ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XII-4/W2-2026
10th International Conference on Smart Data and Smart Cities (SDSC), 29 September–1 October 2026, Sofia, Bulgaria

Democratizing High-Resolution Urban Data: A Cost-Effective Greedy Algorithm for POI Retrieval using Google Places API

Hozumi Kikuchi [1], Takuo Inoue [1], Hiroki Nakajima [1], Hideki Koizumi [1]

[1] Department of Urban Engineering, School of Engineering, the Univ. of Tokyo, Hongo 7-3-1, Bunkyo-city, Tokyo, Japan – hozumi@cd.t.u-tokyo.ac.jp

Keywords: POI Retrieval, Spatial Sampling, Facility Location Problem, Greedy Algorithm, Geospatial Big Data, Smart Cities.

Abstract

In urban analytics and smart city analytics, high-resolution Point of Interest (POI) data is indispensable. While commercial APIs like Google Places offer more comprehensive datasets, their high pay-as-you-go costs create a significant barrier for budget-constrained researchers. Existing spatial sampling strategies, such as the fixed grid approach, face an inherent trade-off: they either incur prohibitive costs due to redundant calls in sparse areas or suffer from reduced recall (data omissions) in dense urban cores. To address this, we reframe the data retrieval process as a capacitated facility location problem with unknown demand. Based on this, we propose an adaptive greedy sampling algorithm rooted in the geometric set cover problem, optimized for the constraints of the Google Places API. By using the distance to the farthest POI as the “effective radius,” this algorithm dynamically excludes areas covered by the feedback, thereby geometrically guaranteeing a 100% retrieval rate. Experimental results from Shibuya City (high density) and Chiba City (medium density) in Japan demonstrate that the proposed method can reduce the total number of API calls by approximately 60% compared to the best-performing conventional method (adaptive quadtree), and by up to 99% compared to high-resolution fixed grid baselines. This study aims to contribute to the democratization of high-resolution urban analysis for researchers and practitioners worldwide by minimizing data retrieval costs. The complete code is released as an open-source Python repository on GitHub.

1. INTRODUCTION

1.1 Background

High-resolution spatial data is the foundational layer for modern urban analytics, digital twins, and smart city infrastructure. Among these data types, Point of Interest (POI) data is widely recognized for its ability to effectively reflect the functional characteristics of diverse geographical locations in real time (Yu et al., 2020). Specifically, large volumes of POIs provide critical insights into urban functional zoning and land use classification (Jiang et al., 2015; Long and Shen, 2015; Liu and Long, 2016; Gao et al., 2017; Yao et al., 2017; Zhai et al., 2019), the delimitation of urban spatial structures and city centers (Lu et al., 2020; Yu et al., 2020), and the analysis of human activities and socio-cultural dynamics (Huang et al., 2016; Sen & Quercia, 2018; Gubert et al., 2025).

To acquire this essential data, researchers initially look to free, crowdsourced platforms like OpenStreetMap (OSM). However, their level of detail rarely meets the requirements of high-resolution analysis because these platforms frequently suffer from uneven spatial coverage and incomplete attribute data, particularly outside major metropolises (Gubert et al., 2025). Consequently, researchers frequently turn to commercial mapping services, such as the Google Places API. This platform is desirable because it has significantly higher penetration rates and broader coverage (Sen and Quercia, 2018; Chen et al., 2024; Gubert et al., 2025), providing comprehensive and up-to-date POI databases worldwide, including developing countries.

1.2 Problem Statement: The "Budget Barrier"

Despite its high data quality, utilizing commercial APIs poses a critical challenge: prohibitive data retrieval costs under a pay-as-you-go pricing model. When attempting to scan large metropolitan areas, many urban studies (Sen and Quercia, 2018; Martí et al., 2019; Gubert and Silva, 2022; Gubert et al., 2025) have conventionally relied on standard spatial sampling methods, such as the "fixed grid" approach.

This static approach fails to account for the dynamic, highly clustered nature of urban facilities, making them fundamentally inefficient. Because they do not adapt to spatial density, researchers are trapped in a methodological dilemma: applying a fine grid to ensure complete data capture leads to massive redundant API calls in sparse suburban areas, driving costs to an unmanageable level. Conversely, applying a coarse grid to constrain costs inevitably exceeds the API's maximum return limit in dense commercial districts, causing severe data omission.

In the context of urban analytics, this foundational data retrieval process has received little methodological attention compared to downstream analysis (Martí et al., 2019). If the initial data collection is flawed—either through massive omissions in dense areas or financially unsustainable redundancies; the validity of subsequent spatial analyses is deeply compromised. Resolving this spatial sampling dilemma is therefore not merely an operational cost-saving measure, but a fundamental scientific requirement to ensure data integrity in urban studies.

