FUNG INSTITUTE
for ENGINEERING LEADERSHIP
UC BERKELEY COLLEGE OF ENGINEERING

Patent-Mapper: Visualization of the geolocation of patents

Yang Shen (Kevin)
Department of Letter and Science, Computer Science
UC Berkeley
kevinshen34@berkeley.edu

College of Engineering
University of California, Berkeley

Fung Technical Report No. 2013.08.05
http://www.funginstitute.berkeley.edu/sites/default/files/PatentMapper.pdf

August 5, 2013

The Coleman Fung Institute for Engineering Leadership, launched in January 2010, prepares engineers and scientists – from students to seasoned professionals – with the multidisciplinary skills to lead enterprises of all scales, in industry, government and the nonprofit sector.

Headquartered in UC Berkeley's College of Engineering and built on the foundation laid by the College's Center for Entrepreneurship & Technology, the Fung Institute combines leadership coursework in technology innovation and management with intensive study in an area of industry specialization. This integrated knowledge cultivates leaders who can make insightful decisions with the confidence that comes from a synthesized understanding of technological, marketplace and operational implications.

Abstract
With thousands of patents emerging annually, and an approximate total of five million patents existing between the years 1975 and 2012, it is helpful to be able to visualize the geographic location of these patents. We introduce a geographic visualization tool, Patent-Mapper, which maps the location of patents within the U.S., given a selected patent application year, patent state, and patent class. This mapper allows the user to zoom down to a city level to obtain useful information on a particular patent and enables the user to see the overall number of patents in a particular city. Because these patents themselves do not have an associated location, we attribute the location of the first inventor of the patent.

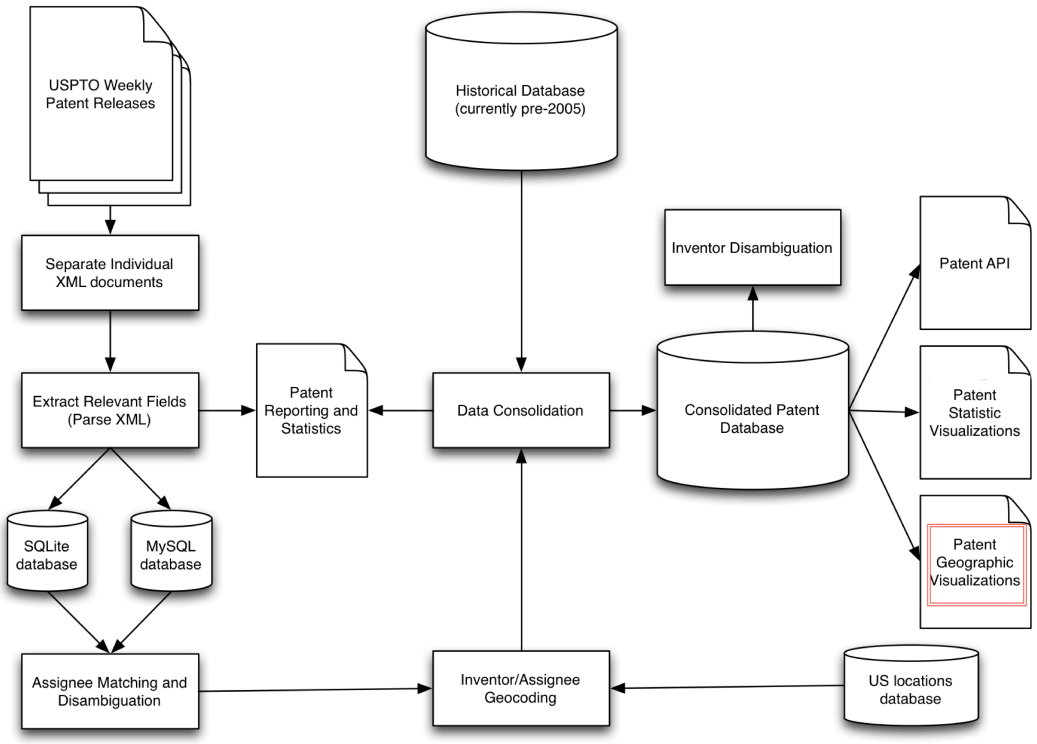

Figure 1: A process flow diagram presenting the process of obtaining the geolocation data.

