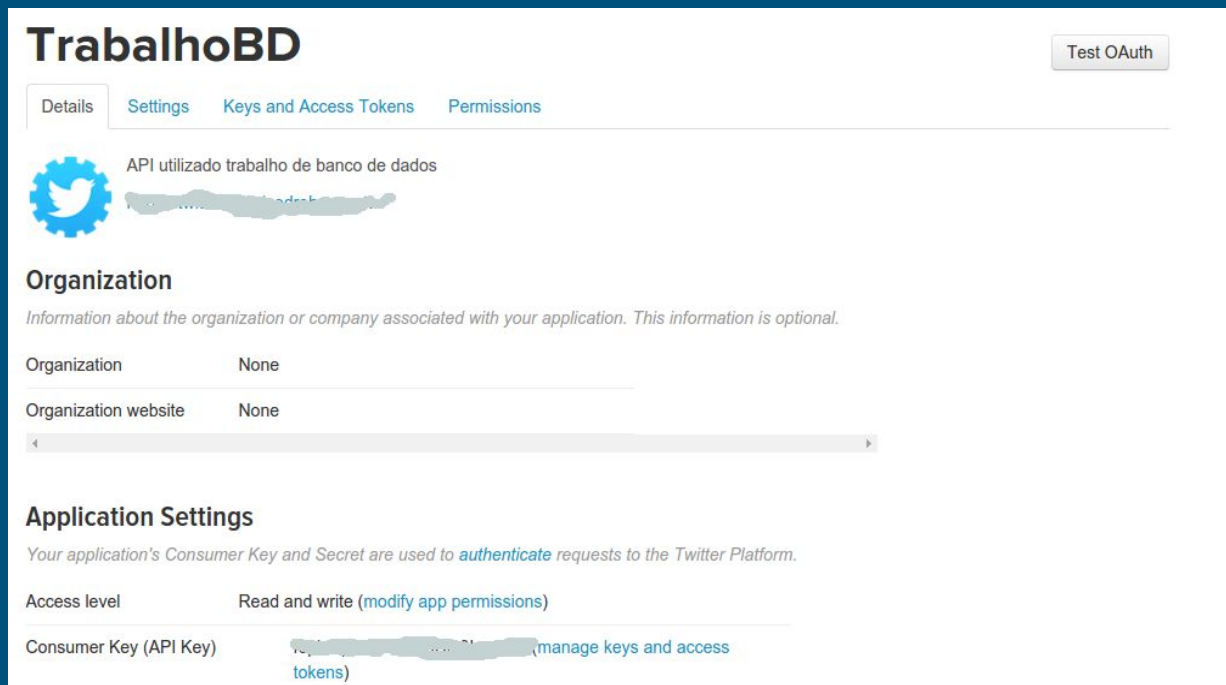
The Rules will evolve along with our ecosystem as developers continue to innovate and find new, creative ways to use the Twitter API, so please check back periodically to see the current version. Don't do anything prohibited by the Rules and talk to us if you think we should make a change or give you an exception.

If your application will eventually reach more than 1 million user tokens, or you expect your application to reach and eventually maintain 10 million daily impressions, you will need to talk to us directly about your access to the Twitter API as you may be subject to additional terms. Furthermore, applications that attempt to replicate Twitter's core user experience (as described in Section 1.1 below) will need our permission to have more than 100,000 user tokens and are subject to additional terms.

☒ Yes, I agree

[Create your Twitter application](#)


Após isso, você será redirecionado para a página de gerenciamento do seu app, como mostra a imagem abaixo.



The screenshot shows the 'TrabalhoBD' app management page in the Twitter Developer Portal. The page has a header with the app name 'TrabalhoBD' and a 'Test OAuth' button. Below the header is a navigation bar with tabs: 'Details', 'Settings', 'Keys and Access Tokens', and 'Permissions'. The 'Settings' tab is selected. The main content area is divided into two sections: 'Organization' and 'Application Settings'. The 'Organization' section includes a Twitter logo icon, a description 'API utilizado trabalho de banco de dados', and a redacted name. It also has fields for 'Organization' (set to 'None') and 'Organization website' (set to 'None'). The 'Application Settings' section includes a description of the Consumer Key and Secret, an 'Access level' dropdown set to 'Read and write (modify app permissions)', and a 'Consumer Key (API Key)' field with a redacted value and a link to 'manage keys and access tokens'.

TrabalhoBD [Test OAuth](#)

[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

 API utilizado trabalho de banco de dados
[Redacted Name]

Organization
Information about the organization or company associated with your application. This information is optional.

Organization

Organization website

Application Settings
Your application's Consumer Key and Secret are used to [authenticate](#) requests to the Twitter Platform.

