Thomas Tregel, Lukas Raymann, Stefan Göbel, Ralf Steinmetz: Geodata Classification for Automatic Content Creation in Location-based Games. In: Mariano Alcañiz, Stefan Göbel, Minhua Ma, Manuel Fradinho Oliveira Jannicke Baalsrud Hauge, Tim Marsh: Proceedings of the 3rd Joint International Conference on Serious Games, p. 212 - 223, Springer, November 2017. ISBN 978-3-319-70110-3.

Geodata Classification for Automatic Content Creation in Location-based Games

Thomas Tregel, Lukas Raymann, Stefan Göbel, Ralf Steinmetz

Multimedia Communications Lab - KOM
Technische Universität Darmstadt, Darmstadt, Germany
{thomas.tregel,lukas.raymann,stefan.goebel,
ralf.steinmetz}@kom.tu-darmstadt.de

Abstract. The ubiquity of smartphones with integrated positioning systems made it possible to develop location-based games playable by most people without requiring additional equipment. Popular location-based games like Ingress or Pokémon Go have demonstrated the public interest in this genre and studies indicate that playing such games has a positive health influence related to the players' increased movement.

A big development challenge for these games is the content creation. With increasing area coverage the amount of potential players increases, which leads to a larger community. In location-based games points of interest for game locations can be either generated semi-randomly, crowdsourced using user-generated content approaches or created by a game designer.

Manually selecting points of interest (PoIs) of the real world for a game is time-consuming, expensive and often results in heterogeneous distributed PoIs with a low concentration in rural areas. It is thus worthwhile to use automatic approaches instead, where PoIs are selected automatically from available geodata services based on their relevance. Relevant PoIs can for example be identified through metadata they are assigned with within the system. Using such an approach gives even small game development studios with limited budget the possibility to develop location-based games that can be played all around the world. This is in particular interesting for the development of serious games, which tend to have a lower budget compared to entertainment-based games.

In this paper we present a system that uses georeferenced data from open available sources to generate a collection of PoIs usable for location-based games of free definable target groups with the goal of providing a comparable game experience everywhere. The content creation algorithm in our approach is fully parametrized allowing for individual configuration for desired PoI criteria.

Keywords: Geodata, Procedural Content Generation, Location-based Games, Points of Interest

1 Introduction

Location-based games have moved into the center of society with the launch of Pokémon Go when it became a global phenomenon. It was released on July 6th, 2016 and quickly built a user base of up to 45 million users worldwide (The Guardian, 2016).

One problematic factor for these types of games is content creation. Because players use their current geographical position as their main input for the game, a player's game experience is dependent from the content in his surroundings. For game developers there are three options of content creation: (i) (semi-)random, (ii) crowd-sourced / user-generated content, (iii) manually created by content creators.

In Pokémon Go's precursor Ingress a crowd-sourcing approach with manual verification was chosen which allowed players to suggest PoIs. This system was suspended in September 2015 (Niantic, Inc.) after having processed 15 million submissions with 24 million submissions in the backlog. For small development teams often present for serious games this amount of work is infeasible. Semi-random location-based content creation has the major downside that content may pose a risk for the player, the surrounding people or the environment by providing content which induces players to trespass on private property or leading players to places which may pose a risk to their health.

Thereby for catering globally available game content a system is required that is based on existing geodata systems, which extracts and transforms the data into location-based game elements, while trying to verify their reachability.

2 Related Work

Due to the ubiquity of smartphones and their incorporation of location sensors the latest location-based games are developed for and run on smart devices. Using a GNSS (Global Navigation Satellite System) like the Global Positioning System (GPS) users can be located with an accuracy of 7.8meters in a confidence interval of 95% using only the service provided for civilians (Grimes, 2008).

