The analysis of architectural discourse in the context of computational public opinion: Data mining of Google map reviews

Hanife Sümeyye Taşdelen[1], Leman Figen Gül[2]

[1] Istanbul Technical University, Istanbul, Turkey
tasdelenh15@itu.edu.tr
[2] Istanbul Technical University, Istanbul, Turkey
fgul@itu.edu.tr

Abstract. New media platforms and mapping tools are created by digital communities, and their representations influence public opinion. Crowdsourcing platforms such as Twitter, Instagram, Google search engines and Maps have no such limitations or boundaries as in the physical space, these platforms are creating new virtual public places. To assess and forecast the accessibility and aesthetic issues of urban spaces in Istanbul, text and image data from Google Maps were employed. In this study, we searched at the reviews of certain public/semi-public places by using text mining and image analytics tools. Recently designed or renovated 11 public buildings and open places were chosen. The main findings of this exploratory study are that; 1) the level of being public can be understood from the crowdsourcing, 2) image analytics of crowdsourced visual data can assist to identify the aesthetic quality, and 3) the accessibility capacity of public spaces can be identified.

Keywords: Data Analytics, Public Spaces, Architectural Criticism, Collaborative Map, Accessibility-Aesthetic issues.

1 Introduction

Mapping of social media platforms have been creating the public opinion that forms the cybernetic world-picture and social formations. Also, the novel architectural and environmental discussions have expanded far beyond their original field of experts, as developed by the public we refer it here as the 'computational public'. Despite the fact that this notion has been used previously (Karami et al., 2018), we further define the term ‘computational public’ as a relatively new concept within the scope of this exploratory study. The purpose of this paper is to methodically demonstrate the possibility of identifying problems in public spaces in terms of aesthetics and accessibility, and to make comparisons of these spaces based on the level of being public. According to the Attoe (1978), every response to the environment is a form of criticism. Architectural criticism can be considered through boundary of academic and professional lenses, but crowdsourcing platforms such as Twitter, Instagram, and Google search engines or maps have no such limitations or boundaries of the physical space. So, crowdsourcing could be useful from the point of incorporating all parts of society for understanding the discourses about the built environment. In architecture and urban planning in general, figuring out where all kind of boundaries (surfaces, walls, social contexts, level of accessibility etc.) between public and private spaces are, is a very important question. For Kimmel, (2018) this is especially true now that computational design technologies allow us shaping of the complex transformable surfaces to tend to blur the lines between virtual and physical spaces. On the other hand, computational design and digital technologies can also create new kinds of virtual public spaces as well as communities. So, these concepts are explained in the literature review briefly. In this exploratory study, we focus on architectural scale discourse and image sharing activities in Google Maps. The paper will investigate architectural scale in the urban context in terms of aesthetics and accessibility, and contribute to current methods for revealing built environment criticisms using the crowdsourced data. Text from Maps' reviews is used to evaluate and predict the aesthetics and accessibility of urban public places. We've focused on 11 publicly accessible urban landscapes with squares or open-air spaces surrounded by novel/contemporary architectural buildings/features in Istanbul. Those spaces with different scale and access were built between 2012 and 2021. We used architectural website Arkiv (Url-1) to select recently designed or renovated public buildings and open spaces that are separated into two sub-groups: one of the group is "semi-public / pseudo public", the other one is "public spaces".

