Leveraging Google Maps Data for Onomastic Research: A Case Study of Javanese Business Names

Eric Kunto Aribowo[1*], Dwi Bambang Putut Setiyadi[1], Joko Sukoyo[2]

[1] Universitas Widya Dharma, Jl. Ki Hajar Dewantara, Klaten, Central Java, Indonesia
[2] Universitas Negeri Semarang, Semarang City, Central Java, Indonesia
erickunto@unwidha.ac.id

Abstract. This paper presents digital onomastics as a novel approach to the study of business naming conventions, focusing on the collection of data through web scraping. Using Leads Snipper to scrape business names from Google Maps across seven cities in Java, Indonesia, the study demonstrates the efficiency of digital tools in building large-scale corpora for onomastic research. The research highlights key ethical considerations, including privacy, data ownership, and consent, which are critical when handling digital data. By analyzing the resulting corpus, the paper identifies patterns in regional and linguistic naming conventions, offering insights into how globalization and local culture influence business names. The findings reveal the potential for digital onomastics to expand traditional research boundaries and create new opportunities for large-scale and real-time analyses of naming systems. Ethical guidelines and best practices for using digital data in onomastics are also discussed.

Keywords: digital onomastics, business names, web scraping, ethics, Leads Sniper, Google Maps.

1. Introduction

Onomastics, the study of names, has traditionally relied on manual data collection methods such as fieldwork, archival research, and population registries[1,2]. These conventional approaches, while rich in detail, often face limitations in terms of geographic scope and scale. As the digital age advances, new methodologies have emerged, expanding the potential for onomastic research. Digital onomastics employs data scraping, Big Data analysis, and other digital tools to collect and analyze large corpora of names, providing a more efficient and expansive approach to studying naming conventions[3].

Recent research highlights the growing trend of utilizing electronic databases and corpus-based methodologies in onomastics. Ainiala & Jantunen[4] discuss the use of electronic corpora to analyze the prevalence and usage of names in textual contexts, including collocations, phraseological relations, and regional or local variations. Similarly, Kersten & Lotze[5] explore the application of digital data in studying usernames, illustrating the expanding scope of onomastics beyond traditional naming conventions. For toponymic studies, Gammeltoft[6] emphasizes the importance of integrating locational

M. I. Sukardi et al. (eds.) *Proceedings of the International Conference Entitled Language, Literary, And Cultural Studies (ICONLATERALS 2024)*, Advances in Social Science, Education and Humanities Research 928,
https://doi.org/10.2991/978-2-38476-404-4_11

information, especially to distinguish between places with identical names, underscoring the necessity for comprehensive contextual data in digital onomastic research.

The adoption of digital tools in onomastics enables researchers to address these knowledge gaps. Tools like Leads Sniper allow the collection of vast datasets from publicly accessible platforms such as Google Maps. This method facilitates the creation of real-time corpora that reflect ongoing cultural, linguistic, and economic changes, particularly in regions like Java, Indonesia, where businesses often blend traditional and modern naming conventions[7]. By analyzing business names across seven cities in Java, this study demonstrates how digital tools can enhance onomastic research and address the aforementioned challenges, including data scalability, contextual richness, and geographic variation.

Furthermore, as Carsenat[3] highlights, the integration of Big Data techniques into onomastics provides new opportunities to uncover patterns across cultural and geographical contexts. However, while much progress has been made in employing corpus-based methods, there remains a gap in demonstrating the utility of digital scraping technologies for onomastic studies, particularly in non-Western contexts. This study contributes to this growing field by utilizing web scraping technologies to explore business names as markers of identity in the Javanese context, offering insights into how traditional and modern naming conventions coexist.

This paper focuses on the methodology of data collection rather than the detailed analysis of the scraped data. By employing Leads Sniper to gather business names from multiple cities, we aim to showcase the potential of digital onomastics in expanding the scale and scope of name studies, while also addressing the ethical considerations that arise in using publicly available data[8]. This research offers a framework for future onomastic studies that seek to leverage digital tools and data mining techniques for larger, more comprehensive analyses, thus advancing the field into new and underexplored areas.

