Transformation of Local Communities from Neighborhoods to Urban Commons in the Production of Social Representations of Space

Aleksandr Antonov(✉), Galina Gornova, Georgii Kontsevik, Leonid Turkov, Vladimir Vorona, and Sergey Mityagin

ITMO University, Birzhevaya Line, 14, Saint Petersburg, Russia
asantonov@itmo.ru

Abstract. The paper explores new types of social interaction between urban communities. These include translocal and hyperlocal identities that reflect the uniqueness of urban areas at different levels. These two types of identities have become more widespread due to the increasing communication of city residents in digital ways, especially through social media. The method proposed in this paper is based on the application of thematic modelling, toponym and keyword extraction, and spatial knowledge graphs. It allows the identification of links between different territories and hotspots of activity resulting from such communication. It also provides the ability to identify the semantic properties of such phenomena, which could be useful in further urban identity research.

Keywords: Local communities · Translocal identity · Urban foundation models · Spatial semantic analysis

1 Introduction

Contemporary studies of sustainable communities are conducted within discourses of locality, translocality, and hyperlocality [1,2]. Within these discourses, the objects of study vary in scale and show different criteria of sustainability [3,4].

A local community is defined as a social group that is united by its place of residence, such as a district or city, and shares similar needs and values. When functioning effectively, a local community produces social capital that supports its sustainability. Social ties within local communities are strong and geographically bound.

Traditional notions of local communities are often linked to neighbourhoods. However, in metropolitan areas, neighbourhoods no longer provide the social support associated with the characteristics of urban lifestyles, changes in the nature of public practices, social ties and ways of accumulating social capital [3]. As a result, communities are transforming from locality to hyperlocality and translocality.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024, LNCS 14814, pp. 436–447, 2024.
https://doi.org/10.1007/978-3-031-64608-9_30

Social media platforms support the development of hyperlocal communities by tailoring their messages to specific geographic contexts. Hyperlocality can encompass a street, a neighborhood, or a district. Public groups on social media serve as a digital medium for communication and public engagement. Social media platforms play a crucial role in supporting weak social ties, which are essential in urban settings for maximizing social support and facilitating the effective transmission of essential information relevant to daily life and professional advancement.

Translocal communities are formed by the interaction of people, ideas, and symbolic resources, located in different spaces but maintaining a local self-identity. They are 'imagined neighbourhoods' supported by digital and networked tools that convey local meanings across territorial boundaries. Translocal connections shape and sustain social representations of the specificity of urban places.

Digital technologies and the evolution of social media have contributed to "reassembling the social" within local communities. Central to this phenomenon are transformations in the dynamics of social connections at multiple levels: the micro level, characterized by hyperlocality; the meso level, associated with locality; and the macro level, encompassing translocality.

Modern cities have retained those features that were reflected in Georg Simmel's classic essay 'Big Cities and Spiritual Life', in the works of Chicago School theorists and a number of other authors. These include individualism and anonymity. The former has a destructive effect on traditional neighbourhoods, while the latter frees the individual from group pressure and excessive social control [5]. Research on social media enables the comprehension of effects engendered by the multiscalar nature of urban space, one of which is the transformation of urban interactions towards commons-based neighboring in the production of social space. Urban commons is the outcome of a series of multiscale processes [1].

The paper consists of the following sections. The Related Works section presents the theoretical developments underlying our study and an overview of digital methods that can be used to extract and analyse social media data. The Method section describes the algorithm of the developed method and its data structure. Experimental Results section presents the practical application of the method to test subjects and presents the results of its use.

