


## WEB SCRAPING AND DATA SCIENCE IN APPLIED RESEARCH IN COMMUNICATION: a study on online reviews

WEB SCRAPING E CIÊNCIA DE DADOS NA PESQUISA APLICADA EM COMUNICAÇÃO: um estudo sobre avaliações online

WEB SCRAPING Y CIENCIA DE DATOS EM LA INVESTIGACIÓN APLICADA EN COMUNICACIÓN: un estudio sobre evaluaciones online

### Marcello Tenorio de Farias

Mestre em Comunicação pela Universidade Municipal de São Caetano do Sul (USCS). Especialista em Gestão de Marketing e Comunicação pela Escola de Comunicações e Artes da Universidade de São Paulo (E. Professor na Universidade São Judas Tadeu. [cello.farias@hotmail.com](mailto:cello.farias@hotmail.com).

 0000-0001-7453-9461


### Alan César Belo Angeluci

Doutor em Ciências com ênfase em Sistemas Eletrônicos pela Escola Politécnica da Universidade de São Paulo (USP), com período sanduíche na University of Brighton (Inglaterra). Possui estudos de Pós-Doutorado na University of Texas at Austin e na Escola de Comunicações e Artes da USP. Docente permanente do PPGCOM e PPGE da Universidade Municipal de São Caetano do Sul (USCS). [angeluci@gmail.com](mailto:angeluci@gmail.com).

 0000-0002-4093-0590

### Brasilina Passarelli

Professora Titular pela Escola de Comunicação e Artes (ECA) da Universidade de São Paulo (USP) e Diretora (2021-2024). Coordenadora Científica do NACE Escola do Futuro - USP. Vice-Presidente da Agência USP de Gestão da Informação Acadêmica - AGUIA. Doutora em Comunicação pela ECA/USP com período sanduíche na Michigan State University. Realizou estudos de pós-doutorado na Universidad Carlos III de Madrid. Foi Diretora de Marketing para a Educação da Apple Computer Brasil. [linapassarelli2@gmail.com](mailto:linapassarelli2@gmail.com).

 0000-0003-2949-7914

Correspondência: Universidade Municipal de São Caetano do Sul, Pró-Reitoria de Pós-graduação e Pesquisa, Rua Conceição, 321 - Santo Antônio - São Caetano do Sul, São Paulo, Brasil - CEP: 09530-060.

Received: 02.16.2021.  
Accepted: 04.16.2021.  
Published: 07.01.2021.

### ABSTRACT

With the spread of access and use of information through the web and social networks, information retrieval in large volumes of data has become unfeasible by manual methods. In this applied study, the contribution of the development and use of a prototype tool for automatic data scraping from online evaluations made on Google Maps – Discovery Stars – was reported. The retrieved data allowed us to investigate how these assessments can have the potential to influence the behavior of the platform's users. Among the results, it was observed that the reading and posting of reviews impact the formation of opinion and motivations of Google Maps users.

**KEYWORDS** Online Evaluation; Data Scraping; Web scraping; Data Science; Communication.

## Introduction

The great advances that have taken place in recent years related to Information and Communication Technologies (ICT), supported by the Internet and by the hyperconnected society have led to the emergence of digital platforms whose main objective is to collect user opinions, in the form of evaluations, about public and private places, business, products and services. These platforms work like a large catalog that gather opinions left by users who use them - reviews - about places, products, or services, and are consulted by other users who seek more information. In this context, a highly collaborative scenario emerges, in which diverse opinions are exposed and consumed at a fast pace, typical of the Internet.

As Passarelli (2007, p. 11) highlighted, "in a globalized network society, permeated by the immediacy and transience of events, some kind of permanence needs to be created in the midst of virtual communities". According to the author, this permanence is manifested through folksonomies: digital footprints of users on social networking sites, wikis and other contemporary digital communication devices that allow users of these networks to identify information that may be relevant to their online experiences - strongly influencing their decision making. The boom in mobile platforms over the last decade has made this phenomenon increasingly striking, prompting methodological adjustments to deal with the challenges of collecting and analyzing massive digital data.

Based on the previous experience of studies on users' behaviors and habits in networks (Angeluci; Rosa; Passarelli, 2020; Passarelli; Angeluci, 2018; Passarelli; Junqueira; Angeluci, 2014), the study presented here aimed to collaborate with the understanding of possible new habits made possible by the behavior of performing and consuming evaluations on these platforms. Among the platforms that allow evaluations to be carried out, the chosen one was Google Maps, as it is a geolocation and maps service that is growing in popularity and with high usage, launched in 2005 by Google.

In addition to being a platform used worldwide, it has undergone several modifications and updates over the years, becoming an environment with a high flow of collaboration and exchange of information on the Internet. By emphasizing reviews on locations that are embedded in the platform, the system allows the user to write a review, share photos, videos, and answer queries about the location in question.

To carry out an analysis of user's behavior of the Google Maps rating platform to understand their motivations for using it, the first challenge to be faced is to extract data on online habits and usage. Netnography is the methodological basis of this study, but the research work also required the application of unconventional strategies in studies in the field of communication that allow the extraction of a massive volume of data from these environments, demanding an effort to innovate in the method. It is understood that this big data, far beyond quantitative issues, reveals a genuine photograph of a phenomenon and needs to be observed from its massive nature.

Therefore, web scraping techniques seem to become relevant in this context: through computational intelligence, it is possible to collect a massive amount of data from digital environments automatically using a robot or algorithm programmed for a certain task. Always observing the ethical and legal aspects, data scraping can be an important strategy in contemporary research that focuses on digital phenomena and needs to be embraced by the field of communication for the benefit of diversity and the improvement of its methodological practices. As well positioned by Santos (2018, p. 20):

It is not defended here sudden changes or any kind of general reformulation, much less the devaluation of the entire academic effort of generations of researchers who, with their work, built a field of Brazilian and legitimate knowledge, but rather the possibility of an expansion, of an airing and freedom to propose and experiment with new connections, with unusual areas until then such as Computer Science, Design, Engineering and Cognitive Sciences, for example.