2. Literature Review

Onomastics has evolved over time from traditional approaches to modern methodologies, incorporating digital tools and Big Data to expand the scope of research. Early onomastic studies relied on manual data collection through oral traditions, written records, and archival materials. For example, Burns[9] demonstrated how field-names in Scotland were collected using spoken interviews as part of socio-onomastic research. These traditional methods provided rich, localized data but were often constrained by the time and effort required for manual collection.

In recent years, digital onomastics has gained prominence by utilizing digital data sources and advanced data collection methods. The integration of Big Data techniques into onomastic research allows for the processing of large datasets, providing a broader and more comprehensive analysis of naming patterns across regions and cultures[3]. As Mansilla[7] explains, the use of digital onomastic vocabulary enables the study of names on a much larger scale, offering new insights into the cultural and linguistic forces shaping name usage.

Further advancements in onomastic methodologies include the development of specialized tools for data gathering, such as Leads Snipper (https://www.leads-sniper.com), which allows researchers to scrape names and associated information from online platforms like Google Maps. This shift toward digital data collection has made it possible to compile extensive corpora of names quickly and efficiently, expanding the geographic and cultural range of onomastic studies[10]. By analyzing business names from multiple regions, researchers can now explore patterns and trends in naming conventions that reflect local customs, economic activities, and global influences.

Additionally, the ethical considerations of collecting and analyzing digital data are increasingly important in onomastic research. Tiidenberg[8] emphasizes the need for researchers to ensure transparency, privacy, and compliance with data protection laws when working with digital datasets. The public availability of information such as business names does not negate the ethical responsibility to handle the data responsibly, particularly when dealing with sensitive cultural and personal contexts.

Onomastics has also intersected with Big Data, offering opportunities to explore naming systems on a global scale. Carsenat[3] suggests that digitalization and data mining techniques can reveal previously hidden trends in name usage across various regions and languages. This modern approach to onomastics not only allows for the study of names but also provides a lens through which broader social, cultural, and economic dynamics can be examined.

In conclusion, the transition from traditional to digital onomastics represents a significant expansion of the field, allowing researchers to explore names in new and innovative ways. By leveraging digital tools and large-scale datasets, onomastic studies are now able to provide a deeper understanding of the linguistic, cultural, and social forces that shape naming conventions across the world.

3. Methodology

This study employed a digital onomastic approach to collect and analyze business names from seven cities in Java—Klaten, Sukoharjo, Surakarta, Semarang, Yogyakarta, Surabaya, and Malang—using the web scraping tool Leads Sniper. The goal was to create a comprehensive corpus of business names, capturing both traditional and modern naming conventions across various regions.

The selection of these cities was motivated by their role in supporting Indonesia's tourism industry through culinary businesses, which serve as a key driver for regional economic activities[11–13]. These cities represent diverse urban and semi-urban settings where Javanese cultural influences intersect with contemporary business practices, making them ideal for studying traditional and modern naming strategies.

The methodology is divided into several stages: data collection, cleaning and preparation, and visualization.

a. Data Collection Process

The data were collected using Leads Sniper, a web scraping tool that extracts business data from Google Maps based on specific keywords and locations. This tool was

selected for its ease of use and efficiency in collecting large datasets without requiring advanced programming skills. The keywords were tailored to different business types (e.g., *rental motor*, *kedai kopi*, *restoran*) to ensure relevant results, while city names were added to focus the search on specific geographical areas. By combining business category keywords with location-specific searches, we were able to gather a targeted dataset of names (see **Fig. 1**).

