OpenPOI: An Open Place of Interest Platform

Grant McKenzie
Department of Geography, McGill University, Montréal, Canada

Krzysztof Janowicz
STKO Lab, University of California, Santa Barbara, USA

Abstract

Places of Interest (POI) are a principal component of how human behavior is captured in today's geographic information. Increasingly, access to POI datasets are being restricted – even silo-ed – for commercial use, with vendors often impeding access to the very users that contribute the data. Open mapping platforms such as OpenStreetMap (OSM) offer access to a plethora of geospatial data though they can be limited in the attribute resolution or range of information associated with the data. Nuanced descriptive information associated with POI, e.g., ambience, are not captured by such platforms. Furthermore, interactions with a POI, such as checking in, or recommending a menu item, are inherently place-based concepts. Many of these interactions occur with high temporal volatility that involves frequent interaction with a platform, arguably inappropriate for the "changeset" model adopted by OSM and related datasets. In this short paper we propose OpenPOI, an open platform for storing, serving, and interacting with places of interests and the activities they afford.

2012 ACM Subject Classification Information systems → Geographic information systems

Keywords and phrases place, point of interest, open data, gazetteer, check-in

Digital Object Identifier 10.4230/LIPIcs.GIScience.2018.47

Category Short Paper

1 Motivation

Gazetteers play an important role in how we understand the world. They facilitate the labeling of geographic space thus forming the foundation of location-based services [3]. Historically, gazetteers have been categorized by scale, resolution, and theme. Some of the more traditional gazetteers are global in scale but at fairly course resolutions focusing largely on geographic features on the macro and meso levels such as airports, populated places, and rivers. Local gazetteers have tended to focus on a specific theme at higher resolution within a limited geographic boundary, e.g., Difangzhi local Chinese histories [4]. With advances in technology and commercial investment, digital gazetteers have quietly taken on new roles, forming the foundation on which a lot of the technology we use today, is built. Anyone who has used a mobile device in the past ten years has benefited from digital gazetteers be it through navigation/wayfinding using Google Maps or photograph tagging on Instagram.

In the last several years, context-based technology has continued to drive commercial investment, as many information technology companies realize the value of location information. This has lead to substantial investments in digital mapping technology [2] as well as the underlying spatial data that drives these platforms [7]. Digital gazetteers are increasingly storing the locations, names, and categories of *local* businesses and venues, today generally referred to as *points of interest* (POI). For instance, the location, name and hours of operation of the *mom-and-pop* shop at the end of your street is now stored alongside millions of other *place* records in a global gazetteer that forms the basis of Silicon Valley's latest mapping

products. Not only are these companies capturing information related to the location and hours of operation, but they are also enlisting citizens to contribute data on everything from menu recommendations to general ambience. Pushing this a step further, the technological scope of many of these commercial entities means that they can determine *popular times* for many of these places based on users' mobile device-reported locations [11]. The amount of auxiliary, or descriptive, data stored about these points of interest arguably contributes to a variation of the POI acronym, namely *Place* of interest. The content contributed about POI are much more than geometric points and really serve to give users of these data an understanding of fine-grained characteristics of a place.

Unfortunately, the high financial cost of developing these place of interest dataset has led to much of these data being siloed within companies, solely being used within (or sold as) their services or products. As is the case with many data silos, the redundancy between gazetteers is high. Companies such as Yelp, Foursquare, and Facebook have all invested heavily in data collection and development of their own proprietary POI datasets, rarely with collaborative interests. This has resulted in multiple academic efforts to match and conflate these datasets [6, 10] and occasional legal and regulatory action [9]. Though many companies offer limited access to their POI datasets for third party application development, recent high-profile events related to data privacy have resulted in tighter restrictions on outside access [1].

