**TaxStats App**

COP 5859 - Semantic Web Programming

Florida Atlantic University

Dr. Shankar

08/01/2015

Authors:

Matthew Sanabria

Kevin Medina

**Abstract**

Property tax information is published annually for each municipality/county in the United States. However, accessing this information can be challenging for some individuals, especially those who are less technically savvy. Therefore, we introduce TaxStats, an application that provides users with the ability to research real property tax information using a simple, user-friendly interface. This application/research tool provides users with statistical analyses of county tax information—to understand the averages, trends, and assessed/market values of real property for any given region. There are many applications for this program for various users, and there are endless opportunities for the future development of this application, including an interactive and color-coded TaxStats map.

**Background**

TaxStats is a research tool that provides users with the ability to conduct brief statistical analyses of county tax information—to understand the averages, trends, and assessed/market values of real property. There are many benefits/uses for this application. Real estate professionals can use this application as a tool for convincing buyers to purchase/not purchase in a particular area based on the tax statistics provided. Prospective and current homeowners can use this application to conduct their own research or due diligence when searching for an area to live. In addition, government officials can use this data for analytical purposes and for building reports.

In this paper, we discuss the (1) methods, (2) output/ results, (3) challenges, and (4) future development of TaxStats.

**Methods**

This section discusses the database used, tools/ languages required, initial configuration, and execution of this application.

**Database**

Obtaining tax information for any specific municipality requires users to access its respective county property appraiser database. This application uses Essex County, NY tax information for its initial release and testing, which can be assessed using the Essex County Treasurer’s Office’s tax query form. This form can be assessed using the following link:

http://www.co.essex.ny.us/Treasurer/FileCreate.aspx

To recreate the data set used in this application and described in this paper, the following steps must be taken:

1. In step 1, select the following fields: (a) Address, (b) Assessment, (c) Town, (d) Amount Due, (e) Tax Year, and (f) Acct Nbr
2. In step 2 , leave all drop down boxes as the default “No Selection”
3. In step 3, select “ALL TOWNS”
4. In step 4, select the following fields: (a) 2014 and (b) 2015
5. In step 6, click “CREATE FILE” to generate a download link on the top of the page
6. A “DOWNLOAD” button should appear next to the “CREATE FILE” button selected in the previous step. Click the “DOWNLOAD” button to begin downloading a CSV file/ the tax dataset

The completed form should appear as shown in **Figure 1** in the appendix. The CSV data set should appear as shown in **Figure 2**, with the exception of the first row column headings (which must be added to recreate the program).

**Tools & Languages**

The tools required for recreating this application are as follows:

1. Protégé: to create a semantic web programming ontology
2. Jena Reasoner: to reason with future datasets
3. Java SE Development Kit 7: as the programming/scripting language
4. Geonames API/ Web Client for Java: to access Geonames’ database
5. Eclipse IDE: to develop/write the program

The code, ontology, dataset, etc. used in this Java program can be found in the GitHub repository link section of the appendix.

**Initial Configuration**

To execute/run this application the user must take the following steps for Eclipse IDE

1. Enable Geonames web services using their Geonames account
2. Download the following .jar files:
   1. GeoNames .jar file

<http://www.geonames.org/source-code/geonames-source-1.1.12.jar>

* 1. JDOM .jar file

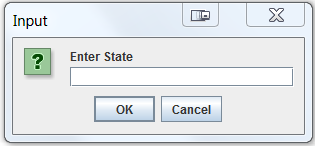
<http://www.geonames.org/source-code/jdom-1.0.jar>

1. Add the .jar files to the build path of the project
2. Enter Geonames username into line 32 of the driver.java file

**Execution**

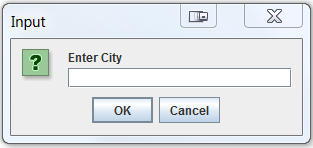
During the execution of this program the user is presented with two pop-up boxes that require input.

The first input box (shown below) prompts the user to enter a state. Note that this program is only written to produce results for the Essex\_County\_Dataset.csv dataset. Thus, the user is required to input “New York,” as this is the only state that will currently work for this application.



State Input Box

The second input box (shown below) prompts the user to enter a city. The user must enter one of the Essex County dataset cities (e.g. Chesterfield, Crown Point, Elizabethtown, Essex, Kenne, Lewis, etc.).