For location-based games this accuracy is relevant as the reachability of PoIs is an important factor. During the time span of Ingress's crowd-sourcing approach the developer Niantic handed out a guideline for high quality location criteria (Niantic, Inc.):

Possible well-suited location candidates are:

- A location with a cool story, a place in history or educational value
- A cool piece of art or unique architecture
- A hidden Gem or hyper-local spot
- Public libraries
- Public places of worship

Excluding locations with one of the follow aspects:

- No safe pedestrian access
- Private residential property
- Candidates that may interfere with the operations of fire stations, police stations and hospitals

Because this game depends on the player input and the submissions of potential PoIs the actual game elements within the game world are distributed heterogeneously. Especially in suburban and rural areas players report a worse game experience due to the lack of sufficient game content both in quantity and in diversity (Hargarten, 2016), which the developer has reacted on (Hoffer, 2016). However there still exist major differences between urban, suburban and rural areas.

Incorporating game content creation and open available geodata has been researched extensively with the focus on providing geographically accurate content elements in combination with the best practices in procedural content generation.

3 Concept

The goal of our concept is to provide a system based upon an open available geodata system that examines and processes the available data to extract a georeferenced set of PoIs, which suffice given quality criteria.

3.1 Data analysis and data availability

We choose OpenStreetMap (OSM) as geodata source system because of the access options to the underlying map data. Data completeness and correctness is an additional factor which OSM tries to manage with its high user base.

Regarding quantity and correctness of OSM data, studies have shown that investigated map providers are comparable in the given aspects. Heterogeneous data distribution has been a problem for all map providers with less data available for rural areas (Ciepłuch, Jacob, Mooney, & Winstanley, 2010) (Neis & Zielstra, 2014). Due to the at least linear increase in active users and tagged locations OSM tries to reach an improved coverage (OpenStreetMap).

In an exemplary analysis we checked the data availability for multiple tag groups in OSM according to the previously described acceptance criteria for well-suited PoI candidates. Tags for historic relevant locations and places of worship appeared to be especially suitable due to their frequent occurrence and worldwide distribution.

For the tags which describe places of worship, over 806.000 locations have been tagged with a distribution depicted in Figure 1. These places are especially valuable for PoI usage, as buildings of religious groups exist in most countries and represent a cultural and often architectural interesting location fulfilling multiple of the previously described acceptance criteria.

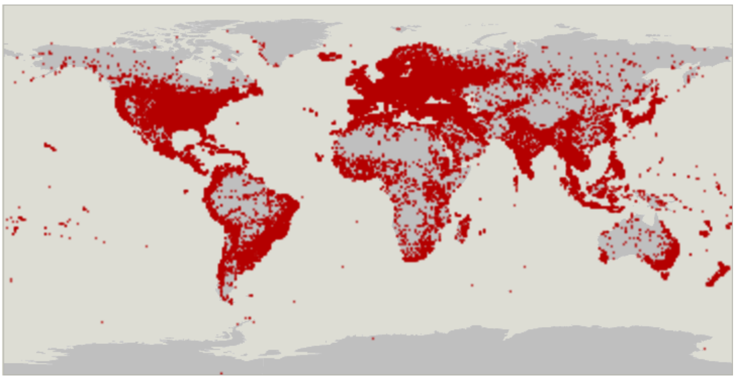

Figure 1: Data distribution using OSM for places of worship.

3.2 Classifier creation

As described in Section 2 the discrepancy between urban, suburban and rural area in current location-based games is clearly perceivable and influences the player experience.

In order to create an ideally balanced distribution of PoIs and thereby meaningful gameplay locations, we selected specific tags or group of tags that are well-known to be available in urban, suburban and rural areas like town halls or educational establishments. Subsequently, we analyzed the tag distributions for candidate tags in order to verify their applicability. Each tag or group of tags, which we henceforth call classifiers, is assigned a numerical value representing its estimated relevance or representativeness for the surrounding area. This was driven by the assumption that e.g. town halls might be more expected to be a PoI in a game, than the public benches located in front of them.

As the frequency of possible PoIs in urban areas is relatively high most of our developed classifiers try to increase the coverage in more rural areas. Nonetheless, tags that are likely to be found in all types of areas are chosen with high priority like bus stop or stations for public transport in general. However we do not claim completeness for the chosen classifiers as they can be extended and modified in order to cover missing areas or corner cases we did not consider yet.

