LocXplore: A System for Profiling Urban Regions

Demo Paper

András Komáromy
Freie Universität Berlin
Germany
andras.komaromy@fu-berlin.de

Paras Mehta
Freie Universität Berlin
Germany
paras.mehta@fu-berlin.de

ABSTRACT

Finding information about regions in a city (e.g., neighborhoods) is important for several applications, particularly where users' interests are broad, thus making recommendation of larger areas instead of specific Points of Interest (POIs) more suitable, or where users would like to learn more about the urban context of specific POIs. In this paper, we present LocXplore, a system for profiling regions of a city to support people planning to move to the city. The system integrates and analyzes data from diverse sources, including location-based social networks, OpenStreetMap and Open Data from government authorities and news agencies, and allows users to search, explore, and compare regions of interest. Our approach is based on iterative user-centered design through user research within our target group and draws from relevant research in the fields of urban computing and urban sociology. The system includes a web-based user interface that provides a comprehensive way to visualize and explore the different regions in a city, and its functionality is demonstrated for the localities in the city of Berlin.

CCS CONCEPTS

• **Information systems → Information systems applications**;

KEYWORDS

regions of interest, location-based social networks, search, exploration

ACM Reference Format:

András Komáromy and Paras Mehta. 2018. LocXplore: A System for Profiling Urban Regions: Demo Paper. In *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '18), November 6–9, 2018, Seattle, WA, USA.* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3274895.3274924

1 INTRODUCTION

Information about regions in a city (e.g., neighborhoods) serve as important data sources for those who are interested not only in certain POIs, but in larger areas where they can best satisfy their needs in multiple nearby POIs of different types, or for those who would like to learn more about the urban context of a specific POI. For example, consider a visitor looking to go for dinner and perhaps a drink might prefer a recommendation for an area containing several restaurants and bars, over a single venue. Similarly, someone searching for an apartment in surrounded by green spaces and close to the lake would also benefit from suggestions for areas that might suit her preferences. In this work, we focus on the scenario of *looking for a neighborhood for living*, taking Berlin and its administrative regions as an example for the application.

Existing research on location recommendation mainly focuses on POIs, such as restaurants and hotels. Although there are some contributions towards defining and retrieving Regions of Interest (ROIs), there is a significant lack of contributions studying the problem of mining region-specific topics to support exploration [3]. Similarly, as discussed in [11], there is related work in the areas of urban computing on discovering and analyzing regions of interest, e.g., finding functional regions in a city [10]. Nevertheless, research in the areas of region profiling and region exploration can benefit significantly from an interdisciplinary approach, integrating research on region profiling from sociology and taking real user requirements into account during the design process.

2 REGION PROFILING

The concept of a neighborhood in an urban context is an ambiguous one. Both in urban computing [3, 7] and in urban sociology [2, 4, 5] various approaches exist to define urban regions. Nevertheless, all definitions have one characteristic in common – they view regions as complex entities with which a collection of features can be associated. In our work, we consider that the administrative regions of a city represent a suitable unit of reference for users looking for an area to live in. Based on this, we focus on the challenge of determining the attributes of regions that are important for the selected use case. It is important to note that focusing on a specific use case is crucial, since the characteristic attributes of regions can vary depending on the application context.

2.1 Features for characterizing neighborhoods

To determine the relevant attributes for our system and to learn more about the users' perspective, we conducted eight semi-structured interviews with potential users based on the guidelines provided in [6]. The structure of the interview sessions followed a so called *funnel model*, which starts with open questions and successively moves towards more specific ones[1]. The interview questions cover a range of topics, starting from the general concept of neighborhoods and scenarios in which the notion of neighborhoods are

[1]The guideline for the interviews and other documents related to the user research can be found under http://userpage.fu-berlin.de/komaromy/

SIGSPATIAL '18, November 6–9, 2018, Seattle, WA, USA

ACM ISBN 978-1-4503-5889-7/18/11.
https://doi.org/10.1145/3274895.3274924

Table 1: Summary of user requirements from interviews.

Labels	Actions	Views
history, reputation facilities (shops, cafés, etc.) prices residents safety guide reviews	explore, search, compare	description
activities	compare, search	description (time-aware)
streets buildings vibe	explore, search	gallery
street network land use infrastructure location within city boundaries	explore, search	map

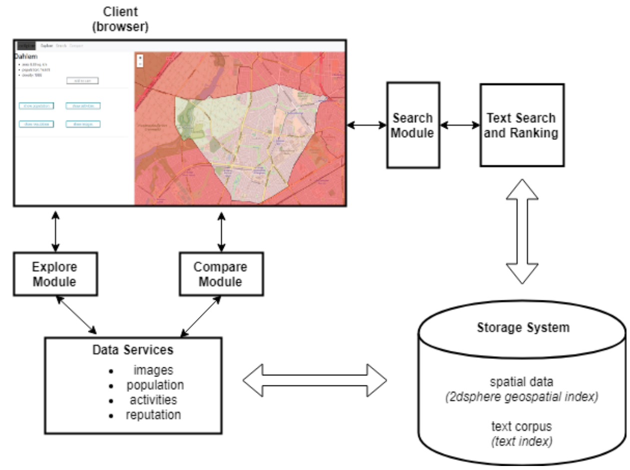

Figure 1: Architecture of LocXplore.

relevant, through the task of finding a place to live in a city (e.g., the strategies and tools used), to the description of the user's own current neighborhood.