1.3 Research Objective

To overcome this trade-off, this study introduces a cost-minimizing, adaptive spatial sampling algorithm that theoretically guarantees the completeness of the retrieved POIs with few exceptions. By treating each API search circle as a capacity-constrained facility deployed over an unknown demand distribution, our algorithm dynamically adjusts the search radius based on API feedback. This minimizes redundant requests without sacrificing a single data point. Furthermore, by releasing this tool as an open-source repository (we will post a link here once the adoption decision is made), we will contribute to the "democratization of data access." This will enable students, independent researchers, citizens, professionals, and, in particular, urban planners in developing countries to overcome budgetary barriers and conduct high-resolution urban analysis at a fraction of the traditional cost.

2. LITERATURE REVIEW

2.1 POI Data in Urban Studies

The rapid growth of urban big data has brought significant changes to urban planning and urban studies (Kitchin, 2013; Batty, 2013), and research utilizing these data has been particularly active these days. While this boom indicates that urban studies have come a long way from the observations by Arribas-Bel (2014), barriers to utilizing these data remain.

Currently, the main barriers lie in data quality and acquisition costs. Volunteer Geographic Information (VGI) and social media check-in features are widely used, but they often come with inherent demographic biases and representational limitations that can distort spatial analysis (Goodchild, 2007; Elwood et al., 2012; Shelton et al., 2015; Martí et al., 2019). To achieve objective and scalable POI mapping, we have no choice but to rely on a comprehensive POI database provided by a private company—namely, the Google Places API—which offers superior spatial coverage and penetration rates.

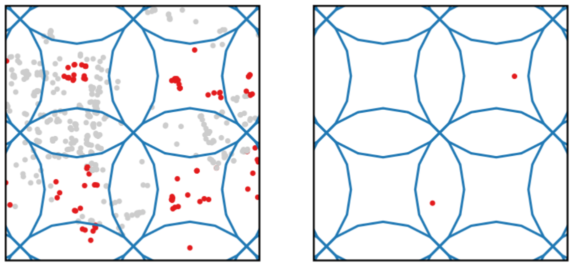
To avoid the high costs associated with commercial APIs, recent research has been exploring alternative data extraction methods. For instance, Kim and Lee (2024) utilized language models to extract POI information from social media text, while Chen et al. (2024) proposed deriving large-scale building functions by applying image processing techniques to downloaded Google Custom Maps. While these ingenious non-API approaches effectively bypass financial barriers, they suffer from severe limitations in attribute depth. They generally yield only basic place types and lack the rich semantic metadata—such as precise coordinates, user ratings, and operational hours—required for comprehensive urban analytics. Therefore, direct retrieval from commercial APIs remains indispensable for high-resolution studies.

Despite this heavy reliance on commercial APIs, the underlying methodology of efficiently retrieving this data has largely been treated as an afterthought. As Martí et al. (2019) precisely articulate, the challenges and limitations associated with the data retrieval process have received scant coverage in the literature, even though the accuracy of these methods is crucial for obtaining valid datasets. This critical methodological gap underscores the urgent need for cost-optimized spatial sampling algorithms like the one proposed in this study.

2.2 Existing Sampling Strategies

Many existing spatial sampling strategies prioritize the final analytical results over optimizing the data retrieval process itself. Consequently, the primary data retrieval methods observed in the literature largely rely on the following two conventional approaches:

1. **Fixed Grid Approaches:** Many studies divide the target area into uniform rectangular meshes (Yu et al., 2015; Wang & Chen, 2018; He et al., 2020; Liu et al., 2021). Since urban density is highly heterogeneous, fixed grids are inherently inefficient, causing a severe mismatch between human-defined static geometries and actual urban agglomeration. As illustrated in Fig. 1, applying identical search radius (blue circles) yields critical inefficiencies. In dense urban contexts (Fig. 1(a)), API limits cause severe under-reporting, leaving many actual POIs unretrieved (grey dots) while capturing only a fraction (red dots). Conversely, in sparse rural contexts (Fig. 1(b)), it results in massive redundancy with multiple empty or underutilized API calls.



(a) Urban context (b) Rural context

Figure 1. Comparison of fixed grid approach results in (a) urban context and (b) rural context.

2. **Adaptive Quadtree Approaches:** Martí et al. (2019) employ an approach known as the quadtree decomposition method (Samet, 1984) to ensure that all available data from the source is retrieved. However, since it still relies on square subdivision, significant geometric redundancy occurs under the constraint of circular searches in the API.