Thereby the classifier table shown in Table 1 is a first prototypical list of classifiers that aim to cover as much area types as possible, while respected corner cases like parks, which have a special characteristic in urban areas.

Classifiers	Priority	Tags
historic places	21	historic = *
places of worship	20	amenity = place_of_worship
places of the categories arts, culture and tourism	19	tourism = information | attraction | viewpoint | museum | artwork | theme_park | zoo | gallery amenity = arts_centre | cinema | community_centre | fountain | planetarium | studio | theatre

		man_made = windmill leisure = water_park
town halls	18	amenity = townhall
libraries and public bookcases	17	amenity = library | public_bookcase
places for picnic or barbecue	16	amenity = bbq | picnic_table tourism = picnic_site leisure = picnic_table shelter_type = picnic_shelter
huts and other shelters	15	amenity = hunting_stand | shelter leisure = bird_hide building = hut tourism = alpine_hut
mountain peaks	14	natural = peak
playgrounds	13	leisure = playground
educational establishments	12	amenity = college | school | university | dancing_school | music_school | language_school building = college | school | university
places for doing sport	11	sport = *
places for food or drinks	10	amenity = cafe | drinking_water | fast_food | food_court | ice_cream | restaurant
product shops and services	9	shop = * craft = * amenity = pharmacy | bank
stations and stops for public transport	8	amenity = bus_station railway = station | halt | tram_stop public_transport = stop_position highway = bus_stop
parks	7	leisure = park
pedestrian walkways	6	highway = pedestrian
benches	5	amenity = bench
public communication	4	amenity = post_box | telephone
waste disposal	3	amenity = waste_disposal | waste_basket | recycling
wells, towers, survey points	2	man_made = survey_point | tower | communications_tower | water_tower | water_well
trees, stones, springs	1	natural = tree | stone | rock | spring

Table 1: The selected classifiers with their priority order and tags.

3.3 Position handling to increase persistency

In location-based games the player's actual position is his main input and control option. According to the player's current position PoIs for the surrounding area could be extracted and processed into game elements. Due to the high amount of data available in certain areas, only a subset of candidate PoIs can be used as game relevant locations, as an oversupply leads to smaller spatial distances and thereby less actual movement between locations.

A first approach would be to define a radius around the player and to filter all available PoIs according to their priority and e.g. choose the two best ones as seen in Figure 2. Due to the non-static approach of the PoI extraction and its filtering a player in the given example that approach the PoI with the number five might see it disappear when movement is respected and PoI with number nine, with an even higher priority comes within range as seen in Figure 2. This behavior could be circumvented by not allowing PoIs to disappear during movement, which however leads to a higher amount of PoIs than intended and does not solve the problem of game world consistency, as players with different starting positions might encounter a different game world.

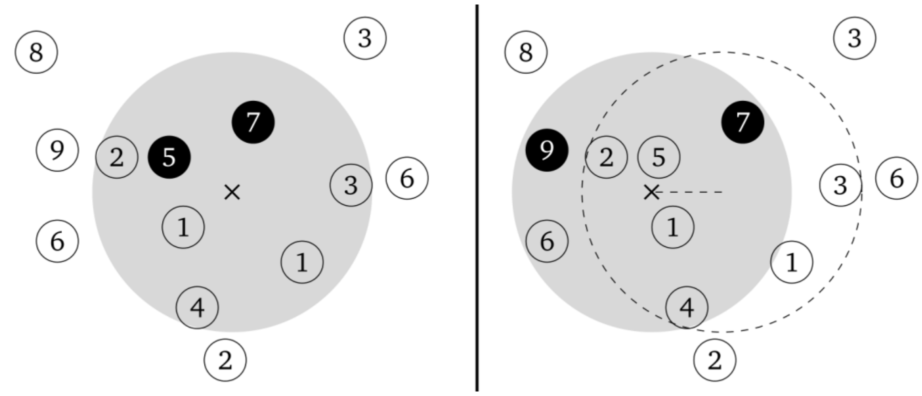

Figure 2: PoI selection example during a horizontal movement.