2 Related Works

The transformation of society in the modern world, associated with the shift from modernity to postmodernity and caused by changes in network communications, is reflected in B. Latour's actor-network theory [6]. The classical concept of M. Granovetter presents the prevalence and effectiveness of weak ties over strong ties, including in the formation of communities of different types. Granovetter argues that weak ties serve as channels through which community members receive socially distant messages and information that are useful for their activities and life [7]. Weak ties form the foundation of hyperlocal and

translocal communities. The term 'hyperlocal' refers to the online content of a small, geographically defined community such as a town, village, or neighbourhood, as defined in a 2012 study by the British Charitable Foundation Nesta when analysing the UK media [4]. A. Appadurai, who coined the term 'translocality', studied the differences between local communities, similar to traditional neighbourhoods, and translocal diasporas [2]. The transformation of communities due to globalisation processes that prevent the reproduction of traditional locality was also examined. The study 'Urban Commons: Moving Beyond State and Market' presents the conceptualisation of urban cohesion in conjunction with multiscale [1].

Researchers from various fields are actively working with the combination of spatial and semantic data. Urban data is analyzed at different scales, including several cities, the city as a whole, districts, neighborhoods, and streets [8,9]. These studies utilize a variety of data, such as social network data, open data, and spatial data [10,11]. Urban studies that utilise spatial semantic data typically concentrate on information agendas and people's activities related to events, as well as thematic modelling of the territory [9]. Additional attention is paid to the spatialisation of objects through textual data [12].

Kim [12] presents a technique for identifying and connecting space objects using their textual features obtained from online sources, including web documents and social. The authors suggest creating graphs with vertices representing particular locations and edges indicating their connections. Semantic and spatial similarities can be discovered by comparing textual descriptions, names, and nearby places. These graphs are compared to determine the most appropriate candidate for location in space. This approach allows the identification of location of objects based solely on text. However, this approach is not suitable for urban identity research. The authors do not conduct a semantic study of topics near identified physical objects.

Capela F. presents a method for identifying unique identities for different cities based on news texts [13]. The process begins with keyword extraction using TF-IDF, followed by the Latent Dirichlet Allocation (LDA) method to generate topic models [14]. The distribution of news texts across cities is then determined in the context of the generated topics. The approach focuses on the inter-city level. Our method uses a similar approach but applies it intra-city.

In the review of topic modelling and analysis methods by Kousis A., two main methods of topic identification are described: LDA and BERTopic [15]. BERTopic is noted as being more advanced. The input data used to create a topic map on smart cities includes news sources, academic articles, and social media comments. It is important to note that post-processing with expert judgement is necessary for BERTopic to generate a list of labels. The inter-city focus is a notable feature of this method.

Bok K. describes an approach for analysing Twitter data to identify event keywords and their relationships [16]. The study generates a graph of keywords associated with certain events by analysing social media data. Keywords are extracted from datasets with Twitter posts, and clusters are built. However, the study does not consider geographical labels in the context of keyword linking. The sampling

is spontaneous, as no specific spatial framework is given for the study. Instead, toponyms are associated as normal words in text without geocoding.

B. Hawelka presents a method for studying interregional migration behaviour of people [17]. The study uses 944 million geo-tagged messages on Twitter to determine The study compared each user's geo-tagged posts over time to determine their place of residence based on the highest number of geo-tagged posts. It also identified places of current residence, such as tourist destinations. The authors then identify the main migration drivers of people's movements in each country.

The study conducted by Tan M.J. and Guan C. utilises Twitter message analysis to evaluate the sentiments of residents towards property prices [9]. The researchers analyse the tone of messages for a specific geographical label (address) to determine the tonal level of the label and create a tonal distribution of the area. Based on the tone level, hotspots are identified, reflecting attitudes towards property transactions on a scale of 0 to 100. The paper analyses property prices and assesses their tone, which is then grouped into a heat map. The study focuses on neighbourhood-level properties. Our research method does not include a sentiment score in the form of a scale that reflects the sentiment of an area. Instead, it conducts semantic analysis of words linked to geocoded address labels.

All of these methods aim to gather information about the city and its processes and patterns. However, these works focus solely on analyzing a specific topic and do not consider other aspects of the territories. The present study employs a distinct approach that considers the major themes in the information agenda.

