Exploring the Impact of Google Discover on Users and Publishers: A Data-driven Study

Artur Strzelecki
University of Economics in Katowice
Katowice, Poland   *artur.strzelecki@ue.katowice.pl*

Mariia Rizun
University of Economics in Katowice
Katowice, Poland   *mariia.rizun@ue.katowice.pl*

Abstract

The paper is dedicated to the analysis of Google Discover recommendation algorithms. The study is conducted on two online stores operating in Poland, for which the data from Google Search Console are available for 16 months. Google Discover is explored from the perspective of web users and web content publishers based on the total number of impressions and clicks and the click-through rate of the two websites. The results allow us to understand that although the users' activity increases the website's performance in Google Discover, its algorithms consider many more factors (like the popularity of other websites with similar content, users' location, and frequency of content updates) and may remove the website from Discover despite a relatively high number of clicks. Additionally, the literature review conducted in the paper revealed a gap in scientific research dedicated to the phenomena of this content recommendation system.

Keywords: Google Discover, recommender system, click-through rate, website performance, search engine recommendation service.

1. Introduction

Google Discover[1] is a content recommendation service that offers personalized feeds of news, articles, videos, and other content to its users [16]. The service was launched in 2018, and it has since become a popular tool for users to discover and consume content on the web [2]. Google Discover is designed to learn about its users' interests and then deliver personalized content recommendations based on their preferences. Google Discover's algorithm considers a variety of factors, such as a user's search history, location, and previous interactions with the platform, to suggest content that is relevant and interesting. One of the key features of Google Discover is its ability to learn about a user's interests and preferences over time [14]. As users interact with the platform, the algorithm analyzes their behavior and refines its content recommendations based on their feedback [1]. This means that over time, Google Discover becomes more accurate in predicting the types of content a user will likely engage with.

To achieve this level of personalization, Google Discover relies on various data inputs. For example, the platform analyzes a user's search history to determine what topics they have searched for in the past, and it uses location data to suggest content that is relevant to their current context. Google Discover also considers users' interactions with content on the platform, such as which articles they have clicked on [9] and the time they have spent reading them [8]. The algorithm that powers Google Discover is based on a machine learning technique called "reinforcement learning", which allows it to learn and adapt gradually, optimizing content recommendations for each user [18]. Reinforcement learning involves training a machine learning model to maximize a reward signal, which in the case of Google Discover is user engagement with content.

There currently is a lack of research on Google Discover, as was highlighted in a scoping review of grey literature and the work of Lopezosa et al. [10]. To fill this gap, the authors set

---

[1] Google LLC, Mountain View, California, USA [www.google.com]

the objective of this paper – to explore the impact of Google Discover on users and publishers based on available data. In this research, we aim to contribute towards a better understating of Google Discover from (1) the user perspective – by determining the level of user interest when using the Google Discover service, and (2) from a publisher perspective – by exploring the potential traffic derived to a website through the use of Google Discover. To achieve the objective of the paper, the authors set the following research questions (RQs):

RQ1: How does users' behavior towards website content affect content publishers' performance in Google Discover?

RQ2: How does the Google Discover algorithm determine which content to recommend to users?

The paper is organized as follows. Firstly, we provide a brief introduction to the background and features of Google Discover (Sections 1 and 2). Later, we present our method for data collection (Section 3), followed by the presentation of our results (Section 4). Finally, we discuss the key findings and implications for users and publishers, conclude with limitations, and make recommendations for future research (Sections 5 and 6).

2.   Background

Google introduced its content discovery platform, Google Feed, in 2017, initially only accessible within the Google app [15]. However, in 2018, Google integrated it into the Google homepage on mobile devices, allowing users to access personalized content recommendations without opening the application. Google Feed also featured "Topic Cards" that displayed the latest news related to a particular topic. In 2018, Google rebranded Google Feed to Google Discover, introducing an updated design and new features, including "Topic Channels" that allowed users to follow specific topics of interest and receive personalized content recommendations. The platform has continued to evolve and improve, with new features and updates regularly introduced, such as the "Web Stories" feature added in 2020 [7].