2.3 Algorithmic Foundation

To address these inefficiencies, we reframe the POI extraction process as a variation of the classical Set Cover Problem (Chvatal, 1979), or the Maximal Covering Location Problem (MCLP) (Resende, 1998; Berman & Krass, 2002). The Set Cover Problem is a fundamental combinatorial optimization problem that asks for the minimum number of subsets, each covering part of a ground set, whose union covers the entire set. It has been widely applied in spatial planning contexts such as optimizing the placement of facilities to maximize coverage under resource constraints.

In our context, the "ground set" is the unknown spatial distribution of POIs across the target area, and each "subset" corresponds to the POIs falling within a single API search circle. Solving for a globally optimal arrangement of search circles is infeasible for two distinct reasons. First, the spatial distribution of POIs is unknown a priori, meaning the algorithm cannot pre-compute an optimal layout and must instead make sequential decisions based on feedback from each API response. Second, even if the distribution were fully known in advance, finding the exact optimal solution to this class of problems remains computationally intractable for large instances, as the problem is NP-hard. A greedy heuristic addresses both constraints simultaneously: at each step, the algorithm selects the next search circle expected to cover the largest remaining uncovered area based only on information observed so far, rather than solving for a globally optimal arrangement. Chvatal (1979) showed that this greedy strategy, despite its simplicity, provides a provable approximation bound relative to the optimal solution even under such sequential, information-limited conditions.

Specifically, we model the API queries as a spatial allocation problem where capacity-constrained search circles must cover an unknown distribution of facilities. Here, each search circle acts as a "facility" with a strict capacity limit of 20 retrievable POIs, and the algorithm's greedy step—targeting the largest remaining unsearched area—mirrors how greedy heuristics iteratively place facilities to maximize incremental coverage. By applying a greedy approach (Jain et al., 2002), which has been successfully utilized in wireless sensor network coverage and proven to provide highly efficient approximations for facility location problems (Cardei and Cardei, 2008), we adapt the algorithm to operate within the specific geometric and limit constraints of the Google Places API.

3. METHODOLOGY

3.1 Characteristics of the Google Places API

To optimize the search process, the algorithm leverages three specific systematic constraints of the Google Places API Nearby Search (New) (Google Developers, 2026a):

1. **Maximum Result Limit:** A single API request returns a maximum of 20 POIs.
2. **Distance-Based Ranking:** The API can be configured to return POIs ordered by their Euclidean distance from the specified search center.
3. **Implicit Effective Radius:** If a requested circular area with radius R contains more than 20 POIs, the API truncates the results. The distance of the 20th POI (d_max) becomes the true "Effective Radius" of that request. Any POI located between d_max and R is omitted.

3.2 Proposed Method: Adaptive Greedy Sampling

We propose an adaptive greedy algorithm characterized by a "continuous subtraction" approach. Instead of pre-defining a mesh, the algorithm dynamically identifies fully explored areas and geometrically subtracts them from the unsearched target polygon.

Process Flow:

1. **Initialization:** The algorithm receives the target area as a multi-polygon and selects a starting point within it.
2. **API Request & Evaluation:** An API call is executed with a maximum theoretical radius. The response yields n POIs and the distance of the furthest POI (d_max).
3. **Coverage Determination:**
   i. If $n < 20$: The entire requested circle (radius R) is confirmed as completely searched.
   ii. If $n = 20$: Only the circle defined by the effective radius (d_max) is confirmed as completely searched.
4. **Geometric Subtraction:** Using spatial operations, the confirmed search circle is subtracted from the unsearched target polygon.
5. **Iteration:** The remaining unsearched polygons are buffered inward and decomposed into fragments smaller than a specified size (e.g., 1000 m²) to identify the next search point. The centroids of these newly generated polygons are selected as the next search points. This process repeats until the unsearched polygon area reaches zero.

Unlike fixed grids that leave geometrical gaps, the proposed greedy algorithm geometrically guarantees completeness (100% coverage) with few exceptions. This algorithm operates in a while loop and does not terminate as long as unexplored polygon fragments exist. Except when POIs perfectly overlap at identical coordinates, the algorithm naturally reduces its radius and adapts to any extreme urban density. The proposed method is implemented in Python 3.12.3 and publicly available via GitHub to ensure reproducibility (Kikuchi, 2026).

