



Coveo Platform 7.0

Twitter v2 Connector Guide

Notice

The content in this document represents the current view of Coveo as of the date of publication. Because Coveo continually responds to changing market conditions, information in this document is subject to change without notice. For the latest documentation, visit our website at www.coveo.com.

© Coveo Solutions Inc., 2016

Coveo is a trademark of Coveo Solutions Inc. This document is protected by intellectual property laws and is subject to all restrictions specified in the Coveo Customer Agreement.

Document part number: PM-160628-EN

Publication date: 1/3/2019

Table of Contents

1. Twitter v2 Connector	1
1.1 Features	1
1.2 Limitations	2
2. Twitter v2 Connector Deployment Overview	3
3. Twitter v2 Connector Requirements	4
4. Obtaining the Twitter API and Access Token	5
5. Configuring and Indexing a Twitter v2 Source	7
5.1 Adding an Explicit Connector Parameter	13
5.2 Modifying Hidden Twitter v2 Source Parameters	16

1. Twitter v2 Connector

CES 7.0.8388+ (June 2016)

Twitter is an online social networking service that enables users to post and read short messages called "Tweets". Registered users can post and read Tweets; unregistered users can only read them. The second generation (v2) of the Coveo connector for Twitter allows Coveo administrators to index and integrate the Twitter content of their managed users into a Coveo unified index. The connector indexes specific content from all managed users, so that any user can easily find the Twitter content he has access rights to through any Coveo search interface.

1.1 Features

The Twitter v2 connector features:

Content indexing

Extraction and indexing of the following Twitter item types:

- Users
- Usernames and names
- Protected users (see [Limitations](#))
- Tweets (normal, mention, reply, retweets, quote)
- Hashtags (#) and cashtags (\$)
- Symbols
- Timelines (user, home, mentions, lists, collections)
- Lists
- Collections

Incremental refresh

Supports incremental refresh to periodically query Twitter for the latest Tweets, newly followed users and newly listed members, keeping index content up-to-date.

Note: Some changes need a full refresh or a source rebuild to be taken into account (see [Limitations](#)).

Multi-threading

Can run multiple threads, which can improve performances considerably (see [Configuring and Indexing a Twitter v2 Source](#)).

Pause/Resume

When indexing Twitter content, connector tasks can be paused and resumed.

1.2 Limitations

- The connector does not support permissions. Permissions need to be defined manually.
- A Twitter API limitation prevents the connector from supporting user mappings.
- During an incremental refresh, deleted items cannot be detected and require a full refresh or rebuild to be updated or removed.
- Incremental refresh cannot be performed if the last rebuild or refresh did not end normally, i.e., the source was stopped, paused, or an exception occurred.
- Direct messages are not indexed.
- Tweets belonging to protected users (protected Tweets) cannot be retrieved from a search query.

Note: Protected Tweets can only be retrieved from a user timeline if the authenticated user is an approved follower of the protected user.

What's Next?

Review the steps to deploy the Twitter v2 connector (see ["Twitter v2 Connector Deployment Overview" on page 3](#)).

2. Twitter v2 Connector Deployment Overview

The following procedure outlines the steps needed to deploy the Twitter v2 connector. The steps indicate the order in which you must perform configuration tasks on both the Twitter and Coveo servers.

To deploy the Twitter v2 connector

1. Validate that your environment meets the requirements (see ["Twitter v2 Connector Requirements" on page 4](#)).
2. From the Twitter server, obtain the necessary API, token, and secret access codes (see ["Obtaining the Twitter API and Access Token" on page 5](#)).
3. On the Coveo server, in the Coveo Enterprise Search (CES) Administration Tool:
 - a. Configure and index a Twitter v2 source.

The connector must know certain details to access and index the Twitter content of your managed users (see ["Configuring and Indexing a Twitter v2 Source" on page 7](#)).

- b. Fine tune connector performance by modifying (if necessary) the default value of hidden source parameters. This step also applies if you encounter indexing issues (see ["Modifying Hidden Twitter v2 Source Parameters" on page 16](#)).

3. Twitter v2 Connector Requirements

For you to use the Twitter v2 connector, your environment must meet the following requirements:

- [CES 7.0.8388+ \(June 2016\)](#)
- Coveo license for the Twitter v2 connector

Your Coveo license must include support for the Twitter v2 connector to be able to use this connector .