Google Discover uses natural language processing (NLP) techniques to identify and categorize content based on topics, entities, and attributes, along with user engagement metrics, to improve the relevance of its recommendations. The platform recommends articles, videos, images, and other multimedia formats from a variety of sources, presented in a visually appealing and user-friendly format, with customization options for users. Compared to social media algorithms used in platforms like Facebook[2] and Twitter[3], Google Discover pulls content from a wider variety of sources based on the users' interests and search history. Discover's algorithmic system is also designed to be more transparent, with options for customizing the feed and providing feedback on the content presented. Moreover, Google Discover prioritizes categorizing and presenting content based on topics and entities rather than solely relying on engagement metrics, providing a more diverse range of content for users. Additionally, Google Discover provides significant traffic and engagement opportunities for publishers.

Google Discover is an important platform for both users and publishers, as it provides a personalized experience that enhances the content discovery process while increasing the reach of publishers' content [12]. From a user perspective, Google Discover allows individuals to access relevant and diverse content based on their interests, search history, and other online activity [11]. The platform's ability to recommend content that aligns with a user's preferences has the potential to reduce the amount of time and effort needed to find content of interest, making it a valuable tool for enhancing the user experience. For publishers, Google Discover offers an opportunity to reach a wider audience by showcasing their content to users who may not have otherwise discovered it. By leveraging Google's machine learning algorithms, publishers can increase their visibility and drive more website traffic, which in turn can increase engagement and revenue [3]. Additionally, Google Discover's ability to recommend content that aligns with a user's interests can lead to higher engagement rates and longer dwell times, which are key indicators of content quality and relevance [19].

Research has shown that Google Discover has become an important source of traffic for publishers. In fact, according to a recent study, "one news domain saw over 30% of its total

---

[2] Facebook, Meta Platforms Inc., Menlo Park, USA [www.meta.com]
[3] Twitter Inc., San Francisco, USA [twitter.com]

web traffic coming from Discover, while other publishers have reported that Google Discover drove more traffic than organic search in some months" [5]. This highlights the growing importance of the platform as a tool for content discovery and consumption. As with any platform that relies on algorithms to deliver content recommendations, there are concerns about the potential for bias and manipulation. Critics have raised questions about the transparency of Google's algorithm and its impact on the diversity of content that users are exposed to [4]. To address these concerns, Google has taken steps to improve the transparency of its algorithm and ensure that users are offered diversified content [13]. For instance, the company has launched a new feature that allows users to provide feedback on content recommendations, which can help improve the algorithm's relevance and accuracy.

Users can interact with the content presented in Google Discover in several ways. They can click on the card to read the full article or watch the video, share the content with others, or save it for later. Additionally, users can provide feedback on the content presented, which helps to refine the recommendations over time. Google Discover policies are designed to ensure that the content presented to users is of high quality and meets specific standards. These policies include guidelines for publishers, which outline the types of content that are eligible for inclusion in Google Discover. For example, the content must be original and of high quality and must not carry misleading or deceptive information. Google Discover policies include guidelines for user privacy and data protection as well. Google uses machine learning algorithms to personalize content recommendations for users, but this is done in a way that protects user privacy and keeps user data secure. Users can also control the types of content presented to them by adjusting their interests and activity settings.

Publishers can employ several strategies to increase the likelihood of their content appearing in Discover. These strategies include (1) creating high-quality content that provides value to readers, (2) optimizing content for mobile devices, (3) using clear and descriptive headlines that accurately reflect the content, (4) focusing on specific topics that are relevant to users, (5) building a strong online presence through social media, email newsletters, and search engine optimization, (6) implementing structured data markup to help Google understand the content, and (7) regularly monitoring the performance of their content in Discover and adjusting their strategy accordingly. By implementing these strategies, publishers can improve their chances of having their content featured in Discover and reaching a larger audience.

