

CS 6315 Project Proposal

Team: Housing Inspection Explorer Team

Title: Housing Inspection Explorer: A Tool For Viewing Inspection Scores Among Public And Multihousing Units

Project Type: Simple Project

Collaborators

- Phillips, Christopher Trevor [ctp180001]
- Nazneen, Amtul [axn180041]

CONCEPT #1 – Housing Inspection Explorer (HIE)

Overview

Housing Inspection Explorer (HIE) provides an insight into the Multifamily and Public Housing Inspection Scores published by the Public Housing Authority (PHA). PHA performs the inspection for metropolitan and micropolitan statistical areas, which are geographical entities defined by the U.S. Office of Management and Budget (OMB). HIE gives access to the published PHA Inspection Scores through interactive dynamic reports and can be used by agencies to publish Federal statistics. HIE also assists potential home buyers and property owners to determine if a given location is decent, safe, sanitary and in good repair.

Prominent Features

Below features are provided for either or both the housing types – Multifamily Housing and Public Housing

- Display all Housing Inspection Scores with appropriate drill-down to State -> City -> County
- Top 5 States with Highest/Lowest Housing Inspection Scores within U.S
- Top 5 Cities with Highest/Lowest Housing Inspection Scores within each State
- Top 3 Counties with Highest/Lowest Housing Inspection Scores within each City
- View all cities with acceptable standard of Housing Inspection Score
- View all the inspections that happened in the given year/range of year
- Estimated next inspection dates for a given county/property

Appealing Factors

This concept is appealing as it provides a graphical outlook into *both* the Multifamily and Public Housing Inspection Scores making it usable to various audiences with affinity towards property. Both the datasets for this concept have semantically well-related data in Housing domain making it comfortable to contrast or collaborate data from these two primary housing types.

Issues/Concerns

There is significantly less variation between the two selected datasets (1258, 1259) as they differ only in the type of Housing – Multifamily or Public Housing. The rest of the attributes in both the datasets share similar kind of data.

CONCEPT #2 – US Library System Viewer (USLSV)

Overview

The US Library System Viewer (USLSV) is a visualization tool that allows users to view the locations of public libraries in the continental US, and to view various metrics related to their operation (such as total books, visitors per year, income and expenditures, etc.) Additionally, information related to the state library agencies that preside over each state is also presented in the tool, and is able to be viewed and filtered on a per-state basis. Thus, the tool provides an intuitive means to determining key metrics related to the operation of libraries and library agencies within the US.

Prominent Features

Among the key capabilities of the USLSV are the following:

- Display locations of all public libraries for the continental US on an interactive map
- Provide metrics/statistics for each individual library, as selected by the user
- Allow filtering on key metrics for libraries (books, visitors, income, expenditures, etc.)
- Show rankings of top/bottom libraries for these metrics
- Provide similar functionality for prior four bullets, but for state library agencies
- Allow simultaneous filtering of state library agency metrics and individual library metrics

Appealing Factors

This concept is appealing in that it allows a user to condense a very large amount of information (in this case, metrics about libraries and state library agencies, as represented by objects in triples) into an intuitive visual tool that succinctly summarizes this information to the viewer. Additionally, there could potentially be an interesting interplay between the libraries in the state and the state library agency that funds them (e.g. do state library agencies with more funding lead to better-equipped libraries), and this tool could be a means to discover these relationships.

Issues/Concerns

The main concern is that the data in both datasets are from different time periods (specifically, from 2006 and 2002, respectively), and so integrating the datasets will likely lead to inconsistencies between them. For instance, income/expenditures for individual libraries will likely not sum to their statewide counterparts in the state library agency dataset, since they are recorded for different years. Secondly, both datasets have different granularities, which would potentially require writing more complex SPARQL queries for the tool. Specifically, the first dataset (353), which contains data on the state library agencies, has a single entry for each state, while the second dataset (613), which contains data on the individual libraries, may have multiple entries for each state.

CHOSEN CONCEPT

We ultimately choose to pursue **Concept #1: Housing Inspection Explorer**, based on the analyses for each concept above. Specifically, given that the housing datasets are (1) very large (giving us ample data to use for our project), and (2) consistent in granularity and time period (allowing us to easily join them, and for interpretations across datasets to be consistent), they avoid the issues associated with the

second concept, and thus make the first concept more appealing overall. Concept #1 is also a “leaner” concept in that there’s fewer quantitative fields to decipher and make available to the user, which allows us to focus more on developing a strong concept that successfully harnesses semantic web technologies. It is for these reasons that we choose **Concept #1: Housing Inspection Explorer** as our concept for our semantic web project.

Expected Results

Prominent features listed under **Concept #1: Housing Inspection Explorer** is expected to be presented in the visualization.

DATA SOURCES

Our project will use the following two datasets, available at the Data.gov Wiki (https://data-gov.tw.rpi.edu/wiki/Data.gov_Catalog_-_Complete):

- Dataset 1258- Public Housing Physical Inspection Scores (1278872 triples)
- Dataset 1259- Multifamily Housing Physical Inspection Scores (1455325 triples)

We will use the SPARQL endpoint at <https://data-gov.tw.rpi.edu/sparql> to access this data, in tandem with the proxy service at <https://logd.tw.rpi.edu/ws/sparqlproxy.php>. This SPARQL endpoint has a total of 176470986 triples available across *all* of its datasets.