

Getting ready

1. go to

https://openhatch.org/wiki/Boston_Python_Workshop_6/Friday

2. check whether you have:

- Python 2.7
- virtualenv (optional)
- pip
- then run `pip install mwclient`

3. open up Python and type `import mwclient`

Using web API client libraries to play with and learn from (meta)data

WikiConference USA 2014

Frances Hocutt

@franceshocutt

frances.hocutt@gmail.com

API = Application Programming Interface

web API = another way to get data from a website

web API client library = code that handles low-level interactions with the API so you don't have to

To get or post data through an API,
you want to know:

Where do I direct my request?

How do I write my request?

How do I send my request?

A client library should answer one or more of these questions for you. It should be easier to use than the plain API.

Client libraries for the MediaWiki API
are listed at [mw:API:Client code](#)

Commentary and a proposed standard
is at [mw:API talk:Client code](#)

Thinking about the MediaWiki API:

Endpoint: where you start. Stays the same as long as you're on the same wiki.

You send it parameters that specify:

- what actions you want to take
- what items you want to act on
- what data you want back
- what format you want the data in
- what you want to happen in special cases

Documentation doesn't always distinguish between types of parameter!

Example

“**query**”: the action you usually use to get data

Its parameters break down to:

- what items you want to act on: **titles, pageids, revids, generator** (choose one)
- what data you want back: **prop, meta, list, export** (choose as many as you want; these often have their own parameters)
- what format you want the data in, helpers: **format, indexpageids**
- what you want to happen in special cases: **continue, redirects, limits**

The API sandbox

API sandbox

Use this page to experiment with the **MediaWiki web service API**. Refer to the [API documentation](#) for further details of API usage. Example: [get the content of a Main Page](#). Select an action to see more examples.

Note that API actions that modify Wikipedia will in fact modify Wikipedia when used from this sandbox. For example, using the "edit" action will edit the specified page.

Parameters

Format json	Action query	Documentation links Returns all links from the given page(s). (read more)	Make request Examples Clear
-----------------------	------------------------	---	---

Parameter for links	Input	Description
pnamespace	(Main) Talk User User talk Wikipedia Wikipedia talk File File talk MediaWiki MediaWiki talk	Show links in this namespace(s) only

Useful information on **query**

Page types [[edit](#) | [edit source](#)]

Page Type	Example	Used in the given page(s)	Which pages have it	List all in the wiki
Page Link	[[Page]]	prop=links	list=backlinks	list=alllinks
Template transclusion	{{Template}}	prop=templates	list=embeddedin	list=alltransclusions
Categories	[[category:Cat]]	prop=categories	list=categorymembers	list=allcategories
Images	[[file:image.png]]	prop=images	list=imageusage	list=allimages
Language links	[[ru:Page]]	prop=langlinks	list=langbacklinks	
Interwiki links	[[meta:Page]]	prop=iwlinks	list=iwbacklinks	
URLs	http://mediawiki.org	prop=extlinks	list=exturlusage	

Take 1:

Using [API sandbox](#) + [/w/api.php](#) (main doc)

and [\[\[mw:API\]\]](#) and [\[\[mw:API:Query\]\]](#)

Let's get:

- the text (“content”) of a page
- a page's edit history
- a user's history of editing talk pages

Take 2:

Using [mwclient](#)/Python

on [\[\[API:Client code\]\]](#), still with [\[\[API:Main page\]\]](#)

Let's get:

- the text (“content”) of a page
- a page's edit history
- a user's history of editing talk pages

Acknowledgements:

Gnome OPW and the WMF
WikiConference USA 2014

Sumana Harihareswara

Tollef Fog Heen

Merlijn Van Deen

Brad Jorsch

Questions?

