Twitter Introduction

The Twitter connector will crawl content from any twitter account.

The Twitter connector is a crawler developed using the Twitter Developer Platform for tweets discovery but relies on the Aspire 3 Connector Framework to handle connections and distributed crawls.

On this page

- Features
- Content Retrieved
- Limitations

Features

Some of the features of the Twitter connector include:

- Authentication using twitter user, consumer key and consumer secret key
- Incremental crawl
- Full crawl

Content Retrieved

The Twitter connector retrieves all tweets related to the twitter user specified. Listed below is some example content information of different tweet types that can be retrieved by this crawler.

- Text tweet
- URL links
- Geo location
- Hashtags
- User mentions entities
- Media entities
- Retweet count

ⓘ This crawler will retrieve any document found linked in the HTML Markup as links (such as PDFs, MS Word, MS PowerPoint, etc).

Limitations

Due to the design implementation, the Twitter connector has the following limitations:

- To retrieve tweets from Twitter, the implementation is using application-only authentication with the getUserTimeline method to extract user tweets.
  For reference, see https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html

  This method can only return up to 3,200 of a user's most recent Tweets.

  Requests / 15-min window (app auth): 1500

  API Calls: ~16-18

  Accounts: ~80

- Once you publish a Twitter update (called a tweet), there is no way to edit it. Your only options are to delete it entirely or to copy the offending tweet before deleting it and then repost a revised version of the tweet.

Anything we should add? Please let us know.