City Input Box

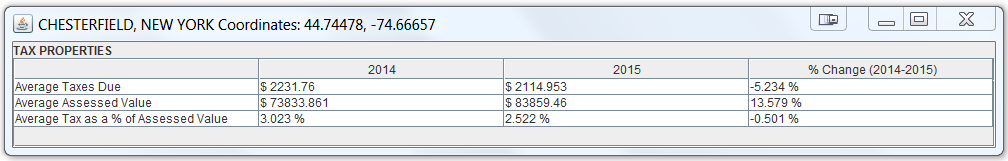
Based on the state and city provided, a pop-up box will appear with the output/results.

**Output/ Results**

This application produces several key statistics based on the user’s state and city inputs. The output appears in the form of pop-up box. The key statistical information/tax properties provided in the output box are:

1. City coordinates (latitude and longitude)
2. Average taxes due for 2014 and 2015
3. Average assessed value for 2014 and 2015
4. Average tax as a percent of assessed value for 2014 and 2015
5. Percent change between 2014 and 2015 for average taxes due and average assessed value.

A sample output for Chesterfield, NY can be seen below, and other sample output can be seen in **Figure 3** in the appendix. Each year this data can be updated and produce an output that provides the current and previous year’s tax data for the given municipality.



Sample Output (Chesterfield, NY)

**Challenges**

There are several key challenges that must be overcome in order to further develop this application in other counties and states.

In general, obtaining the required datasets for this application can be quite challenging and cumbersome. The property tax information for some counties vary in format, accessibility, and/or may be incomplete for the purposes of this application. Some counties do not publish complete tax information for each parcel of land in its tax roll datasets. For example, many counties, such as Broward County, FL do not provide the tax amount for each parcel of land in its downloadable data files; instead, users must look up this information one parcel at a time using whichever search tool is provided. In addition, some counties only publish the data one format (e.g. CSV), while other counties use a variety of formats (e.g. RDF or XML formats).

Hence, Essex County’s tax roll database was chosen to begin the development of this application because it provides users with the ability to customize the data to include in the CSV file for download. Other counties simple provide a download with no customization options, while other counties require the user to purchase tax datasets—this variation makes further development quite challenging.

**Future Development**

Ideally, our team foresees this application being further developed to include more granular statistics using both Geonames and datasets from other databases.  For example, data at the postal code level can be provided, where a user enters a postal code to search instead of a state and city.  In addition, using this database we can add features such as distribution of land use; highest, average, and lowest property tax for each type of land use; list of historical properties, percent of out of state owners per city, etc.

Using the Geonames’ longitude, latitude, and elevation data, this application can eventually be displayed on an interactive map of the United States. This map can allow users to zoom in to each state, county, city, and zip code. Each region can be color-coded to depict the distribution (highest and lowest) property taxes. Therefore, instead of a pop-up output box (as this application currently produces) the user will have a better understanding of the data provided from a visual perspective.

