

# Bay Area Municipal Code Repository

By Kamron Afshar & Michael Schulze

# Goals

The goals of this project were to set up a system that would update a database that contained all municipal codes in the bay area and identify changes to those codes over time.

These municipal codes were being stored across 4 multi-municipal websites and 4-5 other small or independent websites. Since there was no API offered for any of the websites our only option was to scrape the data out of these sites then collect that data as S3 documents and in a redshift database.

# Step 1: Locate the municipal codes

We found this website <https://igs.berkeley.edu/library/california-local-government-documents/codes-and-charters> that located almost all of the municipal codes in their respective websites so we scraped this information and put it in a table so we could identify which municipalities kept their code in which website.

There were four main websites that hosted the majority of the municipalities codes: municode.com, codepublishing.com, qcode.us, and amlegal.com. Each website used a different format for displaying the codes so we had to build four different scrapers, one for each website.

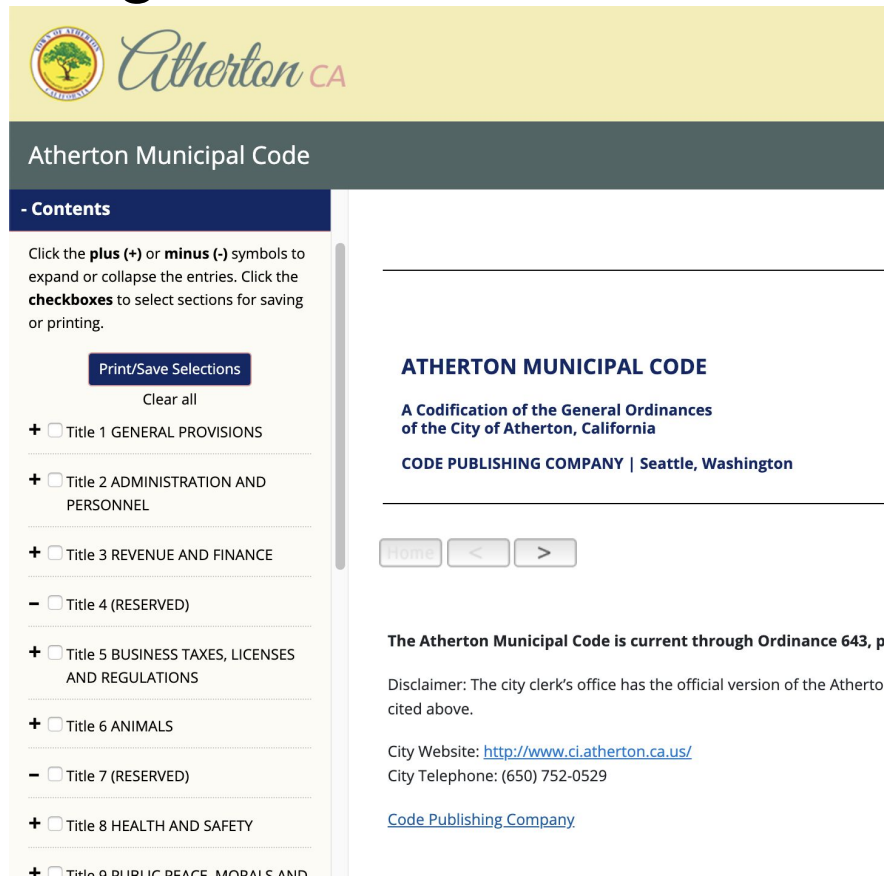
# Step 2a: Scraping municode.com

- Use Selenium to click through each document and extract text.
- Write level 2 documents to S3
- combining level 3 and lower docs into level 2.
- Only crawls page if listed update date has changed

The screenshot shows the official website for the City of Alameda, California, specifically the 'Code of Ordinances' section. The header includes the 'mc' logo and 'Alameda, CA'. A search bar is present in the top right. The left sidebar contains a navigation menu with icons for home, documents, and a list of ordinance categories. The main content area is titled 'Code of Ordinances' and includes 'Supplement 58'. It states that the online content was updated on April 30, 2020, and that the Alameda Municipal Code was codified through Ordinance No. 326. Below this, there is a section titled 'Adopted Ordinances Not Yet Codified' which includes a disclaimer that the listing includes all legislation received by the Municipal Code Commission. Two specific ordinances are listed: Ordinance No. 3279, adopted on 5/5/20, which amends the Alameda Municipal Code to include the Commission on Persons with Disabilities, and Ordinance No. 3278, also adopted on 5/5/20.