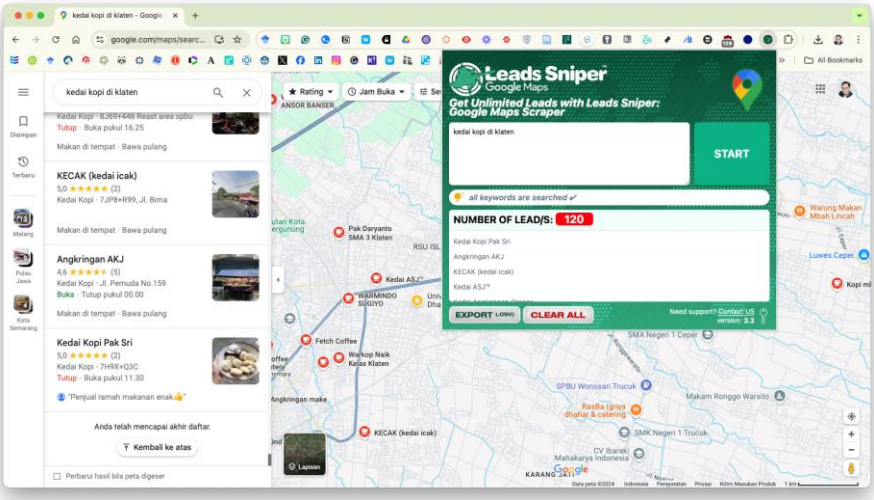

Fig. 1. Screenshot of Leads Sniper interface during keyword input for location-specific business name collection.

The scraping process retrieved the following data points for each business: name, full address, geolocation, and additional details such as contact information when available. The data were exported into CSV format for further analysis.

b. Data Cleaning and Preparation

After scraping, the raw data were imported into Google Sheets for cleaning and preparation. This step involved removing duplicate entries and irrelevant businesses that did not fit the research focus, such as businesses outside the targeted cities or those with incomplete information. Additionally, manual verification was conducted to correct errors in names, especially those written in non-Latin scripts such as Javanese. By standardizing the data, we ensured a consistent format for analysis.

Cleaning the data was critical for maintaining the integrity of the dataset, particularly in the context of comparing naming conventions across regions. Business names were categorized by type and location, enabling us to filter the dataset for specific industries or regions as needed.

c. Data Visualization

Once the dataset was cleaned and prepared, the next step involved visualizing the data using Looker Studio (https://lookerstudio.google.com). To better illustrate the distribution of Javanese business names across Java, we generated an interactive map displaying the precise locations of businesses that utilize Javanese language in their names. The coordinates for each business location were obtained during the web scraping process using Leads Sniper, ensuring geographical accuracy for the plotted data.

The interactive map allows for zooming in and out, providing users with the ability to explore the spatial distribution of businesses more intuitively. Annotations and additional information, such as business type and name, are embedded within the map, making it possible to click on each data point to view detailed information (see **Fig. 2**).