As a preparation for the interviews and as a reference for potential region attributes, we studied attribute lists provided by prominent sociological approaches for describing the nature of neighborhoods [2, 4]. The results of the interviews together with these attribute lists served as a basis for identifying the main candidates for the system components, for the personas, and the user stories.

During the interviews, we used labels to gain a structured overview of users' preferences and mental concept of the domain. Moreover, these labels also serve as candidates for the main components of the user interface and provide an indication on how to structure and organize these components. As a result of this analysis, we provide the following in Table 1:

- the labels representing relevant information that users expect from our system,
- the related actions, i.e., features of the system, that users connect with the aforementioned information,
- and the possible views or components in the application that could satisfy the users needs.

2.2 Data sources

Besides user experience, another important criterion for judging the performance of our system is the quality and amount of data provided. As a result, next we describe the data sources used to meet the defined user requirements.

In our application model, the major geographic object corresponds to a locality of Berlin. It consists of a spatial component, i.e., a well defined geometry, and of a description, i.e., a profile.

2.2.1 Administrative boundaries and demographics. Since all data that the system should provide is related to localities of the city, the spatial representations of these localities are our basic reference. In the case of Berlin, the boundaries of local neighborhoods are provided by the Statistical Office for Berlin-Brandenburg[2]. We also collected official data on demographics to provide an overview of the population of each region. These include population distribution by age, gender, and nationality.

2.2.2 Land usage and user mobility. To provide an overview of the land usage, we utilized two different types of data, namely land usage information from OpenStreetMap and POIs from Foursquare. Moreover, to capture a more dynamic picture of regions, we used data from LBSNs. As Yang et al. [9] show, activity in LBSNs reflect collective user behavior and allow an understanding of behavioral differences between regions by studying mobility patterns over time. We follow a similar approach and use the dataset provided by the authors[3] for characterizing localities. The dataset includes long-term global-scale checkin data collected from Foursquare.

2.2.3 Images. Another important aspect for enabling exploration is to provide users a visual impression of the area. For this, we leveraged the popularity of photo sharing in social networks and extracted geotagged photos lying within the regions from the YFCC100M Flickr dataset [8].

2.2.4 Reputation. Our next task was to infer the reputation and description for the localities by finding texts related to the localities and extracting key concepts automatically. For this, diverse data sources, including user reviews and large text corpora of different genres available online can be used. For our application we first explored the NOW (News on the Web) corpus[4] that contains 5.9 billion words of text from online newspapers and magazines from 2010 to the present time. The main motivating factor for this decision was a high level of neutrality (a formulated requirement in the user research phase) guaranteed by the genre of newspaper articles and the up-to-date nature of the data. A manual check for available documents on the Berlin localities however delivered sparse results with no matches for localities that are not central or popular. A look into other large newspaper corpora (British National Corpus, Times Corpus, Corpus of Contemporary American English) had the

[2] https://daten.berlin.de/datensaetze/geometrien-der-ortsteile-von-berlin-juli-2012

[3] https://sites.google.com/site/yangdingqi/home/foursquare-dataset

[4] https://corpus.byu.edu/now/

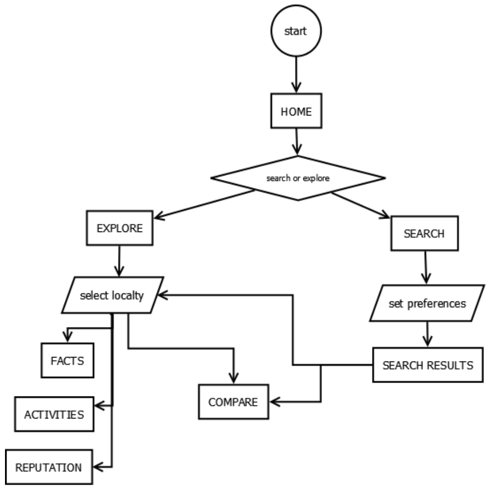

Figure 2: Process chart.