Regarding the locus of this research, the place chosen to analyze the evaluations made on Google Maps was a public square located in the Bela Vista neighborhood, in the central region of the city of São Paulo: Franklin Roosevelt Square. As a high-flow public place, the square's main characteristic is the city's cultural diversity, visible through the intense flow of people of different ages, ethnicities, cultures, and interests. This is already revealed in a pre-observation of the evaluations left on the website, which makes it a complex and full of controversies object of study, with the possibility of analyzing data from different profiles of users of the platform.

This paper is structured as follows: at first, the methods for data collection are presented, starting from Kozinets' netnographic approach, and detailing the development and use of the Python script for web scraping – the Discovery Stars prototype tool. The analysis of more quantitative data is presented below and was supported and discussed based on data obtained through semi-structured interviews. The results show the users' motivations, ethical responsibility, privacy, and opinion formation.

### **Materials and methods: development and application of the web scraping tool in netnography**

As part of the netnographic trail, data collection involved the development of a crawler robot to scrape data from the Franklin Roosevelt Square page on Google Maps. About the web scraping technique, Santos (2015, p. 7) observes:

Automated data collection, also known as scraping or mining is an increasingly common resource (...) it can, in the case of academic work, be used both for the execution of repetitive routines, allowing the researcher more time for more complex tasks, such as identifying patterns and trends in large volumes of information.

Able to perform automated data extraction on Internet sites that do not naturally provide this data for automatic downloading, web scraping enables data extraction

through scripts capable of identifying, collecting, and storing metadata for further analysis.

It is necessary to observe the innovative capacity that web scraping provides to communicational studies, contributing with perspectives that connect with data science, artificial intelligence, and computing, in an exercise of collaboration between areas of knowledge:

It is important to highlight that data mining seeks to extract innovative knowledge, that is, something previously unknown and that has value for the domain in which it is applied. (FACELI, K., LORENA, AC, GAMA, J., and CARVALHO, A., 2011, p. 331)

For the creation of scripts capable of performing web scraping, in general, the Python programming language is used, which writes the algorithms for information retrieval. The extraction, in this research, provided the collection of assessments made about Franklin Roosevelt Square, in Google Maps, and, therefore, it was necessary to develop a specific script in this language so that it was possible to extract the data directly to a structured data sheet.

This script, which can be understood as a set of command codes developed to make data scraping possible, was incorporated into a web scraping tool to perform this function – named Discovery Stars<sup>1</sup> and created by the authors of this study. The code developed used the combination of Python language with Selenium, which works as an integrated development environment for automated test scripts. When run, it allows the script developed in Python language to control a browser, simulating a human action, but with an amount of data that cannot be manually analyzed in a timely manner. This technique is commonly used in scenarios that need agility to perform repeated tasks - as in the case of the "copy and paste" command - in large databases. In this sense, the developed script aims to copy assessments made about a particular location on Google Maps and paste it to a spreadsheet, in an agile way and with minimal occurrence of failures.

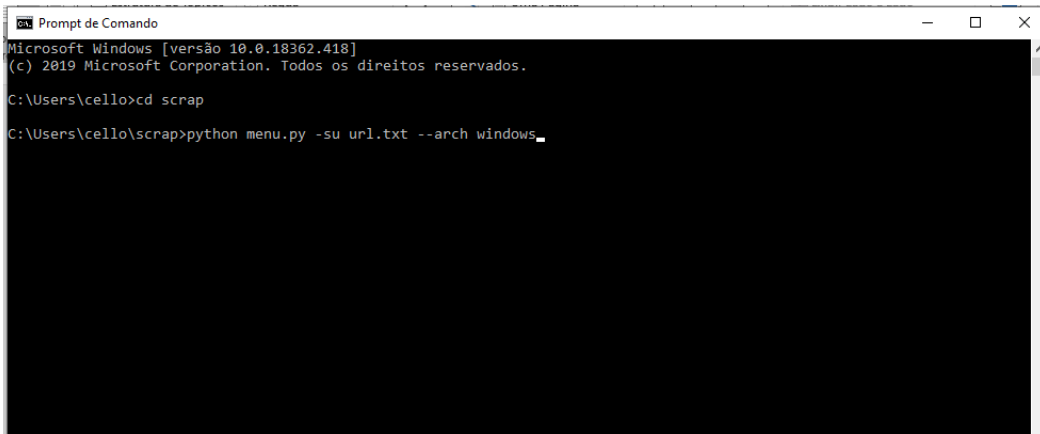
After the script was developed and installed on the computer, the first step was to access Google Maps and find on the platform the desired location where the data would be extracted – the square. After identifying the URL, the address was hosted in a specific file, created in the installed code folder, directing the script to extract data from that particular URL.

---

<sup>1</sup> The tool developed, Discovery Stars, was made available for download, for exclusively academic purposes, at <https://sites.google.com/view/discoverystars/initial-page>. Programming codes can also be requested via the form on the indicated page.

Therefore, to start the tool's operation, the "Command Prompt"<sup>2</sup> of the Windows operating system was run, and it was possible to start the data extraction process, as shown in Figure 1.

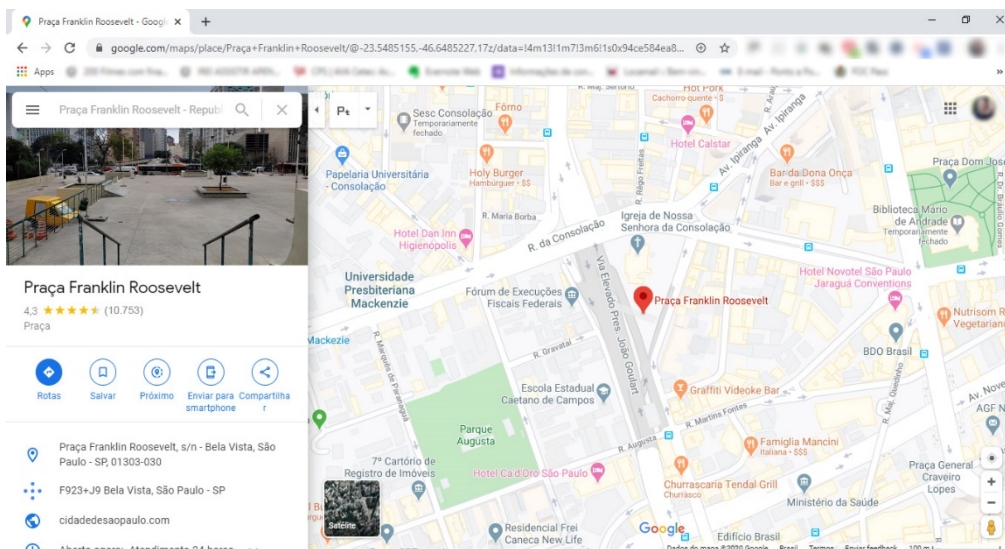
Figure 1 Command Prompt.