- A valid [Twitter user account](#)
- An active corporate [Twitter application](#)

Notes:

- It is strongly suggested that you generate your Twitter account access token upon creation of your account. Doing this will simplify various operations requiring the use of API keys and access tokens.
- It is considered a best practice to setup a Twitter application dedicated to the Coveo connector.

What's Next?

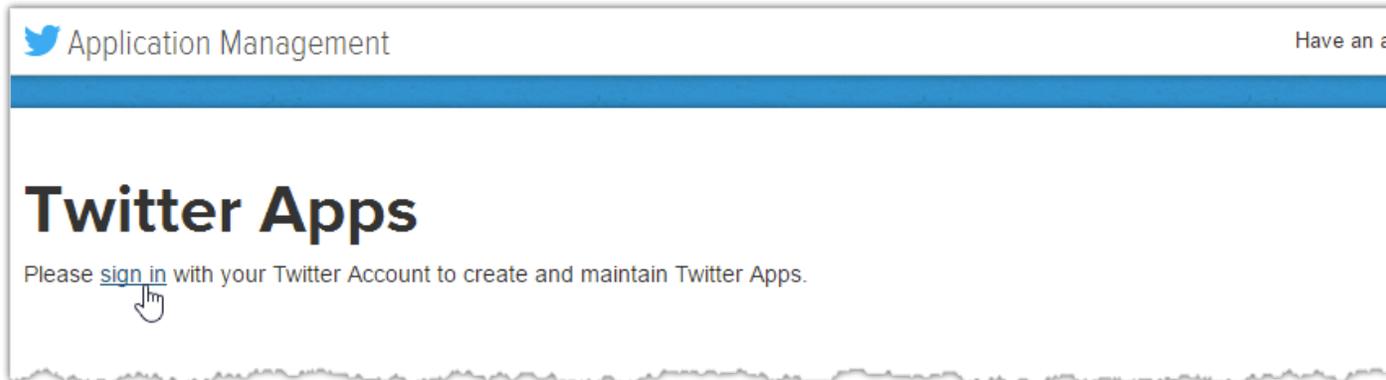
Grant Coveo access to the valid Twitter account and, in turn, to the content of all your managed users (see [Obtaining the Twitter API and Access Token](#)).

4. Obtaining the Twitter API and Access Token

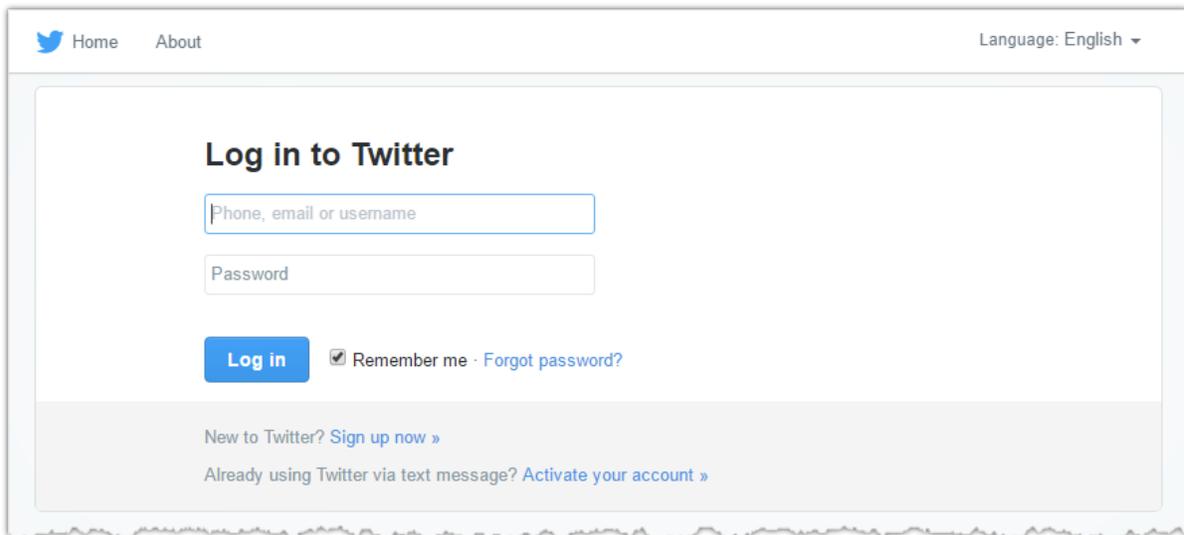
The following procedure outlines the steps needed to obtain the values given by your Twitter application when configuring a Twitter v2 source to index.