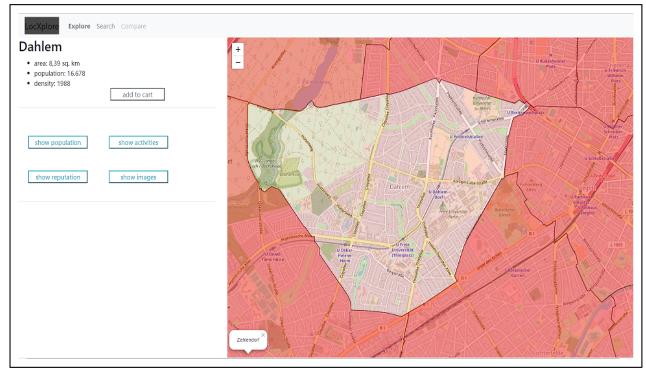

Figure 3: Explore module.

same sparsity with even worse results for more popular localities. The most matches could be found in the Wikipedia corpus which is a dump of Wikipedia entries - containing mostly descriptive information on the localities. The problem with finding an existing corpus that suits our needs is due to the restriction of the language and the genre.

Another challenge is posed by the fact that some results contain information not only on the queried locality, but other localities as well. For this reason it is difficult to relate the texts to the localities automatically.

The problem could also suggest a different computational linguistics approach, namely an analysis which is not based on separate documents but on the occurrences of tokens (in our case, the names of the localities) and their verbal contexts (e.g., +/-6 words). The co-occurrences are computed based on their frequencies in the given context in relation to other contexts. In this case, an appropriate corpus should contain a large number of occurrences, i.e., a lot of mentions of the localities, to be able to get significant results.
As a result of these challenges, our approach is based on a trade-off between quality and quantity. Since the publicly available large corpora with millions or billions of words resulted in sparse data, we chose a qualitative approach and build a small corpus with each document describing the localities. With this decision, the results may not represent collective descriptions, but we are able to retrieve relevant keywords for each of the localities. To this end, we use local descriptions online targeting people moving to the city[5] and Wikipedia entries.

3 SYSTEM MODULES

The architecture for our system is depicted in Figure 1. It comprises a database server for storing and accessing the data, a spatio-textual index to enable search, and a web-based user interface. We collected data from several sources, preprocessed it, built region profiles by associating it with regions and aggregating it, and stored it in a NoSQL database (MongoDB). The textual descriptions and keyphrases for each of the regions were generated automatically from text corpora using an extractive document summarization technique called TextRank [1], which is based on the PageRank algorithm. To support search, we built a geo-textual index on top of this data using ElasticSearch, in order to retrieve and rank relevant results. Angular and Leaflet frameworks were used for frontend development and visualizations. We created several iterations of wireframes for user-centred design of the web interface using the design tool Axure. The stages of the iterative design process can be viewed under http://userpage.fu-berlin.de/komaromy/, whereas the LocXplore system is accessible online under http://locxplore.herokuapp.com/. After two iterations of usability tests, we developed a final design version for the user interface comprising the following functionalities.

At the outset, the user has to choose between **searching** for localities based on personal preferences and **exploring** the localities with the help of the provided data. On the landing page the user gets a short description of these two features and can then navigate to the desired views.

In the **explore module** the user can select a locality by clicking the map, where the names of the localities pop up on hovering over them. After selecting a region, the controls to explore further get activated and it is also possible to add the selected item to the **compare list**. Clicking on the facts, activities, and reputation buttons leads to the respective views that are shown in place of the map component. Clicking on the *compare* button in the header navigates the user to the compare component in which the respective data for all localities appears in a single overview page for comparison.

The **search module** takes advantage of indexing in ElasticSearch and allows users to input their search preferences. The results are shown in a list ranked by relevance to the search criteria. The localities can then be directly added to the compare list or be further explored.

4 DEMONSTRATION

Figure 2 models the possible task sequences and the navigation structure without loops and reverse paths for simplicity. The navigation is provided by the header menu, by the navigation buttons, and by the interactive map.

[5] e.g., https://www.angloinfo.com/how-to/germany/berlin/moving/berlin-file

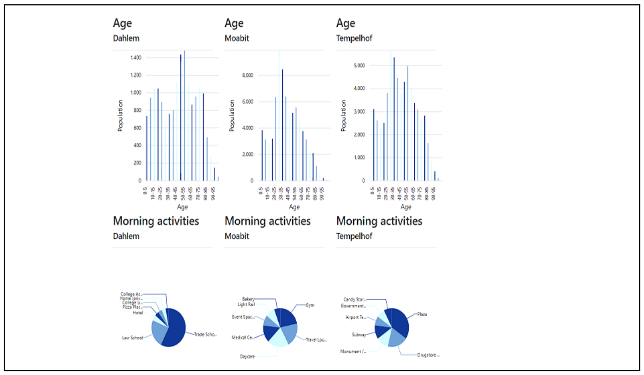

Figure 4: Comparing regions.