2 Reviews on different types of public spaces and place semantics

Public space is a broad term in itself. Urban squares, public parks, passages, memorials, streets, markets, playgrounds, open green spaces, greenways, courtyard of mosques, waterfronts, are all included in this term (Mac Síthigh 2012). Public space's form and function have changed over time. It has always shaped urban fabric with its functions as a market place, gathering point, and media for cultural contacts and political discussions; and with its diverse forms as agoras, forums, piazzas, squares, streets, theaters, parks, libraries, coffeehouses (Carr et al. 1992; Madanipour 2003 as mentioned in Paköz, Sözer, and Doğan 2022). Technology and knowledge have transformed societies towards to digital societies, thus the meaning of public space needs rethinking (Crang, 2000). The computational public, in our definition, are individual citizens forming the digital communities that create virtual spaces as well as create the public opinion using all types of expressions (e.g. likes, dislikes, emojis, texts, multimedia). The computational public expresses their opinion and could create a common voice. Pseudo-public spaces, on the other hand, is defined as any space owned and managed by for-profit enterprises (Langstraat and Van Melik, 2013 as cited in Paköz et al., 2022). Common examples for pseudo public spaces could be plazas and shopping centers. Also, virtual public spaces e.i. social media platforms, online forums or entertainment platforms are refered to the computational public where people could either discuss in groups or be the audience of public discussions (Li, 2010). The character of public space has shifted with the advent of the digital age, the role it once served has been transformed as they privatized, socializing has dwindled, and so on. A blurring of the boundaries between public and private areas has occurred (Paköz et al., 2022). Han et.al (2019) discuss the recognition of publicness as understood by everyday users of public space. They propose that Yang's (2013) notion of "spatial publicness" to overcome the deficiencies in the existing framework. For them, the concept of spatial publicness is crucial to understanding how public spaces are used and understood by end-users. For Sınmaz (2018), with the urban transformation (regeneration) agenda, semi-public spaces, which have been missing from the growth of Turkish cities, especially in the last 30 years, can be made again. In this sense, privately owned public places are not common in Istanbul's building regulation, thus we used the concepts of semi-public, semi-private or pseudo public spaces. Istanbul is high density (3049 person/km2) and a very-crowded city (15.5 million in 2020 based on the data in tuik.gov.tr), and this makes the city a good "candidate for open-source investigations of the place semantics using the geolocation tools. Place semantics research focuses on comprehending the meaning of places through human descriptions and human-place interactions (Hu, 2018). Large-scale surveys can reveal how people interact with built-environment open spaces. Such analysis reveals a city's architectural discourses and tastes. Unstructured natural language text like social media posts, blogs, and Wikipedia entries is also growing quickly. Textual data reveal people's understanding and views of their social and natural settings (Hu, 2018). Unstructured text data can be used to extract place semantics and understand people-place interactions. Thematic, geographical, and temporal frameworks can be used to study place semantics (Hu, 2018). Thus, geotagged social media data could be used for text mining to understand place semantics. Google Maps generates a self-centered/user-centered world map. Each person's particular map or "lens on the world" reduces "our shared world." (Ström, 2020). According to Ström (2017), relatively few works critically investigate the impact of a tool like Google Maps on the production of space and subjectivity. In recent years, proliferation of street view imagery (SVI) data (development of services like Google Street View), breakthroughs in machine learning and computer vision that enable automatic information extraction is well suited for assessing characteristics of the built environments (Biljecki & Ito, 2021). In the same vein, there are some studies that analyses users' Google Maps reviews to quantify the reviews about certain architectural or urban places (Borrego & Navarra, 2020; Munawir et. al, 2019).

3 Methodology

The research objective is to understand whether the level of public quality of the selected public places matches the online map data (photos and reviews). In addition, the objective of the research is the investigation of public (critical) discourses on the existing architectural built environment through the text mining technique of crowdsourcing. Exploratory data analyses are methods for summarizing or revealing features of interest within a dataset that would not be visible through traditional close reading (Saldaña, 2018). In this paper, sentiment analysis, word cloud and topic modeling analysis combined with image analytics are adapted to retrieved texts from the Google Maps reviews. First, data that is reviews and review images is collected on selected publicly accessible buildings. Natural Language Processing (NLP) techniques, such as sentiment analysis, help researchers explore textual data as mention earlier. Orange Data Mining that is a user-friendly tool, extracts topics from text and analyzes the meaning of the expressions. Text sentiment analysis, opinion mining, and topic detection and tracking are new social media-oriented text processing methods (Zong et al., 2021). The objective of sentiment analysis is to quantify the emotional intensity of a text's words and phares. Python based visual programming interface of Orange makes easily possible it to import text format to analyze them with the help of Text Mining add-ons. Image classification and clustering have also been done to compare the Google Maps image groups in general. Image classification is a procedure in computer vision that can categorize an image based on the visual content of the image (Sanghvi, 2020). The workflow in below adapted to figure out general variations about the topics (accessibility, aesthetic) among the levels of being public (semi-public or public) of the selected places (Figure 1).