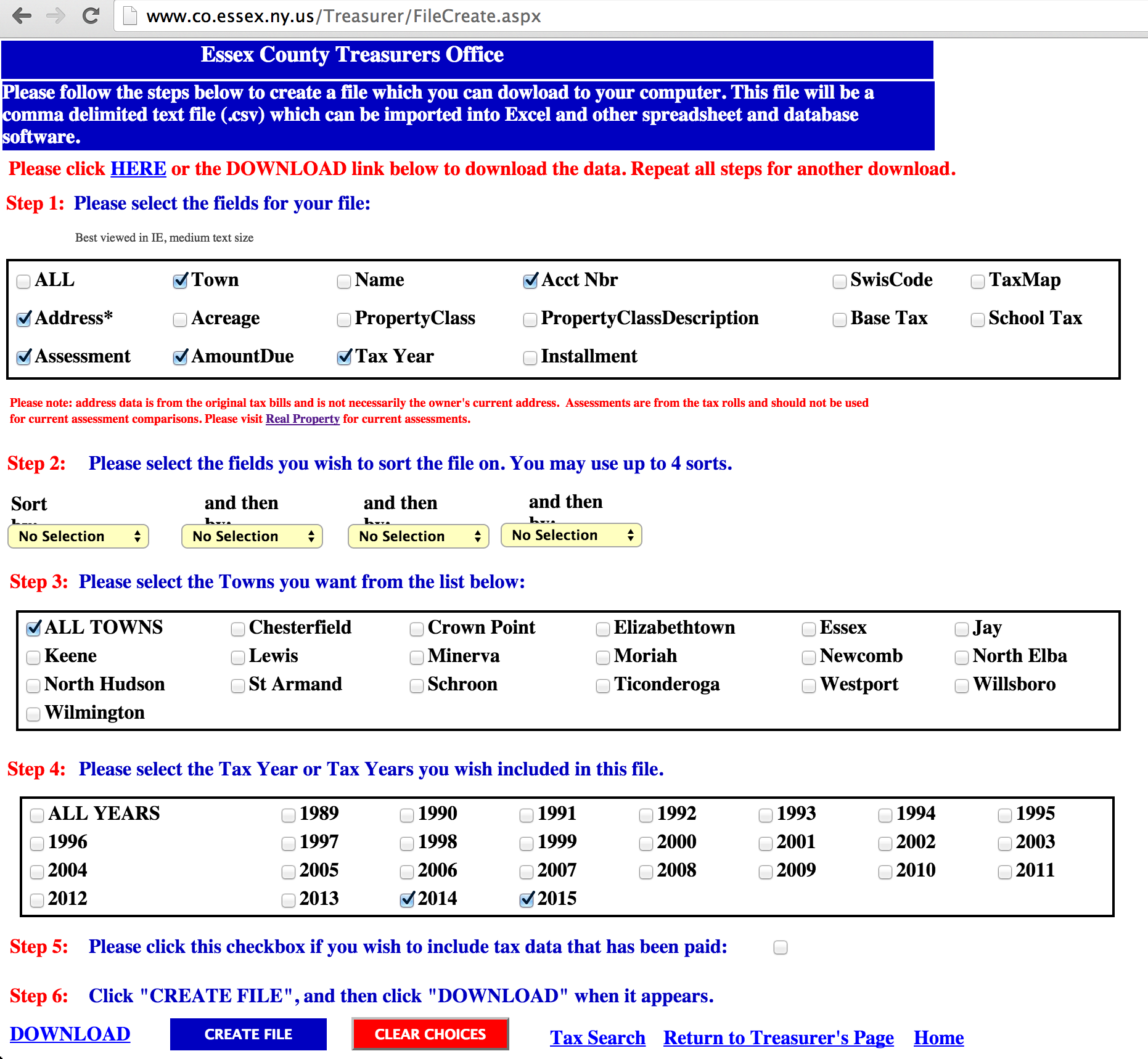
In addition, using the ontology/ taxonomy created with Protégé (**Figure 4**), coupled with a reasoner (e.g. Jena), this application can be used to extract similar properties from other tax databases.  Our goal is to have this application provide the same or more granular statistics for each municipality nationwide.