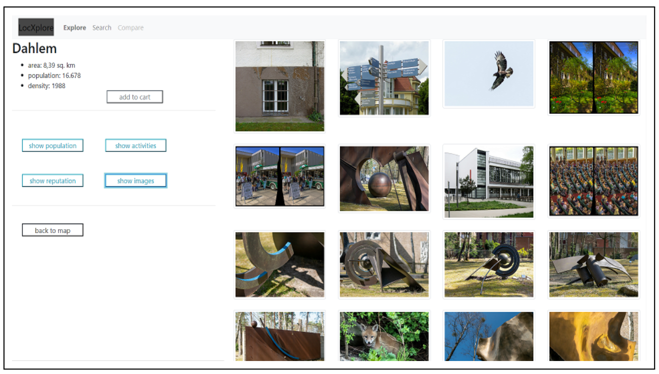

Figure 5: Images for a region (Dahlem).

For demonstration, we outline two typical usage scenarios in which a student moving to Berlin would like to (a) explore three localities of interest, and (b) find a locality according to her preferences.

4.1 Exploring certain localities

After choosing the exploration module, the user has to select her locality of interest by clicking the map. Since the student wants to explore the area where her university is located, she clicks on Dahlem, a locality in the south of Berlin. As shown in Figure 3, the map zooms to the selected locality and shows information on land usage.

Furthermore, the buttons for further exploration get activated enabling the user to retrieve more information on the locality, such as population, typical activities, images, and reputation. Figure 5 shows the actual images of Dahlem from Flickr – giving an impression of Dahlem as a neighborhood with upscale villas, abundant nature, and a vast and green university campus. By clicking the images that catches the user's eye, she can delve deeper into the Flickr galleries.

If the user would like to decide whether to move to Dahlem or to a different neighborhood, e.g., Moabit or Tempelhof, she can add those neighborhoods to her compare list (*cart*) and compare their respective information, as shown in Figure 4 for the population age distribution and typical activities. The user can conclude from the data, that Dahlem is a neighborhood where a larger portion of inhabitants are over the age of 50. Moreover, the distribution of activities shows that Moabit and Tempelhof offer more options for leisure activities and nightlife, whereas Dahlem offers a small town or village character.

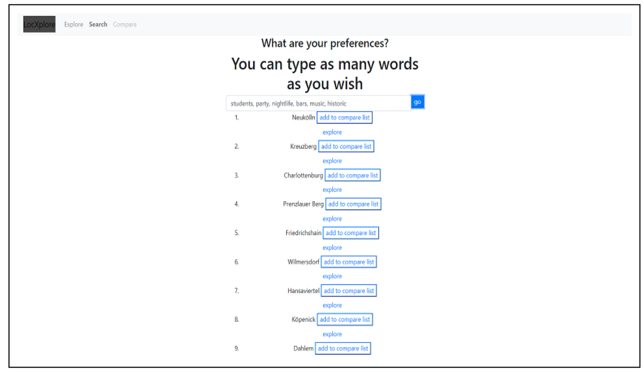

Figure 6: Search results.

4.2 Searching for a suitable neighborhood

If the aforementioned localities do not meet the user's expectations, she can switch to the search module. Figure 6 shows the most relevant search results for the keywords *students, music, bars, historic, nightlife, party*. The user can further explore or compare the search results before deciding.

REFERENCES

[1] Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. 2016. Variations of the similarity function of textrank for automated summarization. *arXiv preprint arXiv:1602.03606* (2016).

[2] Malcolm P. Cutchin, Karl Eschbach, Christine A. Mair, Hyunsu Ju, and James S. Goodwin. 2011. The socio-spatial neighbourhood estimation method: An approach to operationalizing the neighbourhood concept. *Health & Place* 17 (2011), 1113–1121.

[3] Kaiyu Feng, Kaiqi Zhao, Yiding Liu, and Gao Cong. 2016. A system for region search and exploration. In *Proceedings of the VLDB Endowment*. Vol. 9. VLDB Endowment, 1549–1552.

[4] George Galster. 2001. On the nature of neighbourhood. *Urban Studies* 38, 12 (2001), 2111–2124.

[5] John Anthony McGuire. 2014. *Defining and delineating urban neighbourhoods: a case study of housing areas in Brent, North London.* Ph.D. Dissertation. Birkbeck, University of London.

[6] Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering* 14, 2 (2009), 131–164.

[7] Dimitrios Skoutas, Dimitris Sacharidis, and Kostas Stamatoukos. 2016. Identifying and Describing Streets of Interest. In *Proc. 19th International Conference on Extending Database Technology (EDBT)*. Open Proceedings, 437–448.

[8] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.

[9] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. Nation Telescope: Monitoring and visualizing large-scale collective behaviour in LBSNs. *Journal of Network and Computer Applications* 00 (2015), 1–16.

[10] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.

[11] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.