Geolocation data
The database supporting the Patent-Mapper is compiled by data coming from Fung Institute's weekly patent disambiguations. Specific toolchain and data processes have been designed to clean data parsed through raw XML documents. More information regarding extracting and formatting patent data from USPTO XML are provided by Fierro (2013), at http://www.funginstitute.berkeley.edu/sites/default/files/Extracting_and_Formatting.pdf.

Introduction

Visualizing the number of patents in a state on a given year provides useful information in understanding the progress in economic development of a state throughout the years. By providing a city view on the map, we are able to identify the cities that have a large number of patents. For each patent, the information on its primary inventors, assignee and geolocation will also be available, thus providing potential useful information for future citations. For the link to the Patent-Mapper, please see: bit.ly/fung-patent-map

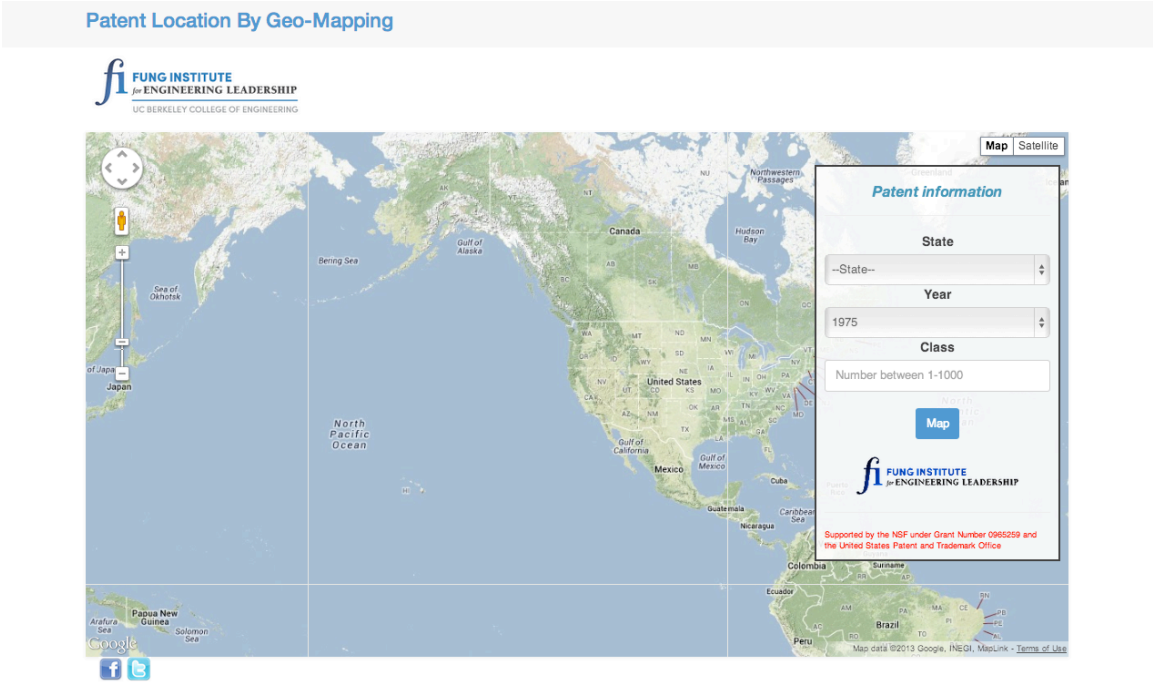

Figure 2: An overview of Patent-Mapper.

Identifying geo-location of patents from the Fung Institute patent database

We compile our location dataset by searching for patents (from 1975-2012) with the keywords: State, Year, and Patent class. When users select this group of criteria, we filter and identify a group of patent that satisfies these keyword searches based on the user's selection. We then identify the city location of the patent based on the primary inventor's geolocation within that state, allowing us to accurately locate these patents. The location of each patent is assessed in the initial stages of the parsing and cleaning processes (see Johnson, 2013, forthcoming).