@franceshocutt

frances.hocutt@gmail.com

Previous “links” example

Putting it together:

<http://en.wikipedia.org/w/api.php>

+

?action=query&prop=links&plnamespace=0&
pllimit=max&title=Organic_chemistry

=

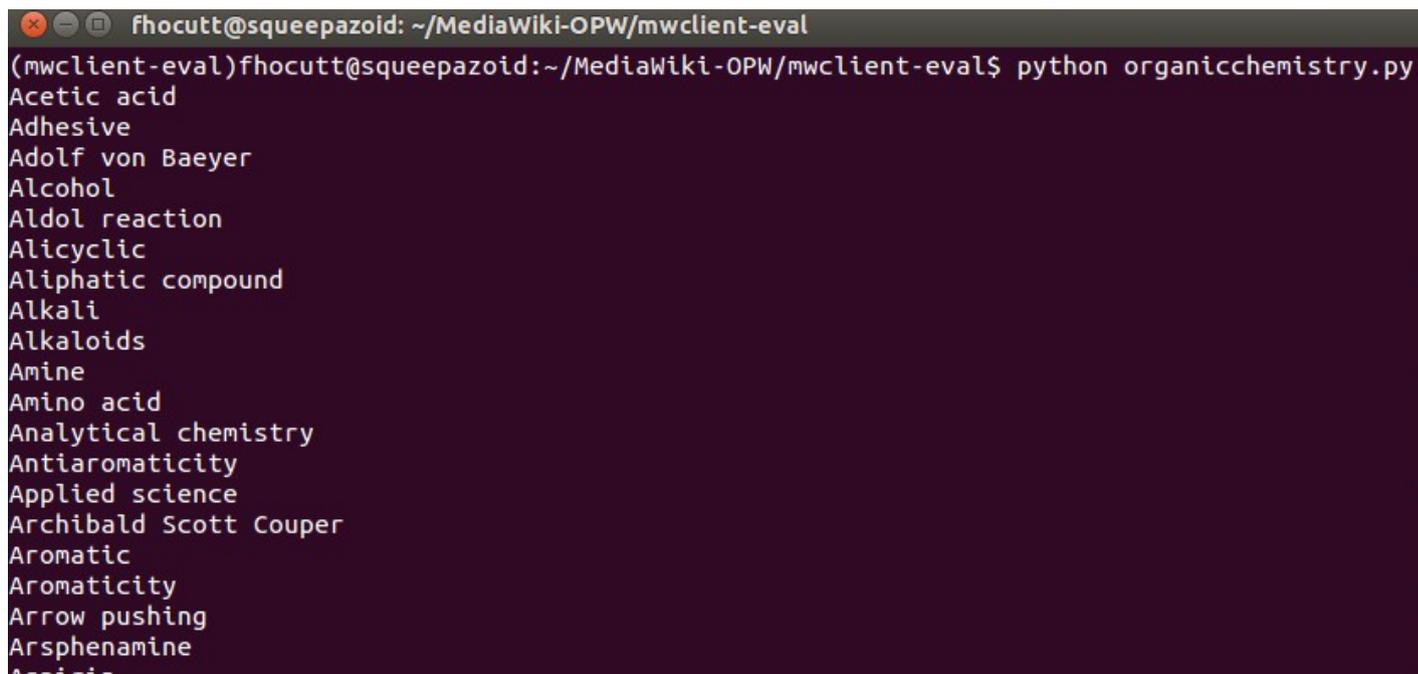
http://en.wikipedia.org/w/api.php?action=query&prop=links&plnamespace=0&pllimit=max&title=Organic_chemistry

Or use a client library:

```
import mwclient

wikipedia = mwclient.Site('en.wikipedia.org')
ochem = wikipedia.Pages['Organic chemistry']
links = ochem.links(generator=False)

for item in links:
    print item
```



```
fhocutt@squeepazoid: ~/MediaWiki-OPW/mwclient-eval
(mwclient-eval)fhocutt@squeepazoid:~/MediaWiki-OPW/mwclient-eval$ python organicchemistry.py
Acetic acid
Adhesive
Adolf von Baeyer
Alcohol
Aldol reaction
Alicyclic
Aliphatic compound
Alkali
Alkaloids
Amine
Amino acid
Analytical chemistry
Antiaromaticity
Applied science
Archibald Scott Couper
Aromatic
Aromaticity
Arrow pushing
Arsphenamine
Aspirin
```