

Getting started with the Kallo API

The Kallo API analyses the HTML of any website and returns structured data back via JSON. Simply send us the URL (to the /get-data-crawl endpoint) or raw HTML (to the /get-data endpoint). Raw html should be 'escaped' or `JSON.stringify(html_string)`. Deserialize the JSON response which contains the tabular data.

To request API Keys for beta access, send an email to: developer@kallo.io

We love feedback. Please send feedback or feature requests via email: feedback@kallo.io

For a visual explanation, view our [brochure](#)

Let's get into it

Note: All requests should have a header including your API key and content-type:

```
header = {
  'x-api-key': 'YOUR_API_KEY',
  'Content-Type': 'application/json' }
```

Base URL

API	Base URL	Version
Kallo API	https://api/kallo.io	v 0.1

Endpoint Summary

1. **/get-data** – get a table e.g. list of product names and prices, from raw HTML
2. **/get-data-crawl** – get a table e.g. list of product names and prices from any URL
3. **/find-element** – detectors to find elements e.g. search bar, from raw HTML

1. [/get-data endpoint](#)

Description

Get a list of items from a web page

Method	Endpoint
POST	https://api.kallo.io/get-data

Request Body

Request Key	Request Value(s)	Description
"action"	"get_items"	The get_items action returns a list of items in the page. Send it a Zillow page and it returns the list of house locations, prices, images and more
"html_string"	[HTML string compatible with json] <i>(see example below)</i>	This is a string of the raw HTML. It should have special characters escaped. E.g. <code>JSON.stringify</code> in JavaScript or <code>json.dumps</code> in Python
"url"	The URL of the html being sent (optional)	

Request Sample

```
{
  "action": "get_items",
  "html_string": "<!DOCTYPE HTML><html><head></head><body><div><div class=\"container\"><a class=\"first\">Item
                    name</a><h3 class=\"aria\">$4</h3></div><body></html>",
  "url": "https://www.sample.com"
}
```

Response

Response Key	Response Value(s)	Description
"item_found"	True or false	Boolean indicating whether useful data was found
"data"	Structured data returned within the JSON	An object of the form: <pre>{ columns: ["col 1", "col 2", "col 3"], index: [0, 1, 2, 3], data: [[r1c1, r1c2, r1c3], [r2c1, r2c2, r2c3], [r3c1, r3c2, r3c3], [r4c1, r4c2, r4c3]] }</pre> <i>Where col1, col2 and col3 are the columns in a table. Index 0-3 means there are 4 rows and each set of [r1c1, r1c2, r1c3] in the data represents a new row with elements of form row-N, column-N</i>
"error"	String describing the error. null if no error	Text string describing the error or null if there's no error

Response Sample

```
{
  "item_found": true,
  "data": "{ \"columns\": [\"Text_Field_1\", \"Text_Field_2\"], \"index\": [0,1,2,3,4,5,6], \"data\": [[\"Item name 1\", \"$4\"], [\"Item name 2\", \"$4\"], [\"Item name 3\", \"$4\"], [\"Item name 4\", \"$4\"], [\"Item name 4\", \"$10\"], [\"Item name 5\", \"$4\"], [\"Item name 6\", \"$4\"] ] }",
  "error": null
}
```

Full Example – Converting HTML to a Python Pandas DataFrame

```
url = 'https://api.kallo.io/get-data' # Edit to desired endpoint
payload={
  "action": "get_items",
  "html_string": raw_html_string, # whether escaped (with \" or \\\") or not escaped
  "url": "www.google.com"
}
my_header = {
  'x-api-key': 'YOUR_API_KEY',
  'Content-Type': 'application/json'
}
response = requests.post(url, json=payload, headers=my_header) # Use the 'json' argument, not 'data'

py_dict = json.loads(response.content) # convert response content to a dictionary in python. response.content is of bytes type
print(py_dict)
if py_dict['item_found'] and not py_dict['error']:
  structured_data_string = py_dict['data'] # Note that py_dict['data'] is still a string of the data we need. Needs conversion to dict
  structured_data = json.loads(structured_data_string)
  df = pd.DataFrame(data=structured_data['data'], columns=structured_data['columns'])
# Some lines expanded for understanding. In 5 commands, you get back a pandas DF from any HTML
```

