

Chorus Tweetcatcher Desktop

The purpose of this manual is to enable users of Chorus Tweetcatcher Desktop (Chorus-TCD) to begin collecting Twitter data for their projects. The document is split into two main parts. The first part is concerned with setting up a Twitter application in order to generate the necessary access codes that enable Chorus-TCD to function. The second part deals with the main features and general operation of the Chorus-TCD.

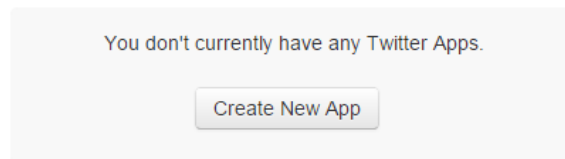
Creating a Twitter Application

In order to get Chorus-TCD working, you need to open a Twitter account (if you don't already have one) and register your intention to use the API on the Twitter developer site. The following Steps describe how to do this.

Step 1: Sign in to the Twitter developer site

Navigate to <https://apps.twitter.com/> and sign in using your usual Twitter credentials. You will then be able to access your **Application Management** area.

Twitter Apps



If you have not yet created an App with this account, you will see the above screen. Otherwise you will see a list of your existing Apps. In either case, you can create your Tweetcatcher app by clicking on the Create New App button.

Step 2: Create your application

After clicking on the **Create New App** button, fill out the required form fields. Just give your application a name and briefly describe what you will be doing with the data. In the example below, we have called our application 'UseTCD' but you can call yours anything you like, but it must be unique to Twitter. A combination of your name and Tweetcatcher (e.g. TimCribbinsTweetcatcher) is most likely to be unique. It also asks you to provide a web address, although in our case we're not working from a web location. You could just use your personal or project webpage here. The callback URL is optional and, again, irrelevant in our case. This is a good opportunity to familiarise yourself with the Twitter Developer Rules of the Road. Once you have read this, check the box and click the **Create your Twitter Application** button at the bottom of the screen.

Home → My applications

Create an application

Application Details

Name: *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description: *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website: *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL:

Where should we return after successfully authenticating? For [@Anywhere applications](#), only the domain specified in the callback will be used. [OAuth 1.0a](#) applications should explicitly specify their `oauth_callback` URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Rules Of The Road

Step 3: Get your consumer key and secret

If successful, you will be transferred to a new page providing details relating to your new application. Select the **Keys and Access tokens** tab to display the required access credentials. In addition to the application name, you will need two kinds of personal codes in order to start working with Chorus-TCD: the **consumer key** and **consumer secret**. You can see in the screenshot below that these are very long, random strings of characters. You'll find it easier to copy and paste these, rather than writing them down as a single wrong character will prevent Chorus-TCD from obtaining authentication. You access this page again later, if you need to, by navigating to the **Application Management** home page and clicking on the name of your application.

SMA924

Details Settings **Keys and Access Tokens** Permissions

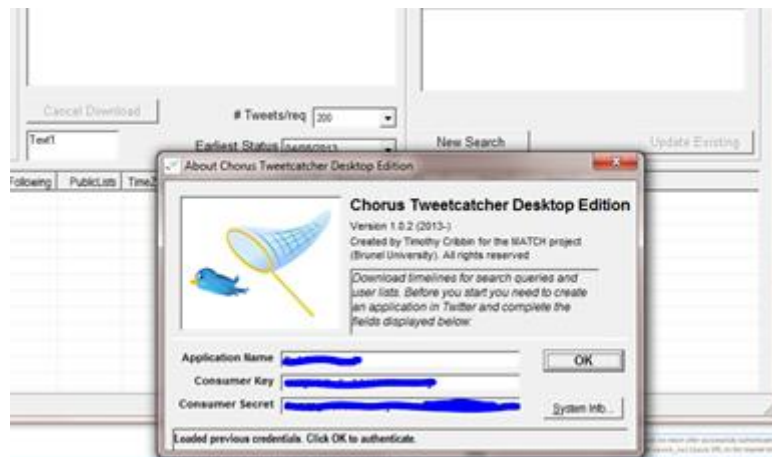
Application Settings

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

| | |
|------------------------------|--|
| Consumer Key (API Key) | svn2F9FAXy5T7ccDsx9bVnzJ2 |
| Consumer Secret (API Secret) | aHTLkeNchvUmFz1cMK7ChQnh7s6EHI4maQ0S3J5HmcSg8OCWQS |
| Access Level | Read-only (modify app permissions) |
| Owner | Chorus_Team |
| Owner ID | 1499166008 |

Step 4: Paste your credentials into Chorus-TCD

The final step is to tell Chorus-TCD about your application. When you start the tool for the first time, you should type or paste the Application name, its Consumer key and its Consumer secret into the three fields provided. When you click OK, if all credentials are correct you should see a notification in the status bar saying that authorisation was successful and you will be returned to the main Chorus-TCD interface.