3. Methodology

Data from this study were collected from the Google Search Console[4] service. Google Search Console is a free tool provided by Google that allows website owners and authorized users to monitor and analyze their website's presence in Google search results. To access data in Google Search Console, certain conditions must be met. Firstly, a website must exist and be verified with Google Search Console. Verification can be done by adding a unique HTML tag to the website's header or by uploading an HTML file to the website's server. Secondly, the website must receive traffic from Google, either through organic search results or by being featured in Google's Discover feed. Once these conditions are met, website owners or authorized users can access a range of data on their website's performance in Google Search, including information on search queries, impressions, clicks, and click-through rates. By analyzing these data, website owners can gain insights into how their site appears in search results and make improvements to optimize their site's visibility and performance.

The authors of the study have access to two domains that are frequently featured in the Google Discover service. However, the data owner did not consent to the disclosure of the names of the domains. As a result, the domains have been anonymized for the purpose of this study – they are further referred to as Domain A and Domain B. These two Domains are online stores that sell various consumer electronic devices and provide users with helpful information about these items.

The following steps were taken to obtain data on traffic reported in the Discover section of Google Search Console. Firstly, the Google Search Console account was accessed, and the

---

[4] [https://search.google.com/search-console]

"Performance" tab was opened. The "Discover" tab was then selected, which displayed information regarding the number of clicks, impressions, and click-through rate (CTR) that the site had received through Google Discover. This data could be filtered by date range, country, pages, and appearance in Discover. Additionally, individual cards in the graph were clicked on to view more detailed information about specific content displayed in Google Discover. The data were exported in Excel format by clicking on the "Export" button in the top right corner of the screen.
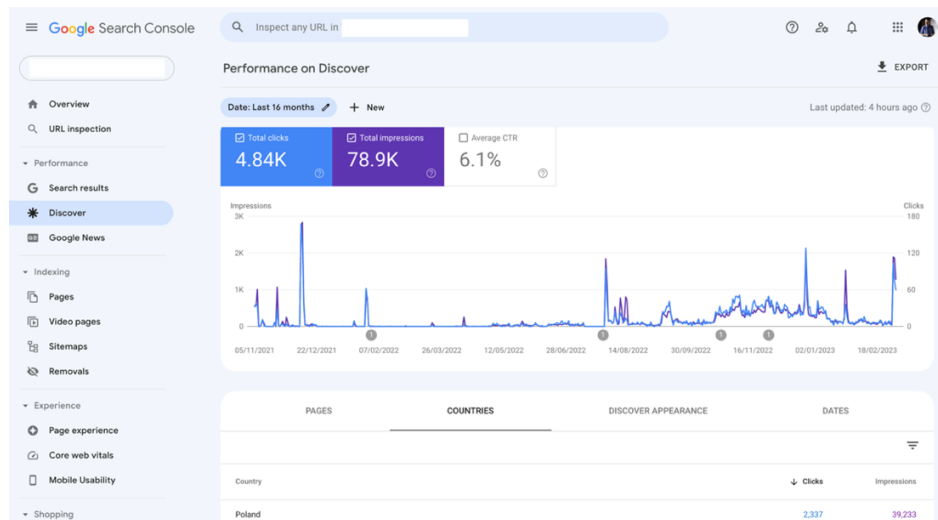

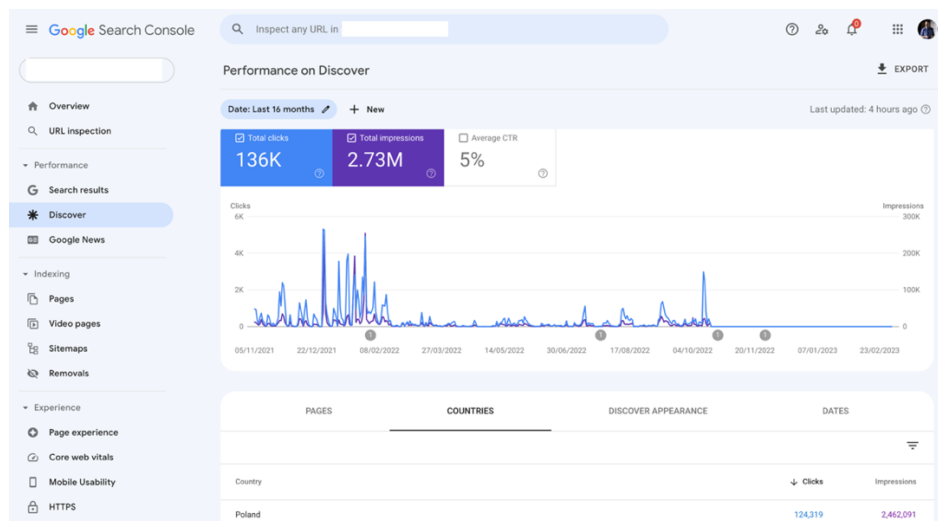**Fig. 1.** Performance on Google Discover – Domain A.