**Conclusion**

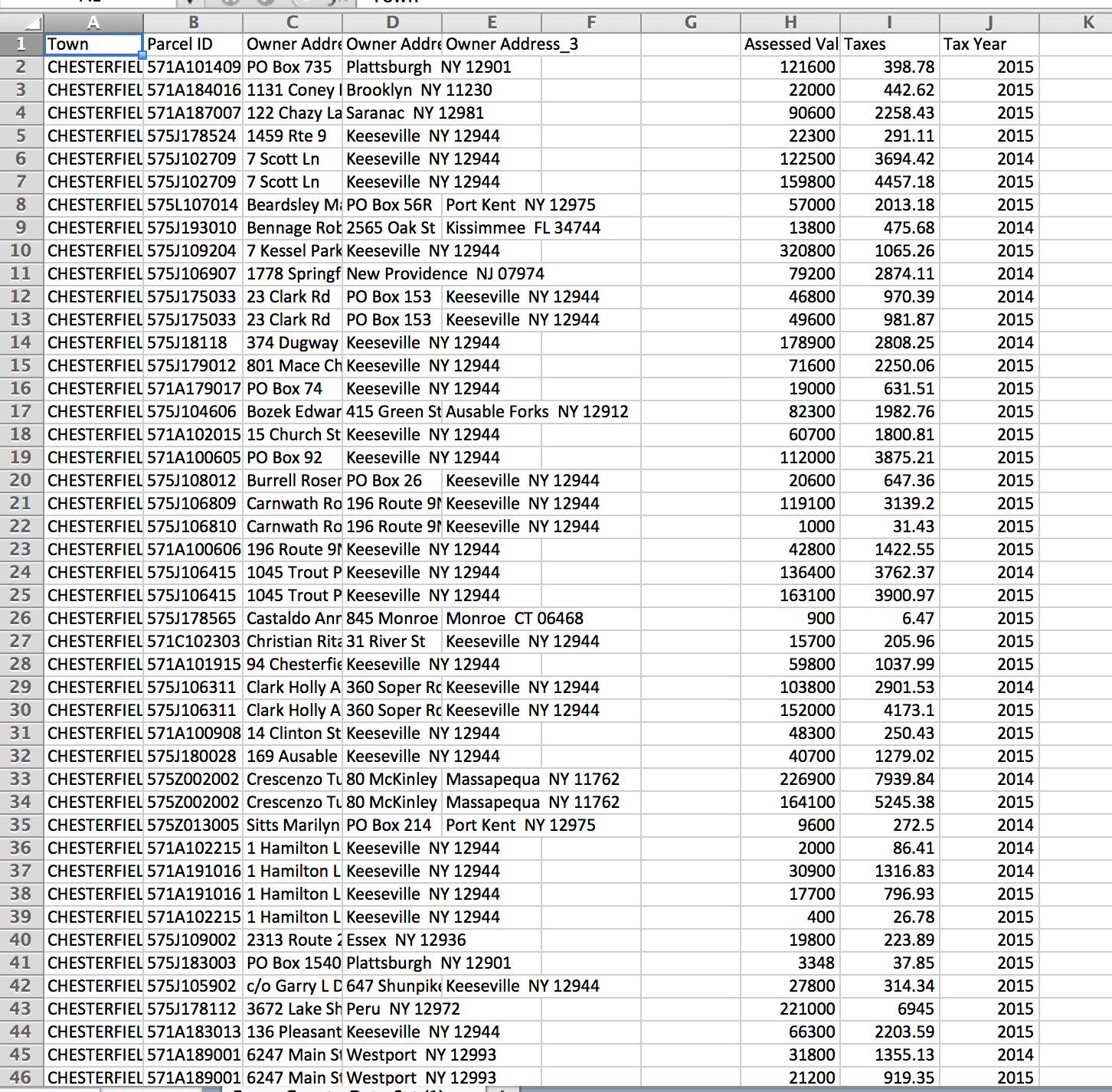
In conclusion, our team believes this application will provide a solution to the cumbersome task of researching tax information for each municipality. In addition, some of the statistics provided by TaxStats are not provided in any other user-friendly application. Therefore, this application will be useful to several different user groups, including real estate professionals, government officials, prospective/current homeowners, etc. In addition, with the opportunities for future development, this application can be incorporated into the websites of each municipality or used on a national level as a research tool for those who subscribe.