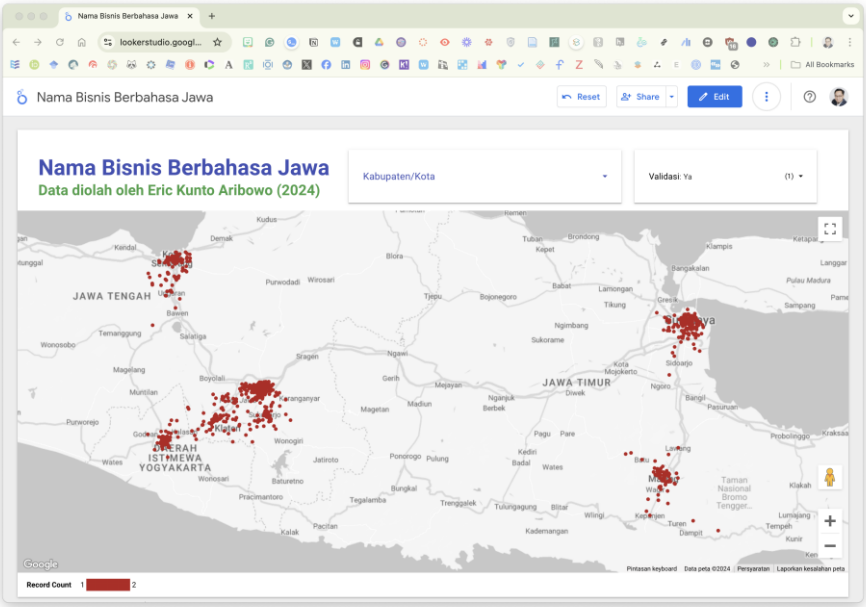

Fig. 2. Interactive map of Java displaying business locations using Javanese language in their names, based on scraped data.

d. Ethical Considerations

As with any research involving digital data collection, ethical considerations were paramount. Since the data were scraped from a publicly accessible platform (Google Maps), no personal information was collected, and the research adhered to ethical standards regarding data privacy. In line with Tiindenberg[8], transparency and data protection were prioritized to ensure compliance with platform guidelines and legal regu-

lations. Anonymization techniques were applied where necessary to protect any sensitive information, and the research remained within the bounds of responsible data usage.

e. Tools and Software

This study utilized a combination of tools and software to streamline the data collection, cleaning, and visualization processes. Leads Snipper was employed to scrape business data from Google Maps, allowing for the extraction of large datasets in CSV format. After scraping, the raw data was imported into Google Sheets, where it was cleaned, filtered, and organized. Google Sheets proved to be an effective tool for managing the dataset, enabling easy manipulation and ensuring the removal of duplicates and irrelevant entries. Finally, the cleaned data was visualized using Looker Studio (formerly Google Data Studio), which provided an intuitive platform for creating dashboards and visual representations of the data. Looker Studio's capabilities allowed the research to identify and highlight patterns in business naming across different cities and industries, offering valuable insights for further onomastic analysis.

4. Data Corpus in Digital Onomastics

The development of a comprehensive data corpus is essential in digital onomastics for analyzing naming conventions across different regions and cultural contexts. In this study, the corpus was built by scraping business names from Google Maps across seven cities in Java: Klaten, Sukoharjo, Surakarta, Semarang, Yogyakarta, Surabaya, and Malang. This method allowed for the creation of a large-scale, dynamic corpus that reflects both traditional and modern naming practices in the region.

a. Structure of the Corpus

The data corpus comprises several critical components that provide a detailed overview of the business naming conventions in Java. The key elements include:

- **Business Name**: The primary focus of the corpus is the collection of business names, which were scraped from Google Maps. These names reveal how local businesses use language to construct their identity, with some incorporating traditional Javanese elements while others opt for more modern or internationalized names.
- **Location**: Each business name is associated with its full address, allowing for a regional analysis of naming patterns. By linking names to specific cities and areas, this research was able to uncover regional differences in the use of Javanese script or modern Indonesian terms.
- **Industry**: The businesses were categorized by industry based on the keywords used during the scraping process. This categorization enables a focused analysis of how naming conventions differ across sectors, such as retail, food services, or traditional crafts.

- **Coordinates**: The exact geographical coordinates of each business, extracted during the scraping process, were included to support location-based visualizations. This allowed for mapping the spatial distribution of traditional and modern business names in Java.

b. Insights from the Corpus

The corpus reveals significant insights into how businesses in Java navigate between traditional and modern naming practices. For instance, businesses in cultural hubs like Yogyakarta tend to use more traditional names. In contrast, urban centers such as Surabaya show a higher prevalence of modern and internationally influenced names. These findings align with previous research suggesting that urbanization and globalization impact local naming practices[7].

Furthermore, the analysis of business names by industry shows that certain sectors, such as tourism and local crafts, are more likely to preserve traditional naming conventions. In contrast, sectors like retail and food services tend to adopt names that cater to broader, often global, audiences[3].

c. Data as a Dynamic Resource

One of the main advantages of digital onomastics is the ability to continuously update and expand the data corpus. As new businesses are registered or existing ones change their names, the corpus can be revisited and expanded. This dynamic nature allows researchers to track real-time changes in naming practices, offering valuable insights into ongoing cultural and economic shifts in Java.