# Step 2b: Scraping codepublishing.com

- Check all the boxes then save the document to our base location.
- Parse document into the separate titles then save those titles (level 2) each as a document in S3.



The screenshot displays the Atherton Municipal Code website. At the top, there is a header with the city seal and the name "Atherton CA". Below this, a dark grey bar contains the text "Atherton Municipal Code". A blue sidebar on the left is titled "- Contents" and provides instructions on how to use the plus and minus symbols to expand or collapse entries, as well as checkboxes to select sections for saving or printing. It includes buttons for "Print/Save Selections" and "Clear all". The main content area on the right lists the titles of the municipal code, each with a plus sign and a checkbox. The titles are: Title 1 GENERAL PROVISIONS, Title 2 ADMINISTRATION AND PERSONNEL, Title 3 REVENUE AND FINANCE, Title 4 (RESERVED), Title 5 BUSINESS TAXES, LICENSES AND REGULATIONS, Title 6 ANIMALS, Title 7 (RESERVED), Title 8 HEALTH AND SAFETY, and Title 9 PUBLIC PEACE, MORALS AND... Below the title list, there is a section titled "ATHERTON MUNICIPAL CODE" which describes it as a codification of the general ordinances of the City of Atherton, California, published by the Code Publishing Company in Seattle, Washington. Navigation buttons for "Home", "<", and ">" are present. A disclaimer states that the city clerk's office has the official version of the code. The website URL and telephone number are also provided, along with a link to the Code Publishing Company.

**Atherton Municipal Code**

**ATHERTON MUNICIPAL CODE**  
A Codification of the General Ordinances  
of the City of Atherton, California  
CODE PUBLISHING COMPANY | Seattle, Washington

Home < >

The Atherton Municipal Code is current through Ordinance 643, p

Disclaimer: The city clerk's office has the official version of the Atherton  
cited above.

City Website: <http://www.ci.atherton.ca.us/>  
City Telephone: (650) 752-0529

[Code Publishing Company](#)

**Table of Contents:**

- + ☐ Title 1 GENERAL PROVISIONS
- + ☐ Title 2 ADMINISTRATION AND PERSONNEL
- + ☐ Title 3 REVENUE AND FINANCE
- ☐ Title 4 (RESERVED)
- + ☐ Title 5 BUSINESS TAXES, LICENSES AND REGULATIONS
- + ☐ Title 6 ANIMALS
- ☐ Title 7 (RESERVED)
- + ☐ Title 8 HEALTH AND SAFETY
- + ☐ Title 9 PUBLIC PEACE, MORALS AND...

# Step 2c: Scraping qcode.us

- Have to travel exhaustively through tree-like structure to get text data.
- Combine all that text data together as I crawl and add each title (level 2) as a separate document to S3

## Burlingame Municipal Code Burlingame, California



Current through Ordinance 1976 and the May 2020 code supplement. For more recent amendments to this code, please contact the city.

This document is provided for informational purposes only. Please read the full [disclaimer](#).

- [Municipal Code Table of Contents](#)
- [Search](#)
- [Ordinance List](#)
- [Statutory References](#)

### Links:

- [Burlingame, California - Home Page](#)