The proposed methodology is focused on identifying multiple thematic areas that influence and shape individuals' perceptions of a particular urban area.

3 Method

3.1 Method Overview

The general scheme of the method is shown in Fig. 1. Blocks marked in blue characterize the information supplied to the input; blocks marked in green characterize the tools used.

Data is collected using the open API of the VKontakte social network, then toponyms (street names) and house numbers are extracted from each text and geocoded. Next, a keyword that describes each message is selected [18]. All keywords are collected into a text corpus and clustered. Thus, each message possesses an address and a semantic class. Next, each cluster is visualized on a map and grouped in a hexagonal grid. The hexagons are clustered by the number of messages that fall into them. Hexagons characterize the spatial proximity of messages of the same class. Thus, the parts of the space that fall into one cluster will show similarity with respect to the activity of residents in the selected category (semantic cluster) and space. The code implementation example and sample data are available in github repository [19].

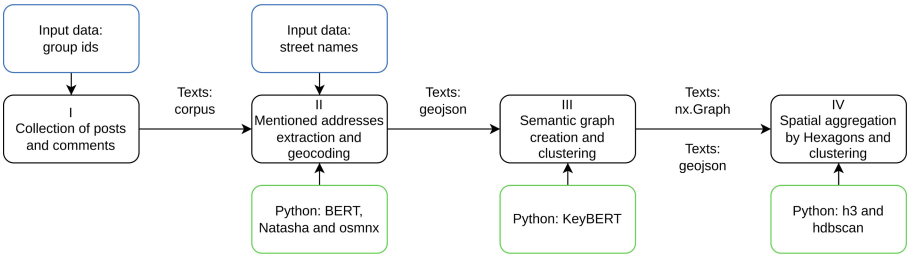


Fig. 1. Scheme of the method

3.2 Data Structure

Each post has comments that refer to it. Accordingly, a post and comments can contain an address (house and street), which is discussed within one post. Comments can have replies to them (discussions) within one post. Such links between comments and between comments and post allows to fill in geo-referencing information for comments within one post without the location mentioned in them.

3.3 Toponym Extraction

Geocoding was implemented according to a method and model similar to the article [20] in order to extract toponyms presented in the text.

3.4 Keyword Extraction

After extracting the toponyms, the text data is prepared for processing. This involves removing numbers, references, extracted toponyms, and words indicating toponyms such as 'street', 'avenue', and 'boulevard'. Additionally, stop words obtained from the nltk library [21] are filtered from the text to increase the relevance of keywords. The KeyBERT algorithm [22], based on the DeepPavlov/rubert-base-cased model [23], is used to extract them one per text. Only keywords with a semantic closeness higher than 0.6 to the source texts are used.

In the next step, the extracted toponyms are linked to the keywords according to the following rules:

1. If the toponym was extracted from a post, the words extracted from all comments that refer to it and the word extracted from the text where the toponym was, are linked to its node.
2. If the toponym was extracted from a commentary, the words extracted from all other commentaries that refer to the toponym, as well as the word extracted from the text where the toponym was, are linked to its node.
3. If the toponym was extracted from a comment that is not referenced by other comments, only the word extracted from the text of that comment is linked to its node.

This set of rules is demonstrated in Fig. 2.

