

Chorus Tweetcatcher Desktop

The purpose of this manual is to enable users of Chorus Tweetcatcher Desktop (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 TCD to function. The second part deals with the main features and general operation of the TCD.

Installing Tweetcatcher

TCD is normally installed as part of the Chorus package, which also includes a visualisation tool called Tweetvis. The installer can be downloaded from the Chorus website¹. The download page also contains instructions for installation, trouble-shooting advice and links to historical and newer software versions.

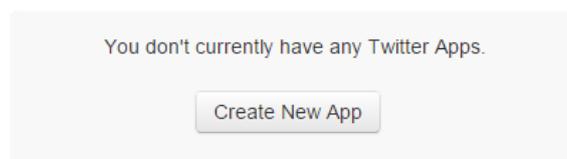
Creating a Twitter Application

Before you can use your new installation of TCD you need to log-in to Twitter and register your intention to use the API on the Twitter developer site by creating an Application. After you have done this, Twitter will provide you a new set of credentials that you will need to enter when TCD starts. The following Steps describe how to do this.

Step 1: Sign in to Twitter

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

¹ <http://chorusanalytics.co.uk/downloads/>

unique to Twitter. A combination of your name and Tweetcatcher (e.g. TimCribbinsTweetcatcher) is most likely to be unique. Although you are not working from a web location, you need to put something in the **Website** field. You could just use your personal or project webpage here. The callback URL is optional and, again, not relevant in this case. This is also 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 TCD: the **consumer key** and **consumer secret**. You can see in the screenshot below that these are very long, seemingly 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 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](#)