Our approach to encounter that problem is using a representation that divides the earth in non-overlapping contiguous polygons of similar size henceforth called cells. Thereby depending on the player's position the cell he is positioned in can be used for content generation. For all eight adjacent cells of similar size this is done accordingly, which leads to a consistent selection of game elements as illustrated in Figure 3.

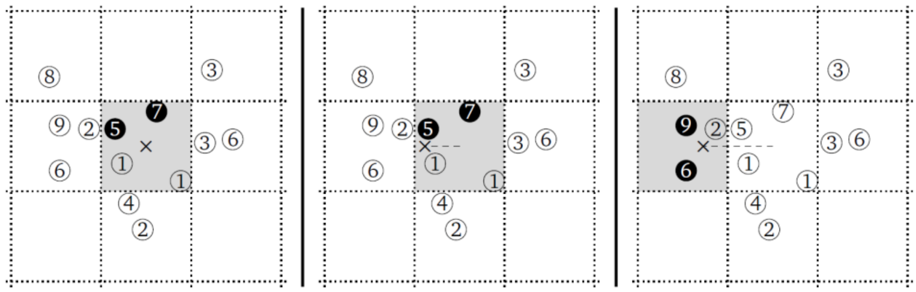

Figure 3: Our approach for the selection of PoIs during player movements.

3.4 PoI clustering

Depending on the chosen cell size (which depends on the specific game), areas can still have sizes of multiple square kilometers, which may lead to a bad distribution within the given cell. This is further enforced by cells that render the transition between different types of areas.

Another problem is the spatial clustering of candidate PoIs like at a university campus. When using a high priority classifier filtering for education-related buildings a high number of PoIs in close proximity would be extracted.

We aim to provide a close to equal distribution of PoIs over all different areas, in order to reduce the discrepancy between area types. In addition, spatial clustering of PoIs leads to central hubs, players may feel obliged to travel to and stay at rather than using the whole content diversity.

For that reason we use a clustering mechanism that finds a given amount of clusters and extract their most relevant representative. That way we eliminate spatially close candidate PoIs to be chosen.

To further increase spatial diversity we separate the given geographic cell into two levels of subcells as seen in Figure 4 and use them as starting positions for the respective clusters. For the final PoI selection we enforce a distribution over the given subcells which we further describe in Section 4.3.

0	3	4	5
1	2	7	6
14	13	8	9
15	12	11	10

Figure 4: Subcells of a cell. The numbers are assigned according to a Hilbert-Curve.

3.5 Metric development

In order to distinguish the quality of the candidate PoIs during the selection process and to compare multiple PoI sets we developed multiple metrics:

1. Number of PoIs: For our approach it is desirable to be able to control the number of PoIs per cell, in order to present the “right” amount of game content to the player for the respective game. As discussed in Section 2 this choice defines the possible playstyles and dictates the pace of the game.
2. Equal distribution of PoIs: Besides the number of PoIs their distribution in the game area is important as discussed in Section 3.4. An unequal distribution might lead to players being able to easily reach a limited number of PoIs before having to travel long distances for more content. The challenge here lies within the data availability as there might be no relevant candidate PoI for a given area due to sparsity.
3. PoI priority focus: During the PoI selection process this metric controls the weight the classifier’s priority has. Enforcing a priority focused approach leads to a result with more meaningful PoIs that can be identified and recognized by players.

4. PoI diversity focus: In order to increase the diversity of selected PoI a metric tries to maximize the number of PoIs resulting from different classifiers. When using this metric alone in the selection process a high classifier diversity can be reached, however resulting into many low priority candidates being chosen.

These metrics can be freely combined and integrated into a weighted sum.

4 Implementation

In order to test and evaluate our concept we implemented both a desktop version for data analysis and a mobile version for smartphones for live location handling. Both versions are based upon the same basic module.

4.1 S2 Geometry

As described in Section 3.3 a cell based approach is chosen in order to support persistency between multiple application cases. Because the targeted cell size may vary depending on the actual application, we chose a hierarchical model. Thereby the cell size can be adapted to the application's needs.