4. EVALUATION / CASE STUDY

4.1 Study Areas and Target Category

To validate the efficiency of the proposed method, we conducted experiments focusing on two structurally distinct urban environments in Japan. As summarized in Table 1, these two areas were selected to represent different examples in urban spatial configuration and density:

a. **Shibuya City, Tokyo:** A highly dense, inner city of the capital characterized by hyper-concentrated commercial facilities and high-rise buildings.
b. **Chiba City, Chiba:** A middle-density, sprawling municipality consisting of multiple urban cores, suburban residential areas, and rural landscapes.

Municipality	**Population**	**Area (km²)**
Shibuya City, Tokyo	243,883	15.11
Chiba City, Chiba	974,951	271.76

Table 1: Profiles of the case study areas[1].

For this case study, we specifically targeted food and beverage establishments (i.e., restaurants and cafes) as the target POI category. Restaurants represent one of the most dynamic, densely clustered, and voluminous facility types in urban environments. Since "POI and associated review data are generally quite available for restaurant establishments" (Torrens, 2022), extracting restaurant POIs provides an ideal stress-test. This extreme spatial concentration, particularly in ultra-high-density areas such as Shibuya, clearly exposes the limitations of conventional sampling methods. Specifically, it frequently triggers API rate limits and leads to data gaps, thereby severely testing the adaptability of the algorithm we propose.

4.2 Evaluation Procedure

The experiment was conducted in two phases. First, we executed the proposed greedy method using the Google Places API (Nearby Search) to retrieve the actual POI datasets for both areas. The API requests were performed on 29th April 2025. As a result, after removing duplicate data between Shibuya City and Chiba City, the number of unique POIs was 9,100 and 4,642, respectively.

Furthermore, after excluding facilities that had been mistakenly included in the request results for Chiba City, the final dataset consisted of 4,055 unique POIs. To assess the representativeness of this data source, we compared it with the official business license data published by Chiba City at the end of the same month. The official data indicates an estimated 5,000 to 5,661 operational facilities, and the cleaned Google Places dataset was shown to cover approximately 70 to 80 percent of actual facilities. This aligns with existing studies indicating that Google Maps has a relatively high penetration rate (Sen & Quercia, 2018; Chen et al., 2024). Since the absolute true number of Google Places POIs in the real world is inherently unknowable due to the API's black-box nature, the

[1] Population and area data are based on the 2020 National Census conducted by the Statistics Bureau of Japan and the spatial surveys in 2025 by the Geospatial Information Authority of Japan.

https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-89-2026

comprehensive dataset retrieved in the first phase was defined as the "total discoverable POIs" for this study.

In the second phase, using the spatial distribution of total discoverable POIs, we simulated the retrieval processes of three conventional spatial sampling strategies to compare their efficiency and cost:

1. **Fixed Grid (50m):** A high-cost, high-resolution setting aimed at maximizing POI recall.
2. **Fixed Grid (250m):** A low-cost, low-resolution setting attempting to minimize API requests.
3. **Adaptive Quadtree:** A dynamic spatial subdivision method adjusting square meshes based on density.

By simulating these methods, we can quantitatively evaluate their comparative retrieval rate against the total number of discoverable POIs and the volume of redundant API calls.

4.3 Quantitative Results (Cost & Efficiency)

The simulation results shown in Table 2 summarize the trade-off between API call volume (cost) and the number of unique POIs retrieved (rate/recall). The cost is set to be 0.035 USD per request according to the price of “Places API Nearby Search Enterprise” (Google Developers, 2026b).

(a) Shibuya City

Method	Fixed grid		Adaptive quadtree	Proposed greedy
	50m	250m		
API calls	6,046	240	2,842	1,067
Retrieved POIs	8,605	3,111	9,100	9,100
Rate (%)	94.6	34.2	100	-
Est. cost (USD)	211.61	8.40	99.47	37.35

(b) Chiba City

Method	Fixed grid		Adaptive quadtree	Proposed greedy
	50m	250m		
API calls	108,655	4,348	1,503	610
Retrieved POIs	4,618	3,600	4,641	4,642
Rate (%)	99.5	77.6	100	-
Est. cost (USD)	3802.93	152.18	52.61	21.35

Table 2: Performance Comparison of Sampling Strategies[2].

The quantitative results highlight a stark contrast between the methodologies. The 250m fixed grid significantly reduced API calls but failed to achieve efficient coverage, particularly in Shibuya City, due to the API's 20-item return limit per request. Conversely, the 50m Fixed Grid achieved near-perfect coverage but generated a massive volume of redundant calls in the sparse regions of Chiba City, driving up the estimated financial cost to an impractical level. The proposed greedy method reduced total API calls by approximately 60% compared to the adaptive quadtree method. Given that the Google Places API offers a $200 monthly free tier, this drastic reduction in request volume physically shifts wide-area POI retrieval from being a prohibitively expensive corporate endeavor to a feasible task well within the free tier or a minimal academic budget.