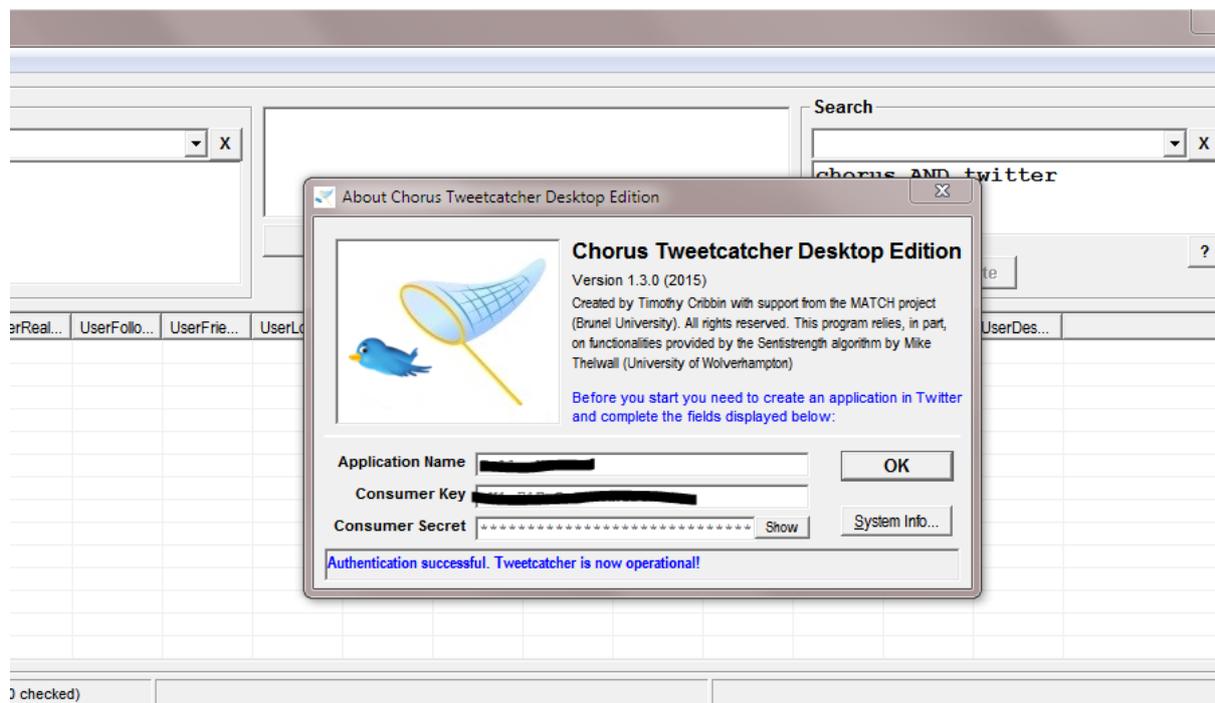
Application Settings

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 TCD

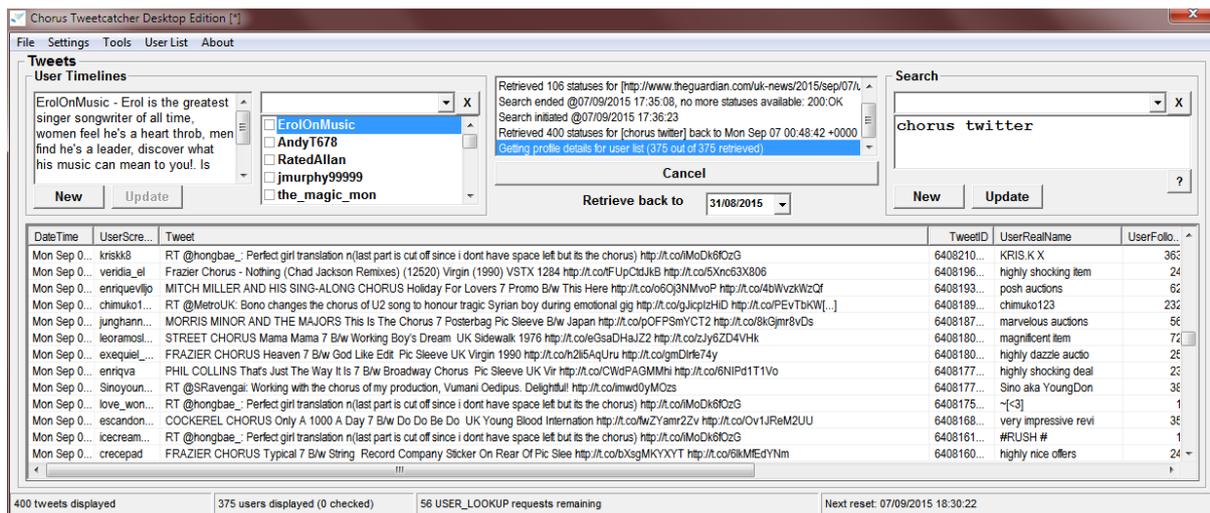
The final step is to tell TCD about your application. When you start the tool for the first time, you should type or paste the Application name, its Consumer key and its Consumer secret into the three fields provided. When you click OK, if all credentials are correct you should see a notification in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



Overview of the interface

All main functions of TCD are accessed from a single window. There are four areas or panes to this interface:

- The top left pane contains controls for retrieving a user timeline based on one or more user screen names.
- The top centre pane contains the data retrieval log plus controls for setting date constraints and cancelling the current retrieval process.
- The top right pane contains controls for retrieving tweets using a query search.
- The bottom half of the window is home to the Tweets view, which displays the most recently loaded or downloaded set of tweets data.