To obtain the API and Access Token values

1. From any browser, [sign in your Twitter application](#).



2. Click on the name of your application and, on the screen that appears, log in with your username and password.

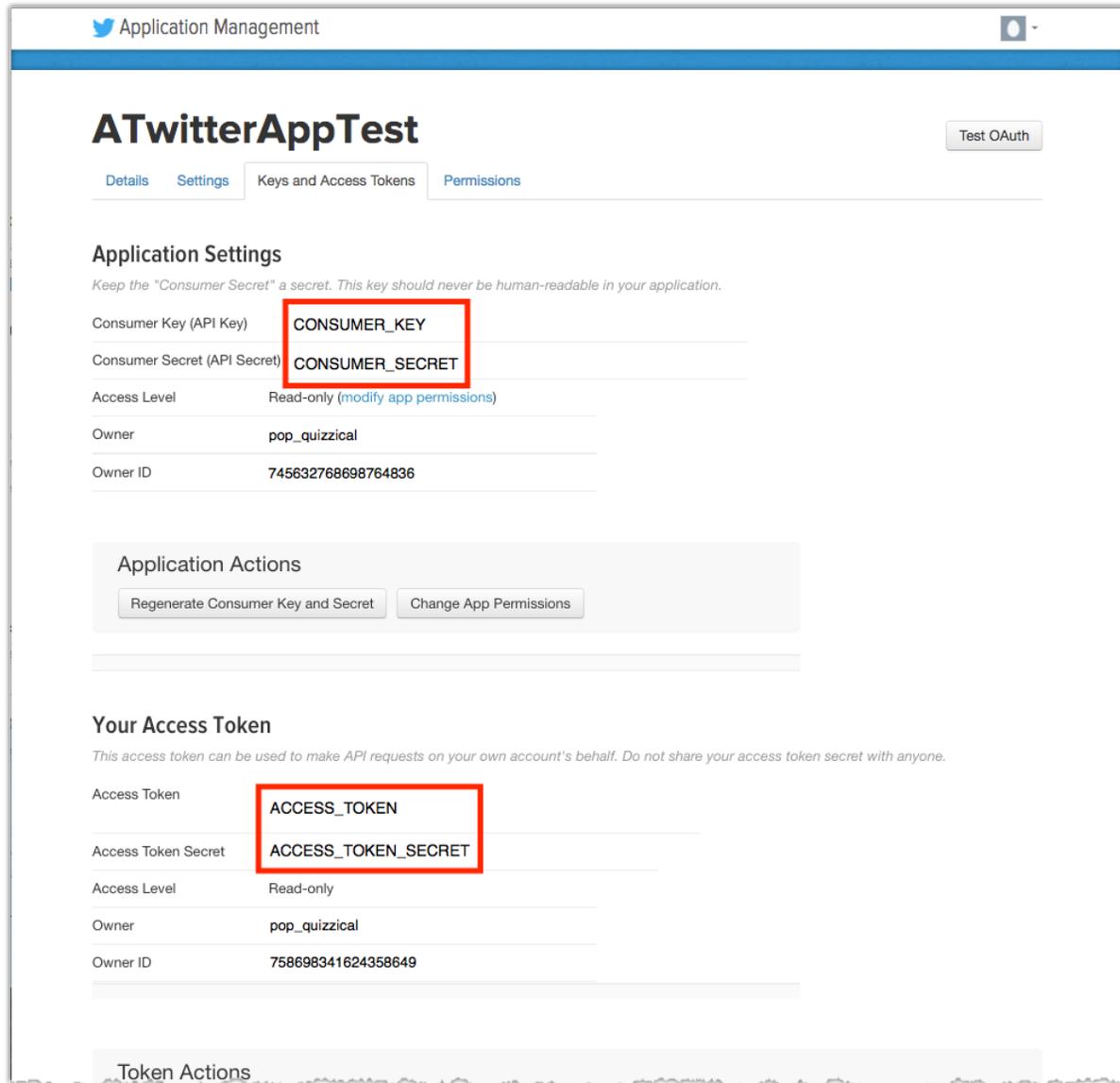


3. In the **Application Management** screen, click **Keys and Access Tokens** under your application name.

On the page that appears, under **Application Settings**, you will find the application **Consumer Key (API Key)** and **Consumer Secret (API Secret)**, and under **Your Access Token**, you will find the application **Access Token** and **Access Token Secret**.