By using tools like Leads Snipper and combining them with modern data visualization platforms like Looker Studio, the corpus becomes a living resource that can adapt to new data and reveal trends that may not be visible through static or smaller-scale data collections.

d. Limitations and Future Directions

While the corpus offers substantial insights, there are limitations to be considered. The reliance on publicly available data from Google Maps means that businesses not registered on the platform are excluded from the dataset, potentially omitting informal or smaller-scale enterprises that play a role in regional naming conventions. Additionally, while the scraping process was largely automated, manual data cleaning was necessary to address inconsistencies[9].

Future research could explore the integration of other data sources, such as social media platforms or local directories from Department of Industry and Trade, to create a more comprehensive dataset. Moreover, expanding the corpus to include longitudinal data would allow for the study of how naming practices evolve over time, particularly in response to socio-economic changes.

5. Opportunities for Onomastic Research Using Scraped Data

The integration of digital tools like Leads Snipper and the use of web scraping techniques offer numerous opportunities for onomastic research that were previously difficult to achieve using traditional methods. Scraped data from publicly available sources such as Google Maps can provide researchers with large-scale corpora of names that reflect a wide range of linguistic, cultural, and social dynamics. These digital datasets enable more comprehensive analyses and open up new avenues for research, particularly in the study of regional variations, linguistic shifts, and the influence of globalization on naming conventions.

a. Large-Scale and Real-Time Data Collection

One of the most significant opportunities presented by scraped data is the ability to collect vast amounts of information quickly and efficiently. Traditional onomastic research often required labor-intensive methods, such as fieldwork, archival research, and surveys, which were limited in both scope and scale[10]. In contrast, digital scraping allows researchers to gather thousands of names across multiple regions and business sectors in a short period of time. This makes it possible to create extensive corpora that represent not only business names, but also place names, and institutional names, all of which can be analyzed for onomastic patterns.

Additionally, the use of digital platforms like Google Maps ensures that the data collected is up-to-date. Businesses often update their information regularly, allowing for real-time analysis of trends and naming practices. This dynamic nature of the data is particularly valuable for studying how external factors such as economic changes, globalization, and urbanization influence naming conventions over time[7]. Researchers can revisit the same regions or industries periodically to observe shifts in naming practices, offering insights into how cultural and social dynamics evolve.

b. Analysis of Regional Naming Conventions

Scraped data enables researchers to explore regional differences in naming conventions, an area of great interest in onomastics. In this study, the dataset collected from businesses across seven cities in Java—Klaten, Sukoharjo, Surakarta, Semarang, Yogyakarta, Surabaya, and Malang—provides a unique opportunity to compare how business names vary across these regions. For example, initial observations suggest that traditional Javanese naming conventions are more prevalent in cities like Yogyakarta and Surakarta, while more modern or globally influenced names are commonly found in urban centers like Surabaya. To illustrate, businesses in Yogyakarta and Surakarta frequently incorporate Javanese words, reflecting cultural identity and local heritage. Examples of such names include *Omah Wedangan*, *Omah Bakul Buah*, *Kopen*, *Pantes*, *Marem*, and *Nglaras Ngopi* from Surakarta, and *Sego Brakot*, *Sekeco*, *Klambi Apik*, *Monggo*, *Omah Uti*, *Kembang Gula*, *Gemah Ripah*, and *Sekar Jagad* from Yogyakarta.

These examples highlight the strong presence of Javanese linguistic and cultural elements in certain regions, which contrast with more modern or internationalized naming practices observed in larger cities like Surabaya. This preliminary analysis demonstrates the effectiveness of the proposed methodology in capturing and analyzing naming conventions across different urban and cultural landscapes.