User Timelines

This panel allows you to retrieve tweets and associated metadata posted by a specific set of Twitter users. The main element is the User List on the right side of the panel. This contains screen names of users sourced in one (or more) of the following ways through the *User List* menu:

- A pre-defined list compiled in a text-editor. This can be pasted into the User List using the *Paste* option from the menu.
- A list of users associated with a tweets dataset. This could be copied from Tweetvis and pasted into the list or extracted from the current tweets table using the menu item *Get Users From Tweets*
- The followers of a specified user specified via *Get User Followers*
- The friends of a specified user specified via *Get User Friends*
- Manual addition of one or more users specified via *Add User*

Clicking on a user item (screen name) will cause that user's profile details to appear. In the screenshot above, for instance, ErolOnMusic has been selected and we can see that he says his bio says he is "the greatest singer songwriter of all time...". Scrolling down this text box will reveal further details, including the user's friend, follower and tweet counts. Check boxes to the left of User list items allow users to be selected for retrieving their timelines. Users can be selected manually by clicking the boxes or all users in the list can be checked at once by clicking **Select All**.

To retrieve the timelines of all selected users, click **New**. By default, the tool will attempt to retrieve all available statuses made by each user in the list back to the **Retrieve back to** date. Twitter places a limit of the most 3200 statuses per user, so it is may not be possible to retrieve the entire timeline of

high volume users or those with older accounts. Make sure you set the date well into the past (e.g. two or three years) if you want to retrieve the maximum number of available tweets.

The **Update** button allows users to update timeline data that have been collected during an earlier session. The drop-down list control above the user list keeps a record of the most recently saved user timeline data files. Alternatively a data file can be opened from the *File* menu. After loading the existing data, you can populate the user list by either clicking the menu item *Get Users From Tweets* or by pasting the user list into the list. Select all required users and click the **Update** button. TCD will then attempt to append the existing dataset with all recent status updates made by the selected users.

Search

The search panel enables the user to retrieve all recent statuses matching a keyword query. This query will match both tweet content (words and hashtags) and metadata associated with a tweet such as user name. There are a range of operators that can be used to make a query more specific. A guide to using these operators can be found on Twitter's site² or by clicking on the ? button in the Search panel.

A **New** search will retrieve as many statuses that are available that match the query from the current point in time. Any tweets displayed in the Tweets table will be deleted, so it is important to save these (if required) before commencing a new search. The number of statuses that can be retrieved is normally limited to the last 7 days' worth of updates. However, for high volume topics the limit may be more recent than this. If you want to collect data over a longer time period, you will need to save your data and do an Update operation in a future session.

Update will retrieve all the available matching statuses from the most recent status in the loaded timeline. If the timeline was saved in a recent session, the query and the name of the file should appear in the drop-down history list. Loading a file in this way will also populate the query box with the correct query. Alternatively, you can load a timeline data file using the File menu. If you do this, you will need to re-enter the query yourself. TCD does not save your query as part of the dataset, so if you are performing a new search that you plan to update later, it's a good idea to write the query down somewhere safe before ending the TCD session.

To avoid the problem of gaps appearing in your timeline, you should try to run an update session at least once a day and certainly no more than three days since the previous one. Although the normal limit is seven days, Twitter may reduce this limit without warning, particularly if the topic suddenly goes viral.

Understanding the data output

TCD generates a data file with the following fields which appear in the following order both in the tweets table displayed on the user interface and in any saved data files:

- DateTime – the date and time of that the tweet was sent
- UserScreenName – the author's Twitter handle (e.g. TimCribbin)
- Tweet – the actual status content

² <https://dev.twitter.com/rest/public/search>

- UserRealName – as declared by the author (e.g. Tim Cribbin)
- UserFollowers – the number of followers of the author
- UserFriends – the number of users followed by the author
- UserLocation – as declared by the author in their profile (e.g. London, UK)
- UTCOffset – the time difference in hours from universal standard time (UTC/GMT), as specified by the author
- GeoCoords – a string describing the longitude and latitude of the author location at the time the tweet was sent. Note that this depends on the user allowing their location to be disclosed and thus is frequently empty or null.
- UserTweets – the number of status updates made by the author since opening their account
- Retweets – the number of times that the original status had been retweeted at the time the data was collected. Multiple instances of the same retweet will tend to have the same count value if they were collected during the same session.
- LinkURL – a list of any web links embedded within the status text
- SentPos - a rating of positive sentiment ranging from 1 to 5 (e.g. a happy/excited status might score four or five whilst a purely negative or neutral one might score just one)
- SentNeg - a rating of negative sentiment ranging from -1 to -5 (e.g. a sad/angry status might score four or five whilst a purely positive or neutral one might score just one)
- UserBio – the bio or self-description as declared by the author in their profile

TCD saves data in a tab-delimited format. It is designed to be compatible with Chorus-TV, but is also easily read into other applications such as Excel, SPSS or any good text-editor.

