8kun

Table of Contents

About the Platform:

8kun (see: https://8kun.top/index.html), previously named 8chan, infinitechan, and infinitychan, is an anonymous online imageboard. 8kun is referred to as an imageboard which "lets you post text and images anonymously (without a name) about any topic or no topic at all. Unlike forums, imageboards do not hold or store old content permanently and old threads are pruned as new ones are created" (8kun.top). 8kun has been linked

to hate groups such as white supremacism, neo-Nazism, hate crimes, and multiple mass shootings (Mezzofiore & O'Sullivan, 2019; Roose, 2019; Wong, 2019). On 8kun, there are boards containing threads where users can either make a new thread under a particular board or reply to a preexisting thread. "On most boards, you are required to upload an image, but if you do not have one, you can also draw a picture" (8kun.top). Each board has individual owners who moderate each board with little interaction from the administration of 8kun. Similarly, individual boards have their own rules. For example, on the board named Coronavirus General #57, the following rules apply: "Off-topic, bait/spam/feeding bait, hoaxnigger, one liner, kiketube embeds subject to deletion at discretion. Thread is for virus related news, discussion and debate. Thread is not for: rehashing the same burger oil/politics, nuke x, survival, decon or PPE and /k/ related questions every bread - please take that to /cvg/ survival bread [thread]. All other coronavirus threads will be deleted. This isn't an indictment of the thread or any questions therein. It's just that we have a general for it" (8kun.top/pnd). The Coronavirus General #57 thread can be found under the politics, news, and debate board on the homepage.

1) What data is available from this platform?

This platform will allow us to collect data on content (actual posts to the boards, including threads and comments) and timestamps. There are 20,000 boards on 8kun. For the project's purpose, we will use COVID-19 related keywords to return a list of boards. This will not provide a full list of boards because, for example, 8ch.net (8kun.top) has about 20,000 boards, so getBoards() will return only the top 20 boards list. The data will return "true" if an imageboard has a "get *all* boards" API endpoint that's distinctive from the regular "get boards" API endpoint.

2) What are the restrictions/limitations of the data?

The major limitation is not being able to access historical data as the site does not store old threads. In fact, old threads are removed as new ones are created.

3) What variables are available? What meta-data (i.e., the data of the data) is included? For example, what information is available about user profiles (e.g., username, demographic data, number of followers/friends)? What information is linked to the content/post/blog/tweet (e.g., timestamps, re-tweets, location, device/platform that content was shared from, number of likes)?

- imageboard – A board where content is posted
- comments – Responses to a board post (With a comment limit)
- timestamps – Date and time of the post
- id – This is available to identify users. Most users post anonymously, but this number provides a way to pull the data and associate it with a generated ID number.

- title – The title of the board
- createdAt – The time that the post was created
- commentsCount – The total comments received on a board/thread
- attachmentsCount – The total attachments added to a board/thread
- comments – The content of the comments
- replies – The number of replies to a post
- attachments – These are video and picture attachments. This variable gives you access to the actual attachments.

4) Is there an Application Programming Interface (API) available for this platform? If yes, please provide the link for the platform.

We have located three separate APIs for collecting data from 8kun. Our team went with the first API for this project, as it proved to be more thoroughly documented and intuitive than the other two.

1. **A Python wrapper for 8kun on GitHub (bibanon/py8chan)**:
   https://github.com/bibanon/py8chan

2. **A JavaScript wrapper for 8kun on GitHub (catamphetamine/imageboard)**:
   https://github.com/catamphetamine/imageboard

3. **The Socrata Open Data API for 8kun**:
   https://dev.socrata.com/foundry/internal.open.piercecountywa.gov/r76p-8kun

We also provide additional links that we think this research audience may find helpful:

1. https://github.com/OpenIB/OpenIB – OpenIB is a security fork of the imageboard Infinity focused on user security.

2. https://gitlab.com/catamphetamine/captchan – This is an imageboard client by @catamphetamine.

3. https://captchan.surge.sh/8ch – This offers an alternative use of the 8kun site.

5) Does the API provide a way to limit the search based on geolocation?

There may be a way to search data based on "country tags," but this was not clear from the API documentation.

6) Does the platform use any kind of filters to remove the content on their sites?

Although our team utilized the py8chan Python library for accessing the 8chan API, GitHub user @catamphetamine's wrapper for the imageboards' API offers a solution for censoring certain words contained within imageboard comment titles and overall content that may be deemed offensive to users. A censoredWords option can be passed to the imageboard function to censor these particular words. Persons using this functionality have the option to:

- **Filter a string** (i.e., "A lowercase two-letter language code used to generate a regular expression for splitting text into individual words" such as "en" for English).

- **Filter an array of string word patterns using standard regular expression syntax** (i.e., "^ meaning word start, $ meaning word end, etc."). "The patterns are applied to each individual word and if there's a match then the whole word is censored."

Censored words identified in a parsed comment's content are ultimately replaced with: "{ type: spoiler, censored: true, content: "the-word-that-got-censored" }". Censored words identified in a parsed comment's title do not result in its replacement, but instead, a new titleCensored property is generated for the word(s). This is done to ensure that title is maintained as a string data type, unlike content which is edited in-place.

References

8kun. (2020). Retrieved October 22, 2020, from https://8kun.top/index.html.

bibanon. (2020). *bibanon/py8chan*. GitHub. https://github.com/bibanon/py8chan.

Github. (2020). Retrieved October 22, 2020, from
https://github.com/catamphetamine/imageboard.

Mezzofiore, Gianluca; O'Sullivan, Donie (August 5, 2019). "El Paso shooting is at least
the third atrocity linked to 8chan this year". CNN. Archived from the original on
August 20, 2019. Retrieved August 5, 2019.

Roose, Kevin (August 4, 2019). "8chan Is a Megaphone for Gunmen. 'Shut the Site
Down,' Says Its Creator". The New York Times. ISSN 0362-4331. Archived from
the original on August 5, 2019. Retrieved August 5, 2019.

Socrata. (2020). Retrieved October 22, 2020, from
https://dev.socrata.com/foundry/internal.open.piercecountywa.gov/r76p-8kun.

Wong, Julia Carrie (August 4, 2019). "8chan: the far-right website linked to the rise in
hate crimes". The Guardian. ISSN 0261-3077. Archived from the original on
August 21, 2019. Retrieved August 5, 2019.