This regional comparison can reveal how local cultures, languages, and economic activities shape the way businesses are named. In addition, the corpus allows for a granular analysis of the impact of religious and cultural factors on naming conventions, particularly in cities with strong local identities or specific cultural heritages[9]. This approach can provide valuable insights into how regional identity is preserved or transformed through business names, offering a rich field of study for onomasticians.

c. Linguistic Innovation and Borrowing

The use of digital corpora also facilitates the study of linguistic innovation and borrowing in names. Businesses often adopt names that reflect not only local languages but also international influences, particularly in more globalized urban areas. In Indonesia, this phenomenon is evident in cities like Surabaya and Semarang, where businesses frequently use English or hybrid names that combine elements of Indonesian, Javanese, and foreign languages. Hybrid names often emerge as a creative blend, utilizing foreign words as markers for products or commodities being traded.

For example, in Surabaya, business names such as *Kedai Sanga*, *Omah Kopi Kota*, *Mbois Barber*, *Sekar Sari Florist*, and *Morodadi Furniture* reflect a combination of Javanese words with Indonesian or English. Similarly, in Semarang, examples like *Gondrong Arts Studios*, *Moro Juice*, *Guyub Rent*, *Rorojonggrang Art*, and *Kembangdjati Workshop* showcase the interplay between local linguistic identity and globalized influences. This trend highlights the influence of globalization on naming practices, as businesses seek to appeal to a broader, often international, customer base. Hybrid names, in particular, reflect a strategic linguistic innovation, where traditional and modern elements are combined to create a distinctive identity while catering to diverse audiences[14].

By analyzing the scraped data, researchers can track the frequency and patterns of linguistic borrowing in business names. This opens up opportunities to explore how global cultural and economic forces interact with local linguistic traditions. It also provides insights into the strategies used by businesses to navigate between maintaining a local identity and appealing to international audiences.

d. Expanding the Scope of Onomastic Studies

Scraped data from digital sources allows onomastic research to extend beyond traditional focus areas, such as personal and place names, into new domains like business and institutional names. As noted by Mansilla[7], the study of business names offers a unique perspective on cultural and economic activities within a region, providing a

complementary dimension to the study of personal names. Business names, in particular, reflect the identity and branding strategies of their owners, as well as the sociocultural values they wish to convey to their customers.

Moreover, this approach opens up interdisciplinary opportunities, allowing onomastic research to intersect with fields such as marketing, sociolinguistics, and cultural studies. The large datasets gathered through web scraping enable more robust statistical analyses, allowing researchers to identify correlations between naming conventions and various demographic or economic variables.

e. Future Research Directions

The potential of digital onomastics is vast, and the use of scraped data can support a wide range of future research projects. For example, researchers could expand the scope of their studies by incorporating additional regions or exploring other business sectors to identify broader trends in naming practices. Additionally, scraped data could be used in combination with traditional onomastic methods, such as ethnographic research, to provide deeper insights into the motivations behind specific naming choices.

The ability to analyze naming trends over time also presents exciting opportunities for longitudinal studies. By revisiting the same businesses or regions at different points in time, researchers can observe how naming conventions evolve in response to social, cultural, and economic changes. This would provide valuable data for understanding the long-term impact of globalization, urbanization, and technological advancements on naming practices.

In conclusion, the use of scraped data in onomastic research opens up new possibilities for large-scale, real-time analysis of naming conventions. It allows researchers to explore regional differences, linguistic innovation, and the broader socio-cultural implications of names in various contexts. By leveraging digital tools and methodologies, onomastic studies can continue to expand and offer new insights into the dynamic relationship between language, culture, and identity.

6. Conclusion

a. A Subsection Sample

This study demonstrates the value of digital tools such as Leads Snipper for expanding the scope of onomastic research. By using web scraping to collect business names across Java, researchers can efficiently build large-scale corpora that capture both traditional and modern naming conventions. This approach offers significant opportunities for analyzing regional differences, linguistic innovation, and the broader socio-cultural dynamics reflected in names.