The sentiment ratings are computed using the Sentistrength algorithm created by Mike Thelwall's team at the University of Wolverhampton, UK³.

User Menu

File

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.

Settings

The **Settings** menu has options to **Sort Tweets Table** (normally best left checked), **Set Search Language** and **Save Tweets with Headers**. Setting the search language requires you to specify the relevant ISO-639-1 code. This would be *en* for English or *fr* for French, for example. Saving the tweets file with headers is useful if you are exporting to a data analysis tool like SPSS or Excel. However, if you intend to use your data in Tweetvis, you should save without headers.

Tools

The **Tools** menu provides various other useful functions. **Convert DateTime** provides sub-menus to convert the date-time string into different formats. The default Twitter format is quite verbose and,

³ 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.

whilst easy to read by a human, is not so useful for machine processing tasks. The first option, **to Numeric format**, will convert to a more common format that is recognised by Excel and most other applications/programming languages. The exact format will vary depending on your Region/Date settings, but would be *DD/MM/YY HH:MM:SS* for UK based users. The second option, **to SPSS compatible**, is for SPSS users as SPSS requires a slightly different format to Numeric: *DD-MM-YYYY HH:MM:SS*. The third option, **back to Twitter format**, simply returns the date-time string back to its original format, regardless of its current format. Also in in this menu are functions to **Remove Retweets** (all posts beginning with 'RT @'), **Filter Tweets** (containing or not containing some specified word or phrase) and **Copy Log**, which simply copies the contents of the request history list to the clipboard.

User List

The **User List** menu includes functions relating to the user cases listed in the **User Timeline** panel. It can also be accessed as a context menu by right-clicking when the mouse pointer is over the user list. Items **Copy** and **Copy Metadata** copy, to the clipboard, the current list of user names or the complete table of user metadata respectively. Various metadata about users in the list is provided in this table. Headers on the first row of the table describe the content of each column. Paste will replace the contents of the current user list with a list that is currently on the clipboard. **Select All** will check all users in the current list and changes to **Deselect All** if one or more items are already checked. If one or two items are already selected but your intention is to select all, just click twice to enable Select All. **Clear** will clear the contents of the user list. **Add user** allows you to manually add an individual user name to the list, whilst **Delete user** will remove an existing one. These functions work slightly different if accessed from the Tweets table using the context menu where they will add or delete the author of the currently selected tweet.

Get User Followers and **Get User Friends** 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 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)⁴ 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. Note that either of these functions will replace the existing contents of the user list.

Tweet view context menu

A context menu is also available for the Tweet view by right-clicking when the mouse pointer is over this area. This menu is dynamic and is driven by the data in the currently selected list item. The first item, **Go To http...** will open a web page that is cited by the currently selected tweet. **Find...on Twitter** will open the author's Twitter homepage. **Add/Delete...to User List** will add/delete that user to/from the user list. Finally **Delete tweet** will delete the currently selected row in the Tweets

⁴ 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.

table, whilst **Delete all Tweets by...** will search the tweets tables for all tweets posted by the author of the selected tweet and delete all of them.

Other UI elements

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. 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** button 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.

Final words

Whilst we (and many others) have found 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 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 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 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, we hope you find TCD useful and thank you for choosing Chorus.

Any constructive feedback is welcomed: team@chorusanalytics.co.uk