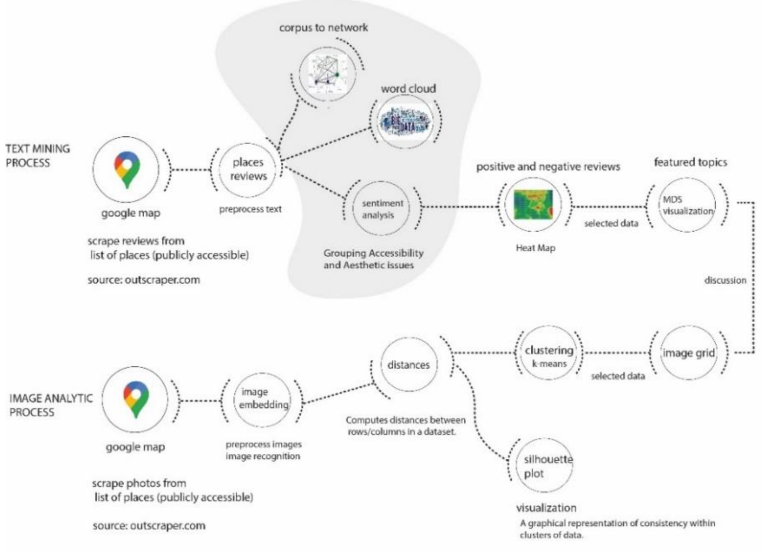


Figure 1. Workflow of text mining and image analytics, prepared by authors.

By this, one of the aims is to investigate the relations of the computational public with an architectural scale in different urban places on online platforms. The chosen places in Istanbul are shown in the map below (Figure 2).

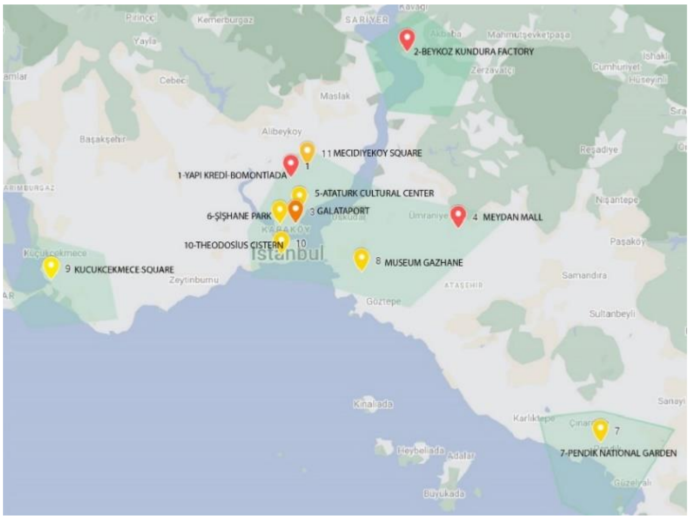


Figure 2. The location of selected places on the Istanbul Google Maps view (the color oranges are semi-public/pseudo public, the color yellows are public spaces).

Due to the significant differences in the total number of reviews made for places, we retrieved the reviews made between the years 2021-2022 into analysis. For this, we selected a subset of most new 250 reviews of 11 places from two different urban categories in Istanbul (Table 1).