[2] This “rate” is defined as the proportion of POIs detected by each method against the total number of POIs retrieved by the proposed method.

4.4 Qualitative Results (Visualization)

Visualizing the spatial distribution of the API requests further proves the algorithmic superiority of the proposed method (Fig. 2 and Fig. 3). The red dots are search points, and the blue circles show the search radii for each request. The visualizations clearly expose the geometric limitations of conventional strategies. In the fixed grid approach, the inherent mismatch between static square cells and the circular search areas of the API inevitably generates extensive overlapping boundaries (redundant requests) and interstitial blind spots. Similarly, while the quadtree method attempts spatial adaptation, it still defines facility agglomerations using rigid rectangular subdivisions. This geometric constraint forces the algorithm to generate an excessive number of redundant API calls along the irregular fringes of high-density districts.

In contrast, the proposed method dynamically deploys massive search circles (e.g., up to the maximum allowed radius) in empty suburban tracts of Chiba City, completing vast areas in a single request. Simultaneously, it organically packs tightly condensed, miniature search circles around major transit hubs like Shibuya Station. This mapping visually confirms that the algorithm automatically senses the "density" of urban commercial agglomeration and adapts its sampling geometry without any prior information.

5. DISCUSSION

5.1 Algorithmic Superiority in the Context of Urban Agglomeration

The superiority of the proposed method is deeply rooted in the fundamental nature of urban geography. Urban facilities, particularly commercial entities like restaurants and retail stores, do not distribute uniformly; they are highly clustered according to the principles of spatial economics and agglomeration, compared to residential locations (Jiang et al., 2015).

Because of this clustering, the Fixed Grid approach is fundamentally flawed: it forces a uniform spatial assumption onto a non-uniform reality. Although Adaptive Quadtree attempts to solve this problem by subdividing high-density regions, it still suffers from geometric mismatches when fitting a circular API search radius into a square bounding box. As a result, it generates nearly three times as many redundant API calls as the method we propose. The Greedy method transcends this limitation by rigorously exploiting the circular geometry of the API's own output, perfectly adapting to the irregular and asymmetric spatial distribution and density of urban functions.

5.2 Interpretation of the "Adaptive Radius" and Applications in Facility Location

Furthermore, the spatial sampling strategy proposed in this study can be fundamentally reinterpreted as a variation of the classic Facility Location Problem (FLP) and Maximal Covering Location Problem (MCLP) (Berman & Krass, 2002). Specifically, it represents a complex scenario where the underlying user demand (i.e., the spatial distribution of POIs) is entirely unknown a priori. In this analogy, each API search circle acts as a deployed "facility" with a strict capacity constraint (a maximum limit of 20 returnable POIs).

To efficiently distribute these capacitated facilities across an unknown landscape, our algorithm dynamically targets the largest remaining unsearched area. At first glance, this strategy

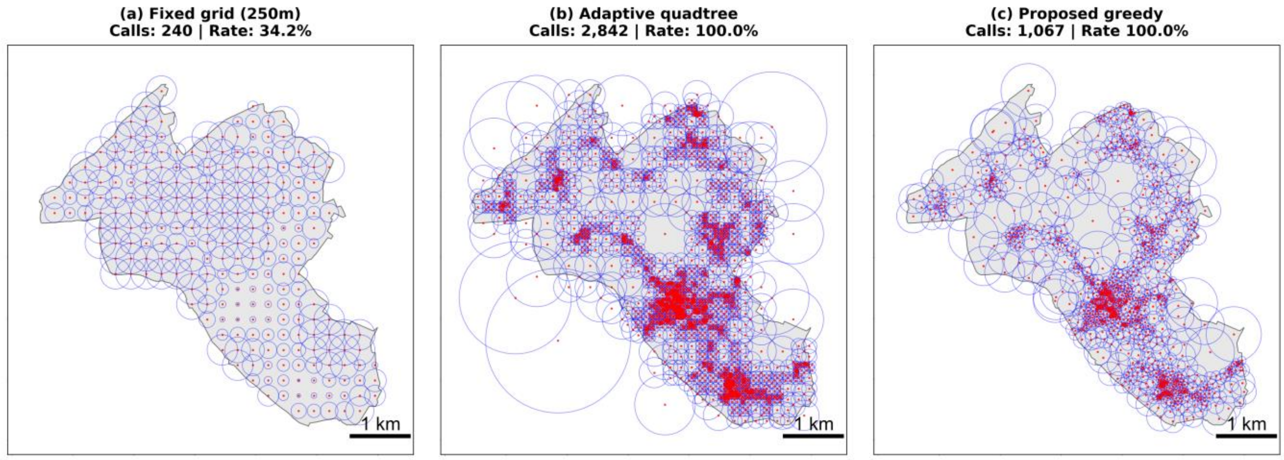


Figure 2. Spatial distribution of API requests in Shibuya City.