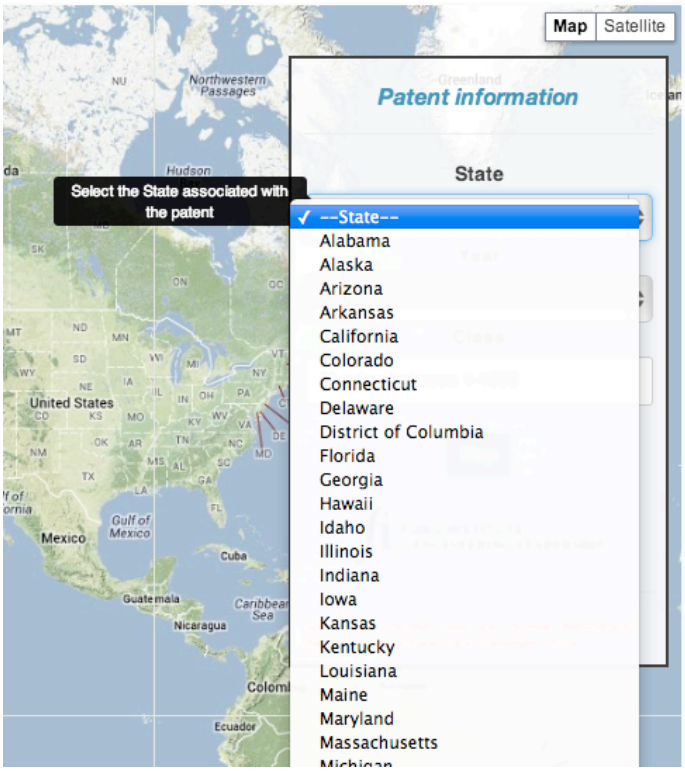

Figure 3: State filter allows user to filter patent by state.

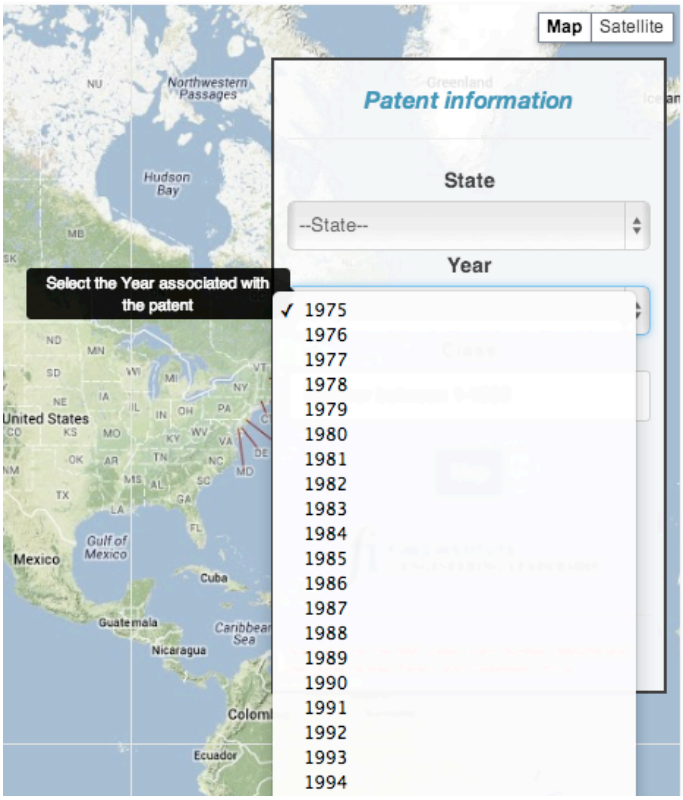

Figure 4: Year selection filter allows user to filter patent by patent's application year.