One concern related to the construction of these POI datasets is the reliance on volunteered contributions. Most of the data stored in these proprietary data silos were contributed by individuals not employed by the companies. Users are actively choosing to share, or are coerced [5] into sharing, often personal information with these platforms which are in turn monetized and sold back to those same users. While there is an argument to be made for the value added by these companies through their services and platforms, the reality is that most users no longer have legal rights, or even digital access, to the data that they've contributed to these platforms. These silos also hurt the research community as they limit access to attribute data needed for modern recommender systems and work on geographic information retrieval more broadly. Hence, we see this paper and the data and services it introduces as a *research enabler* for the community. Such dataset papers play an important role in geoinformatics research and are gaining importance in many other communities [8].

2 OpenPOI

Considering this, we introduce the *OpenPOI platform*, a dataset and service for storing, sharing, and interacting with a common set of places of interest. Following the open and user-contributed, geo-data model proposed by OpenStreetMap (OSM) and others, OpenPOI aims at supplying highly descriptive content related to local places of interest and enable a high level of interaction and sharing. Both of these approaches sit outside the mission of the OSM community and are likely not suitable for the *changeset* model and validation approach adopted by OSM. Through the OpenPOI platform, users can share recommendations, opinions, and place-specific information as well as *check-in*,[1] post photographs, or access any form of information they would like concerning a place. The purpose is to enable free and open access to *platial* information that is owned and shared by the community. This should be appealing to place-based social media users as the project is completely transparent, allowing everyone open access to all data and code associated with the platform. For researchers, it

[1] Check-in, in this case, refers to the social act of publishing one's presence at a location.

offers a valuable resource on which to study human activity behavior and a place of interest dataset that can be used as the basis for any application, study, or research project.

As this platform is in a prototype phase, we give a very brief overview of the components, the current state of the platform, and some directions for moving forward. Currently, there are three components to the OpenPOI project: The dataset, the web service, and the mobile application.

3 Dataset

OpenStreetMap nodes are the source for all POI in this current version of the platform. As OSM is user-contributed and regularly updated, it provides the most extensive coverage of non-proprietary POI in the world. As the OpenPOI user-base grows, new POI may be added and existing POI updated or removed, branching the dataset from the OSM community while still maintaining links through original OSM node IDs. Future versions of the platform will ingest changes from OSM and publish changes, with basic attributes, back to OSM, following community best-practices and appropriate validation. In this prototype version of the platform, country specific OSM PBF files were downloaded for the United States and Canada. After thoroughly testing the platform using these data, global OSM planet files will be used. The *OsmPoiPbf* POI extraction script[2] was used to extract point of interest nodes from the raw OSM files resulting in a series of CSV files that were automatically inserted into a PostGIS-enabled PostgreSQL database. Once in the database, a duplication check was made before building a spatial index on the point geometries.

This PostGIS-enabled *PostgreSQL* relational database is used to store all point geometries in the OpenPOI dataset. Data related to users, check-ins, and tags are all stored in *MongoDB*, a document-oriented database system, often classified as *NoSQL*. The primary reason for the two different storage formats is to keep the POI geometry data spatially indexed and separate from the application-level data. The rate at which POI geometries are changed is far less than that of descriptive content, tags, and check-ins. As MongoDB was developed with consistency across database replicas in mind, it serves this purpose well. The current version of the database stores user profile information, time stamps and locations of check-ins as well as collections of tags and attributes assigned to a specified POI. Data extracts are available for each of these data collections or access to the data is available via the OpenPOI application programming interface (API).

4 Web Service

The current version of the OpenPOI API allows for basic interaction with the underlying OpenPOI dataset. Again, the API, including source code, is freely available and accessible via creative commons license. The API forms the basis for the OpenPOI mobile application and forthcoming data extraction tool. *NodeJS*[3] in combination with the *Express*[4] framework supply the foundation for back-end development. A set of public API endpoints are now available and currently permit the following data requests:

- Provided a latitude and longitude, return an array of nearby POI objects.
- Provided a User ID or POI ID, return an array of check-in objects.

[2] https://github.com/MorbZ/OsmPoisPbf
[3] https://nodejs.org/
[4] https://expressjs.com/

- Provided a POI ID, return an array of tag objects for the specified POI.

