4chan

Table of Contents

About the Platform:

4chan (see: https://www.4chan.org/) is a popular message board with a dedicated community of anonymous users, including Internet activists known as Anonymous (Knuttila, 2011). This platform was created in October of 2003 and contains 49 themed boards in the English language. No log-ins are required, thus, most users post under the default name of anonymous. Some sources indicate there were 18 million unique site visits and 730 million page views a month (Tsotsis, 2011; Cha, 2010). Many themed boards focus on fandom cultures such as Japanese animation, literature, and photography. However, the site also hosts a random board where nearly half of the 800,000 daily posts can be found (Cha, 2010; Dibbell, 2010). The site has "a boards" for animation and hosts "b boards" for "everything else." The b boards are again what

account for 30% of the sites total traffic (Bilton. 2010). The social media site has 17 global rules and then specific rules for each board. These rules can be found here: https://www.4channel.org/rules. While the site is well known for its creative memes, it likewise has a reputation for hate speech, hacking, harassment, violence, and terrorism. What sets this platform apart from other social media sites is most others are focused on sharing aspects of one's personal identity while 4chan emphasizes the importance of posting anonymously.

Figure 1. An example of 4chan posts from the politically incorrect board (/po/).

1) What data is available from this platform?

Users' posts, usernames, dates of posts, serial numbers for posts, comments, photos, gifs, message board threads, page numbers and quotes are all types of data that can be collected from 4chan. Each thread can be up to 10 pages. From the API, researchers can collect user images, country flags, thumbnails, posts, user identification numbers, page number, number of replies, etc. The full list of available data is provided in the Data Directory section below.

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

First, data that has been deleted or lost cannot be retrieved from 4chan.org. In response to a frequently asked question, "Can I retrieve an old post/image?" The site responds with, **"**No. Content that has expired is removed from our servers. Please do not contact us requesting content that has been pruned, for it is impossible for us to assist you. Instead, try placing a request on the /r/ board."

Second, user identity information cannot be retrieved. In response to the statement, "Can't register a username." The site replies that only site administrators of 4chan can view personal IP addresses and these will only be shared "under a court order or from law enforcement."

The API rules are that you cannot make more than one request per second, the thread updating should be set to a minimum of 10 seconds, use if-modified-since when doing requests, and make API requests using the same protocols as the app. There are also 6 terms of services for using the API. This information can be found here: https://github.com/4chan/4chan-API.

Though the site does not explicitly suggest a rate limit, our team used a small number of requests and timeouts as recommended by the API.

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)?