The integration of big data techniques and interactive visualization tools enhances the ability to explore naming trends geographically and over time, providing new insights into how names evolve in response to globalization and local identity. While digital onomastics presents numerous opportunities, ethical considerations surrounding data privacy and responsible use remain paramount.

In conclusion, digital onomastics opens new avenues for research, enabling more dynamic and large-scale analyses that enrich our understanding of the linguistic and cultural forces shaping names.

Acknowledgments. This research was funded by the Ministry of Education, Culture, Research, and Technology under the 2024 Basic Research Grant with Master Contract Number: 108/E5/PG.02.00.PL/2024 and Derivative Contract Numbers: 017/LL6/PB/AL.04/2024 and 002/F.01.01/LPPM/VI/PENELITIAN/2024.

Disclosure of Interests. The Authors declares that there is no conflict of interest.

References

1. Aribowo, E. K. & Herawati, N. Pemilihan Nama Arab sebagai Strategi Manajemen Identitas di antara Keluarga Jawa Muslim. in *International Seminar Prasasti III: Current Research in Linguistics* 270–277 (2016).
2. Aribowo, E. K. & Herawati, N. Trends in Naming System on Javanese Society: A Shift From Javanese to Arabic. *Lingua Cultura* **10**, 118–122 (2016).
3. Carsenat, E. Onomastics for Business: Can Discrimination Help Development? https://www.paristechreview.com/2013/09/17/onomastics-business/ (2013).
4. Ainiala, T. & Jantunen, J. H. Korpusonomastinen tutkimus slanginimistä hesa ja stadi digitaalisissa diskursseissa. *Sananjalka* **61**, 57–79 (2019).
5. Kersten, S. & Lotze, N. Anonymity and authenticity on the web: towards a new framework in internet onomastics. *Internet Pragmat.* **5**, 38–65 (2022).
6. Gammeltoft, P. Place-Name Databases: A Spatio-Temporal Mess. in *Digital Spatial Infrastructures and Worldviews in Pre-Modern Societies* (eds. Petrulevich, A. & Skovgaard Boeck, S.) 135–158 (Amsterdam University Press, 2023). doi:10.1515/9781802700794-008.
7. Mansilla, F. R. The Onomastic Vocabulary as the Cultural Lacunarity Marker. in *Synergy of Languages & Cultures: Interdiscipilinary Studies* 184–196 (St. Petersburg State University, 2023). doi:10.21638/2782-1943.2022.15.
8. Tiidenberg, K. Ethics in Digital Research. in *The SAGE Handbook of Qualitative Data Collection* 466–479 (SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road London EC1Y 1SP, 2018). doi:10.4135/9781526416070.n30.
9. Burns, A. Field-Names of North-East Scotland: A Socio-Onomastic Study. (University of Glasgow, 2015).
10. Rutkowski, M. Metodologie W Onomastyce a Kultura. *Onomastica* **66**, 37–50 (2022).
11. Ardriyati, W. & Wiwaha, J. A. An Exploratory Study of Typical and Traditional Culinary Arts in Surakarta and Semarang as Cultural Heritage to Support Indonesian Tourism Industry. *Asian Cult. Hist.* **8**, 132 (2016).
12. Jayanti, C. T. Struktur Diksi dalam Penamaan Rumah Makan Pedas di Malang Raya. *Ranah: J. Kaji. Bhs.* **13**, 116 (2024).
13. Wijayanti, A. Wisata Kuliner sebagai Strategi Penguatan Pariwisata Di Kota Yogyakarta, Indonesia. *Khasanah Ilmu - J. Pariwisata Dan Budaya* **11**, 74–82 (2020).

14. Butler, J. O. Name, Place, and Emotional Space: Themed Semantics in Literary Onomastic Research. (University of Glasgow, 2013).