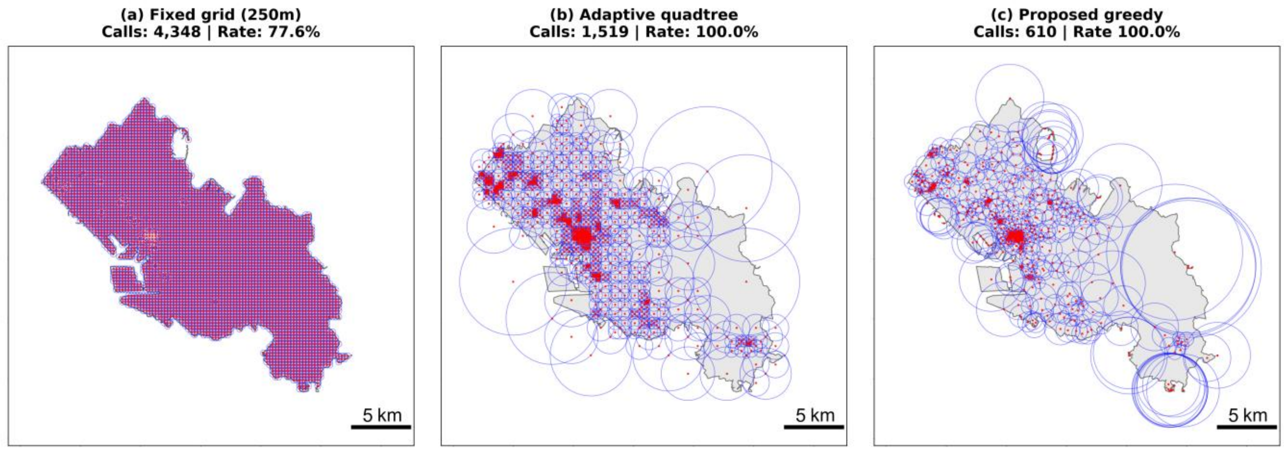


Figure 3. Spatial distribution of API requests in Chiba City.

might appear ad-hoc. However, operations research literature demonstrates otherwise. Greedy approaches provide robust and mathematically sound approximations for complex facility location problems (Resende, 1998; Jain et al., 2002; Cardei and Cardei, 2008). Our results align with these established findings. Even when treating the underlying demand distribution as a black box, our method maximizes spatial coverage while strictly respecting capacity limits. This demonstrates that classical greedy heuristics can also be highly effective for contemporary online spatial data extraction.

5.3 Democratization of Urban Analytics

The most profound contribution of this research lies in its potential to contribute to the democratization of spatial data science. Historically, exhaustive urban POI retrieval via commercial APIs such as the Google Places API incurred costs reaching hundreds or thousands of dollars per study, effectively monopolizing high-resolution and wide-area urban analytics to well-funded institutions. By optimizing the Google Places API retrieval process to maximize the utility of every single request, this tool can significantly reduce the financial burden to just a few dozen dollars. This potential increase in accessibility could empower students, independent researchers, citizens, and practitioners, especially urban planners in developing nations, to conduct robust, data-driven smart city research previously hindered by the "budget barrier."

Furthermore, the potential reach of this democratization extends beyond the Google Places API. Distance-ranked, capacity-limited responses to radius-based proximity queries represent a common design pattern across commercial POI data providers, suggesting that the proposed framework is not inherently API-specific. Even for APIs that support pagination, such as the Yelp Fusion API and the Foursquare Places API (Foursquare, 2026; Yelp, 2026), the proposed greedy algorithm offers a promising alternative when conducting wide-area searches that risk exceeding pagination or radius constraints, or when minimizing the total number of API requests is prioritized over issuing a single large-radius request encompassing the entire study area. Empirical validation of the proposed framework on these alternative APIs remains a direction for future work.

5.4 Limitations and Future Directions

Despite its high efficiency, the proposed method has structural limitations tied to the Google Places API's mechanics and dynamic nature. First, the algorithm relies entirely on the API's "distance-based ranking" feature. Any future deprecation or modification of this feature by the data provider would render the geometric subtraction logic invalid.