4. Stay on this page until you have input these keys and tokens in the appropriate fields when configuring the

Twitter v2 connector in the Administration Tool (see [Configuring and Indexing a Twitter v2 Source](#)).



What's Next?

Configure and start indexing Twitter accounts and, in turn, the Twitter content of all your managed users (see [Configuring and Indexing a Twitter v2 Source](#)).

5. Configuring and Indexing a Twitter v2 Source

A source defines a set of configuration parameters for a specific Twitter account.

To configure and index a Twitter v2 source

1. On the Coveo server, access the Administration Tool.
2. Select **Index > Sources and Collections**.
3. In the **Collections** section:
 - Select an existing collection to which you want to add the new source.

OR

 - Click **Add** to create a new collection.
4. In the **Sources** section, click **Add**.

The **Add Source** page that appears is organized in two sections.

5. In the **General Settings** section of the **Add Source** page:

The screenshot shows the 'Add Source' configuration page. The breadcrumb trail is: Sources and Collections > Top Results > Ranking > Thesaurus > Mirrors > Remote Indexes > Search Scopes > Result Clustering > Advanced. The page title is 'COLLECTION: TWITTER V2 - ADD SOURCE' with a subtitle 'Defines the parameters of the source of documents.' and a 'Help' button. The 'General Settings' section contains the following fields:

Name	Twitter v2
Source Type	Twitter v2
Addresses	
Rating	Normal
Document Types	Default
Active Languages	Default
Fields	Default Scheme
Refresh Schedule	Every day

- a. Enter the appropriate value for the following required parameters:

Name

Enter a descriptive name of your choice for the connector source.

Example: `Twitter v2`

Source Type

Select the connector used by this source. In this case, select **Twitter v2**.

Note: If you do not see **Twitter v2**, your environment does not meet the requirements (see "[Twitter v2 Connector Requirements](#)" on page 4).

Addresses

Starting addresses are source locations in your Twitter account. The Twitter v2 connector supports four types of starting addresses. Enter all required starting addresses:

- **users:** a list of comma-separated usernames.

Example: `users:twitterapi,coveo,support`

- **list:** a list's name and its creator's username in the format `{username}/{listname}`. These values can be obtained in the list's URL (`https://twitter.com/{username}/lists/{listname}`).

Example: `list:twitterapi/meetup-20100301`

- **query:** a search query. Only Tweets posted within the last 7 days will be retrieved with this starting address, and the results are not comprehensive (some Tweets posted in the last week might be missing).

Examples:

- `query:#coveo`
- `query:$GE`
- `query:puppy filter:media`

- **collection:** a collection's ID, which can be obtained from the collection's URL (`https://twitter.com/{username}/timelines/{ID}`).

Example: `collection:719516944244817923`

- Review the value for the following and seldom-modified parameters:

Rating

Change this value only when you want to globally change the rating associated with this source relative to the rating of other sources.

Example: When a source replaces a legacy system, you may want to set this parameter to **High** so that, in the search interface, results from this source appear higher in the list compared to those from legacy system sources.

Document Types

If you defined a custom document type set for this source, select it.

Active Languages

If you defined custom active language sets, make sure that you select the most appropriate for this source.

Fields

Coveo Enterprise Search (CES) 7.0 does not require the presence of custom field sets to extract metadata from Twitter. However, if you elected to use a specific field set, you can select it in the **Fields** drop-down list. Otherwise, leave it at **Default Scheme**.

Refresh Schedule

Time interval at which index content is automatically updated. By default, the **Every day** option instructs CES to refresh the source every day at 12 AM. Since the incremental refresh takes care of updating the source, you can select a longer interval such as **Every Sunday**.