Table 1**.** The selected places for text mining and image analytics, retrieved from Google Maps.

no	Name of place	type	Review rate	Total images	Total reviews	Built date
1	Yapıkredi Bomonti ada	Semi-public	4,5	51	4403	2015
2	Beykoz Kundura Factory	Semi-public	4,4	13	798	2018
3	Galataport	Semi-public/pseudo public	4,6	20	6711	2022
4	Meydan Mall	Pseudo public	4,4	26	12566	2010
5	Atatürk Cultural Center	Pseudo public	4,6	25	951	2021
6	Şişhane park	Public	4,2	18	226	2013
7	Pendik National garden	Public	3,9	22	86	2021
8	Museum Gazhane	Public	4,7	23	3633	2021
9	Küçükçekmece Square	Public	4,0	19	3230	2012
10	Theodosius Cistern	Public	4,6	19	5255	2018
11	Mecidiyeköy Square	Public	4,4	16	257	2021

As seen in the table, there are two different predetermined group; public places consist of Pendik National Park, Küçükçekmece Square, Şişhane park, Museum of Gazhane, Mecidiyeköy Sqaure; semi-public/pseudo public places consist of Galataport, Beykoz Kundura, Meydan Mall, Theodosius Cistern, Ataturk Cultural Center, Bomontiada. When Google map reviews were examined closely, the new design regulations and newly acquired public spaces in the City were praised. The VADER scale was used for quantitative analysis in Sentiment Analysis. With this scale (Hutto &Gilbert, 2014), it was easy to identify the most negatively reviewed subjects. This scale considers neutral all statements except clear ones. Thus, implied or ironic negative statements require a closer look. Using Orange, general keywords were first revealed by Word Cloud analysis then semantic analysis was used to examine the topics. Image analysis used image embedding and silhouette analysis to group similar images. Photos of public spaces are also analyzed to determine where they were taken most often. We can categorize them as follows: indoor (café, restaurant), courtyard/garden/landscape, facades of buildings, view, activity, and action (drink/dinner). The first stage of the exploratory study can reveal some general differences between public and semi-public places, but general inferences cannot be made. As public space activities and actions diversify, photo sharing increases. In this regard, the fact that Bomontiada is an attraction hub and a building island with relatively more opulent locations also influences the motivation to share photos from that region. As the quality of the spaces improves, so does the citizens' motivation to share. In below, Image grid of two distinct groups together are shown (Figure 3). Public group consist of 5 places, while semi-public group consist of 6 places. Even though they are not equally divided, this is not a mandatory metric for our explanatory research.

3.1 Image Analytics and Text mining

The photographs with similar elements are placed together as shown in Figure 3. When we look at the image grid of semi-public places more closely, we see that it divides the images into four categories: the exterior, the interior, the street view of historical sites, and the place images obtained during the night activity. In the public place image grid, wide-angle and completely open-air images, such as green-sea view and sunset-city silhouette, are grouped together. In general, photographs were shared from aesthetically meaningful points that make the spaces unique. This shows that there is an opportunity to use citizen data as a parameter in the analysis of urban aesthetics for further research.

Figure 3. Showing that similar features of images (such as sunset or sea views, indoor and activity) by grouping with image embedding, applied by authors.

Image embedding could be used to identify similar images which is hard to interpret by humans. After embedding images, we can arrange them on the grid in the Image Grid widget so that resembling images are close to one another (Godec, 2022). In a visual representation of this data, you can see how many clusters there are by plotting points' proximity to other points in neighboring clusters. Then, we adapt as a first one NLP tool: word cloud, and Corpus network analysis and then sentiment analysis. Corpus network (Sulis et al., 2020) and word frequency is to quantify the most frequent words of texts and to understand general review topics about places. Word Cloud shows tokens from the corpus, with the size of the token indicating how often the word appears in the corpus or an average bag of words count when the bag of words features are at the widget's input. In the widget, words are listed by how often they are used (weight). The widget makes documents with tokens from the word cloud that has been chosen (Url-2). For every place reviews, the process was repeated. Sample results of process is shown in Figure 4.