4chan has users' posts, a username, date of post, a serial number for the post, comments, photos, gifs, message board threads, page number and quotes. Each thread goes up to 10 pages and the oldest posts are deleted in a couple of days. Researchers can also collect data on boards and catalogues, all of which are listed below in the Data Directory section.
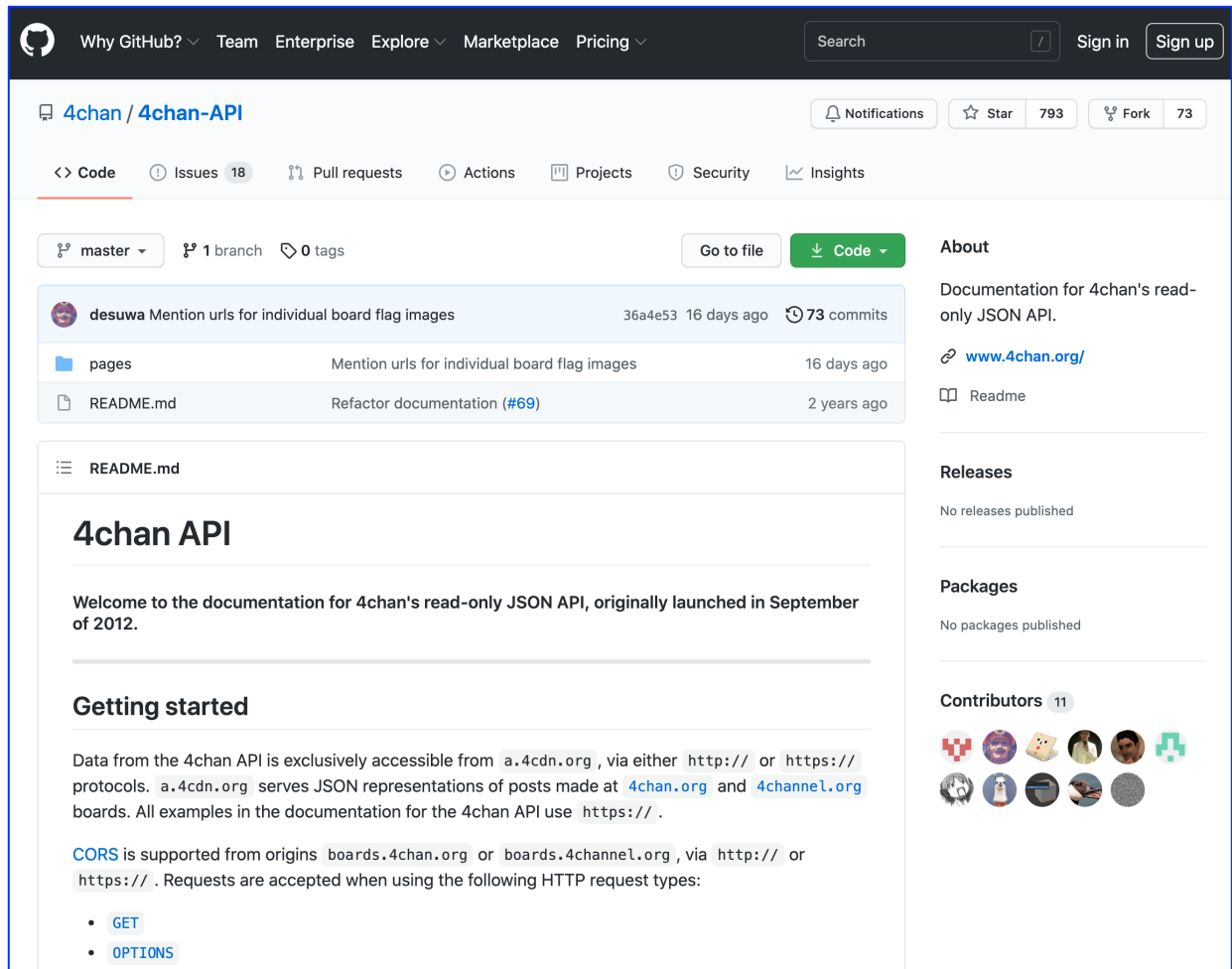
Figure 2. The GitHub repository for the 4chan API.

Data Dictionary

(See: https://github.com/4chan/4chan-API)

Boards

- board: where the board is, type-string; the board can include multiple threads and each of those threads can contain comments.
- title: the title of the board, type-integer
- ws_board: whether or not the board is safe for workspace (e.g., no porn), type-integer
- per_page: threads on one page, type-integer
- pages: how many pages on a board, type-integer
- max_filesize: non.webm attachments, max size allowed, KB, type-integer
- max_webm_filesize: .webm attachments, KB, type-integer
- max_comment_chars: total amount of characters allowed in a comment on a post, type-integer

- max_webm_duration: seconds, maximum amount of time for a .webm attachment, type-integer
- bump_limit: number of replies before the thread is bumped, bumping-when the comments are pruned or deleted, ensures fresh content, type-integer
- image_limit: number of image replies on a thread before the image replies are deleted, type-integer
- cooldowns: no definition, type-array
- meta_description: HTML that has a brief summary of the page, type-integer
- spoilers: whether or not a post has spoilers (e.g., spoiling the ending of a tv show, movie, book etc.,), type-integer
- custom_spoilers: how many custom spoilers a board has, type-integer
- is_archived: is a post archived, type-integer
- troll_flags: whether or not a thread has troll flags, troll-defined by urban dictionary as, "Someone who deliberately pisses people off online to get a reaction" type-integer
- country_flags: whether or not the poster's country is available on the board, type-integer
- user_id: is the user id tag available on the board, type-integer
- oekaki: whether the user can use the app oekaki to submit a drawing, type-integer
- sjis_tags: (Shift Japanese Industrial Standards), used for the Japanese language, can users use the sjis tag to submit sjis drawings, type-integer
- code_tags: board support, under the code tag, type-integer
- math_tags: support for math TeX (a math formula thing), and eqn tag (probably equation tag), type-integer
- text_only: can you post an image? type-integer
- forced_anon: are names disabled on this board? type-integer
- webm_audio: is audio allowed? type-integer
- require_subject: does the original poster (OP) require a subject for the board? type-integer
- min_image_width: pixels, how small can an image be, type-integer
- min_image_height: pixels, how short can an image be, type-integer