In our implementation we use the S2 Geometry Library because it provides an efficient approach to identify the relatedness between cells and their subcells due to their similar ID-prefix.

4.2 Overpass query creation and limitations to avoid trespassing

OSM offers the Overpass-API for custom queries, which enable the selection of specific tags and the specification of a target area using bounding boxes. The data is specified using the keywords node, way and relation.

While nodes can be directly translated into candidate PoIs by their respective coordinate, ways need specific handling. A way can either represent a real way with a start- and end-point or an area enclosed by the given polygon. For these ways the Overpass-API defines the area type, which can be used to obtain information about e.g. the area's center. For our implementation an area is treated as a PoI because further investigation would be needed to calculate the reachability of the area's center like for e.g. lakes or non-public buildings.

In order to exclude PoIs that are located in area that are either dangerous or private property we apply a filter that uses tags indicating the landuse for military or private matters, as trespassing can become a serious problem for location-based games.

4.3 Specifying the classifiers

To further reduce the amount of data sent for the mobile implementation we implemented a dynamic reloading approach that widens the range of classifiers from rare to common for areas or subareas that contain an insufficient amount of PoIs. The classifiers' placement into each rarity category has been done according to their assigned priority.

Additionally each classifier can assign a subpriority to their individual tags to e.g. indicate the higher priority of cathedrals over chapels for all places of worship.

The developed metrics are all incorporated into the PoI selection process. Hereby the number of PoIs (1) is used as an input value for the clustering algorithm leading to the desired outcome. By using subcells of each cell the equal distribution (2) is tackled, by enforcing the coverage of these subcells. Regarding metrics (3) and (4) a weighted approach was implemented that aims to maximize the weighted sum score of both approaches. Due to the wide range of possible application scenarios these weights are mapped onto the main configuration UI.

5 Evaluation

To evaluate the implementation of our concept we have applied it to 24 positions in 16 different countries and examined the results in regard of the metrics introduces in section 3.5. The places were grouped into three categories:

1. places with more than 100,000 residents
2. places with 5,000 - 20,000 residents
3. places with less than 2,000 residents

Comparing the results of the categories showed that the average number of PoIs determined in an area is much greater in places with many residents. Figure 5 shows the amount of elements detected by the classifiers for the different places. After applying the clustering algorithm the amount of PoIs was as depicted in Figure 6.

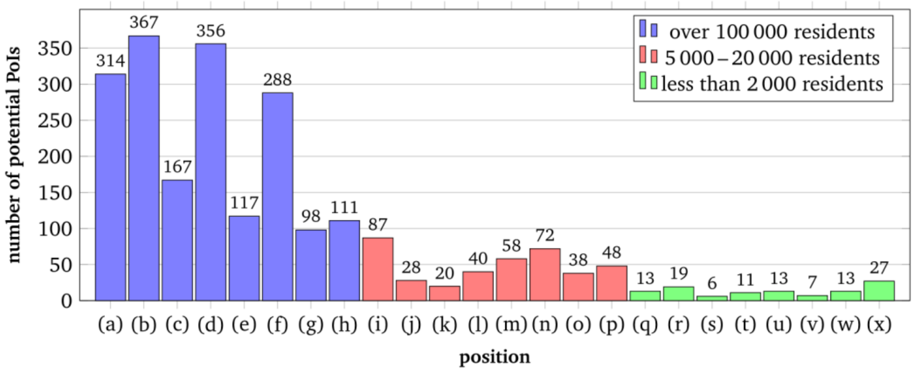

Figure 5: The amount of potential PoIs for the examined places.

