**Thesis idea: web tool for a ‘local government alert’**

**What:** a tool that searches local government City Council agendas, packets, minutes and letters to the council for a keyword that you provided, and then emails you the documents that contain that keyword. Could run daily, weekly or on push – searches the newest additions for you.

**For whom:** journalists who work for a local newspaper who have to cover multiple beats/cities and as such don’t have the time and resources to comb through all the city council documents manually. Experience with programming is not needed.

**How:** a website where you enter your keyword, the type of documents that you are interested in, the city or cities that you’re interested in, the frequency with which you want to receive alerts, and your email address.

**Existing tools:** the only other website that does something similar is Sqoop, but they are not focused on local government. They do have a nice ‘database search’ feature, but that is currently outside the scope of what I can accomplish in the next couple of months.

Sqoop’s tagline is “News before it’s news”. It allows journalists to receive email alerts when ‘new matching documents’ are published. The data sources they provide are SEC filings, patents, court dockets (from federal district courts, federal bankruptcy courts, courts of federal claims, courts of international trade, and partially circuit courts of appeals), press releases from the DOJ and USAO and ‘business stories’, provided by corporate users of the Sqoop platform.

Sqoop is, based on the type of documents they provided, mainly focused on business journalists who cover national/federal affairs. The website is also quite technical and requires some query language experience from the end users.

**What’s needed:** In order to build the script that would enable this tool to work, I need the following:

* Code that scrapes websites from local governments for their publicly available records (start with City Council agendas / Palo Alto + another city
* Code that downloads all the available public records and saves them in an archive (could possibly use the DocumentCloud API to do this)
* Code that searches all the documents (either only the newest additions or the complete archive) for the specified keyword (could use the DocumentCloud API to do this)
* Code that sends the requester the files that contain the requested keyword, on a daily/weekly/push basis (could use the WeTransfer API to do this)
* A web domain with a simple interface that allows the requester to specify a keyword, the type of documents, the city or cities, the frequency and their email address. (could be expanded with an ‘advanced search’ options that would allow the requester to only search in the title, metadata, or other elements)