Source: Authors' elaboration.

Thus, the code inserted automatically ran the Google Maps screen with the location that was previously chosen, starting the automatic copy of each of the comments, as shown in Figures 2 and 3.

Figure 2 Google Maps location in browser.



Source: Authors' elaboration.

<sup>2</sup> Tool available on Windows system that allows running commands entered from executable code.

Figure 3 Script extracting comments.

```

C:\> Prompt de Comando - python menu.py -su url.txt --arch windows
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements HTTP/1.1" 200 643
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element {"using": "css selector", "value": "html body.keynav-mode-off.screen-mode-jsl div#app-container.vasquette.pane-open-mode div#content-container div#pane div.widget-pane.widget-pane-visible div.widget-pane-content.scrollable-y div.widget-pane-content-holder div.section-layout.section-layout-root div.section-layout.section-scrollbox.scrollable-y.scrollable-show"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element HTTP/1.1" 200 88
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value {"text": "\ue00f", "value": ["\ue00f"], "id": "377be584-fb15-42f6-a01c-f3f8eb5d6601"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value HTTP/1.1" 200 14
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements {"using": "css selector", "value": "div.section-review.ripple-container div.section-review-content"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/elements HTTP/1.1" 200 643
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: POST http://127.0.0.1:53957/session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value {"text": "\ue00f", "value": ["\ue00f"], "id": "377be584-fb15-42f6-a01c-f3f8eb5d6601"}
2019-10-31 12:14:00 [urllib3.connectionpool] DEBUG: http://127.0.0.1:53957 "POST /session/32166e6b-5081-4e4a-a53a-e7e6340bea1c/element/377be584-fb15-42f6-a01c-f3f8eb5d6601/value HTTP/1.1" 200 14
2019-10-31 12:14:00 [selenium.webdriver.remote.remote_connection] DEBUG: Finished Request

```

Source: Authors' elaboration.

Finally, after the data scraping was completed by the script, all comments were extracted and saved in CSV format, as shown in Figure 4. The extracted data were separated by columns, which contain "Name", which is the name used by the profile who performed the comment; "Star", being the grade that the user gave to that location, rating from 1 to 5; "Rating", which means the amount of comments that user has already made with its profile on Google Maps; and "Text", which is the content of the assessment written by the user. This organization of information in metadata and attributes comes from the informational architecture of Google Maps

Figure 4 Data in CSV format.

Name	Star	Rating	Text
4	5	40	Ambiente tranquilo e calmo para dar uma caminhada ao ar livre ou praticar esportes como skate e patinação! Também cai bem levar o dog para dar um passeio por lá!
6	5	47	Quem nunca foi , tem que ir, e quem foi nunca mais vai deixar de ir ; é uma vibe muito louca varios tipos de gente, vários níveis sociais, mais em único sentido que curti a noite Paulista e a Praça.
7	5	49	Só precisa de mais segurança ????
8	5	59	Uma das principais praças e espaços de lazer do centro de São Paulo. Wi-fi livre e fácil acesso pra Consolação e Augusta. Ótima opção de lazer pra quem deseja conhecer a região do centro, mais por lado da Consolação.
9	4	59	Ideal para publico jovem ,moderno e desencanado, pessoas co mundo todo passa por aqui. Ideal para varias acoes, esportiva , ponto de encontro, tem os barzinhos ao redor, lugar massa!
11	4	29	Praça simples, porém enorme, com pessoas interessantes, alternativas... skatistas e tal. Ponto de passagem importante. Há presença da policia e algumas lojinhas.
11	3	28	Espaço multiuso, mas as vezes é impossível passar pelo local, pois são tantos cachorro correndo e pulando em você que mais parece um canil.

Source: Authors' elaboration.

Through netnography, according to Kozinets (2014, p. 95), these extracted data can be classified as archival data:

Data that the researcher copies directly from preexisting computer-mediated communications of members of an online community, data whose creation or stimulation he is not directly involved in.

On August 18, 2019, 4,008 assessments of the Franklin Roosevelt Square left on Google Maps were extracted until that moment, which brought a large sample of data with the variables of the assessments made. It is noteworthy that Discovery Stars performed this task of scraping over 4,000 data in a few minutes with great precision. This collection would certainly take weeks if done uninterruptedly by humans and would probably be subject to a high error rate.

### **Extracted data: interviewing information**

After extracting data from comments made about Roosevelt Square on Google Maps using Discovery Stars, the classic data science itinerary followed. The database was cleaned up, fixing special characters, and reorganizing the data by the platform's own category attributes, as mentioned before (Name / Star / Rating / Text). Then, through basic descriptive statistics, possible analytical cuts were defined with the extracted data and what relationships could be established in this scenario to understand the behavior of users who evaluate on the platform.

Within the available variables, it was possible to analyze the relationship between the number of evaluations already made by the user (rating) and the grade he gave (star). It was inferred whether users with a higher rating, that is, with more ratings in their history, usually have a behavior of evaluating places with higher or lower scores and analyzing users with lower ratings. This approach seemed to be promising for understanding the influence of online ratings on other users' behavior.

Thus, all users who gave score 1, score 2, score 3, score 4 and score 5 were separated. Then, an analysis of ratings within these profiles was performed, carrying out an investigation about how many ratings users, divided by scores, have in their profiles. In other words, it was possible to visualize within each grade given to Franklin Roosevelt Square how many ratings users had in their histories.

To perform this analysis, three different intervals of assessments were tested: 10 out of 10, 50 out of 50 and every 100 out of 100. The first attempt, every 10 out of 10, brought an analysis with a lot of data and with very large intervals. On the other hand, the attempt of 100 out of 100 brought an inaccurate analysis, where it was not possible to understand

the impact of the intervals. Finally, the spacing chosen for this study was the division with an interval of 50 in 50 ratings, which gives better indications about which are the highest concentrations of ratings in the users' history. This way, it was possible to understand if most users who gave a score 1 had less than 50 ratings in their history, for example. This was done for profiles that also scored two, three, four and five. A graphical visualization of this information is explored in the following sections.