Fig. 2. Performance on Google Discover – Domain B.

The downloaded data in Excel format include a list of specific URL pages that have appeared in Google Discover, along with the total number of impressions and clicks, and click-through rate (CTR) for each URL page. It also includes the total sum of clicks, impressions, and CTR for each country and day.

4.   Results

The data collected from Google Search Console for Domain A include 32 specific URLs that appeared in Google Discover, with 2 337 clicks, 39 233 impressions, and an average CTR of 5.96%. The data present clicks made from one country. For Domain B, the collected data include 514 specific URLs that appeared in Discover, with 127 189 clicks, 2 561 639 impressions, and an average CTR of 4.97%. In the case of this Domain, the data show clicks made from 21 different countries. During data collection, we noticed some inconsistency

between the data available for download and the data displayed on charts in the Google Search Console service. The numbers on the chart were slightly higher than the data available for download.

The data collected from Domains A and B were approached from two perspectives: the publisher of the content (i.e., the website owner) and the reader of the content (i.e., the user/client of a website).

4.1. Content analysis – publisher perspective

The URLs collected from both domains were analyzed regarding the type of content available on these URLs. Both Domains are online e-commerce stores. The content analysis allowed us to distinguish six content types: news, products, guides, landings, categories, and promotions. News mainly consists of blog posts with news information. Products are pages with products (or services) available for purchase on the online store. Guides are expanded articles about a particular topic that are much longer than typical news articles. Landings are pages prepared for advertising purposes to sell a product or service. Categories are pages that list products belonging to a particular category. Promotions are pages that contain offers with promotional prices.

Domain A

The main topics on the list of news are sales and promotions, with a focus on discounts on LEGO sets and Black Friday/Cyber Monday deals. There is also one article about new LEGO sets for Spring 2022. The major topics within "product" URLs are exercise equipment, kitchen appliances, and, among others, LEGO sets. Some of the LEGO sets listed include the Batmobile Tumbler, Ferrari Daytona SP3, and the Star Wars AT-TE walker. There are also exercise equipment items such as the Kingsmith Walking Pad R1 Pro and the Yesoul S3 spinning bike. Kitchen appliances such as the DeLonghi F32110CZ fryer and the SodaStream Terra Black water carbonator are also included. Lastly, there are other LEGO sets, such as the Bugatti Bolide and Halloween Owl.

Finally, in the list of guides, the key topics include buying guides for various household items, such as an iron, an electric toothbrush, a men's shaver, a water filter pitcher, and a lawn mower. There are articles about specific products, such as automatic toothbrush heads and hair dryers. Additionally, there is a guide on how to properly maintain a Philips coffee maker, as well as an article about the benefits of water relaxation. This allows concluding that the website is aimed at providing helpful information to Polish consumers looking to purchase household items and appliances.

Domain B

The main topics in the list of news are (1) promotions and discounts on electronics and computer hardware (including laptops, PC components, monitors, and accessories) from various brands such as Green Cell, Logitech, Huawei, MSI, Samsung, Razer, and Gigabyte; (2) special events and sales, including Black Friday, Cyber Monday, and Black Week; (3) new product releases in the gaming and technology space, such as gaming chairs, headphones, and graphics cards from brands like NVIDIA and Palit; (4) gift guides and ideas for the holiday season, including Christmas and Saint Nicholas Day; (5) reviews and recommendations of technology products from experts. There are also other topics related to technology, such as Wi-Fi extenders, networking equipment, and computer cooling systems.

