

Bus trips and building

Author : Tim Sherratt

Date : January 24, 2013

Last week I took my daughter to Sydney so she could attend a girls-only Minecraft workshop at the Powerhouse Museum (they created [some wonderful things](#)). It was a 3½ bus journey each way, so to keep myself occupied I set myself the challenge of trying to build something en route. I made a fair bit of progress, but ultimately failed. I had to steal a few extra hours this week to get it to the point where people might find it useful.

The Australian WWI Records Finder

So here it is — a (sort of) [aggregated search interface](#) to records about Australian First World War service personnel. Give it a name and it will search:

- [service records](#) held by the National Archives of Australia
- the [Roll of Honour](#) at the Australian War Memorial
- the [First World War Embarkation Roll](#) at the Australian War Memorial
- the [Red Cross Missing and Wounded files](#) at the Australian War Memorial
- the [Honours and Awards database](#) at the Australian War Memorial
- the [Commonwealth War Graves Commission database](#)

It's 'sort-of' aggregated because it's really just a series of separate searches presented on the one page. But even this should make it easier for people to match up records across the different data sets.

Using

Type in a family name and, optionally, a given name or a service number. Hit search. Wait. Wait a bit more. The National Archives' RecordSearch database can often be pretty slow. Eventually though, each of the databases will be queried in turn and the results added to the page.

Once the results have loaded, click on a title and the little spinny thing will start up again as more details are retrieved from the database. In this 'detail' view, all the other results from the database are hidden. This makes it a bit easier to compare records across databases. Just click on the title again to go back to the 'list' view.

If your search returns lots of results, you can use the 'next' and 'previous' links to explore the complete set. They'll all load in the current page via the magic of AJAX.

It's not obvious from the interface, but you can feed query parameters directly via the url. For example try http://wraggelabs.com/ww1-records/?family_name=wragge. Why is this useful? Perhaps you've got your own database of names on the web. Using this you could easily create links from each name that looked for relevant records in the Finder.

That's about it. It's just a quick, bus-trip-inspired experiment, so there are many limitations and future possibilities.

Limitations

I'm just using the standard search interfaces of the various databases and screen-scraping the results. Unfortunately they all work slightly differently. For example, the AWM databases don't distinguish between family names and given names, so if you search for the family name 'Smith' you'll also get results like 'Jones, Bruce Smith'. The CWGC database, on the other hand, will only match an other name if it comes first, while RecordSearch (or more strictly NameSearch) will also match the names of next-of-kin. Fun fun fun.

I figure anything is better than nothing, but if you're not getting the results you expect head off to the original interfaces and try your luck there. I'm making no promises.

You'll also notice that the maximum number of results for each data source varies. The CWGC returns 15 results, while the AWM hands over a whopping 50. These are just the default settings for the original search engines. I could've fiddled with the settings, but it didn't really seem worth it.

And oh yeah... screen scraping... inherently fragile... might fall over and die at any minute.

Possibilities

As you may have guessed from previous posts, I rather like making connections. This experiment grew out of the work I'm doing on the ['Doing Our Bit'](#) project with the Mosman Library. I've been building a series of forms that will make it easy for contributors to link people in the Mosman project to any of these databases. Just paste in a url from RecordSearch and the system will automagically retrieve all the file metadata and also check for an entry in [Mapping our Anzacs](#). It's pretty nifty. But of course it made me think about having a way to search across all these different databases.

And then what?

Having found a series of records for an individual it would be good if they could then be permanently linked. If I had the time and money to do more work on this, I'd want to allow people to save the connections they find. And of course then expose these connections as Linked Open Data. It wouldn't be difficult.

There's probably also a lot more that could be done with machine matching of records. Perhaps someone's already working on this for the centenary — it seems like an obvious point of attack. It would be good if the forthcoming centenary commemorations resulted in something that brought

all these datasets together and exposed identifiers that could be easily used by community projects like 'Doing Our Bit'.

Details

Yes, I cheated. I had already done a lot of work on the screen-scrapery bits of this pre bus trip. I've been working a [RecordSearch client](#) on and off for a while to use with projects like [Invisible Australians](#). The [AWM](#) and [CWGC](#) scrapers I wrote for 'Doing Our Bit'. Feel free to grab the code and play.

The actual application was built using the Python micro-framework [Flask](#). I'm a big fan of Django, but there's a lot of overhead involved if you just want to throw together a simple app. I've been wanting to try Flask for a while and was pleased to find just how quick and fun it was to get something up and running.

To make the whole thing as responsive as possible, the search results are retrieved using AJAX calls to simple APIs I built in Flask on top of my screen scraper code. There's actually [very little code](#) in the Flask app itself. The downside of this is that the Javascript is a bit of a mess. Ah well.

Next

I don't know whether I can put any more time into this at the moment — too many other projects competing for my time and no more bus trips coming up. But if you think it's useful or worthwhile please let me know and I'll see what I can do.

At the very least it shows how with just a little impatience and ingenuity we can find fairly simple ways to integrate records from a variety of sources. We don't have to wait for some centralised solution.

Share this:

- [Click to email this to a friend \(Opens in new window\)](#)
- [Click to print \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Share on Facebook \(Opens in new window\)](#)
- [Click to share on Google+ \(Opens in new window\)](#)
-

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).