Overview of the interface

The main interface consists of three main parts. In the top left are the controls for retrieving a user timeline. In the top right are the controls for retrieving a keyword query search. In the bottom half is a table view of the results of either of the above search modalities.

| DateTime | ScreenN... | Tweet | TweetID | RealName | Followers | Following | PublicLists | TimeZone | GeoCoords | UserTwe... |
|--------------|--------------|---|------------|--------------|-----------|-----------|-------------|-------------|-----------|------------|
| Sun Feb 0... | SpacePic... | A Digital Opportunity Rover on Mars. http://t.co/LGKZVFXAw | 5618597... | Outer Sp... | 6976 | 7674 | 0 | Baghdad | null | 15345 |
| Sun Feb 0... | ASHWINI... | NASA Jet Propulsion Laboratory Considers Building Mars Helicopter Rover for Next Mission - Pioneer ... | 5618582... | ASHWINI... | 407 | 413 | 0 | null | null | 14207 |
| Sun Feb 0... | uglyone | Mars Fossils? Curiosity Rover Team Questions Report on Potential Microbe Traces http://t.co/yCNITGN... | 5618570... | Rui de Br... | 1712 | 1998 | 0 | Lisbon | null | 6818 |
| Sun Feb 0... | mosambro | The rover #61-Curiosity takes a moment away from Mars exploration 2 search 4 a date on grindr. | 5618538... | Morgan ... | 5 | 69 | 0 | null | null | 188 |
| Sun Feb 0... | kennethd... | UFO Crash Photographed By Mars Curiosity Rover. http://t.co/E6q7zzSDMU via @YouTube | 5618537... | kenneth c... | 12 | 22 | 0 | null | null | 1554 |
| Sun Feb 0... | mattocha | Curiosity Mars rover back in the groove http://t.co/23MGsyJmpU | 5618533... | Mat Wille... | 120 | 298 | 0 | Tokyo | null | 5189 |
| Sun Feb 0... | iasishidd... | RT @ScottMechlowicz: At NASA JPL watching the Mars Curiosity rover landing live on Mars. http://t.co/6... | 5618526... | lasisLee... | 426 | 278 | 0 | Kuala Lu... | null | 9897 |
| Sun Feb 0... | appegate... | Mars Rover Finds Perfect Rectangle NASA Stays Quiet. http://t.co/aatzhGk8t | 5618523... | synthelep... | 49 | 185 | 0 | null | null | 407 |
| Sun Feb 0... | MaytonD... | Interesting photonNASA Mars Rover Finds New u2018Ancient Structure, u2019 Evidence Of u2018Intelli... | 5618479... | Dave Ma... | 1298 | 1976 | 0 | null | null | 8932 |
| Sun Feb 0... | watress4t... | @markbuckley23 @LouiseWalters12 @alisonbarrow also little space-bot things on Mars. #rover. Oh! An... | 5618453... | Emma Ho... | 2477 | 185 | 0 | null | null | 1061 |
| Sun Feb 0... | mars_fasn | NASA Mars Rover Finds New u2018Ancient Structure, u2019 Evidence Of u2018Intelligent Speciesu201... | 5618435... | Mars Fans | 193 | 933 | 0 | Tehran | null | 490 |
| Sun Feb 0... | HelenM17... | Curiosity rover clobbers Mars rock on http://t.co/4lHrzEUsc | 5618403... | Helen Mt... | 180 | 294 | 0 | null | null | 26667 |
| Sun Feb 0... | ushio_590 | RT @yuuka_518: 'Human Shadow' Visible In Mars Rover Photo http://t.co/ezmpzU7 http://t.co/aGbtGRK... | 5618365... | lu00e3lu... | 4 | 0 | 0 | null | null | 864 |

User Timeline

The user can import a list of user names (screen names) by copying a column from a spreadsheet or a line-delimited list from a text file and clicking the **Paste List** button. If you have a recent version of

Tweetvis, it is also possible to select and copy user names from the **Term Statistics** panel. All the user names should appear within the adjacent list window. User names can be selected manually by checking the boxes. Alternatively, all users in the list can be selected by clicking **Select All**.

To retrieve the timelines of all selected users, click **Retrieve Selected**. By default, the tool will attempt to retrieve all available statuses made by each user in the list back to the **Earliest Status** date. Twitter places a limit of 3200 statuses per user, so it is may not be possible to retrieve the entire timeline of certain high volume users. Make sure you set the Earliest Status date well into the past if you want to retrieve complete timelines wherever possible.

The **Update Existing** button allows users to update timeline sets that have been collected in the past and re-loaded. On clicking this button, Chorus-TCD will attempt to append the existing dataset with all the status updates made by the selected users since their most recent update in the existing file. Make sure you save your list of users somewhere safe as Chorus-TCD does not save this list as part of the dataset. You will need to reapply (paste) the list on each subsequent update session.

Search

The search panel enables the user to retrieve all statuses matching a keyword query. This query will match both free text and hashtags.

A **New Search** will retrieve as many statuses that are available that match the query from the current point in time. The number of statuses that can be retrieved at any point in time is limited, normally to the last 7 days. However, for high volume topics the limit may be more recent. If you want to collect data over a longer time period, you will need to save your data and reload it later to do an update operation.