Access level

Consumer Key (API Key) [Redacted] ([manage keys and access tokens](#))

Vá até a aba “API Keys” para acessar a página de gerenciamento das chaves de acesso do seu app ao serviço REST do twitter. Nela, clique no botão “Create my access token” (em vermelho na imagem acima) para que seja criado o token de acesso do app ao serviço

TrabalhoBD

[Details](#)[Settings](#)[Keys and Access Tokens](#)[Permissions](#)

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	[REDACTED]
Consumer Secret (API Secret)	[REDACTED]
Access Level	Read and write (modify app permissions)
Owner	pedrohe_costa
Owner ID	805742105381523456

Your access token

You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.

Token actions

Create my access token

Após isso, você será redirecionado para a página de gerenciamento do seu app, como mostra a imagem abaixo.

The screenshot displays the 'API Keys' management interface. It features a top navigation bar with tabs for 'Details', 'Settings', 'API Keys', and 'Permissions'. The 'API Keys' tab is active.

Application settings

Keep the "API secret" a secret. This key should never be human-readable in your application.

API key	[Redacted] 1
API secret	[Redacted] 2
Access level	Read-only (modify app permissions)
Owner	stummjr
Owner ID	67708679

Application actions

[Regenerate API keys](#) [Change App Permissions](#)

Your access token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access token	[Redacted] 3
Access token secret	[Redacted] 4
Access level	Read-only
Owner	stummjr
Owner ID	67708679

Acessando os recursos de busca da API

Exemplo de extração de tweets

```
def get_twitter(query, api) :  
    for tweet in tweepy.Cursor(api.search, q=query, lang='pt').items(15000):  
        idtweets = tweet.id;  
        mensagem = tweet.text;  
        data = tweet.created_at;  
        usuario = tweet.user.screen_name;  
  
        if (tweet.place != None) :  
            latitude = tweet.place.bounding_box.coordinates[0][0][0];  
            longitude = tweet.place.bounding_box.coordinates[0][0][1];  
        #fim if...  
  
        if (tweet.coordinates != None) :  
            latitude = tweet.coordinates['coordinates'][0];  
            longitude = tweet.coordinates['coordinates'][1];  
        #fim if...  
  
        if (tweet.place != None or tweet.coordinates != None) :  
            if (not conect_db_postgre.check_tweets(str(idtweets))) :  
                lst = [str(idtweets), mensagem, str(data), usuario, latitude, longitude];  
                print(lst);  
  
                conect_db_postgre.insert_tweets(lst);  
            #fim if...  
        #fim if...  
    #fim for...  
#fim def...
```


Parâmetros adicionais API.search

API.search(q[, lang][, locale][, rpp][, page][, since_id][, geocode][, show_user])

Parâmetros:

1. q- a seqüência de consulta de pesquisa
2. lang - Restringe tweets para o idioma dado, dado por um código ISO 639-1.
3. locale - Especifica o idioma da consulta que você está enviando. Isso é destinado a clientes específicos de idioma e o padrão deve funcionar na maioria dos casos.

1. Rpp - O número de tweets a devolver por página, até um máximo de 100.
2. Page - O número da página (a partir de 1) para retornar, até um máximo de —aproximadamente 1500 resultados (com base na página rpp *.
3. Since_id - Retorna somente status com um ID maior que (ou seja, mais recente que) o ID especificado.
4. Geocode - Retorna tweets por usuários localizados dentro de um determinado raio da dada latitude / longitude. A localização está preferencialmente a partir da API de Geotagging, mas voltará ao seu perfil do Twitter. O valor do parâmetro é especificado por "latitude, longitude, radius", onde as unidades de raio devem ser especificadas como "mi" (milhas) ou "km" (quilômetros). Observe que você não pode usar o operador próximo através da API para geocodificar locais arbitrários; No entanto, você pode usar esse parâmetro de geocódigo para pesquisar perto de geocódigos diretamente.

Construção query de busca

Suas query de podem ficar ainda mais poderosas para extração de informações.

Exemplos:

1. query = "imóvel OR veículo" Contendo "imóvel " ou "veículo" (ou ambos).
2. query = Que contêm a hashtag "veículo".

A uma variedade de combinações que podem ser encontradas:

<https://dev.twitter.com/rest/reference/get/search/tweets>

Link Úteis

Campos retornados nas chamadas dos métodos

1. <https://dev.twitter.com/overview/api/tweets>
2. <http://tweepy.readthedocs.io/en/v3.5.0/api.html#list-methods>
3. <https://dev.twitter.com/overview/api/tweets>
4. <https://dev.twitter.com/rest/reference/get/search/tweets>