Second, a critical edge case arises from the Google Places API's 20-result limit when multiple POIs share the exact same geographical coordinates. If a cluster of POIs shares identical coordinates exactly at the perimeter of the search radius, and only a portion of them fits within the 20-result allowance, the

https://doi.org/10.5194/isprs-annals-XII-4-W2-2026-89-2026

algorithm will register that location as "searched." Consequently, the remaining unretrieved POIs at that exact coordinate will be permanently omitted. While Google Maps has been gradually enriching indoor mapping and floor-aware geocoding in major urban centers, POIs within high-rise commercial buildings and multi-level structures are still occasionally plotted at precisely the same geographic coordinates (e.g., the building's centroid). To address this potential data loss, the robustness of the algorithm can be enhanced by refining the determination of the effective radius. Specifically, setting the effective radius to "the distance to the farthest POI located within a distance shorter than the distance to the 20th POI" serves as an effective improvement. This reliably prevents POI clusters sharing the same coordinates from being split at the search boundary.

Third, as shown in Fig. 3, when the boundary of the target polygon is extremely complex or jagged, the algorithm will perform slightly redundant calls at the edges of the search area. However, this minor inefficiency can be easily mitigated by performing boundary smoothing as a preprocessing step.

Finally, due to the advanced adaptability of the sampling process, the total number of API calls required—and consequently the precise financial cost—cannot be accurately determined prior to execution. Although Section 4.3 demonstrates that the final number of API requests is significantly lower than other methods, researchers must begin retrieval without a guaranteed final budget. To mitigate this unpredictability, exploring whether OpenStreetMap (OSM) data density can be utilized to estimate the required number of API requests represents a promising future direction. While prior studies have demonstrated that OSM data coverage and completeness vary significantly across different regions (e.g., Gubert et al., 2025), investigating the correlation between OSM point density and the final API query count remains highly valuable for improving cost predictability before execution.

6. CONCLUSION

Comprehensive POI data is a prerequisite for advanced urban studies and smart city development. This study addressed the critical trade-off between spatial coverage and API retrieval costs by proposing an adaptive greedy spatial sampling algorithm optimized for the Google Places API. By dynamically utilizing API feedback to perform geometric set subtraction, the proposed method completely bypasses the inefficiencies of traditional fixed grid and quadtree sampling, while guaranteeing 100% retrieval rate with few exceptions. Experimental evaluations in Shibuya and Chiba Cities achieved an approximately 60% reduction in API calls compared to the adaptive quad-tree method, the best-performing conventional approach evaluated in this study, and by up to 99% compared to high-resolution fixed grid baselines.

Ultimately, this research provides an approach that reduces the cost of utilizing commercial geospatial data. By releasing the complete methodology as an open-source Python tool (Kikuchi, 2026), we hope to democratize high-resolution POI retrieval, fostering reproducible, wide-scale urban analytics for researchers and practitioners worldwide.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP25KJ1003. The authors also sincerely thank the anonymous reviewers for their valuable comments that helped improve this manuscript.

References

Arribas-Bel, D., 2014. Accidental., open and everywhere: Emerging data sources for the understanding of cities. Applied Geography, The New Urban World, 49, 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012

Batty, M., 2013. Big data, smart cities and city planning. Dialogues in Human Geography, 3(3), 274–279. https://doi.org/10.1177/2043820613513390

Berman, O., & Krass, D., 2002. The generalized maximal covering location problem. Computers & Operations Research, 29(6), 563-581.

Cardei, I., & Cardei, M., 2008. Energy-efficient connected-coverage in wireless sensor networks. International Journal of Sensor Networks, 3(3), 201–210. https://doi.org/10.1504/IJSNET.2008.018475

Chen, W., Zhou, Y., Stokes, E. C., & Zhang, X., 2024. Large-scale urban building function mapping by integrating multi-source web-based geospatial data. Geo-spatial Information Science, 27(6), 1785-1799.

Chvatal, V., 1979. A greedy heuristic for the set-covering problem. Mathematics of Operations Research, 4(3), 233–235. https://doi.org/10.1287/moor.4.3.233.

Elwood, S., Goodchild, M. F., & Sui, D. Z., 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. Annals of the association of American geographers, 102(3), 571-590.

Foursquare, 2026. Place Search. Foursquare Developer Documentation. https://docs.foursquare.com/fsq-developers-places/reference/place-search (1 July 2026)

Gao, S., Janowicz, K., & Couclelis, H., 2017. Extracting urban functional regions from points of interest and human activities on location ‐ based social networks. Transactions in GIS, 21(3), 446-467.

