

I Tracked Everything I Read on the Internet for a Year

Published: 2022-09-08

Updated: 2022-09-08

Tags:

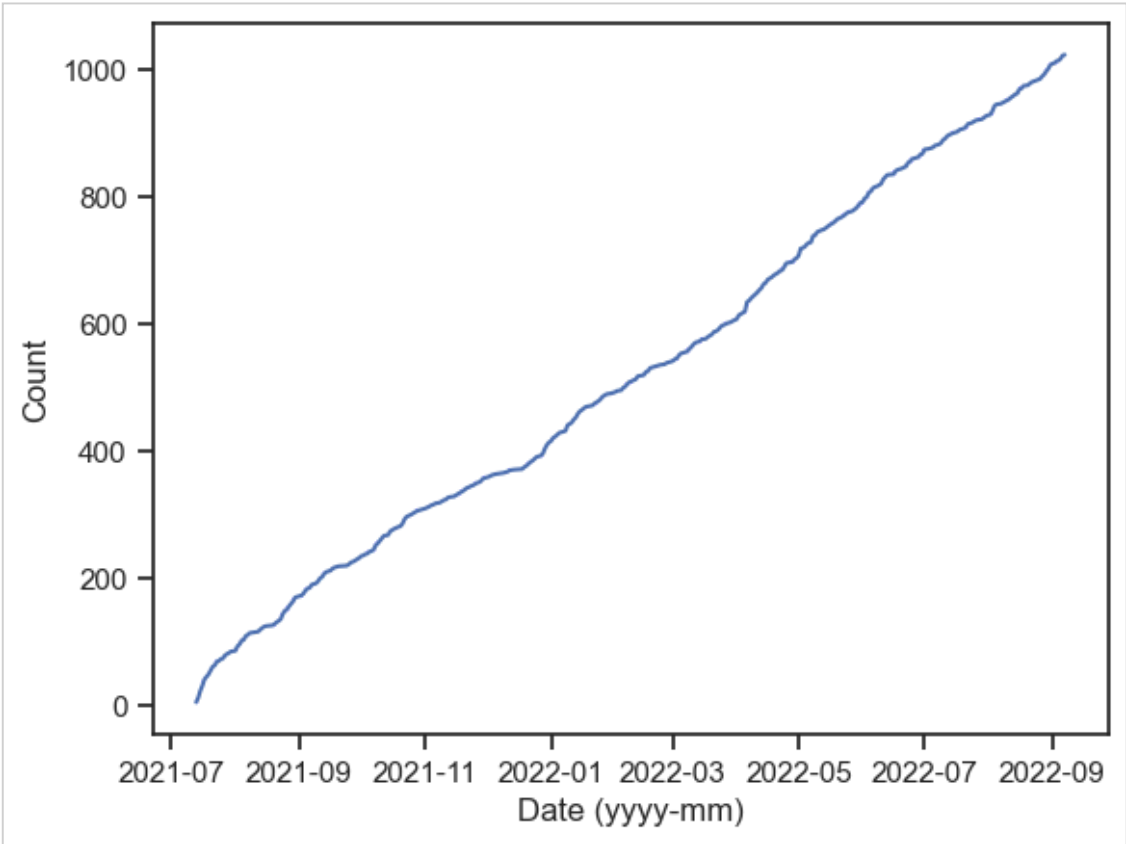
[website](#) [readingList](#) [favourite](#)

Since early July 2021, I've been keeping [a record of every article I read on the internet](#), which at the time of writing, is *just a few months* over a year's worth of articles.¹ Every article I read, I put on the list - and as it turns out, I read a fair amount. I thought I'd take a moment to do some stats, talk about how the list works and why I do it.

Some stats

At the time of writing, there are 1024 articles in my list. On average, I've read ~2.4 articles in each of the 421 days since I started the list. The majority of those articles come from the Hacker News digest email I get every morning or the long list of RSS feeds I subscribe to², but I also find articles in a smattering of other places, like Reddit or a news app of some sort.