The most occurred device types in product URLs are smartphones (Xiaomi Poco F4 and Samsung Galaxy Z Flip4 5G are the most common, with six different models in total); smartwatches (Huawei Watch GT 3 Pro and Samsung Galaxy Watch 5 are the most common with nine different models in total); headphones (Razer Viper V2 Pro, SteelSeries Arctis Nova Pro Wireless, and Sony H9 Inzone are the most common with two different models in total each); laptops (Huawei MateBook D 16 and Apple MacBook Air M2 are the most common with 1 model each); desktop computers (Acer Predator Orion 3000 is the only desktop in the

list); tablets (Huawei MatePad Paper 4 is the only tablet in the list); other devices (camera stabilizer DJI RS 3 Pro Combo and computer case Be Quiet Pure Base 500 FX are the only other devices in the list).

The main topics on the list of guides appear to be technology-related buying guides and advice, covering issues such as choosing the USB drive, memory card, headphones, power bank, graphics card, processor, and soundbar, and utilizing the full potential of Nvidia's GeForce RTX 30 graphics cards. There is also a guide on devices and accessories for cyclists.

4.2. Behavior analysis – user perspective

To analyze the behavior of users of Domains A and B, we have taken into consideration the following indicators: (1) amount of content (presented as "content pieces" – the number of specific URLs of a domain that appeared in Google Discover); (2) clicks for each of the specific URLs; impressions for each of the specific URLs (the number of times the URL was shown); (3) click-through rate (the ratio of URL clicks to its impressions).

Domain A

Table 1 presents the number of clicks for 32 content pieces in Domain A, the number of impressions for each type of content, and their CTR. The content in Domain A is presented by three groups, and in the table, they are sorted by the CTR value.

Table 1. Number of clicks and impressions for each content type displayed in Google Discover for Domain A.

Content type	Content pieces	Clicks	Impressions	CTR
Guides	9	1 176	15 268	7.70%
News	9	519	10 291	5.04%
Products	14	642	13 674	4.70%

It can be seen that various guides presented in Domain A have drawn the most attention of its users. Even though the number of specific URLs with products is around 1.5 times higher than that of the guides (14 compared to 9), the guides' CTR is 3% higher (7.7% compared to 4.7%). As stated before, Domain A offers its users interesting guides on various topics connected with household items. This allows claiming that informative guides are rather valuable for the users, while they can find products for purchase on other websites (for instance, with a better offer, better prices, better quality, or all at once).

But this is a moment to remember that the analyzed data come from a content recommendation service. So, the analyzed clicks were users' interactions with the links to the content that Google Discover selected for recommendation at a certain moment. Later, new content from Domain A was recommended to users already based on their previous clicks, as well as on the number of clicks the other users made. That is, the clicks on interesting content most probably led to more clicks on that content.

Domain B

In Table 2, we see the number of clicks for 514 content pieces in Domain B, the number of impressions, and the CTR for each type of content. In this Domain, the content is presented by six groups, sorted by the CTR value in the table.

Table 2. Number of clicks and impressions for each content type displayed in Google Discover for Domain B.

Content type	Content pieces	Clicks	Impressions	CTR
Guides	9	2 166	30 233	7.16%
Categories	2	207	2 924	7.08%
News	238	51 519	917 528	5.61%
Products	261	72 682	1 591 508	4.57%
Promotions	1	13	317	4.10%
Landings	3	602	19 129	3.15%

For Domain B, we observe a situation similar to that of Domain A – the URLs with guides have the highest CTR among all six content types, although the number of content pieces is not very high (only 9). At the same time, the "Guides" type does not stand out by the number of clicks and impressions. The "News" and "Products" types have many more clicks (51 519 and 72 681 respectively) and impressions (917 528 and 1 591 508 respectively). Nevertheless, the ratio of quite a few clicks (2 166) and quite many impressions (30 233) put the "Guides" in the first place. It looks as if the overall situation for this Domain is the same – the guides it offers are of higher interest to its users than the products it offers. The CTR for products for Domain B (with 261 content pieces) is 4.70%, as well as for Domain A (with 14 content pieces).