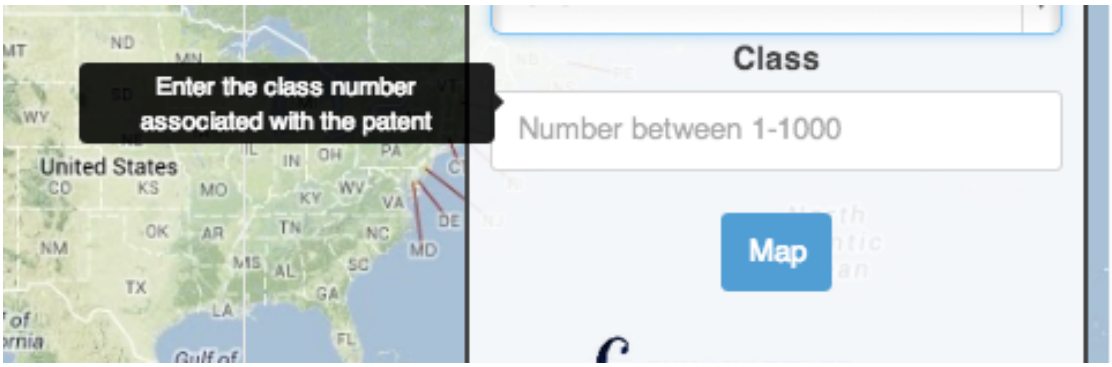

Figure 5: The class filter allows user to filter patent by class. Some class numbers do not exist in the database, though if a patent is found, it will be displayed.

Geolocation of the patents


Figure 4: Single marker

Each patent will be displayed as a small marker when it is being viewed at the city level. Currently, the location of the patent is based on the primary inventor's address, but the specific street level address for each patent is not always available. Because all the patent geolocations are only available at a city level, all patents within a city are shown in one single cluster, and the city with only a single patent will project the patents as a single marker. From a higher-level view of the state, these small markers will be grouped into a large marker-cluster, which contains thousands of patents. Information on each city's patents can be downloaded to a .csv file.

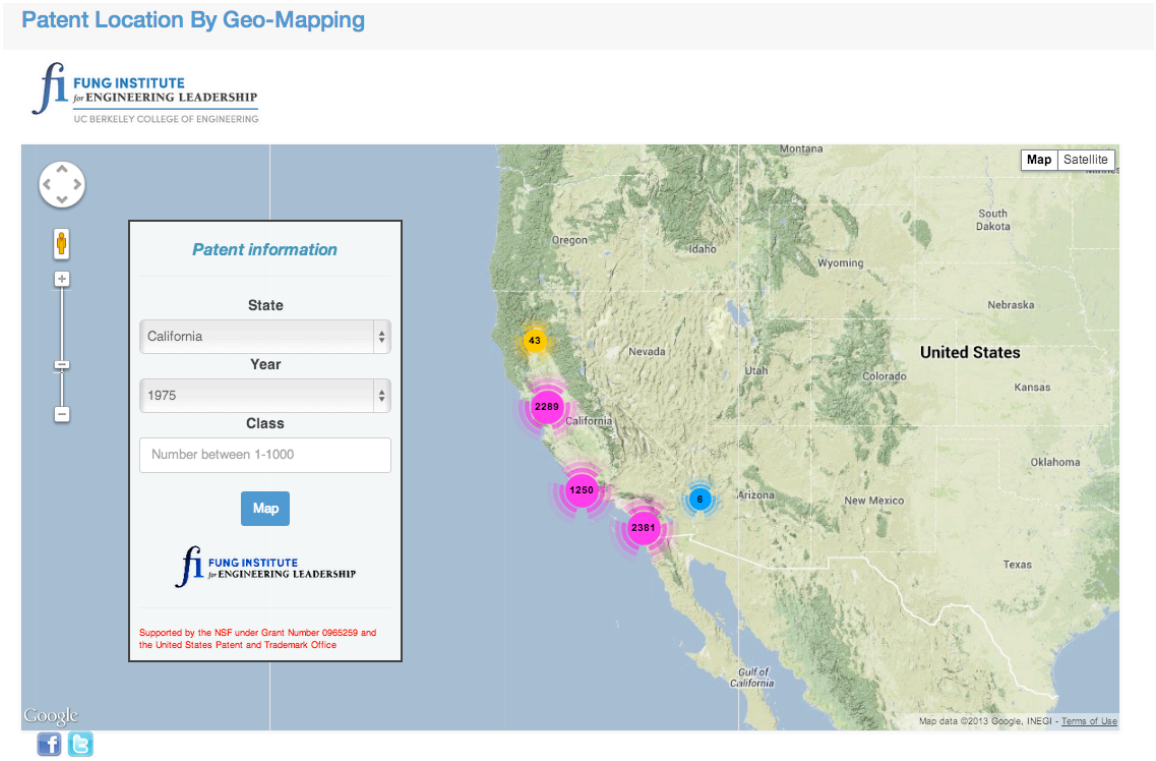

Figure 6: All smaller markers will be grouped into multiple large marker-clusters from a high-level state view.