Note: Since deleted items are not taken into account during normal incremental refresh operations (see "[Limitations](#)" on page 2), it is recommended to force a full refresh, or even a source rebuild, to update the status of those deleted items.

6. In the **Specific Connector Parameters & Options** section of the **Add Source** page:

Specific Connector Parameters & Options

API Key	<input type="text" value="CXkaZRMVgJsUstRW2ua1EclMa0"/> ?
API Secret	<input type="password" value="....."/> ?
Access Token	<input type="text" value="737748624214503430-nS18F39tsTzia0TMA"/> ?
Access Token Secret	<input type="password" value="....."/> ?
Number of Refresh Threads	<input type="text" value="2"/> ?
Mapping File	<input type="text" value="Coveo.CES.CustomCrawlers.Twitter2.Mappi"/> ?
Index Users' Followings	<input type="checkbox"/> ?
Index Mentions	<input type="checkbox"/> ?
Index Retweets	<input checked="" type="checkbox"/> ?
Index Replies	<input checked="" type="checkbox"/> ?
Keep old items	<input checked="" type="checkbox"/> ?
Parameters	Add Parameter ?
Option	<input checked="" type="checkbox"/> Index subfolders ? <input type="checkbox"/> Index the document's metadata ? <input type="checkbox"/> Document's addresses are case-sensitive ? <input checked="" type="checkbox"/> Generate a cached HTML version of indexed documents ? <input type="checkbox"/> Open results with cached version ?

Save
 Save and Start
 Cancel

- a. Enter the previously obtained values. These values allow the Twitter v2 connector to access the content of your managed Twitter users.

Note: These values are long series of alphanumeric characters.

API Key

The API key provides the identity of your corporate Twitter application when crawling the data of managed Twitter users whose Twitter account needs to be indexed.

API Secret

The API secret is used to authenticate the API key above.

Access Token

The access token is used to make API requests on behalf of your corporate Twitter application.

Note: This access token is used to retrieve a new OAuth 1.0a access token.

Access Token Secret

The access token secret is used to authenticate the access token above.

- b. The default value for the following parameters often does not need to be changed:

Number of refresh threads

Determines the number of threads that the connector can refresh simultaneously. The default and recommended value is 2.

Mapping File

Indicates the path to the default mapping file that defines how the connector handles metadata. The installer creates the file indicated in this field.

- c. Select additional content to be indexed using the following options:

Index User's Followings

Indexes the Tweets of users that a managed user is following. With this check box selected, users followed by users specified in `users:user1,user2,...` starting addresses will be retrieved. By default, this check box is not selected.

Index Mentions

Indexes the Tweets mentioning a user specified in a `users:user1,user2,...` starting address. By default, this check box is not selected.

Note: Only Tweets from the last 7 days will be indexed. Mention Tweets are retrieved by searching for "@username", so parameters set for search query results will also affect the mention Tweets retrieved.

Example: if `SearchResultIsoLanguageCode` is set to "fr", only mention Tweets written in French will be retrieved.

Index Retweets

Indexes all retweets when retrieving Tweets from a user timeline. By default, this check box is selected.

Index Replies

Indexes all replies when retrieving Tweets from a user timeline. By default, this check box is selected.

Note: When the **Index Retweets** and **Index Replies** check boxes are not selected, the number of Tweets retrieved per page will be up to the number specified with the `TweetsPerTimelinePage` parameter, because results are filtered after being retrieved.

Keep old items

With this check box selected, items are not deleted, even if they are not recrawled. By default, this check box is selected. Items are not deleted unless this check box is cleared.

Note: Old items can be removed by manually removing a folder or an item from the index, clearing this check box or deleting the source.

- d. (Optional) Click **Add Parameter** when you want to show and change the value of advanced source parameters (see ["Modifying Hidden Twitter v2 Source Parameters" on page 16](#)).
- e. The **Option** check boxes generally do not need to be changed:

Index Subfolders

When selected, CES indexes subfolders recursively. By default, this check box is selected.

Index the document's metadata

When selected, CES indexes all the document metadata, even metadata that is not associated with a field. Orphan metadata is added to the body of the document so that it can be searched using free text queries.

When this check box is cleared (default), only system and custom field values for which the **Free Text Queries** attribute is selected will be searchable without using a field query.