### **Semi-structured interviews: better understanding user behavior**

Different profiles that made evaluations showing strong opinions about the Franklin Roosevelt Square were selected from the processed data, and a search of the respective names on Facebook was carried out. The goal was to try to find users and contact them. After finding them by similarity of names, contact was made through Messenger, confirming that it was the same person and then inviting to conduct semi-structured interviews by video call on the platform itself. About the semi-structured interview technique:

We can understand a semi-structured interview, in general, as one that starts from certain basic questions, supported by theories and hypotheses, that are of interest to the research, and which then offer a wide range of questions, the result of new hypotheses that emerge as if they receive the answers from the informant. In this way, the informant, spontaneously following the line of his thought and experiences within the main focus placed by the investigator, begins to participate in the elaboration of the research content (TRIVINOS, 1987, p. 146).

For the interviews, a question script was previously made with the purpose of relating the objectives of this research with the data collected by the participants during the scraping and analysis of the data. Conducting the video interview via Facebook facilitated contact with users and was in line with the assumptions of netnography. About this technique, Kozinets points out:

Using Skype or a similar medium for an online interview can mean you have face-to-face contact with someone. In many circumstances, this can be almost as good as a face-to-face interview in relation to reading and recording social indicators such as body language (KOZINETS, 2014, p. 107).

Among the selected users, those with the highest number of evaluations in their history on the platform were chosen (data called Rating) and others with few evaluations in their history, aiming to select characters with different behavior when using Google



Maps. After this prior selection, users with the most unusual names were searched on Facebook, which would consequently be easier to find.

After the individuals were found, they were added to Facebook, and then introduction messages were sent, explaining that they were found because they left ratings on Google Maps and, finally, asking if they would accept to participate in a survey, which would be carried out by through video interview, with questions about your interactions with Google Maps.

Thus, the technique used to define the sample for the collection of interviews was that of saturation (Glaser and Strauss, 1967). The interviews were carried out until saturation occurred, which happened to eight interviewees. Of the eight, six interviews were selected for analysis and transcription. Two were discarded for not bringing significant contributions.

The six profiles chosen for the analysis (Table 1) bring a plurality of reports, as they have different characteristics of age, region where they live, results in different numbers on the platform and different ways of using Google Maps. It is noteworthy that respondents agreed to use their anonymized data for research purposes.

**Table 1** Profiles of interview participants.

Pseudonym	Sex	Place	Age	Profession	Level on Google Maps <sup>3</sup>
Interviewee 1	Male	São Paulo resident (SP)	33 years old	Transport professional	Level 8
Interviewee 2	Female	Mogi Mirim resident (SP)	60 years old	School principal and Youtuber	Level 8
Interviewee 3	Female	São Paulo resident (SP)	32 years old	E-commerce professional	Level 4
Interviewee 4	Male	São Paulo resident (SP)	29 years old	Cultural producer	Level 5
Interviewee 5	Male	São Paulo resident (SP)	45 years old	Historian and technician in museology	Level 6
Interviewee 6	Male	Rio de Janeiro resident (RJ)	29 years old	Chemical engineer	Level 7

Source: Authors' elaboration.

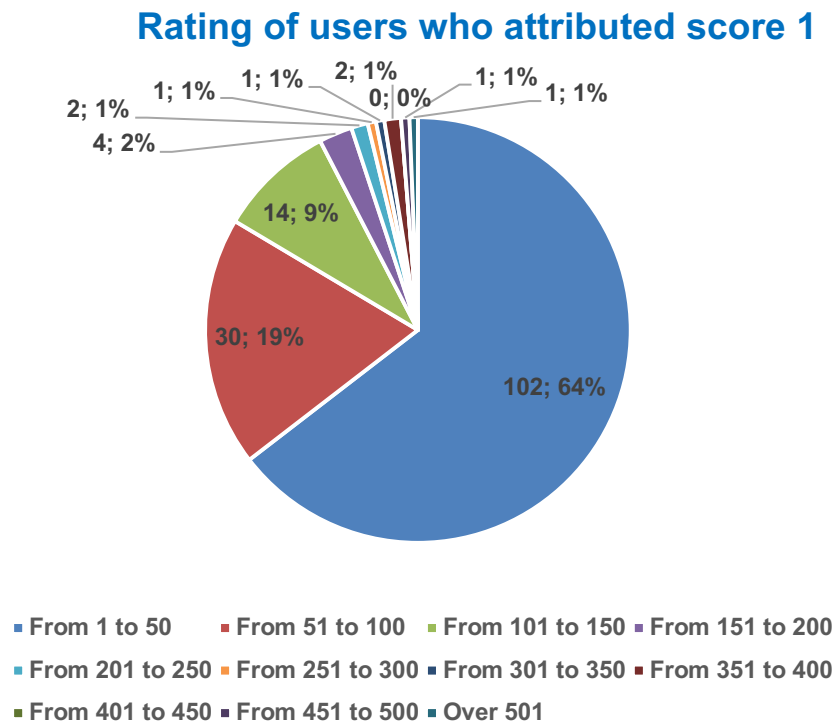
## Opinion formation

<sup>3</sup> Refers to Local Guides points, meaning users who receive points for contributing to the Google Maps location rating. Its scale is from 1 to 10. More information at: <https://support.google.com/local-guides/answer/6225851?hl=pt-BR>.

One of the objectives of this study was to investigate how Google Maps influenced the opinion formation of its users. Thus, the analyzed data sought exploratory evidence on this behavior.

Based on quantitative data, an analysis was made on the rating of the users who evaluated the Franklin Roosevelt Square and the rating they attributed (star), to identify possible relationships. Thus, the scores were divided, from 1 to 5, so that it was possible to analyze the users' rating range, considering an interval of 50 to 50. This analysis can be seen in Graphs 1, 2, 3, 4 and 5.