In addition to requesting data related to POI, users can also interact with the data through submission of various types of content, namely,

- Add a check-in object to a POI given the POI ID and User ID.
- Add an array of tags (hash-pairs) to a POI given the POI ID and User ID.
- Create a new user object.

These endpoints form the core of the OpenPOI platform functionality with additional endpoints being added as development continues. Documentation including example requests and required parameters is available at `https://github.com/ptal-io/OpenPOI-Server`. The current version of the API does not require authentication, nor does it limit requests, though authentication will be required in future versions of the platform.

5 Mobile Application

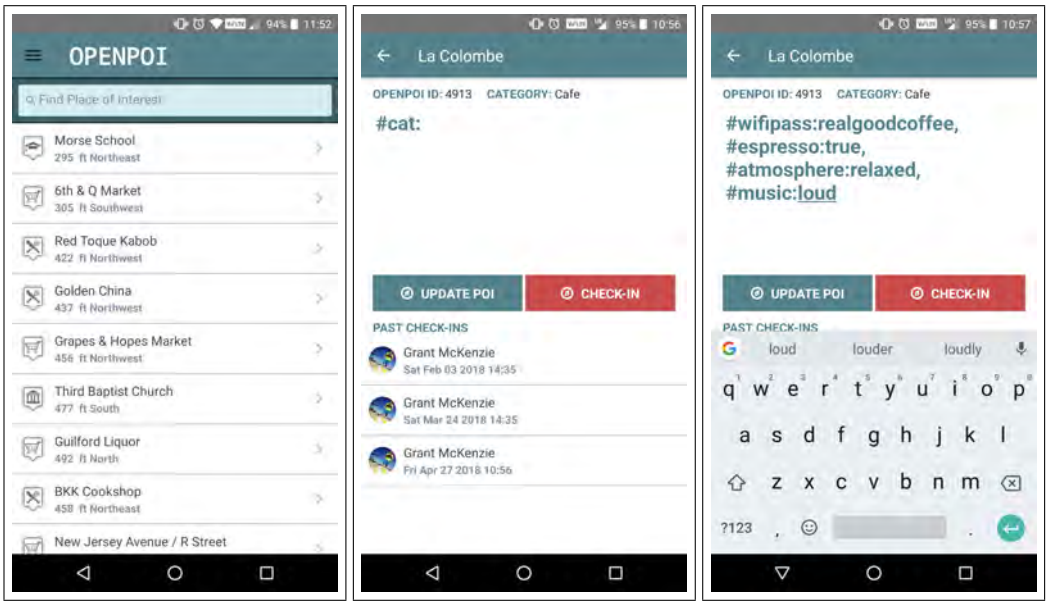
The OpenPOI mobile application sits as a front-end interface through which the OpenPOI dataset is accessed and updated. The mobile application communicates through the aforementioned public API endpoints keeping the entire project modular. Anyone can build a mobile, desktop, or web-based front-end that interacts with the data through this API. The mobile application presented here is one possible interactive window into the dataset.

The OpenPOI mobile application[5] is currently in development using the *React-Native* framework.[6] React-Native allows developers to use the JavaScript scripting language in combination with the React library to create mobile apps that are compiled into platform-specific applications. The current release of the mobile application has been compiled for use on an Android mobile device and can be downloaded for testing at `http://openpoi.org`. An iOS version of the OpenPOI application is forthcoming. The prototype version of the application is limited in functionality to a few core interactions, but serves the purpose of demonstrating the value of such an application.

Upon logging in and ensuring the location services are enabled, a list of nearby places of interest are shown to the user ranked by proximity (Figure 1a). The *list view* shows the name, category icon,[7] distance and direction from the device's current location. After selecting a POI from the list, the user is presented a screen listing descriptive information and permitting two forms of interaction. Users can check in to the POI by selecting the *check-in* button, in which case they are added to the database and list of previous check-ins to the specified POI (Figure 1b), or they can choose to update the POI with attribute information (Figure 1c). Virtually any type of descriptive textual information can be added to the POI on this screen using a *key-value pair*. Through this method, users can specify a *key* term such as *wifipass* by prepending a hashtag symbol. This term is then followed by a colon and the value associated with the key term. Attribute information is separated by these hashtags allowing for free text entry of any information. Currently, the application prompts user to update the category of the POI, but future versions will suggest potential key terms that may be most useful for the specified POI.