Example: A document has the following metadata:

- `LastEditedBy` containing the value `Hector Smith`
- `Department` containing the value `RH`

In CES, the custom field `CorpDepartment` is bound to the metadata `Department` and its **Free Text Queries** attribute is selected.

When the **Index the document's metadata** option is cleared, searching for `RH` returns the document because a field is indexing this value. Searching for `hector` does not return the document because no field is indexing this value.

When the **Index the document's metadata** option is selected, searching for `hector` also returns the document because CES indexed orphan metadata.

Document's addresses are case-sensitive

Determines if, when two addresses differ only by their casing, they are considered as two different addresses. This check box is cleared by default.

Generate a cached HTML version of indexed documents

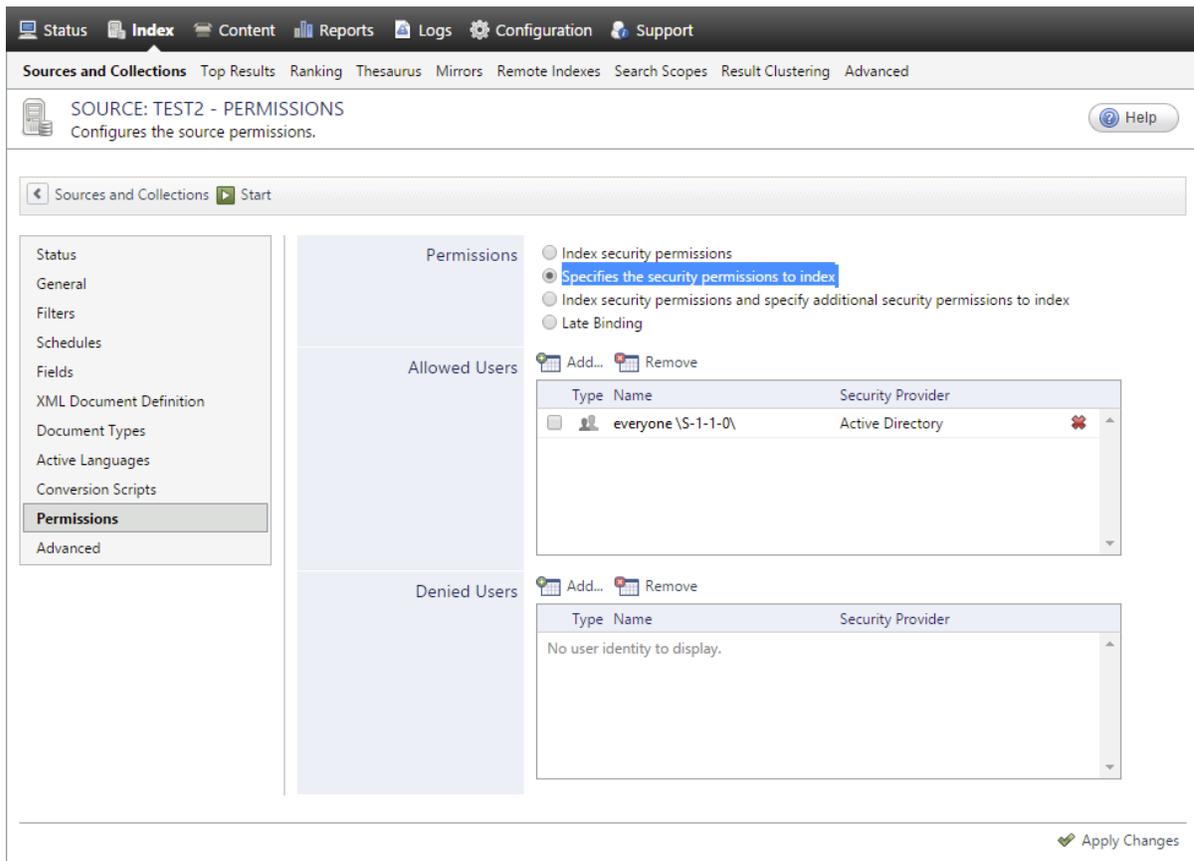
At indexing time, CES creates HTML versions of indexed documents. In search interfaces, users can then more quickly review the content by clicking the **Quick View** link rather than opening the original document with the original application. By default, this check box is selected. Consider clearing this check box only when you neither want to use **Quick View** links nor save resources when indexing the source.