**Appendix**

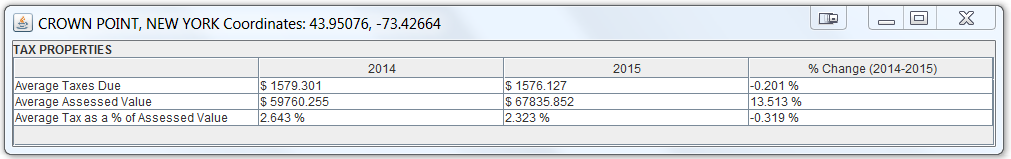
**Figure 1: Completed Tax Query Form**

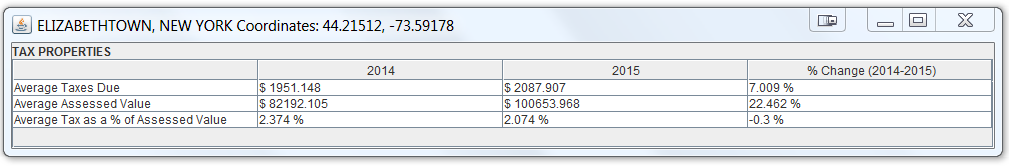
****

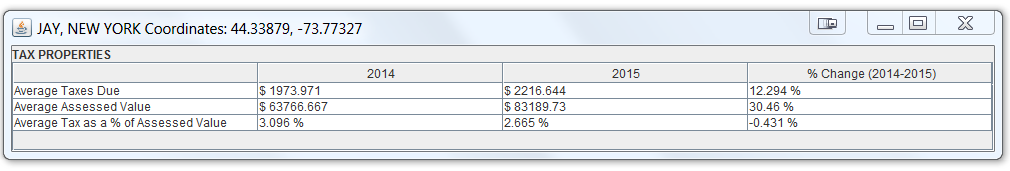
**Figure 2: CSV Dataset**

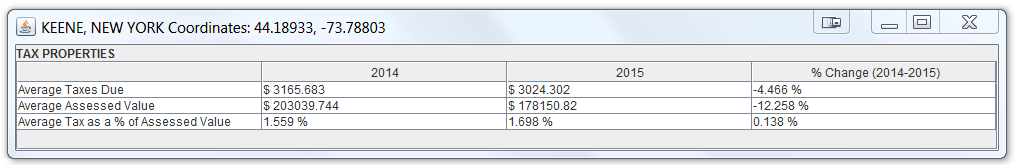
****

**Figure 3: Sample Output**

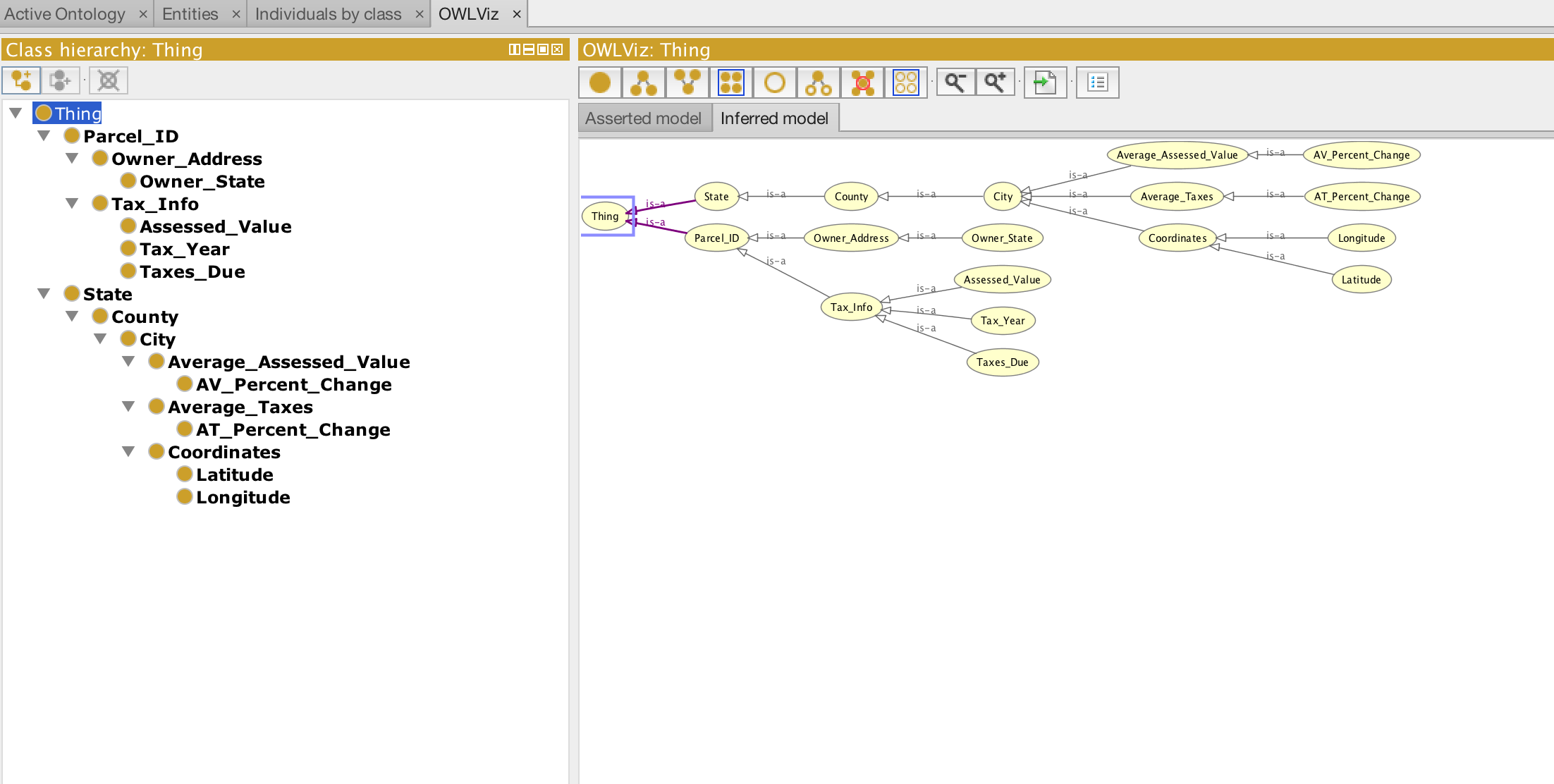








**Figure 4: Ontology/Taxonomy/Class Hierarchy using Protégé**

****

**GitHub Repository Link**

https://github.com/Semantic-Web/Matthew-S/tree/master/Final%20Project