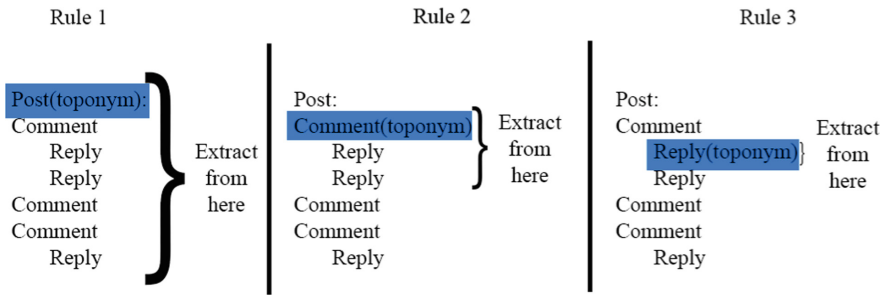


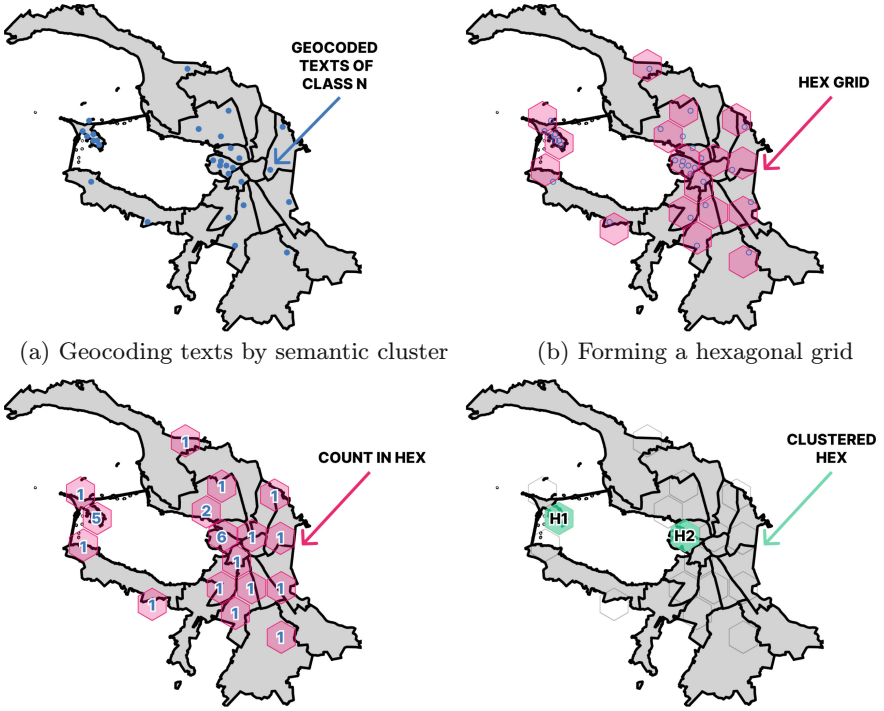
Fig. 2. Rules for toponym extraction from social media posts and comments

3.5 Graph Generation

Thus, the necessary set of nodes for the knowledge graph is obtained. The nodes consist of toponyms and keywords from related texts, connected by edges. As different texts may contain the same keywords, several edges may lead to one node with a keyword. This is how links between toponym nodes are formed. To enhance the number of links, we assess the semantic similarity of all keywords by encoding them into embeddings and calculating the cosine distance between them using the DeepPavlov/rubert-base-cased model. If the similarity of words exceeds a certain threshold (in this case, 0.75), we include the relationship between words in the graph as the edge. To map the graph in space, we add the coordinates of the centroids of the corresponding streets to the attributes of the toponym nodes.

3.6 Spatial Aggregation

Spatial aggregation as such was chosen because we need to consider parts of the city and not just the points themselves. The hexagonal grid is interpretable, in this paper we use level 7 of h3 hexagonal grid, which corresponds to a scale of 1 km for each edge of the hexagon. The use of the hexagonal grid rather than the boundaries of administrative units allows us to consider vernacular areas of the city. The process of spatial clustering is shown in Figs. 3a-3d.



(a) Geocoding texts by semantic cluster (b) Forming a hexagonal grid

(c) Count the number of texts in each cell (d) Clustering cells by number of texts

Fig. 3. Spatial analysis steps

4 Experimental Results

4.1 Study Area

In order to identify similarities between community identities in different parts of the city, two experiments were conducted in St. Petersburg and Vyborg. Both cities are nuclei for their nearest agglomerations and are located in the north within drive accessibility from the border with Finland.