Open results with cached version

Leave this check box cleared (recommended) so that, in search interfaces, the main search result link opens the original document with the original application. Consider selecting this check box only when you do not want users to be able to open the original document but only see the HTML version of the document as a **Quick View**. In this case, you must also select **Generate a cached HTML version of indexed documents**.

7. Click **Save** to save the source configuration. You are brought back to the **Source** page.
8. In the **Permissions** section of the **Source** page, select **Specifies the security permissions to index** and click

Apply Changes. You are brought back to the **Source** page.



9. When you are ready to start indexing the Twitter v2 source, click **Start**.
10. Validate that the source building process is executed without errors:
 - In the navigation panel on the left, click **Status**, and then validate that indexing proceeds without errors.
 - OR
 - Open the CES Console to monitor source building activities.

What's Next?

- Set an incremental refresh schedule for your source.
- Consider modifying some hidden source parameters to try resolving other issues (see ["Modifying Hidden Twitter v2 Source Parameters"](#) on page 16).

5.1 Adding an Explicit Connector Parameter

Connector parameters applying to all sources indexed using this connector are called explicit parameters.

When you create or configure a source, the Coveo Enterprise Search (CES) 7.0 Administration Tool presents parameters with which you can configure the connector for most setups. For many connectors, more advanced and

more rarely used parameters also exist but are hidden by default. CES then uses the default value associated with each of these hidden parameters.

You can however choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value.

To add an explicit connector parameter

1. On the Coveo server, access the Administration Tool.
2. Select **Configuration > Connectors**.
3. In the list on the **Connectors** page, select the connector for which you want to show advanced hidden parameters.
4. In the **Parameters** section of the selected connector page, click **Add Parameter** for each hidden parameter that you want to modify.

Note: The **Add Parameter** button is present only when hidden parameters are available for the selected connector.

5. In the **Modify the parameters of the connector** page:

The screenshot shows the 'MODIFY THE PARAMETERS OF THE CONNECTOR' page. The breadcrumb is 'Connectors - Active...'. The form fields are:

- Type: String
- Name: [Text Input]
- Default Value: [Text Input]
- Label: [Text Input]
- Quick Help: [Text Input]
- Option:
 - Optional parameter
 - Sensitive information
 - Validate as an email address
- Maximum length: [Text Input]

Buttons: Save, Cancel

- a. In the **Type** list, select the parameter type as specified in the parameter description.
- b. In the **Name** box, type the parameter name exactly as it appears in the parameter description. Parameter names are case sensitive.

- c. In the **Default Value** box, enter the default value specified in the parameter description.

Important: Do not set the value that you want to use for a specific source. The value that you enter here will be used for all sources defined using this connector so it must be set to the recommended default value. You will be able to change the value for each source later, in the **Add Source** and **Source: ... General** pages of the Administration Tool.

- d. In the **Label** box, enter the label that you want to see for this parameter.

Example: To easily link the label to the hidden parameter, you can simply use the parameter name, and if applicable, insert spaces between concatenated words. For the **BatchSize** hidden parameter, enter `Batch Size` for the label.

Note: To create multilingual labels and quick help messages, use the following syntax: `<@ln>text</@>`, where *ln* is replaced by the language initials—the languages of the Administration Tool are English (en) and French (fr).

Example: `<@fr>Chemin d'accès du fichier de configuration</@><@en>Configuration File Path</@>` is a label which is displayed differently in the French and English versions of the Administration Tool.

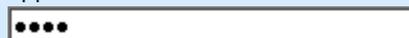
Tip: The language of the Administration Tool can be modified by pressing the following key combination: `Ctrl+Alt+Page Up`.

- e. Optionally, in **Quick Help**, enter the help text that you want to see for this parameter when clicking the question mark button  that will appear beside the parameter value.

Tip: Copy and paste key elements of the parameter description.