**Graph 1** Rating range of users who attributed score 1.

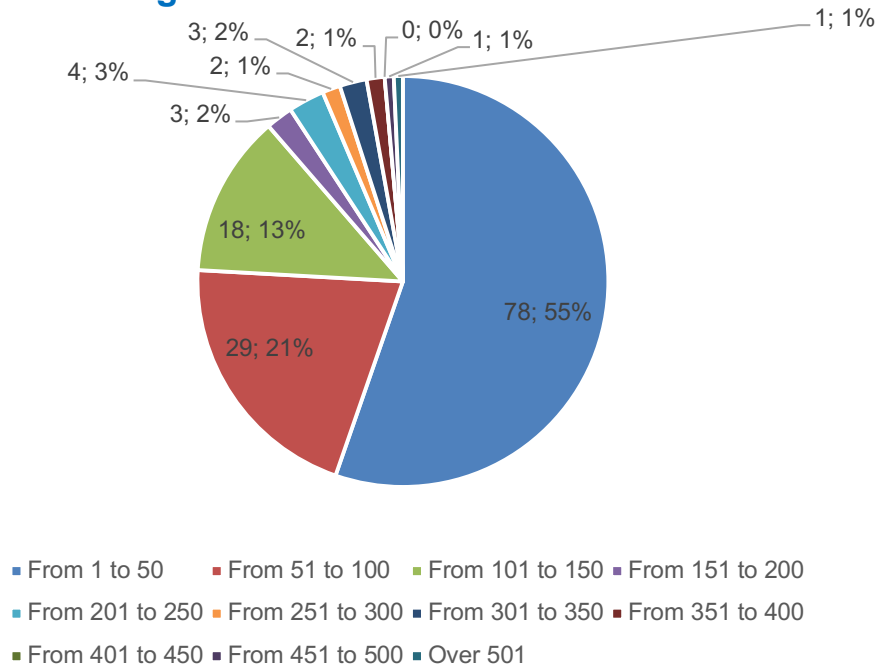


Source: Authors' elaboration.

According to this analysis, it was possible to observe that, of users who gave a score 1, 64% have a rating from 1 to 50, that is, they have already made 1 to 50 ratings on their profile, followed by 19% with a rating from 51 to 100, 9% with ratings from 101 to 150, and from there, with ratings ranging from 0% to 2% up to ratings above 500.

**Graph 2** Rating range of users who attributed score 2.

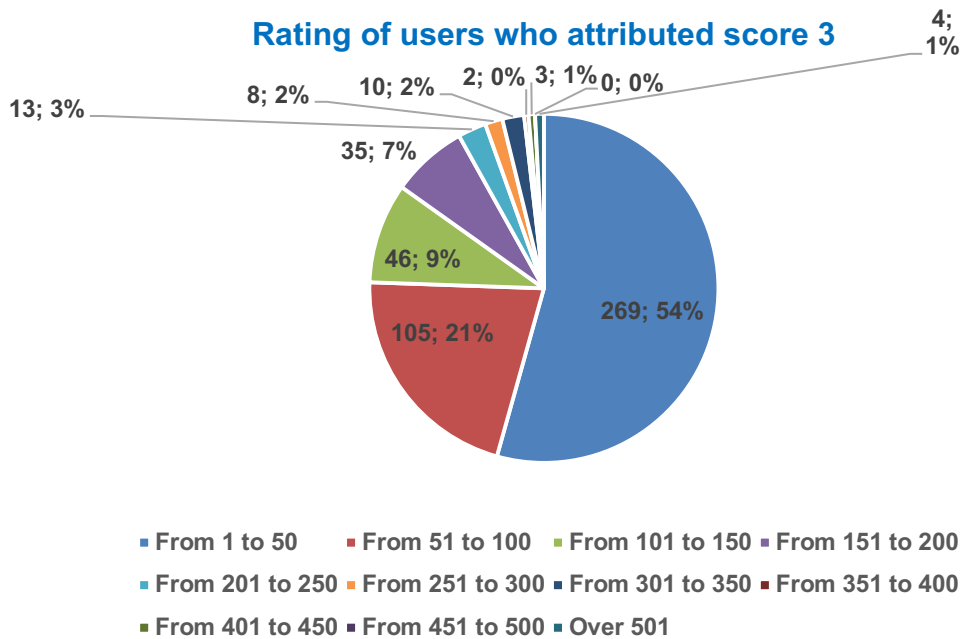
### Rating of users who attributed score 2



Source: Authors' elaboration.

On the other hand, users who gave Franklin Roosevelt Square a score 2, the majority, 55%, had a rating from 1 to 50, 21% obtained a rating from 51 to 100, 13% from 101 to 150, and from that, they also presented variations from ratings ranging from 0% to 3% to ratings above 500.

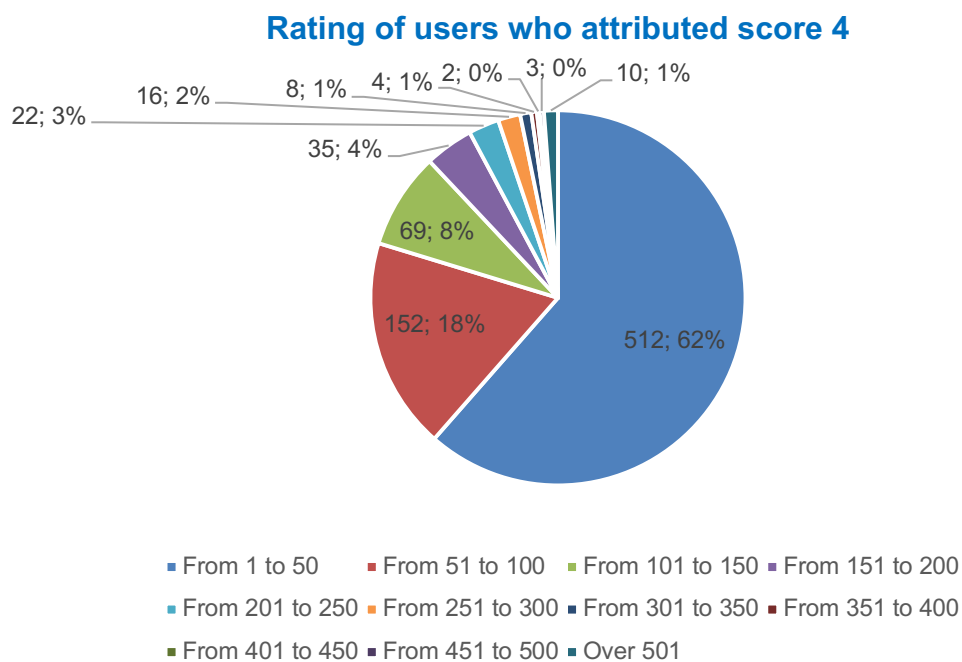
Graph 3 Rating range of users who attributed score 3.



