

Summary:

Wayback machines API returns the url which includes the snapshot archived on the closest date to the date requested to the API. Since the requirement is to provide date and url for the archive of the website ("[pillpack.com](https://www.pillpack.com/)") nearest to the first of each month in 2016, I implemented a program with python which sends the request to Wayback machines API twelve times, from January to December in 2016.

Briefly, my code has four steps.

Step 1. Create a timestamp which is set to the first of a month in 2016 (01012016 ~ 12012016) and generate a url with the timestamp for an API call.

Step 2. Call the API with the url created in step 1 and retrieve the data returned from Wayback machines API.

Step 3. Parse and store an archived date and a url.

Iterate from step 1 to step 3 for twelve times for twelve months.

Step 4. Write a CSV file.

Below is the screenshot of the CSV file generated from my code.

date	archived_date	url
20160101	20151230	http://web.archive.org/web/20151230152834/https://www.pillpack.com/
20160201	20160130	http://web.archive.org/web/20160130000639/https://www.pillpack.com/
20160301	20160301	http://web.archive.org/web/20160301033409/https://www.pillpack.com/
20160401	20160320	http://web.archive.org/web/20160320185003/https://www.pillpack.com/
20160501	20160513	http://web.archive.org/web/20160513020746/https://www.pillpack.com/
20160601	20160603	http://web.archive.org/web/20160603013900/https://www.pillpack.com/
20160701	20160618	http://web.archive.org/web/20160618012248/https://www.pillpack.com/
20160801	20160618	http://web.archive.org/web/20160618012248/https://www.pillpack.com/
20160901	20161030	http://web.archive.org/web/20161030150241/https://www.pillpack.com/
20161001	20161030	http://web.archive.org/web/20161030150241/https://www.pillpack.com/
20161101	20161030	http://web.archive.org/web/20161030150241/https://www.pillpack.com/
20161201	20161203	http://web.archive.org/web/20161203125305/https://www.pillpack.com/

Figure 1. CSV file generated from the code

- date : requested date in the API call
- archived_date : archived date of the website which is the closest date to the requested date
- url : link to the archived snapshot in the Wayback Machine on the archived date