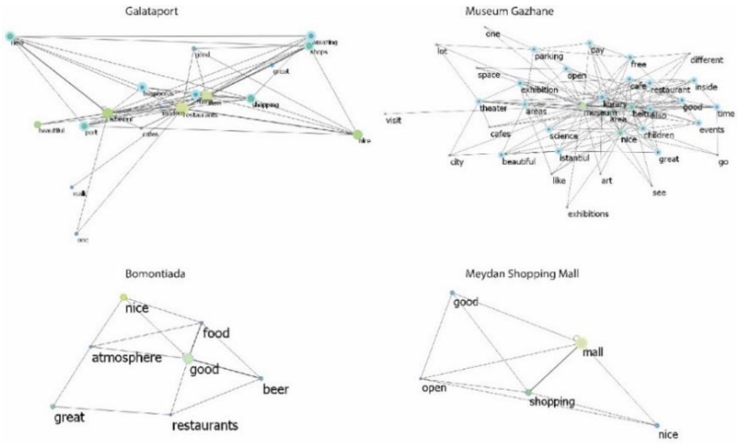


Figure 4. Sample results of corpus to network, applied by authors.

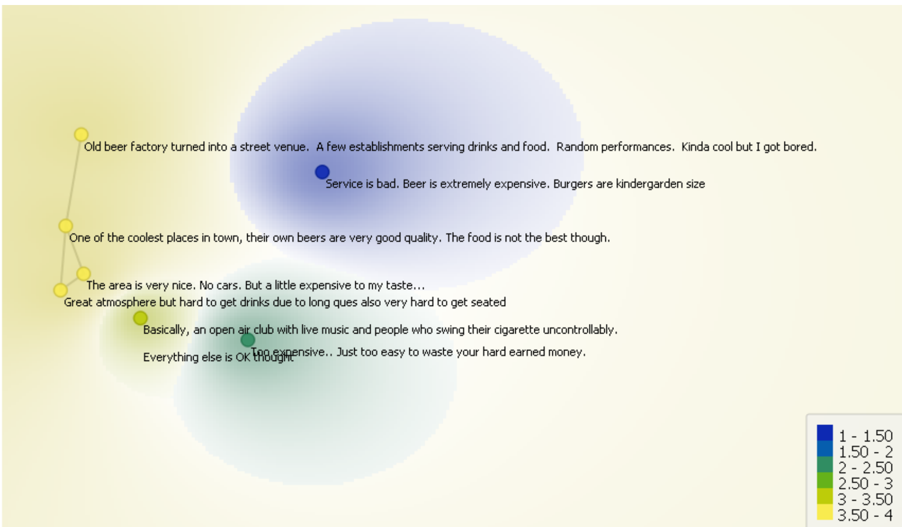


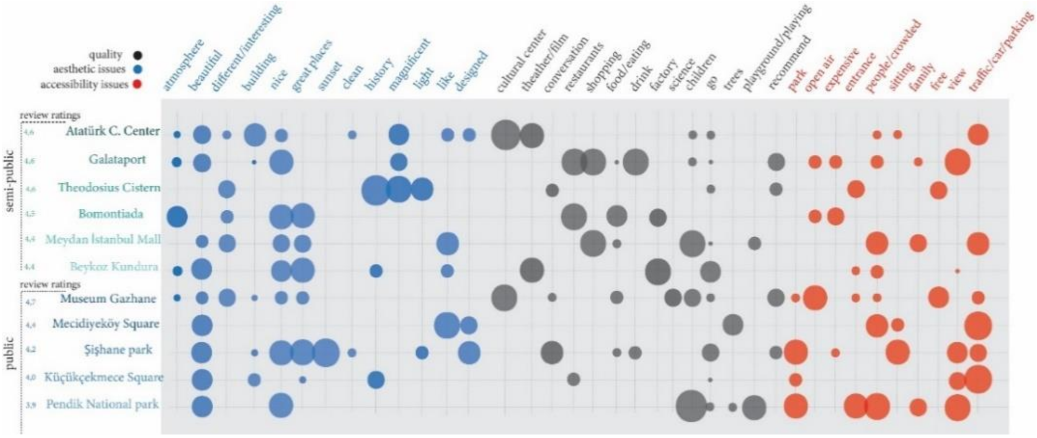
Figure 5. Sample results of negative sentiment analysis Vader module with review rates for Bomontiada, applied by authors.