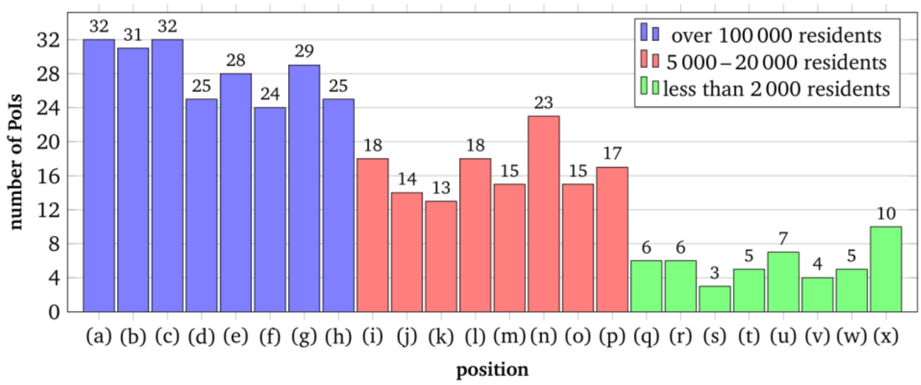

Figure 6: The amount of PoIs for the examined places after the clustering was applied.

The amount of PoIs can easily be customized by increasing the amount of clusters and of simultaneously used classifiers. Figure 7 shows an example of such an increase for the center of Darmstadt. While the left side shows the result for the standard values with 32 clusters, the right side demonstrates the result when 256 clusters are used and all classifiers are applied at once.

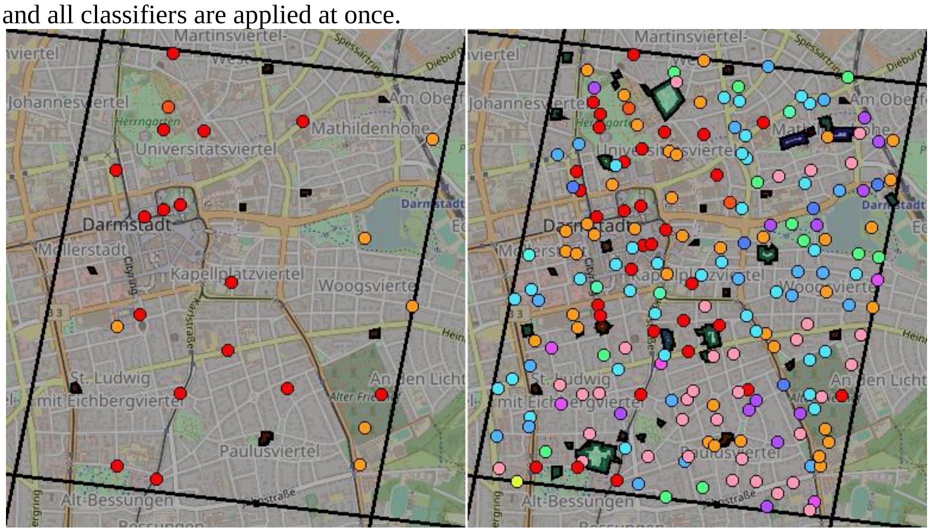

Figure 7: PoIs in Darmstadt before and after the increase of the amount of clusters.

To examine the distribution of the PoIs we constructed a minimal spanning tree for each cell with the PoIs being the nodes. By doing this we could estimate how far players would approximately walk to get from one PoI to the next one. The cells have a size of between 3.31km^2 and 6.38km^2. As depicted in Figure 8 the distance becomes larger for places with fewer residents. For places with less than 2,000 residents when examining the standard deviation the PoI distance is within the [213m, 862m] interval. For cities with 5,000-20,000 residents the interval changes for the upper limit to [194m, 538m] and for cities with over 100,000 residents to [215m and 444m]. Thereby the results are stable for the minimum distance, which is the result of our clustering and selection approach. For the upper limit the distance decrease with higher PoI availability, hence more PoIs are available in larger cities.