St. Petersburg is located in north-western Russia on the coast of the Gulf of Finland and at the mouth of the Neva River. It is bordered by the Leningrad Oblast to the south and has maritime borders with Finland and Estonia to the west. St. Petersburg has a population of over 5 million people and spans across 10 administrative districts, covering a total area of 1,439 square kilometers.

According to Federal State Statistics Service the most densely populated districts of Saint Petersburg by number of inhabitants are [24]:

1. Primorsky district (699,243 people)
2. Nevsky District (547,896 people)
3. Vyborgsky District (541,590 people)
4. Kalininsky District (536,794 people)
5. Krasnoselsky District (431,546 people)

Vyborg is located in the Leningrad Oblast on the coast of Vyborg Bay, which is located in the north-eastern part of the Gulf of Finland. Vyborg is smaller in size and population, and is considered part of the St. Petersburg agglomeration due to its good connectivity via public transportation (high speed trains connection). Vyborg is the second most populous city in the Leningrad Oblast, with a population of approximately 71,772 residents.

4.2 Data

The text data were obtained using open API of VKontakte social network [18]. Geospatial data for geocoding was obtained from OpenStreetMap.

For the experiment, we collected two sets of posts and comments. The first set covers the entire city of St. Petersburg, and includes posts and comments from 38 groups in 14 administrative districts for the period from 29.06.2022 to 12.02.2024. The second set pertains to the city of Vyborg, and includes texts from 5 groups for the period from 01.03.2023 to 01.03.2024.

4.3 Hypothesis

This paper examines the hypothesis that different parts of the city may be similar in terms of the projection of their identity through social media activity.

4.4 Results

The method was tested on a small scale in the city of Vyborg near St. Petersburg, using the dataset described in the previous section. The results are presented in Fig. 4.

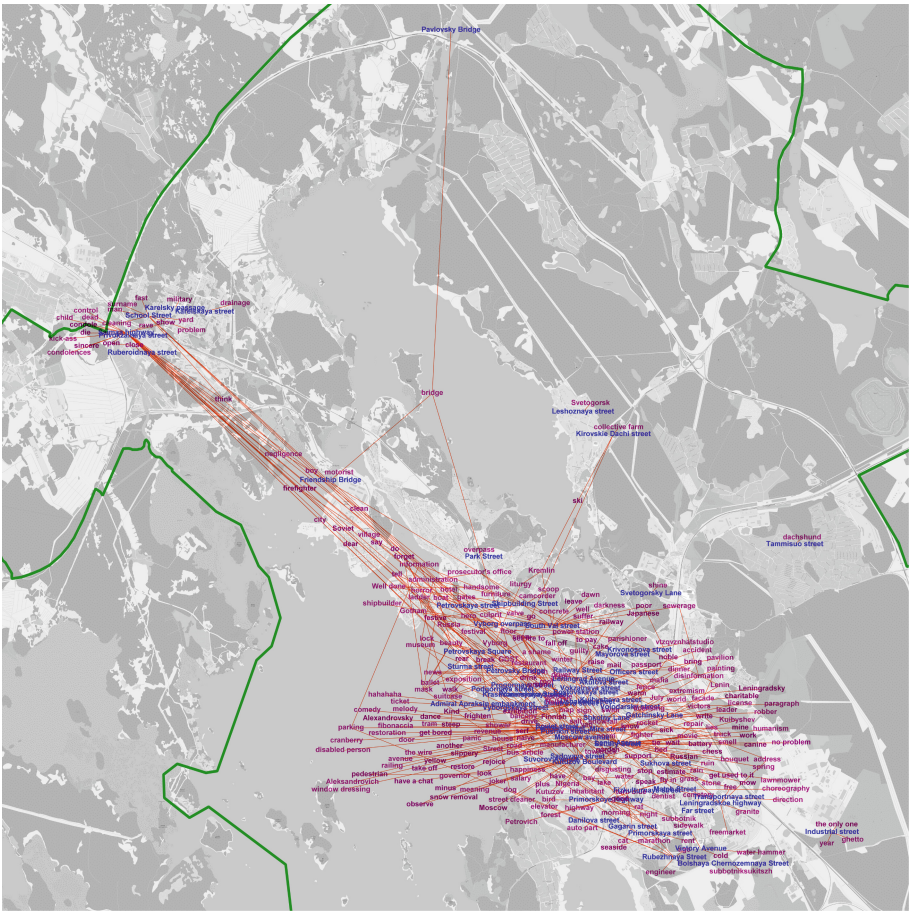
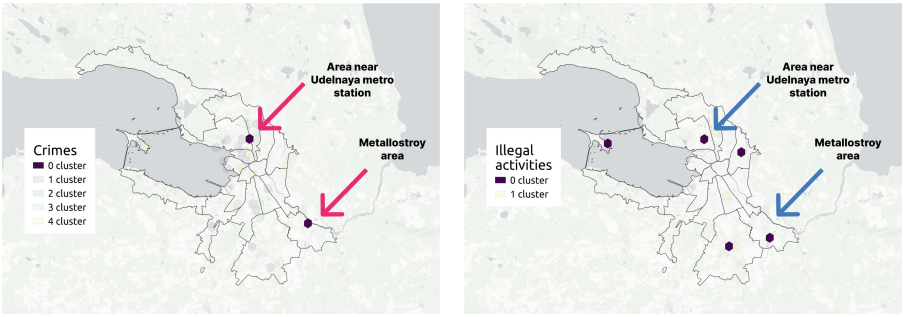


