

Fetching API

Variables

api.web.service.addr = the IP or HTTP address (with port) of our API web service

your.web.service.addr = the IP or HTTP address (with port or default 80) to send our callbacks to

Request for fetching

User programs can send us a request for fetching web pages by means of HTTP POST to api.web.service.addr/parse (e.g. <http://apiserp.com:88/parse>) containing a JSON such as the following:

```
{
  "ApiKey": "YourApiKeyHere",
  "CallbackUrl": "TheUrlAtWhichYouReceiveCallbacks",
  "Pages": [
    { "Src": "TheUrlYouRequestHere", "FileName": "TheFileNameToUseForThisPage",
    "UserAgent": "UserAgentStringToPassToTheWebsite", "Cookies": "Cookies":
    "CookieStringToPassToTheWebsite", "Locale": "LocaleStringToPassToTheWebsite" },
    { "Src": "https://www.bing.com/search?
    q=blowzier+Ted+conversationize+LDMTS+Biblicism+Vereshchagin+choux", "FileName":
    "0.html", "UserAgent": "Chrome", "Locale": "en-GB", "Cookies": "tool=curl; fun=yes;" }
  ]
}
```

Above, you can see a request for two pages, where in the first page we put a description and in the second page we put an example. You must not send us the description page: it is only there for your convenience. We recommend sending 500-1000 different pages per block to optimize your user experience.

Typically, the latency for a block is up to 10 minutes, depending on our success of querying the websites. So if you need more than 500-1000 pages fetched, you can send us several blocks one after another, each 500-1000 pages in size, because the throughput of our system is millions of pages per hour. We just depend on the other systems due to latency.

Our API web service responds with JSON with the following:

```
{ "BlockToken": "128934ygieh89y13" } or { "Error": "ErrorDetails" }
```

depending on whether your request is well-formed, contains a valid API key, your page balance allows sending such a request, whether our system was able to handle the request at this time or not, etc. In the latter case, you can retry your request in ~30 seconds or use our alternative IP/HTTP address in case our main IP/HTTP address is undergoing maintenance.

Callbacks

For callbacks, you should expose a web service your.web.service.addr/your/path e.g. <http://example.com:8000/fetcher/callback> that is able to handle a POST request without any CSRF token, with the following parameters in the POST body (please, note that if your script uses low-level HTTP, you should keep in mind that these parameters are URL-encoded):

BlockToken - the block token you received when sending us this fetching request

In the case that your script has handled the callback successfully, you must respond to us in plain text "OK" word (without quotes). Otherwise, we will try to deliver the callback to you two more times with a delay of 60 seconds.

Download of Blocks

After you have received a callback notifying that your block of pages is ready, you can download the block as a .zip archive using a GET request to:

api.web.service.addr/download?BlockToken="YourBlockTokenWithoutQuotes"

E.g. <http://apiserp.com:88/download?BlockToken=128934ygieh89y13>