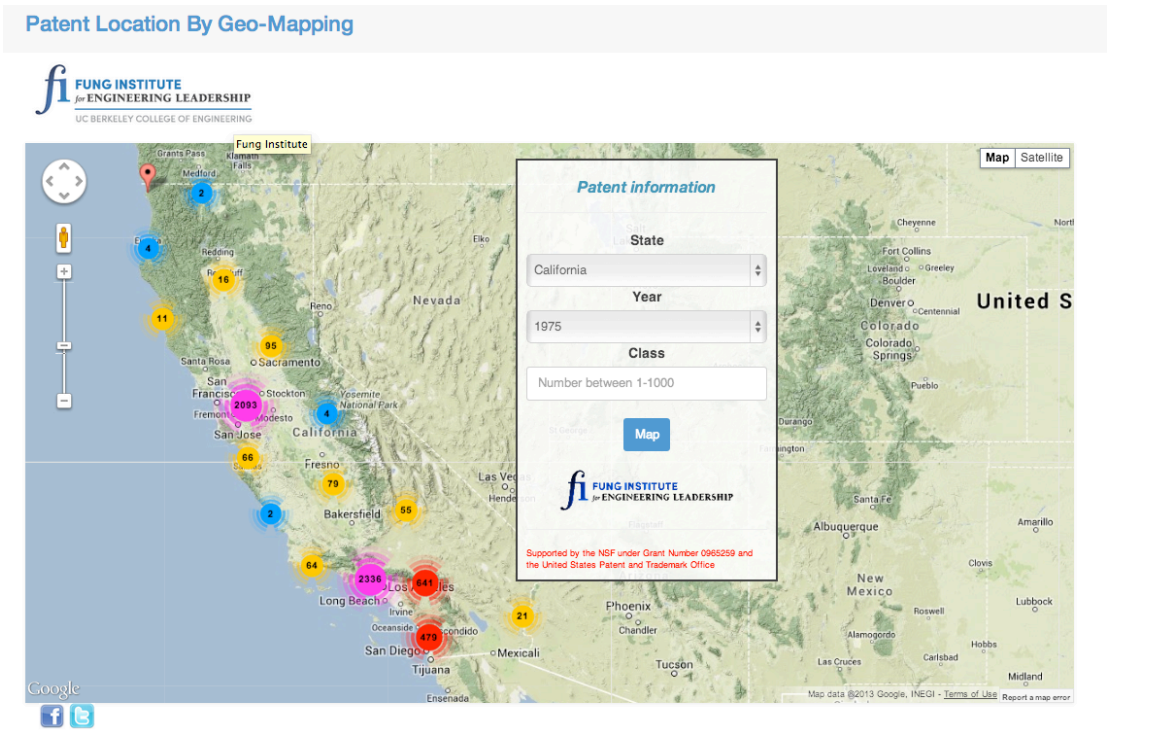

Figure 7: Larger marker-cluster zoomed into multiple smaller marker-clusters and markers.