### Contact:

City Clerk: (650) 558-7203

## Burlingame Municipal Code

<a href="#">Title 1 GENERAL PROVISIONS</a>
<a href="#">Title 2 ADMINISTRATION</a>
<a href="#">Title 3 OFFICERS, BOARDS, COMMISSIONS AND PERSONNEL</a>
<a href="#">Title 4 REVENUE AND FINANCE</a>
<a href="#">Title 5 (RESERVED)</a>
<a href="#">Title 6 BUSINESS LICENSES AND REGULATIONS</a>
<a href="#">Title 7 (RESERVED)</a>
<a href="#">Title 8 HEALTH AND SANITATION</a>
<a href="#">Title 9 ANIMALS</a>
<a href="#">Title 10 PUBLIC PEACE, MORALS AND SAFETY</a>
<a href="#">Title 11 TREES AND VEGETATION</a>
<a href="#">Title 12 STREETS AND SIDEWALKS</a>
<a href="#">Title 13 VEHICLES AND TRAFFIC</a>
<a href="#">Title 14 (RESERVED)</a>
<a href="#">Title 15 WATER AND SEWERS</a>
<a href="#">Title 16 (RESERVED)</a>
<a href="#">Title 17 FIRE</a>
<a href="#">Title 18 BUILDING CONSTRUCTION</a>
<a href="#">Title 19 (RESERVED)</a>
<a href="#">Title 20 (RESERVED)</a>
<a href="#">Title 21 HISTORIC RESOURCE PRESERVATION</a>
<a href="#">Title 22 SIGNS</a>
<a href="#">Title 23 SWIMMING POOLS, HOT TUBS AND SPAS</a>
<a href="#">Title 24 RESERVED</a>
<a href="#">Title 25 ZONING</a>
<a href="#">Title 26 SUBDIVISIONS</a>

View the [mobile version](#).

# Step 2d: Scraping amlegal.com

- Have to expand and collapse the different “codes” (level 2) then select all parts of that level two so I can save the document to our base location.
- Upload each level 2 as its own document to S3.

The screenshot displays the American Legal Publishing Corporation website. At the top, the logo and name "AMERICAN LEGAL Publishing Corporation" are visible. Below this, a navigation bar shows "CA > San Francisco > Overview". The main content area is titled "San Francisco, California" and includes text about the current ordinance (57-20, File No. 191183) and the online version of the San Francisco Municipal Code. A sidebar on the left lists various codes under "San Francisco Overview", including Administrative Code, Business and Tax Regulations Code, Campaign and Governmental Conduct Code, Charter, Municipal Elections Code, Environment Code, Fire Code, Health Code, Park Code, Planning Code, Zoning Maps, Police Code, Port Code, and Public Works Code. At the bottom, there are buttons for "Order Code", "New Legislation", and "GET HELP", and a "Website" section with a "Back to Code Library" link and a "Previous Doc" link.

AMERICAN LEGAL  
Publishing Corporation

2020 S-71 (current)

CA > San Francisco > Overview

**San Francisco, California**

Current through Ordinance 57-20, File No. 191183, approved April 1, 2020.

This online version of the San Francisco Municipal Code is current as of the date of the last amendment.

The Transportation Code was last amended by Resolution No. 2019-01, approved November 1, 2019.

\*The Charter was last amended by provisions approved November 1, 2019.

\*\* The online version of the Building Inspection Commission Code is current as of the date of the last amendment.

Administrative Bulletins approved January 1, 2020.

[Order Code](#) [New Legislation](#) [GET HELP](#)

**Website**

[Back to Code Library](#) [Previous Doc](#)

## Step 3: S3 document

- Level 2 documents are written to S3 when crawled or pieced together
- Documents are identified by an S3 key or filename
- Once written, S3 keys are included in the redshift table to allow access



# Step 4: Redshift

- Columns:
  - Name of municipality
  - Date the code was updated
  - Title of the level 2 section
  - Did this document contain information on zoning
  - The calculated “diff” score
  - The S3 key that links to the level 2 document

## Step 5: Repeat

We need to run this executer script in order to update the data and the interval that this occurs at can depend on how frequently one expects the data to change. Our recommendation is to run this script monthly.

## Step 6: In the future...

- Use NLP text parsing to further subdivide our level 2 documents in S3 so we can narrow our text search.
- Reverse engineer scrapers so they travel to a designated subsection that the user wants to look at in the code.
- Create a more sophisticated text difference metric to better identify how much the document is actually changing.
- Do NLP analytics on the documents to discover interesting patterns in zoning policy or legal phrasing. How do different cities word and treat housing concerns
- Parallelize scrapers