2. /get-data-crawl endpoint

Description

Get a list of items from a URL. No need to provide raw HTML

Method	Endpoint
POST	https://api.kallo.io/get-data-crawl

Request Body

Request Key	Request Value(s)	Description
"action"	"get_items"	The get_items action returns a list of items in the page. Send it a Zillow URL and it returns the list of house locations, prices, links
	"get_element"	Returns the first occurrence of an element. Must also use the "playwright_locator" key below in your request
"url"	The URL of the html being sent	-
"playwright_locator" <i>(optional)</i>	A string used to find an element e.g. "text=Sign up"	Uses Playwright locators to find an element. Refer to these docs . Text, CSS, xpath locators are compatible
"wait_after_window_load" <i>(optional)</i>	Number of seconds to wait after the window.load event is fired	Integer value. Maximum of 5 seconds wait. Defaults to 2 if not supplied. <i>(Kallo may change this default in future)</i>

Request Sample

Ex 1:

```
{
  "action": "get_items",
  "url": "https://www.sample.com",
  "wait_after_window_load": 2
}
```

Ex 2:

```
{
  "action": "get_element",
  "url": "https://www.sample.com",
  "playwright_locator": "xpath=//html/body/div/article[0]",
  "wait_after_window_load": 2
}
```

Response

Response Key	Response Value(s)	Description
"item_found"	True or false	Boolean indicating whether useful data was found
"data"	Structured data returned within the JSON	An object of the form: <pre>{ columns: ["col 1", "col 2", "col 3"], index: [0, 1, 2, 3], data: [[r1c1, r1c2, r1c3], [r2c1, r2c2, r2c3], [r3c1, r3c2, r3c3], [r4c1, r4c2, r4c3]] }</pre> <i>Where col1, col2 and col3 are the columns in a table. Index 0-3 means there are 4 rows and each set of [r1c1, r1c2, r1c3] in the data represents a new row with elements of form row-N, column-N</i>
"error"	String describing the error. null if no error	Text string describing the error or null if there's no error

Response Sample

Ex 1:

```
{
  "item_found": true,
  "data": "{\\"columns\\": [\\"Text_Field_1\\", \\"Text_Field_2\\"], \\"index\\": [0,1,2,3,4,5,6], \\"data\\": [[\\"Item name 1\\", \\"$4\\"], [\\"Item name 2\\", \\"$4\\"], [\\"Item name 3\\", \\"$4\\"], [\\"Item name 4\\", \\"$4\\"], [\\"Item name 4\\", \\"$10\\"], [\\"Item name 5\\", \\"$4\\"], [\\"Item name 6\\", \\"$4\\"]]}",
  "error": null
}
```

Full Example – Converting a URL to a Python Pandas DataFrame

This is like the /get-data endpoint. However, you only need to specify an action and a URL

```
url = 'https://api.kallo.io/get-data-crawl' # Edit to desired endpoint
payload={
  "action": "get_items",
  "url": "www.google.com"
}
my_header = {
  'x-api-key': 'YOUR_API_KEY',
  'Content-Type': 'application/json'
}
response = requests.post(url, json=payload, headers=my_header) # Use the 'json' argument, not 'data'

py_dict = json.loads(response.content) # convert response content to a dictionary in python. response.content is of bytes type
print(py_dict)
if py_dict['item_found'] and not py_dict['error']:
  structured_data_string = py_dict['data'] # Note that py_dict['data'] is still a string of the data we need. Needs conversion to dict
  structured_data = json.loads(structured_data_string)
  df = pd.DataFrame(data=structured_data['data'], columns=structured_data['columns'])
# Some lines expanded for understanding. In 5 commands, you get back a pandas DF from any HTML
```

3. [/find-element](#)

Description

Gets a particular element extracted from raw HTML

Method	Endpoint
POST	https://api.kallo.io/find-element

Request Body

Request Key	Possible Request Value(s)	Description
"action"	"find_search_bar" "find_search_button" <i>(more detectors to come – email us)</i>	Specify what element you want to find in the HTML
"html_string"	[HTML string compatible with json] <i>(see example below)</i>	This is a string of the raw HTML. It should have special characters escaped. E.g. JSON.stringify in JavaScript or json.dumps in Python

"url"	The URL of the html being sent (optional)	-
-------	---	---

Request Sample

```
{
  "action": "find_search_bar",
  "html_string": "<!DOCTYPE HTML><html><head></head><body><div><div class=\"container\"><a class=\"first\">Item
                name</a><h3 class=\"aria\">$4</h3></div></body></html>",
  "url": "https://www.sample.com"
}
```

Response

Response Key	Response Value(s)	Description
"item_found"	True or false	Boolean indicating whether useful data was found
"element_html"	The returned element searched for	Raw HTML of the element searched for
"x_path"	XPATH of the element within the entire HTML string that was sent	-
"error"	String describing the error. null if no error	Text string describing the error or null if there's no error

Response Sample

```
{
  "item_found": true,
  "element_html": "<input id=\"search\" placeholder=\"search\"/>",
  "x_path": "",
  "error": null
}
```

For issues, bugs or technical assistance, reach out to: developer@kallo.io