Fig. 4. Spatial-semantic graph of Vyborg

Based on visual and expert assessments, the city is divided into two enclaves of the most frequently mentioned areas, separated by a strait. The bottom right enclave refers to the older historic city, while the top left enclave refers to the newer Kalininsky district. Both areas are distinguished from other less active areas, such as the private residential areas near Kirovsky Dachas, Tammisuo and the industrial areas around Vyborg, by the presence of multi-storey apartment buildings. Additionally, they are connected by common semantic links that may indicate markers of translocal identity. These connections include ‘fireman’, ‘hotel’, ‘city’, ‘soviet’, ‘prosecutor’s office’, and ‘clean’, which have some common meaning, typical for a small tourist town.

The application of proposed method on larger scale was performed for the area of St. Petersburg. The method was applied including additional aggregation

of social network texts on a hexagonal grid and clustering of hexes. The results are presented in Fig. 5a and Fig. 5b.



(a) Cluster "Crimes" (b) Cluster "Illegal activities"

Fig. 5. Spatial clusters of messages by semantic clusters

Of particular interest are the topic clusters with the highest emotional load, among a number of others. In this instance, clusters related to security threats of varying degrees, such as 'murder', 'theft', and 'vandalism' were selected. The analysis identified two territories belonging to the most active cluster in this context. These territories are located in the settlement of Metallostroy (bottom right of the figure) and near the Udelnaya metro and railway station (top of the figure). These two areas are morphologically similar as they both contain green spaces and industrial areas. They are often characterised by few regular visitors and may be perceived as unsafe by residents.

5 Conclusion

The article outlined theoretical approaches to the study of identity, as well as digital tools that facilitate the study of identity in the information space. It also demonstrated experiments relying on the data from the VKontakte social network with the method based on text analyses approach and presented the interpretation of the received results of the work [18]. This research presents a method for identifying spots of high intensity of individual thematic clusters in a limited territory as manifestations of hyperlocal identity. As the semantic connection between some clusters concentrated in remote areas can be estimated by the semantic similarity of text topic representations, it is reasonable to conclude that some else of these areas' identities can be called translocal. As an example of such a process, the case of districts of Vyborg and two locations in St. Petersburg in Russia is shown. It means that citizens of two or more remote areas can be concerned about the same problems and processes depending on the found topic.