Goodchild, M. F., 2007. Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Google Developers, 2026a. Places API Nearby Search (New). Google Developers. https://developers.google.com/maps/documentation/places/web-service/nearby-search (8 March 2026)

Google Developers, 2026b. Google Maps Platform core services pricing list. Google Developers. https://developers.google.com/maps/billing-and-pricing/pricing (8 March 2026)

Gubert, F. R., & Silva, T. H., 2022. Google places enricher: A tool that makes it easy to get and enrich google places api data. In Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia) (pp. 91-94). SBC.

Gubert, F. R., Santos, G. H., Delgado, M., Silver, D., & Silva, T. H., 2025. Urban Cultural Signature with Web Data: A Case Study with Google Places Venues. Journal of Internet Services and Applications, 16(1), 388-406.

Huang, Y., Xiong, H., Leach, K., Zhang, Y., Chow, P., Fua, K., ... & Barnes, L. E., 2016. Assessing social anxiety using GPS trajectories and point-of-interest data. In Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing (pp. 898-903).

Jain, K., Mahdian, M., & Saberi, A., 2002. A new greedy approach for facility location problems. In Proceedings of the thiry-fourth annual ACM symposium on Theory of computing (pp. 731-740).

Jiang, S., Alves, A., Rodrigues, F., Ferreira Jr, J., & Pereira, F. C., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. Computers, Environment and Urban Systems, 53, 36-46.

Kikuchi, H., 2026. python-adaptive-google-places-poi-sampler Software, version 1.0.1. Zenodo. https://doi.org/10.5281/zenodo.21328864.

Kim, H., & Lee, S., 2024. POI GPT: Extracting POI Information from Social Media Text Data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4-W10-2024, 113–118. ISPRS WG IV/9
8th International Conference on Smart Data and Smart Cities (SDSC) - 4–7 June 2024, Athens, Greece. https://doi.org/10.5194/isprs-archives-XLVIII-4-W10-2024-113-2024.

Kitchin, R., 2013. Big data and human geography: Opportunities, challenges and risks. Dialogues in Human Geography, 3(3), 262–267. https://doi.org/10.1177/2043820613513388.

Liu, X., & Long, Y., 2016. Automated identification and characterization of parcels with OpenStreetMap and points of interest. Environment and Planning B: Planning and Design, 43(2), 341-360.

Long, Y., & Shen, Z., 2015. Discovering functional zones using bus smart card data and points of interest in Beijing. In Geospatial analysis to support urban planning in Beijing (pp. 193-217). Cham: Springer International Publishing.

Lu, C., Pang, M., Zhang, Y., Li, H., Lu, C., Tang, X., & Cheng, W., 2020. Mapping urban spatial structure based on poi (point of interest) data: A case study of the central city of Lanzhou, China. ISPRS International Journal of Geo-Information, 9(2), 92.

Martí, P., Serrano-Estrada, L., & Nolasco-Cirugeda, A., 2019. Social Media data: Challenges, opportunities and limitations in urban studies. Computers, Environment and Urban Systems, 74, 161–174. https://doi.org/10.1016/j.compenvurbsys.2018.11.001.

Resende, M. G., 1998. Computing approximate solutions of the maximum covering problem with GRASP. Journal of Heuristics, 4(2), 161-177.

Samet, H., 1984. The quadtree and related hierarchical data structures. ACM Computing Surveys (CSUR), 16(2), 187-260.

Sen, R., & Quercia, D., 2018. World wide spatial capital. PloS one, 13(2), e0190346.

Shelton, T., Poorthuis, A., & Zook, M., 2015. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. Landscape and Urban Planning, 142, 198–211. https://doi.org/10.1016/j.landurbplan.2015.02.020.

Torrens, P. M., 2022. Data science for pedestrian and high street retailing as a framework for advancing urban informatics to individual scales. Urban Informatics, 1(1), 9. https://doi.org/10.1007/s44212-022-00009-x.

Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., & Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. International Journal of Geographical Information Science, 31(4), 825-848.

Yelp, 2026. Business Search. Yelp Developer Documentation. https://docs.developer.yelp.com/reference/v3_business_search (1 July 2026).

Yu, L., Yu, T., Wu, Y., & Wu, G., 2020. Rethinking the identification of urban centers from the perspective of function distribution: A framework based on point-of-interest data. Sustainability, 12(4), 1543.

Zhai, W., Bai, X., Shi, Y., Han, Y., Peng, Z. R., & Gu, C., 2019. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. Computers, environment and urban systems, 74, 1-12.