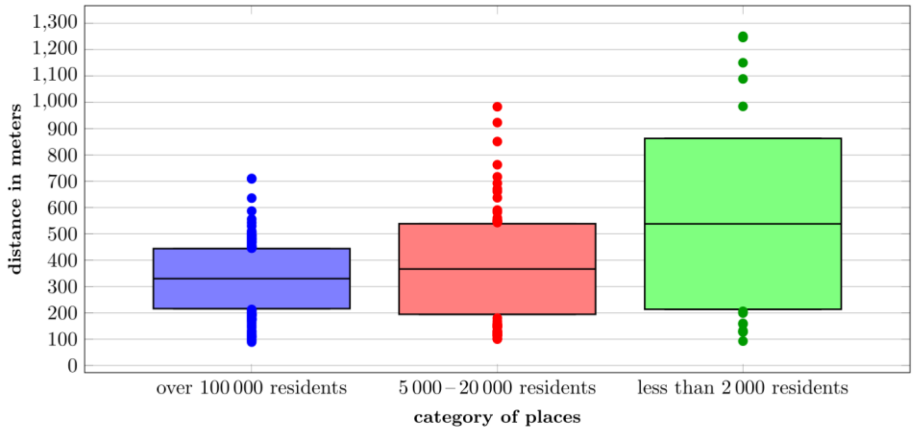

Figure 8: Distances between PoIs in the minimal spanning tree, as well as the standard deviation from the mean value.

In the next step we investigated the PoIs' priority values for each of the places. PoIs in places of the first category had an average value of 20.52, PoIs in the second category an average value of 14.61 and PoIs in the third category an average value of 17.12. The high last value can be explained by the fact that some of the classifiers with high priority values are also effective for rural areas, resulting e.g. in PoIs for touristic places or churches.

Finally we examined the amount of used classifiers. On average places of the category 1 used 6.25 classifiers, places of the category 2 used 8.25 and places of the category 3 used 3.88. The reason for the first value being lower than the second is that in cells of category 1 most of the times fewer queries are needed to get enough data to fill the cell, so that fewer classifiers are applied to such cells.

6 Conclusion

In this paper we presented a concept for automatically determining relevant PoIs for location-based games. Publicly available geodata has been analyzed using a set of developed tag groups. The 21 tag groups have been designed to value the usefulness and interestingness among the PoIs and to reach high area coverage among heterogeneous areas. A clustering algorithm is applied to the respective data, selecting the highest priority PoIs as the representative location within the cluster. These clusters are then to be used in location-based games of different scopes of application.

Our evaluation showed the concept is applicability for various cities and towns in different countries as well as the customization options regarding the amount of desired PoIs.

Further research will focus on optimizing the list of classifiers to find more PoIs in very sparsely populated areas and modifying the classifier hierarchy accordingly.

Classifiers based upon meta-data for fallback scenarios can prove to be useful in areas without enough available data or for aspects not actively tagged within the data like street crossings.

References

Ciepłuch, B., Jacob, R., Mooney, P., & Winstanley, A. C. (2010). Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. *Proceedings of the 9th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (pp. 337-340). University of Leicester.

Grimes, J. G. (2008, September). Global Positioning System Standard Positioning Service Performance Standard. Washington, D.C., USA: U.S. Department of Defense.

Hargarten, J. (2016, August 3). *Why some places in Minnesota are better for playing Pokémon GO than others.* Retrieved November 19, 2016, from StarTribune: http://www.startribune.com/urban-vs-rural-why-playing-pokemon-go-in-the-suburbs-could-be-hurting-your-game/388985262

Hoffer, C. (2016, 12. 17.). *Pokemon Go Increases Spawns in Rural and Suburban Areas*. Retrieved April 26, 2017, from WWG: http://wwg.com/pokemon/2016/12/17/pokemon-go-increases-spawns-in-rural-and-suburban-areas/

Neis, P., & Zielstra, D. (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, pp. 76-106.

Niantic, Inc. (n.d.). *Candidate Portal criteria*. Retrieved April 26, 2017, from https://support.ingress.com/hc/en-us/articles/207343987-Candidate-Portal-criteria

OpenStreetMap. (n.d.). *Stats*. Retrieved April 26, 2017, from http://wiki.openstreetmap.org/wiki/Stats

The Guardian. (2016, April 23). *Pokémon No: game's daily active users, downloads and engagement are down*. Retrieved April 26, 2017, from https://www.theguardian.com/technology/2016/aug/23/pokemon-go-active-users-down-augmented-reality-games