Sentiment Analysis forecasts the sentiment of every document inside a corpus. It uses NLTK's (Natural Language Processing Toolkit) Liu & Hu and Vader sentiment modules, Data Science Lab's multilingual sentiment lexicons, Arthur Jacobs' SentiArt, and Walter Daelemans' LiLaH sentiment (2020) (Url 3). VADER sentiment modules were used. Some sentiment analysis tools consider punctuation and emoticons. Sentiment analysis tools process text (a sentence, a paragraph, a book, etc.) and produce quantitative scores or classifications indicating whether the algorithm believes the content conveys positive or negative feeling. Some tools measure a text's positivity or negativity (Saldana, 2018). Sample semantic analysis with Vader modules demonstrates problematic concerns and positive qualities of places (Figure 5). Figure 5 shows negative-positive Multidimensional scaling reviews.

4 Results

The research method allows us to address differences-similarities and issues-problems in aesthetics and accessibility in culturally and publicly diverse urban areas. Text mining and image analytics helped us see prominent issues in reviews. We focused on 2021-2022 reviews for this exploratory research. Expanding the sample size will improve future study results. The first corpus-to-network workflow was customized for 11 public and semi-public locations. In table 2, we group the most common review text topics into aesthetic concerns, accessibility issues, and place quality/description. Public and semi-public area review ratings determined the order. Table 2 shows that semi-public places frequently discuss aesthetic terms and topics. Accessibility issues discussed in public places more. The analysis revealed accessibility problems. For instance, reviews of Pendik National Park note the closed entrance and ramps. Ataturk Cultural Center's parking problem is very visible. “View” is also considered an accessibility issue because public spaces need to be visible from all sides. As seen in the table, view issues have been positively discussed in public. The issue of view also seems also important for Galataport, but it has been discussing as negatively because of the cruise at the port blocking the Bosphorus view. Therefore, it was necessary to control with semantic analysis data for each topic and keyword.

Table 2. Data visualization of word cloud and corpus network analysis, prepared by authors.



5 Conclusion

The contribution of the paper will be providing insights of investigating the discourse about architectural scale in the urban context in terms of aesthetics and accessibility and will contribute to current methods for revealing criticisms about the built environment using crowd source data. This study shows that the aesthetic and functional issues change by the level of publicness of urban spaces in the result of crowdsource reviews data mining, this phenomenon is detected on Google Maps Reviews data. In this sense, urban place crowdsource (or crowdsourcing urban data) research might improve by integrating urban areas with varying levels of public space that would be chosen after literature review. With the increase in the number of urban places to be included in the study, the aesthetic and functional issues change related to the level of publicness would be seen. This study demonstrates that online map reviews and location-based photographs could be used to determine the level of public of urban spaces, particularly pseudo public spaces from the perspective of end-user experience of public spaces. As volumes of data are available, online reviews provide valuable information for assessing public places. Also, different social network platforms could include to the analysis process. People (citizen) who experience the selected places and share their experiences in various seasons and various time intervals can provide information on accessibility and quality problems, as well as the aesthetic dimension of their spatial experience and the reasons why certain places have higher aesthetic values than others in terms of place semantics.

References

Attoe, Wayne, Architecture and Critical Imagination, Wiley and Sons, Chichester, 1978.

Borrego, Á. and Comalat Navarra, M. (2021), "What users say about public libraries: an analysis of Google Maps reviews", Online Information Review, Vol. 45 No. 1, pp. 84-98.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. Landscape and Urban Planning, 215(November 2020), 104217. https://doi.org/10.1016/j.landurbplan.2021.104217.

Carr, S., M. Francis, L. Rivlin, and A. Stone. 1992. Public Space. Cambridge: Cambridge University Press.

Daelemans, Walter; et al., 2020, The LiLaH Emotion Lexicon of Croatian, Dutch and Slovene, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1318.

Han S, Kim JW and Kwon Y. (2019), Contemporary Spatial Publicness: Its New Characteristics and Democratic Possibilities. *Sustainability*., 11(17):4729. https://doi.org/10.3390/su11174729

Hu, Y. (2018). Geospatial Semantics. Comprehensive Geographic Information Systems, 2017, 80–94. https://doi.org/10.1016/b978-0-12-409548-9.09597-x.