There's nothing interesting about the graph of the number of articles I've read over time. It's a straight line.



On this page

[Some stats](#)

[How it works behind the scenes](#)

[The endpoint](#)

[The static site](#)

[But... why do all of this?](#)

Y

If you look closely, you can see where I hit my "personal best" of 15 articles in a day on 2022-04-06. My calendar says that day in particular was during a holiday, so I can only imagine the 15 articles was the result of a few lazy hours in bed one morning.

What's a little more interesting is the data about the sites I was actually reading.

So far, I've read pages on 596 unique websites.³ I only read articles on 116, or 19.46%, of those websites more than once.

The top three most-read websites in my list are:

1. the BBC (www.bbc.co.uk), 79 articles (!)
2. Xe Iaso's blog (xeiaso.net), 26 articles
3. Ars Technica (arstechnica.com), 24 articles

Some other notable mentions are rachelbythebay.com (9th, 9 articles), simonwillison.net (14th, 7 articles), and jvns.ca (16th, 6 articles).

I wish there was an easy way to filter by "independent websites", but there isn't. There's quite a lot of commercial domains on the top end of my most-visited list, I'd like to see a ranking of my most visited independent websites.

How it works behind the scenes

Somehow, I have to add an article to my reading list. The easiest way I came up with to do this was with a JavaScript bookmarklet.

It's pretty simple - it gathers the URL, page title and meta description and image, then modifies `window.location.href` to redirect me to an endpoint on this website. That endpoint processes the input data then redirects me back to the original article. If you're interested, [the source code can be found here](#).

It's done this way because some websites restrict what HTTP requests can be made with JavaScript on their site. I originally tried to use the `Fetch` (or was it `XMLHttpRequest`?) API to save the data, but quickly ran into the a dead-end and pivoted to using the current redirect-based method.

Since I use Firefox on both the desktop and my Android phone, my bookmarklet auto-syncs between my devices. It's fairly easy to add to my reading list on my phone, which is pretty handy as that's where I do a large portion of my reading.

The endpoint

All the data for my reading list is stored in a big CSV file in [a Git repository](#). It used to be that the endpoint on my website would clone the repository, make a new commit and push the changes, but now all of that is handled completely within GitHub Actions. [I wrote about making this change in my last post](#).

The static site

[The static site that displays the list](#) is generated using a small static site generator I wrote in Golang and deployed on GitHub Pages. When a new item is added to the list, the generator automatically runs and deploys using the magic of someone else's computer.

The static site uses a variation of this website's stylesheet, just with tweaked colours. Otherwise, everything is identical.

If you've so much as looked at the reading list site, you'd probably agree that it's a bit of a disaster. It's a single *huge* page with every entry in it, making it difficult to navigate and near impossible to search. I'd like to fix this - at some point, I'm going to enhance the reading list to split the content over a few pages and attempt to add a client-side search, but that's a project for the future.




But... why do all of this?

Honestly? I don't remember why I started doing this in the first place, but I certainly like having it around.

Often I'll find myself wanting to share an article I read a few weeks ago with someone, and with my reading list, I can do that with a quick Ctrl+F. It's really handy, and for that reason alone, I'm going to keep doing it. There's not really another, practical use for it.

That said, having a list of all the articles I've read and being able to fiddle around with that data whenever I want to very much satisfies my inner nerd.

I guess I keep up with my reading list, in part, because I enjoy it.

-
1. May it be known that I had every intention to write this article earlier than I eventually did. 
 2. These are also served to me in email form. I get an email every morning which is generated by [Walrss](#), an AGPL-licensed, Docker-ready app that you can self-host if you really want to. 
 3. [The graph of unique domains over time](#) is also an uninteresting, near-straight line 
-

This article is referenced elsewhere on the internet:

- [Hacker News](#)
- [Automating a Reading List](#) :: zanshin.net

Thoughts? Corrections? Questions?

Comment via email!