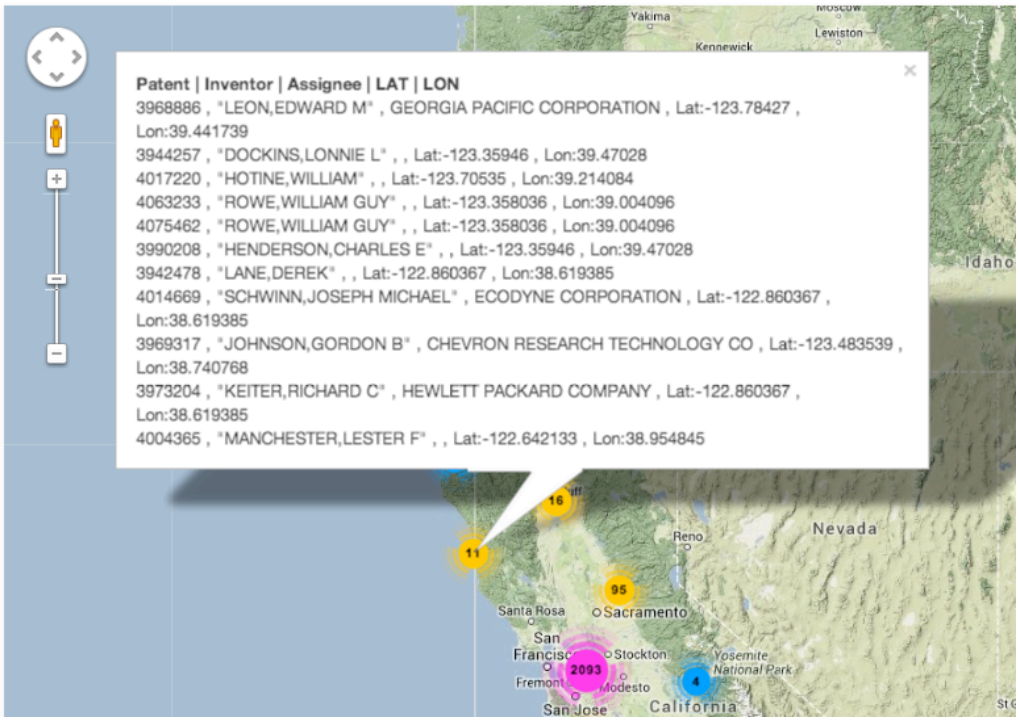

Figure 8: Each marker is clickable. When user click on a marker, an info-window will display all the patents and their corresponding information from that particular geolocation.

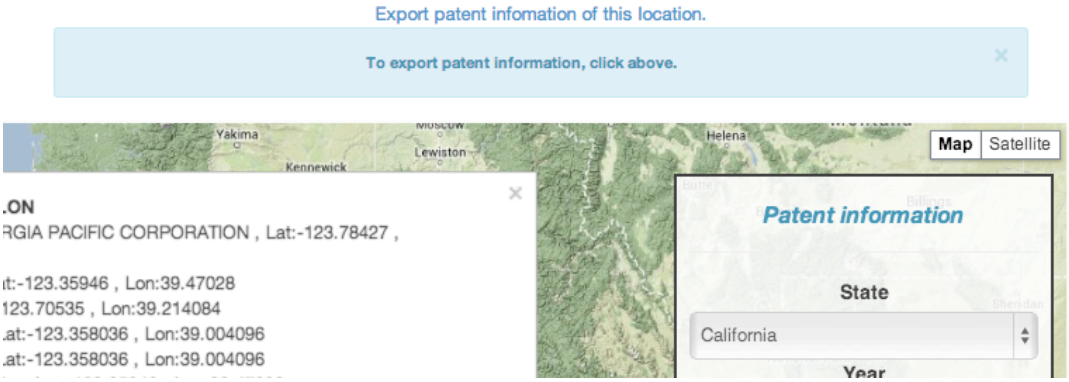

Figure 9: All patent information can then be exported into a CSV file.

Database query

Sqlite3 database queries are being used to obtain all the information from the database. Below is a sample of the query.

"select Patent, Longitude, Latitude, Lastname, Firstname, Assignee from invpat where State = "CA" AND AppYear = 1975 And Class like "128";"

By running this query, the database will return all patents from California from 1975 that have a class number of 128. Along with the patent, the Longitude and the Latitude will also be provided for geolocation purposes. Other information such as inventor and assignee are useful in characterizing each patent.

Acknowledgements

I would like to thank the Fung Institute for Engineering Leadership for supporting this research, and Guan-Cheng Li in particular for his technical assistance. This work is funded by the National Science Foundation under grant 1064182.

References

https://developers.google.com/maps/documentation/javascript/reference
http://flask.pocoo.org/snippets/56/

Johnson, K. (forthcoming). "Geocoding Patent Data." Fung Institute Technical Report.

Appendix: Data Sources and Code Repository

The NBER data is available at http://www.nber.org/patents/.

The DVN data is available at http://dvn.iq.harvard.edu/dvn/dv/patent/faces/study/StudyPage.xhtml;jsessionid=fd8595bd5c692dce0bef4ed95108?globalId=hdl:1902.1/15705&studyListingIndex=0_fd8595bd5c692dce0bef4ed95108.

The USPTO data is available at http://www.google.com/googlebooks/uspto-patents-grants-text.html.

Links to the merged database can be found at https://github.com/funginstitute/downloads.

The source code of Patent-Mapper can be found at:
https://github.com/kevshin2/Patent_mapper