While the unrestricted ability to add any type of character-based content to a POI will undoubtedly lead to noisy data, the purpose of this application is not to restrict what people

[5] Source Code: `https://github.com/ptal-io/OpenPOI-App`
[6] `http://www.reactnative.com/`
[7] Currently based on OSM's amenity category taxonomy



(a) Nearby places of interest. **(b)** POI details and check-in screen. **(c)** Adding *key-value hashtag pairs*.

Figure 1 Three screens of the mobile OpenPOI application.

can or cannot enter, it is to get as much content as possible contributed to the application so that users can decide for themselves which information they care about and researchers have access to a wide variety of data. In today's area of *big data* and machine learning, it is much easier to clean, organize and extract meaning from large, noisy data than to work with a very limited supply of well structured content. Along these same lines, the underlying motivation for this application is not financial, meaning that clean, curated, and validated data is a secondary thought after free and open access to a large, heterogeneous, POI-specific dataset.

6 Summary & Next Steps

As digital gazetteers and POI datasets becomes increasingly silo-ed behind commercial firewalls, additional efforts must be made to ensure continued access to these types of geospatial information. In this short paper we introduced the OpenPOI platform and provide a brief overview of the components, functionality, and motivation for its development. We believe that such a dataset and services will be of value for the research community and act as a *research enabler* for many researchers in a wide range of disciplines.

Next steps for this platform will focus on three primary areas. First, a robust automated work flow for the extraction and merging of global places of interest from OpenStreetMap is in development. This process will merge the latest updates and changes from the OSM community with the rich attribute information and check-ins added through the OpenPOI platform. Additional effort will focus on inclusion and conflation of other data sources. Second, further functionality for interacting with the OpenPOI dataset, e.g., adding new places and updating geometry, are in the works along with associated API documentation. Last, further development on the mobile application will focus on rigorous testing of the core features, addition of a mapping screen, and overall interface development.

References

1 Josh Constine. Facebook restricts apis, axes old instagram platform amidst scandals. *TechCrunch*, April 2018. Online; 04-04-2018.

2 The Economist. The battle for territory in digital cartography. *The Economist*, June 2017. Online; 08-06-2017.

3 Michael F Goodchild and Linda L Hill. Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10):1039–1044, 2008.

4 James M Hargett. Song dynasty local gazetteers and their place in the history of difangzhi writing. *Harvard Journal of Asiatic Studies*, 56(2):405–442, 1996.

5 Grant McKenzie and Krzysztof Janowicz. Coerced geographic information: The not-so-voluntary side of user-generated geo-content. In *Eighth international conference on geographic information science*, 2014.

6 Grant McKenzie, Krzysztof Janowicz, and Benjamin Adams. A weighted multi-attribute method for matching user-generated points of interest. *Cartography and Geographic Information Science*, 41(2):125–137, 2014.

7 Greg Miller. The huge, unseen operation behind the accuracy of google maps. *The Economist*, December 2014. Online; 08-12-2014.

8 Jennifer C Molloy. The open knowledge foundation: open data means better science. *PLoS biology*, 9(12):e1001195, 2011.

9 Jack Nicas. Google rival yelp claims search giant broke promise made to regulators. *The Wall Street Journal*, September 2017. Online; 11-09-2017.

10 Tessio Novack, Robin Peters, and Alexander Zipf. Graph-based matching of points-of-interest from collaborative geo-datasets. *ISPRS International Journal of Geo-Information*, 7(3):117, 2018.

11 Google Support. Popular times, wait times, and visit duration. Technical report, Alphabet Inc., 2016. Accessed: 28-04-2018.