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

Godec, P. (2022, 4 May), Image Analytics for Everyone: Image Embeddings with Orange, retrieved from https://towardsdatascience.com/image-analytics-for-everyone-image-embeddings-with-orange-7f0b91fa2ca2.

Karami, A., Bennett, L. S., & He, X. (2018). Mining Public Opinion about Economic Issues. International Journal of Strategic Decision Sciences, 9(1), 18–28. https://doi.org/10.4018/ijsds.2018010102

Kimmel, L. (2018). Possibility of critical practice in computational design: applications on boundaries between public and private space. Cahiers de La Recherche Architecturale, Urbaine et Paysagère, 1. https://doi.org/10.4000/craup.361.

Langstraat, F., and R. Van Melik. 2013. Challenging the 'End of Pub-lic Space': A Comparative Analysis of Publicness in British and Dutch Urban Spaces. Journal of Urban Design 18 (3): 429–448.

Li, S. 2010. The Online Public Space and Popular Ethos in China. Media, Culture & Society 32 (1): 63–83.

Mac Síthigh, D. (2012). Virtual walls? The law of pseudo-public spaces. International Journal of Law in Context, 8(3), 394–412. https://doi.org/10.1017/S1744552312000262.

Madanipour, A. 2003. Public and Private Spaces of the City. London: Routledge.

Meneghetti, F. R. P. M. V. and F. K. (2020). Is it Crowdsourcing or Crowdsensing? An Analysis of Human (Vol. 26, Issue 2).

Munawir, Koerniawan MD and Dewancker BJ. (2019), Visitor Perceptions and Effectiveness of Place Branding Strategies in Thematic Parks in Bandung City Using Text Mining Based on Google Maps User Reviews. *Sustainability*.; 11(7):2123. https://doi.org/10.3390/su11072123

Paköz, M. Z., Sözer, C., & Doğan, A. (2022). Changing perceptions and usage of public and pseudo-public spaces in the post-pandemic city: the case of Istanbul. Urban Design International, 27(1), 64–79. https://doi.org/10.1057/s41289-020-00147-1.

Saldaña, Zoë Wilkinson (2018), "Sentiment Analysis for Exploratory Data Analysis," Programming Historian 7, https://doi.org/10.46430/phen0079.

Sanghvi, Kavish (2020, 8 May), Image Classification Techniques [Medium post], retrieved from https://medium.com/analytics-vidhya/image-classification-techniques-83fd87011cac.Sınmaz, S, (2018), Türkiye'de Kentsel Planlama ve Dönüşüm Sürecinde Eksik bir halka: Yarı Kamusal Mekanlar, Temmuz-ağustos, M. M. O. (n.d.). The Branch Mimarlık ve Tasarım Müzesi.

Ström, T.E. (2017), Mapping Google Maps: Critiquing an Ideological Vision of the World, Doctore of Philosophy, Western Sdyney University.

Ström, T.E. (2020), Journey to the center of the world: Google Maps and the abstraction of cybernetic capitalism, Cultural Geographies 1-19.

Sulis, E., Humphreys, L., Vernero, F., Amantea, I. A., Di Caro, L., Audrito, D., & Montaldo, S. (2020). Exploring network analysis in a corpus-based approach to legal texts: A case study. CEUR Workshop Proceedings, 2690, 27–38.

Yang, S. (2013), Investigation into Shanghai Spatial Publicness. Ph.D. Thesis, Technische Universität München, München, Germany.

Zong, C., Xia, R., & Zhang, J. (2021). Text Classification. In Text Data Mining. https://doi.org/10.1007/978-981-16-0100-2_5.

Url-1 < https://www.arkiv.com.tr/ >, accessed at May, 2022.

Url-2<https://orange3-text.readthedocs.io/en/latest/widgets/wordcloud.html>, accessed at June 2022.

Url-3<https://orange3-text.readthedocs.io/en/latest/widgets/sentimentanalysis.html>, accessed at June 2022.