Update Existing will retrieve all the available matching statuses from the most recent status in the loaded timeline. Make sure that you type the query exactly as you did on the previous session, unless you have some reason to change it. Note that Chorus-TCD does not save your query as part of the dataset so write it down somewhere safe whilst you are running your first search. You should try to run an update session at least once a day to avoid the problem of gaps appearing in your timeline i.e. when the oldest accessible tweet is not the next one to occur after the newest tweet in your existing dataset. This can happen even for non-viral topics and sudden surges in the volume of statuses matching your query can occur, unexpectedly, for a variety of reasons.

Understanding the data output

Chorus-TCD generates a data file with the following fields:

- DateTime – the date and time of that the tweet was sent
- ScreenName – the author's Twitter handle
- Tweet – the actual message content
- RealName – the full provided name of the author
- Followers – the number of followers of the author
- Following – the number of users followed by the author
- PublicLists – the number of public lists the author belongs to
- TimeZone – the timezone in which the author has declared that they reside
- Geocoords – the longitude and latitude of the author at the time the tweet was sent

- UserTweets – the number of status updates made by the author
- Retweets – the number of times that the status update has been retweeted
- LinkURL – any links embedded within the message
- Sent+ - a rating of positive sentiment (1 to 5)
- Sent- - a rating of negative sentiment (-1 to -5)
- NetSent – the sum of both Sent+ and Sent-

Chorus-TCD saves this data in a tab-delimited format, which is compatible with Chorus-TV. It is also easily read into other applications such as Excel, SPSS and various text-editors.

The sentiment ratings are computed using the Sentistrength algorithm created by Mike Thelwall's team at the University of Wolverhampton, UK¹.

Other features

Twitter API methods encourage users to work in a retrospective fashion, gathering the most recent statuses first then working backwards to the earliest available data point. Chorus-TCD dutifully follows this model. However, we recognise that there might be situations where users do not need to go back beyond a certain point in time. For this, there is the **Retrieve back to** date selector. This is particularly useful for user timeline retrieval sessions where it might be desirable to limit data collection for all users to a limited period rather than the last 3200 posts (which might go back several years in some cases). By default, the date is always set to one week prior to the current date, which is the normal limit for Twitter keyword searches.

The **Cancel Current Download** menu item (**Actions** menu) is provided to enable you to stop a session early. This might be useful, for instance, if you are experimenting with queries and wish to stop early and reformulate.

The **File menu** provides users with a means of loading and saving data. **Open** and **Save** work in a predictable way. However, there is one other menu item called **Save & Send to TV**. This allows users who have installed the full Chorus package to seamlessly load the current dataset into Chorus-TV for analysis.

The **Actions** menu provides various other useful functions. **Convert DateTime** provides sub-menus to convert the date-time string to different formats. The first, **to DD/MM/YY HH:MM:SS**, is a more common format that is recognised by Excel and many other applications and programming languages. The second option, **to SPSS compatible**, is for SPSS users as SPSS requires a slightly different format to the above: DD-MM-YYY HH:MM:SS. The third option, **back to Twitter format**, simply returns the date-time string back to its original format. Also in **Actions** are functions to **Remove Retweets** (all posts beginning with 'RT @') and **Copy Log**. The latter function simply copies items from the request history list to the clipboard.

The **User List** menu includes functions to **Get User Followers** and **Get User Friends**. This allows you to retrieve timelines of users sharing a common relationship (i.e. all users who follow a specified account or, in the latter case, all users followed by a particular account). In each case it is necessary

¹ Thelwall, M., Buckley, K., Paltoglou, G., & Cai, D. (2010). Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

to enter a single valid screen name to start the process. Once retrieved the follower/friends screen names will be copied to the Users list box in the User Timelines frame. Summary details for the selected user are displayed in adjacent text box. More detailed metadata are copied as a table to the clipboard. See Brooker et al. (2014)² for an example of the user driven retrieval strategy in action. We are currently in the process of developing more sophisticated support for building, analysing and refining user lists.

Final points

Whilst we (and many others) have found Chorus-TCD to be a useful tool and have applied it extensively to in-house projects, users still need to remember that it is currently an experimental (and free to use) prototype and not an ‘industrial strength’ data collection service. It is also wholly dependent on the Twitter API service for its functionality – any disruption to this service (or your internet connection) may cause data loss or produce other unexpected results. Before you employ Chorus-TCD for serious projects, you are advised to experiment initially, on a pilot basis, to ensure that you understand its limitations. It is not recommended to use Chorus-TCD for ‘mission critical’ projects where failure to gather the expected data might result in material loss, financial otherwise. If this is the case, you are advised to use an approved Twitter syndication service.

Likewise users are reminded that they should use Chorus-TCD, and any data retrieved by it, in a fair and responsible manner. This means adhering to Twitter’s Rules of the Road along with any applicable ethical codes of practice and data protection laws.

That said, have fun with Chorus-TCD.

² Brooker, P., Barnett, J., Cribbin, T., Lang, A., & Martin, J. (2014). User-driven data capture: Locating and analysing Twitter conversation about cystic fibrosis without keywords, *SAGE Research Methods Cases*. London: Sage Publications Ltd.