- f. When **Predefined values** is selected in the **Type** parameter, in the **Value** box that appears, enter the parameter values that you want to see available in the drop-down parameter that will appear in the Administration Tool interface. Enter one value per line. The entered values must exactly match the values listed in the hidden parameter description.
- g. Select the **Optional parameter** check box when you want to identify this parameter as an optional parameter. When cleared, CES does not allow you to save changes when the parameter is empty. This parameter does not appear for **Boolean** and **Predefined values** parameter types.
- h. Select the **Sensitive information** check box for password or other sensitive parameter so that, in the Administration Tool pages where the parameter appears, the typed characters appear as dots to mask them. This parameter appears only for the **String** type.

Example: When you select the **Sensitive information** check box for a parameter, the characters typed appear as follows in the text box:



- i. Select the **Validate as an email address** check box when you want CES to validate that the text string that a user enters in this parameter respects the format of a valid email address. This parameter appears only for the **String** type.

- j. In the **Maximum length** box, enter the maximum number of characters for the string. This parameter appears only for the **String** type. When you enter 0, the length of the string is not limited.
 - k. Click **Save**.
6. Back in the **Connector** page, click **Apply Changes**.

The hidden parameter now appears in the **Add Source** and **Source: ... General** pages of the Administration Tool for the selected source. You can change the parameter value from these pages. Refer to the documentation for each connector for details.

Note: When you want to modify a hidden source parameter, you must first delete it, and then redefine it with the modified values.

5.2 Modifying Hidden Twitter v2 Source Parameters

The **Add Source** and **Source: ... General** pages of the Administration Tool present the parameters with which you can configure the Twitter v2 connector for most setups. More advanced, or seldom used parameters, are hidden. You can choose to make one or more of these parameters appear in the **Add Source** and **Source: ... General** pages of the Administration Tool so that you can change their default value. Consider changing hidden parameter values when you encounter issues, or when you want to fine tune connector performance.

The following list describes the advanced hidden parameters available with Twitter v2 sources. The parameter type (integer, string, etc.) appears between parentheses following the parameter name.

NumberOfRetries (Integer)

The maximum number of times a failing request will be retried. The default value is 5.

RequestTimeout (Integer)

The maximum amount of time (in seconds) an HTTP request can be executed before being canceled. The default value is 60.

RetrieveProtectedUsers (true, false)

Whether or not to retrieve protected users. If this parameter is set to `true`, protected users and their Tweets will be accessible to everyone. The default value is `false`.

FollowedUsersPerPage (Integer)

The number of users per page when retrieving users followed by a user, between 1 and 200. The default value is 200.

NumberOfUsersPerPage (Integer)

The number of users per page when retrieving users from a list, between 1 and 5000. The default value is 5000.

TweetsPerTimelinePage (Integer)

The number of Tweets per page when retrieving user timelines, between 1 and 200. The default value is 200.

TweetsPerSearchPage (Integer)

The number of Tweets per page when retrieving Tweets from a search query, between 1 and 100. The default

value is 100.

TweetsPerCollectionPage (Integer)

The number of Tweets per page when retrieving Tweets from a collection, between 1 and 200. The default value is 200.

SearchResultsType (recent, mixed, popular)

The type of results retrieved with a search query, including mention Tweets. Can be either recent, mixed or popular. The default value is `mixed`.

Note: Popular usually returns around 15 Tweets.

SearchResultsIsoLanguageCode

The [ISO 639-1 language code](#) to filter Tweets retrieved with a search query, including mention Tweets (does not work when retrieving Tweets from a user). The default value is `none`.

To modify hidden Twitter v2 source parameters

1. Refer to "[Adding an Explicit Connector Parameter](#)" on page 13 to add one or more Twitter v2 source parameters.
2. To modify parameter values, from the Administration Tool:
 - *For a new Twitter v2 source:*
 - a. Select **Index > Sources and Collections**.
 - b. Under **Collections**, select the collection in which you want to add the source.
 - c. Under **Sources**, click **Add**.
 - d. In the **Add Source** page, edit the newly added advanced parameter value.
 - *For an existing Twitter v2 source:*
 - a. Select **Index > Sources and Collections**.
 - b. Under **Collections**, select the collection containing the source you want to modify.
 - c. Under **Sources**, click the existing Twitter source in which you want to modify the newly added advanced parameter.
 - d. In the **Source: ... General** page, edit the newly added advanced parameter value.
3. Rebuild your Twitter v2 source to apply the changes to the parameters.