

# Congressional District Geocoder

## Description

This algorithm converts the given **5-digit standard U.S. ZIP codes** into its congressional districts and geographical coordinates (latitude and longitude). Download the most recent version of the plugin [here](#).

## Pros & Cons

1. The algorithm is using a local database mapping with a large file size. It will increase the application size dramatically. So it is build as an external plugin
2. For first execution in the same application window, the plugin required 5 seconds to load the database. The consequent execution will not required the pre-loading phase.
3. A previous version of this plugin supported more accurate 9-digit ZIP+4 codes, but the version supporting the 113th Congress only supports 5-digit ZIP codes.
4. Congressional district might be varied by each election. The database would need to be maintained and updated on a regular basis.

## Applications

This plugin only support U.S. ZIP codes. It convert 5-digits ZIP codes to their belonging congressional district. It is an external plugin since the data size is so large. The dataset is based on the 2012 election (113th Congress).

## Implementation Details

*Words for developers: Please do take a look at the ZIP code wiki at [here](#) to have a better understand on how U.S. ZIP+4 code system works. The first 5-digits number in ZIP code is called Uzip. The last 4-digits number in the ZIP+4 code is Post Office box number which can refer to [here](#).*

The challenge of the implementation is the design of the mapping model that used to look up congressional districts from ZIP codes. To understand the metadata file (provided by GovTrack), create a mapping model with constant  $O(1)$  look up time and easy to managed. The implementation detail is documented in the source code.

The following will provide a high level view of the design.

1. The algorithm is facilitated by the Model-View-Controller idea
2. Model
  - The core of this implementation. Formed by *ZipCodeToDistrictMap*, *PostBoxToDistrictMap* and *DistrictRegistry*.
  - *ZipCodeToDistrictMap* hold a map of uzip to *USDistrict* and a map of uzip to *PostBoxToDistrictMap*.
  - *PostBoxToDistrictMap* hold a map of postBox to *USDistrict* and a map of wildcard to *USDistrict* map.
  - *DistrictRegistry* contains non-duplicated of *USDistrict* objects. It holds entire U.S congressional districts information.
  - *USDistrict* contains district label and geolocation. The class is imported from *edu.iu.scipolicy.model.geocode* package
3. View - *ZipToDistrictAlgorithmFactory* contains all the view setup implementation, including title, windows and options
4. Controller
  - *ZipToDistrictAlgorithm* prepares the model; parses the input ZIP codes to *USZIPCode* objects; performs the district look up, handles exceptions and saves the result to a CSV file.
  - The **Look up** is performed through *ZipCodeToDistrictMap*. If there isn't found a direct match of uzip to *USDistrict*, it will performed a look up through *PostBoxToDistrictMap* that holds by the uzip. Return *USDistrict* in success while throws *ZipToDistrictException* if no matched found
5. Dependency: *dist2geolocation.txt* and *zip4dist-prefix.txt*

The output table contains all columns of the input table with three new columns (Congressional district, latitude and longitude).

## Usage Hints

Here is a four steps guide to use the plugin:

1. Load your input data file that contains 5-digit U.S. ZIP codes to be geocoded.
2. Select **Analysis > Geospatial > Congressional District Geocoder** from menu bar. A window will be pop up
3. Choose place name column that represents the ZIP code field in your data file.
4. Press Ok button to start the geocoding

5-digit ZIP codes with multiple congressional districts, empty entries and invalid ZIP codes that failed to be geocoded will list in warning messages on the console.

The output of this algorithm is the original input table with additional 3 columns (Congressional district column, latitude column and longitude column). ZIP codes that failed to be geocoded will have blank entries.

Our benchmark is 50,000 ZIP codes per second.

## Geomap the congressional districts

1. Firstly, you might want to aggregate your data based on congressional district. To do this, you can follow user hints at [here](#).
2. You are ready to plot your aggregated result to geomap. It is recommended to plot the congressional district results on a country map due to some U.S. districts are located outside of the America Continents. To geomap the congressional districts, please follow the user hints at [here](#).

Enjoy!!!

## Links

- Source Code: [Link](#)
- External Package: [Link](#)
- Home Page: [Link](#)
- WATCHDOG.NET: [Link](#)
- ZIP code's wiki: [Link](#)

## Acknowledgments

The geocoding algorithm was authored, implemented, integrated and documented by Chin Hua Kong. Many thanks to the Sprint team for providing advices and suggestions. Many thanks to GovTrack that provides ZIP to district mapping data and district's geolocation information. Thanks to Carl Malamud and Aaron Swartz, that make the data available on WATCHDOG.NET for GovTrack.

## Contributed Comments

It is interesting to work on this algorithm from zero knowledge of ZIP codes and congressional district. A lot exploring works and analysis are done during development which have caused the design and preparation period in Sprint longer than expected. There is a lot of mapping databases available for sale. However, we are lucky to found the GovTrack that provide all free data and web service for the mapping. A lot of revise and improvement were done during the development which make the plugin in better and accurate. It is fun and worth for the knowledge I gained. Now I have better ZIP code system knowledge and congressional district concept. V!

## Data Source Update for 113th Congress (updated December 2013)

The data used to power this plugin was originally sourced from the [GovTrack.us](#) website. As of the 113th Congress, they no longer support or update the district to geolocation or zip code to district data. We have recently updated the data to reflect the most current 113th Congress data using the following sources:

- US Census Congressional District data: [Congressional Districts and ZIP Code Tabulation Areas](#)
- 2013 US Census Gazetteer files: [113th Congressional Districts](#)
  - (Under 113th Congressional Districts, download the zip, then use the .txt file inside)

The data from this site must be parsed correctly before being used in this tool. We used the following Python scripts to parse this data before including it in the plugin. Refer to the script comments for documentation:

- [parseGeocoord.py](#)
- [parseZip.py](#)

For the convenience of users, we have already pulled this new data, parsed it, and included it in the most recent build. For anyone who wants to use legacy data for the 112th Congress, however, those data files may be found here:

- [dist2geolocation.txt](#)
- [zip4dist-prefix.txt](#)