Source: Authors' elaboration.

Users who gave a score 3 also show similar behavior: 54% are also rated from 1 to 50, 21% are rated from 51 to 100, 9% from 101 to 150, 7% from 151 to 200, and from that, too it has rating variations ranging from 0% to 3% up to ratings above 500.

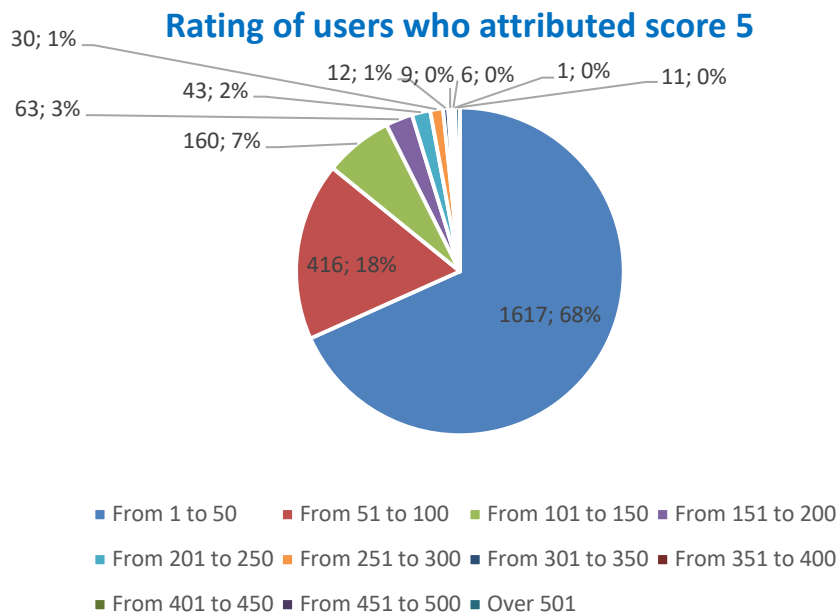
**Graph 4** Rating range of users who attributed score 4.



Source: Authors' elaboration.

Analyzing users who gave a score 4, the percentage of users who have a rating from 1 to 50 increases, approaching users who have a rating of 1: 62% have a rating from 1 to 51, 18% have a rating of 51 to 100, 8% from 101 to 150, and from there, it also presents rating variations ranging from 0% to 4% up to ratings above 500.

**Graph 5** Rating range of users who attributed score 5.



Source: Authors' elaboration.

Finally, analyzing users who attributed a score 5, there is a significant increase in users who have a rating from 1 to 50, representing 68% of them. Next, 18% rated from 51 to 100, 7% rated from 101 to 150, and from there, with ratings ranging from 0% to 3% up to ratings above 500.

Therefore, it is possible to notice a 4% variation between users rated 1 to 50 who gave a score of 1 and a score of 5, which may indicate a slight inclination towards the investigation that users with fewer reviews in their history tend to rate with a higher score for the location, fostering the suspicion that users who assess negatively are more critical and therefore assess more places.

This analysis of the critical view of users who use the platform was deepened in semi-structured interviews, which proved that some believe that Google Maps has expanded their perception and opinion about the places they visit after they started using the platform:

I think so. Because Google Maps asks a lot of questions, it usually asks "does this place have parking?", "can you do happy hour?", "this place, can you go with a child?", "is it pet friendly?", "is this, is it that?", so after you've answered several, when you enter somewhere you already map everything in your mind. It asks this bunch of questions, so you're already conditioned to... when you look at the place, whether it already meets those criteria or not, what criteria it fits. So yes. I think I became a little more discerning. (Interviewee 3, verbal information)

Too much, too much. I've already done this self-assessment [...] it's such a thing... that I already go [to places] with that look, you know? [...] I became much more critical. A lot more. (Interviewee 2, verbal information)

From this point of view, it is important to emphasize that the use of the platform and its questioning mechanism can be directly related to the construction of a new look at places users frequent. This seems to broaden their worldview, as they become more questioning in relation to several points that the platform questions and that, intrinsically, become questions that they reflect on their lived experiences. Google Maps, therefore, can become an interference factor in the construction of the opinion of its users, making them perceive spaces with a new look: more critical, more questioning, and more curious.

### **Motivations, ethical responsibility, and user privacy**

The Google Maps ratings platform tends to encourage users to use it through gamification techniques, which reward users with points for each activity they perform, making them evolve in level and earn "stamps" - techniques that aim to encourage the engagement of users to remain active on the platform. However, in addition to this factor, users had their own motivations to write reviews on the platform, bringing personal aspects. In this sense, we can observe other motivations that occur in this behavior, influenced by the social change that the Internet results. Hjarvard helps us to understand this context by pointing out that "contemporary society is permeated by the media in such a way that it can no longer be considered as something separate from cultural and social institutions" (2012, p. 53).

When asked about their motivations for using the platform, it was first common among interviewed users to point out their willingness to contribute with other users and to help, in a collaborative way, other people. We can see in the speech of Interviewee 1:

[I do evaluations] with the intention of helping, not to want to defame a place, but for someone who is looking for some information about the place to go without being in the dark, to know how it is. Because I take pictures of various places, so I want to make the person experience what it is like before they go, so that the person has a parameter. (Interviewee 1, verbal information)

Interviewee 5 also demonstrated, in his way of analyzing the motivations for carrying out evaluations, arguments in accordance with the same reasoning presented:

I think the possibility of being able to talk about the place gives us a sense of participation, of an importance so that other people can have parameters. I thought my reviews weren't read, until several questions started popping up on my cell phone and Google started sending me monthly saying "well, you had 15 thousand views", and I said "but what do you mean, I had 15 thousand views, right? (Interviewee 5, verbal information)

Castells (2015), when using the concept of conscious communication, reflects that it is extremely important to our experience and is what makes us human. Thus, thinking about the responsibility that one must communicate is essential to reflect the importance of this act in society that is increasingly connected and mediatized, after all, the opinion exposed in the network goes beyond spheres and has an unimaginable reach. Considering this, Castells states that "conscious and meaningful communication is what makes us human. Therefore, any major transformation in technology and in the organization of communication is of great relevance for social change" (CASTELLS, 2015, p. 20).