Catalogues

- no: post identification, type-integer
- resto: for the OP-this is 0, this is the id of the thread being replied to, type-integer
- sticky: if the thread is pinned to the top of the page, type-integer
- closed: if the thread doesn't allow replies, type-integer
- now: date down to the minute, type-string
- time: when the post was created, type-integer
- name: name on post, default is anonymous, type-string
- trip: tripcode, used to determine if the post is authentic or not, format is !tripcode or !!securetripcode, type-string

- id: posters id, 8 characters, type-string
- capcode: identifier for a post, type-string
- country: ISO 3166-1 alpha-2 country code, type-string
- country_name: name of country, type-string
- sub: original posters subject text, type-string
- com: comments, HTML, type-string
- tim: timestamp of post or microtime for image, type-integer
- filename: as appears on posters' device, type-string
- ext: file type, type-string
- fsize: bytes, filesize, type-integer
- md5: 24 characters, type-string
- w: width of image, type-integer
- h: height of image, type-integer
- tn_w: thumbnail width, type-integer
- tn_h: thumbnail height, type-integer
- filedeleted: was the file deleted? type-integer
- spoiler: Was the image a spoiler? type-integer
- custom_spoiler: id, type-integer
- omitted_posts: # of replies - # of previewed replies, type-integer
- omitted_images: # of image replies - # of previewed image replies, type-integer
- replies: replies on a thread, type-integer
- images: image replies on a thread, type-integer
- last_modified: time of last modification which includes post added, modified, deleted, thread closed, sticky setting modified, type-integer
- tag: .swf upload, type-string
- semantic_url: SEO URL for thread, type-string
- since4pass: when the year 4chan pass bought, type-integer
- unique_ips: # of unique posters in a thread, type-integer
- m_img: Mobile optimized image exists for post, type-integer
- last_replies: most recent replies, type-array

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

Yes, two APIs are available: an API for both 8kun and 4chan and an API that is exclusively for 4chan. The shared API is here: https://github.com/catamphetamine/imageboard and the API exclusive to 4chan is here: https://github.com/4chan/4chan-API.

We opted not to use the shared API because it was written in JavaScript. Our team preferred using the same programming language (i.e., Python) across all ten of our programming scripts for this project.

The 4chan-only API was a more attractive option, as it is compatible with multiple programming languages, thus allowing for use of Python. All outputs from the API are in JavaScript Object Notation (JSON).

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

The site offers "country flags," such as mentions or indicators of "place" which may offer a way to assess aspects of geolocation.

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

While the site does not offer a set of filters, it does have general rules for 4chan and more specific rules per board. The general rules include: refraining from doing things that are illegal by United States law, being above the age of 18, not spamming the site, and more. The comprehensive list of rules can be found here: https://www.4channel.org/rules.

References

Bilton, N. (2010, March 19). *One on One. Christopher Poole, Founder of 4chan (2019).* The New York Times. https://bits.blogs.nytimes.com/2010/03/19/one-on-one-christopher-poole-founder-of-4chan/

Cha, A. E. (2010, August 10). *4chan users seize Internet's power for mass disruption (2021).* The Washington Post. https://www.washingtonpost.com/wp-dyn/content/article/2010/08/09/AR2010080906102.html

Knuttila, L. (2011). User unknown: 4chan, anonymity and contingency. *First Monday*.

Tsotsis, A. (2011, May 25, 2011). *4chan has 18M a month, canvas participation is optional (2011).* TechCrunch.