By implementing this method in the decision-making process authorities and other actors can acquire a better understanding of the current urban processes in the agglomerations and give a basement for more precise prediction of different project results in urban development and management. It is essential to pay attention to all possible actors concerned with any changes in city life.

References

1. Dellenbaugh-Losse, M., Bieniok, M., Kip, M., Müller, A.K., Schwegmann, M.: Urban Commons: Moving Beyond State and Market (2015)
2. Appadurai, A.: The production of locality. In: Fardon, R. (ed.) Counterworks. Managing the Diversity of Knowledge, pp. 204–225. Routledge (1995)
3. Ganugi, G., Prandini, R.: Fostering social cohesion at the neighbourhood scale: the role of two social streets in Ferrara and Verona. SN Soc. Sci. 3 (2023). https://doi.org/10.1007/s43545-023-00688-6
4. Radcliffe, D.: Here and now: UK hyperlocal media today (2012)
5. Gornova, G.V.: Philosophy of the City. FORUM Publ., INFRA-M Publ., Moscow (2014)
6. Latour, B.: Reassembling the Social: An Introduction to Actor-Network-Theory. Oup Oxford (2007)
7. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**(6), 1360–1380 (1973)
8. Yu, Z., Zhu, X., Liu, X.: Characterizing metro stations via urban function: thematic evidence from transit-oriented development (TOD) in hong kong. J. Transp. Geogr. **99**, 103299 (2022)
9. Tan, M.J., Guan, C.: Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using twitter data. Appl. Geogr. **132**, 102474 (2021)
10. Chen, Y., Niu, H., Silva, E.A.: The road to recovery: sensing public opinion towards reopening measures with social media data in post-lockdown cities. Cities **132**, 104054 (2023)
11. Gao, Y., Chen, Y., Mu, L., Gong, S., Zhang, P., Liu, Y.: Measuring urban sentiments from social media data: a dual-polarity metric approach. J. Geogr. Syst. **24**(2), 199–221 (2022)
12. Kim, J., Vasardani, M., Winter, S.: Similarity matching for integrating spatial information extracted from place descriptions. Int. J. Geogr. Inf. Sci. **31**(1), 56–80 (2017)
13. de Oliveira Capela, F., Ramirez-Marquez, J.E.: Detecting urban identity perception via newspaper topic modeling. Cities **93**, 72–83 (2019)
14. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003). submitted 2/02; Published 1/03
15. Kousis, A., Tjortjis, C.: Investigating the key aspects of a smart city through topic modeling and thematic analysis. Future Internet **16**(1), 3 (2023)
16. Bok, K., Kim, I., Lim, J., Yoo, J.: Efficient graph-based event detection scheme on social media. Inf. Sci. **646**, 119415 (2023)
17. Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C.: Geo-located twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. **41**(3), 260–271 (2014)
18. VKontakte: about VKontakte. https://vk.com/about (Year not provided)

19. Sloyka: a sample repository. https://github.com/GeorgeKontsevik/sloyka (2024)
20. Antonov, A., Kontsevik, G., Natykin, M., Mityagin, S.A.: Feedback2event: public attention event extraction from spontaneous data for urban management. Procedia Comput. Sci. **229**, 138–148 (2023). https://www.sciencedirect.com/science/article/pii/S1877050923020057, 12th International Young Scientists Conference in Computational Science, YSC2023
21. Project, N.: NLTK library. https://www.nltk.org/. Accessed 27 March 2024
22. Issa, B., Jasser, M.B., Chua, H.N., Hamzah, M.: A comparative study on embedding models for keyword extraction using keybert method. In: 2023 IEEE 13th International Conference on System Engineering and Technology (ICSET), pp. 40–45 (2023)
23. Kuratov, Y., Arkhipov, M.: Adaptation of deep bidirectional multilingual transformers for Russian language (2019)
24. Rosstat: Number of permanent population of the Russian federation by municipalities as of 1 January 2023. https://rosstat.gov.ru/storage/mediabank/Bul_MO_2023.xlsx (2023)