In line, the highlighted speech of Interviewee 1 and Interviewee 5 exemplify the importance of conscious communication by demonstrating their motivations in helping different users, demonstrating that they consider, while they are evaluating, the willingness to help other people who access the platform too. Along with this note, we can characterize that the main motivations of users come from gamification and the desire to contribute with other people who use Google Maps to consult ratings, which portrays the concern with conscious communication.

From the point of view of awareness and ethical responsibility in creating evaluations, Interviewee 3 explained how she feels when evaluating a place, considering the responsibility she has with this action:

I feel in two very different ways: I feel like I'm judging something, so I'm thinking "how is the owner of this place going to feel when he reads this"? Importantly, word choice is important because it's all about people. Even though you're in a digital environment, despite not having that much, face to face with each other, you're dealing with people, it's very difficult for us to forget that. I think a lot about how the owner of the place, or the person who runs that place, is going to feel, whether they think it was cool or not, but I think a lot more about who is researching to go there. I think a lot more, about 80% about this guy [who is researching], and about 20% about the administrator. (Interviewee 3, verbal information)

Contributing to this view, some interesting aspects appear when asked about the types of places that user's rate. Interviewee 2 pointed out the responsibility of evaluating places, bringing information about accessibility for people with disabilities:

[I evaluate] everyone, everyone. [...] because, if you think about it, today a person with a disability, right? He has... he is a citizen, and he has the right to a social life, so that's all, right? It's everywhere, from a store I go to, an ice cream shop I go to, a supermarket. (Interviewee 2, verbal information)

In sequence, after mentioning that he understands the act of evaluating as a great responsibility, Interviewee 4 raised the question about how other users use the platform:

I think that most evaluations are not made with this type of parameter, you know? I can't take into account people I know... bro, there are people who sometimes, you see that the person is doing an assessment for doing, like "I've been in this place, blah blah blah, very cool I recommend it", you see which are superficial things, and then I think "bro, I'm trying to make something that adds a little more, so I can't get bitten for it". That's my thinking, being very honest. (Interviewee 4, verbal information)

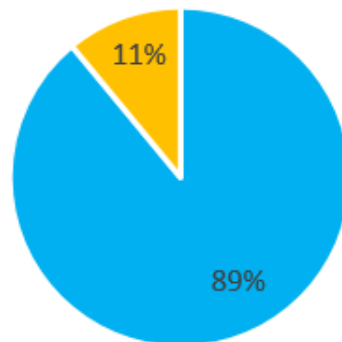
According to the exposed view, Interviewee 4 questioned how other users use the platform, subjectively inquiring about the individual values and beliefs of each Internet user. This happens because, in the mediatized society, individuals tend to find a way to express their opinions and perceptions of the world on the Internet, in a simplified way and often with high exposure. Interviewee 4's questioning, then, brought reflection on the Internet as a meeting space for different opinions and perceptions, and instigates thinking about how people expose them.

About the exhibition, an analysis was made of the data extracted from the evaluations of Franklin Roosevelt Square on Google Maps, with the objective of verifying the number of users who use false or unidentifiable names at the time of carrying out an evaluation. The result of this analysis can be seen in Graph 6.



Graph 6 Analysis of user identification.

## User's identification



- Users with identifiable names
- Users with unidentifiable names

Source: Authors' elaboration.

It is possible to observe that among the analyzed sample, 11% of users use nicknames or unidentifiable names. Thus, the 11% of users who wrote reviews, using names that make them difficult to find, may be protecting their privacy. On the other hand, this practice tends to generate written evaluations without ethical responsibility: false, without relevant information, disconnected or non-contributory. In any case, users, when writing a review, are exposed to the digital world: it is the moment when they expose their perception for countless other people to read, whether intentionally biased or not. Remember that unidentifiable names are only shielded from a first layer of privacy, as individuals can be tracked if the platform so wishes, such as in cases of court order.

On the one hand, having the full name explicit in the expression of an opinion about a certain place, for anyone who accesses that content, shows a certain legitimacy to what is written, as it denotes that it was a real user who wrote it and who can be identified and easily confronted. On the other hand, it makes the user's privacy vulnerable, since it is possible to find him on other social networks, as a trail left. Also, the use of a recognizable name can be done in an unauthorized manner by someone else. Therefore, caution in identifying the subject is necessary.

Thus, Kerckhove reflects on privacy in the digital world:

But of course, we are already totally imprisoned by the data we leave behind in our tracks. We still don't know the extent to which only law or technology can protect privacy. But is it really about privacy? (KERCKHOVE, 2016, p. 4).

Thus, the act of writing reviews online is not just about writing reviews on the Internet. It is clear that this behavior goes beyond issues of ethical responsibility, privacy and belonging to the social environment in which the user proposes to interact and evaluate: when a user writes an online evaluation, he expresses his worldview, his experience, the way he analyzed and processed the experience of getting to know a certain place, and narrates, in a personal way, what it meant for him.

It is important to emphasize the fact that these users understand their responsibility, and that they realize that this is not just a comment, but a way to contribute to other people who are looking for this information, which in a way is related to an ethical and citizen responsibility, through conscious communication. On the other hand, they also understand that the platform runs into ethical limits, which would need to be better analyzed in its mechanisms.

### **Final remarks**

This study, in applied research domain, revealed several perceptions that permeate different areas related to contemporary digital environments and their uses from the users' point of view. The Internet is responsible for transforming society's way of relating, bringing changes in our behavior, reflections that were evident in the results presented in this research and what was exposed by quantitative and qualitative analyses.

The developed tool made it possible to extract data that were able to build a netnographic methodological path capable of analyzing quantitative and qualitative data, through graphic analysis and interviews, on the different perceptions of users when using the Google Maps platform. The exploration of techniques from data science, such as web scraping, contributed significantly to the methodological innovation present in the study. This practice made it possible to extract massive data in an automated way. Based on these data, it was possible to carry out essential quantitative analyzes for the generation of graphs and analyzes that allowed relating hypotheses and raising new questions. It also made it possible to find characters to carry out the semi-structured interviews, which were decisive for the investigation.

From the Discovery Stars tool, comments that were left on Google Maps about Franklin Roosevelt square, a public place in São Paulo, were extracted and analyzed. Based on these data, a 4% difference was identified when comparing users rated 1 to 50 who

gave a score of 1, represented by 54%, and users rated 1 to 50 who gave a score of 5, represented by 68%. This data, although not conclusive, reinforces the inference that there may be a tendency for users with fewer reviews in their history to have a predisposition to rate the location with a higher score; also, that users who rate with lower scores tend to be more critical, and therefore rate more places.

When comparing these data with the interviewees' speech, it was noticed that the frequent use of Google Maps contributes to the perception that users build, based on what they read and what they write on the platform. Users pointed out that they became more critical after using the platform, which showed that it was able to broaden its users' opinion about the spaces they frequent. It is possible to reflect that, from the use of the Google Maps online assessment tool, users began to perceive several aspects about the places they visited, such as factors related to accessibility, location, price, service, and environment, developing a more critical view that accompanies them in the places they visit, in addition to the platform.

This research also revealed that the motivating factor of users when writing reviews is the feeling of contribution. Respondents reported that they felt useful when they gave their opinions on the platform, as they felt they were helping other users who would access and consult what they wrote there. The importance of this help for them became even more evident when they report that when they received notifications from Google Maps notifying how many people had already viewed their comments, they were even more motivated and felt they were really contributing to other people.

When analyzing the data contained in Graph 6, where it was possible to identify the relationship of users with the way they display their names and identifications, it was noticed that only 11% of them somehow hide their real identity. It is possible to infer, with some caution, that the majority, represented by 89%, was not concerned with exposing their full name. This behavior can be justified by users who may not be concerned with the traceability of their comments, or even if they have not been aware that they can be found from the opinion left.

Finally, this study contributed, from the data collected through the Discovery Stars web scraping tool, with new perceptions about the behavior of Internet users who carry out online assessments, pointing out characteristics of this group, prompting other questions about the impacts that this habit causes in their experiences, and opening space for future studies on the same theme, which has an outlet to be more in-depth and discussed.

## References

- ANGELUCI, A. C. B.; ROSA, B. B.; PASSARELLI, B. Podcasts sobre Covid-19: o projeto# MDDFcontraocorona. *Comunicação & Educação*, v. 25, n. 1, p. 186-199, 2020.
- CASTELLS, M. The impact of the internet on society: a global perspective, 2015. Disponível em: <https://http://www.bbvaopenmind.com/en/article/the-impact-of-the-internet-on-society-aglobal-perspective/?fullscreen=true>. Acesso em: fev. 2021
- SANTOS, M. C. Métodos digitais e a memória acessada por APIs: Desenvolvimento de ferramenta para extração de dados de portais jornalísticos a partir da WayBack Machine. *Revista Observatório*, v. 1, n. 2, p. 23-41, 8 dez. 2015.
- SANTOS, M. C. Pesquisa Aplicada em Comunicação - O estranhamento da interdisciplinaridade que nos assombra. *Comunicação & Inovação*, v. 19, n. 41, 2018.
- FACELI, K., LORENA, A. C., GAMA, J., e CARVALHO, A. (2011). Inteligência artificial: Uma abordagem de aprendizado de máquina. *Livros Técnicos e Científicos*.
- GLASER, B.G.; STRAUSS, A.L. The discovery of grounded theory: strategies for qualitative research. New York: Aldine de Gruyter; 1967.
- HJARVARD, S. Mídiação: teorizando a mídia como agente de mudança social e cultural. *Matrizes: São Paulo*, ano 5, n. 2, p. 53-91, jan./jun. 2012.
- KERCKHOVE, D. Ética de transparência na era do Big Data. *Comunicação, Cultura e Mídias Sociais*. São Paulo, 2016.
- KOZINETS, R. V. *Netnografia: Realizando pesquisa etnográfica online*. Porto Alegre: Penso, 2014.
- PASSARELLI, B. *Brasilina-Interfaces Digitais na Educação: @ lucin [ações] consentidas*. São Paulo: Escola do Futuro da USP, 2007.
- PASSARELLI, B.; ANGELUCI, A. C. B. Conectividade contínua e acesso móvel à informação digital: jovens brasileiros em perspectiva. *Informação & Sociedade*, v. 28, n. 2, 2018.
- PASSARELLI, B.; JUNQUEIRA, A. H.; ANGELUCI, A. C. B. Os nativos digitais no Brasil e seus comportamentos diante das telas. *Matrizes*, v. 8, n. 1, p. 159-178, 2014.
- TRIVINOS, A. N. S. *Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação*. São Paulo: Atlas, 1987.

**RESUMO:**

Com a difusão do acesso e uso da informação pela web e redes sociais, a recuperação de informação em grandes volumes de dados tem se tornado inviável por métodos manuais. Neste estudo aplicado, relatou-se a contribuição do desenvolvimento e uso de um protótipo de ferramenta de raspagem automática de dados de avaliações online feitas no Google Maps – a *Discovery Stars*. Os dados recuperados permitiram investigar como

**RESUMEN:**

Con la difusión del acceso y el uso de la información a través de la web y las redes sociales, la recuperación de información en grandes volúmenes de datos se ha vuelto inviable por métodos manuales. En este estudio aplicado, se informó la contribución del desarrollo y uso de una herramienta prototipo para el raspado automático de datos de evaluaciones en línea realizadas en

essas avaliações podem ter potencial para influenciar comportamentos dos usuários da plataforma. Dentre os resultados, observou-se que a leitura e postagem de avaliações impactam na formação de opinião e nas motivações dos usuários do Google Maps.

**PALAVRAS-CHAVES:** Avaliações Online; Raspagem de Dados; Web scraping; Ciência de Dados; Comunicação.

Google Maps - Discovery Stars. Los datos recuperados nos permitieron investigar cómo estas evaluaciones pueden tener el potencial de influir en el comportamiento de los usuarios de la plataforma. Entre los resultados, se observó que la lectura y publicación de reseñas impacta en la formación de opinión y motivaciones de los usuarios de Google Maps.

**PALABRAS-CLAVES:** Evaluaciones en línea; Raspado de datos; Web scraping; Ciencia de los datos; Comunicación.