To understand the users' behavior, it is necessary to compare results for Domains A and B. For that purpose, Figures 1 and 2 were shown. In Figure 1, Domains A and B are set side by side from the perspective of three content types (guides, news, and products), which are presented in Google Discover for both Domains. We see that in these two cases, the "Guides" content attracted much more attention from the users than the other two types, with a slight difference between the Domains: the CTR is 7.7% for A and 7.16% for B. Further, "News" users' interests prevailed in Domain B – CTR is 5.61% here, while for Domain A, it is 5.04%. Finally, the content of "Products" takes third place in this rating, with relatively similar values for both Domains – 4.7% for A and 4.56% for B.
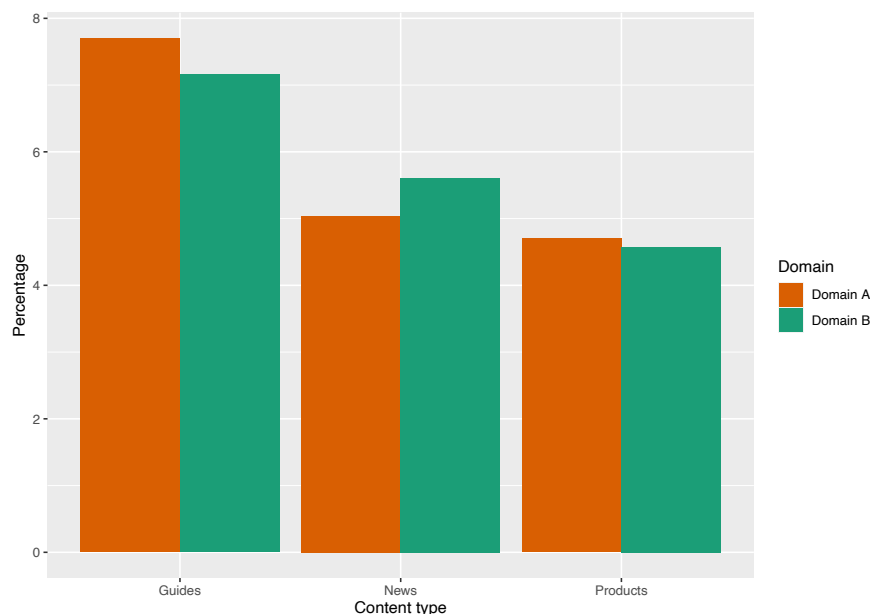

Fig. 1. Click-through rate values for the same content type displayed in Google Discover.

Such results allow us to build an overall picture of users' behavior when visiting the analyzed Domains A and B. Since the authors did not conduct any survey of users' opinions, we can only make suggestions about their motivation based on the information we have about these Domains. The online market presently contains quite many stores that sell computers, smartphones, and lots of household electronic devices. Internet users do not face the problem of not finding an offer that would interest them and fit their taste or wallet. On the other hand, in addition to purchasing, the users sometimes require reliable information, detailed instructions, trustful comparisons, etc., about the devices they possess or are going to buy. That is why well-written guides and actual news from the world of technology may be of higher interest to the users since not all online stores provide such information.

Figure 2 presents the distribution of CTR values for Domains A and B in time – from November 2021 to March 2023 (the time when this study is conducted). In this figure, Domain A is presented by a somewhat uneven fluctuation of the CTR level within the considered period. In total, there were 171 days (of 485 days analyzed) when the CTR was 0%. These zero levels mostly happened at the end of 2021 (November and December), then remained almost unchanged from January to April 2022. After that, there were a few days with 0% CTR in different months, with a row of 16 days in August 2022. However, it can be claimed that such

less "successful" days might happen for any online store, so these 16 days are much less interesting for our analysis than the period from November 2021 to April 2022. In all, the CTR values between 20% and 50% were observed 13 times during the whole period of analysis; there were 36 days when the CTR was between 10% and 16.98%; 249 days showed the CTR between 2.25% and 9.76%; and finally, the CTR between 0.78% and 1.79% happened 14 times. The zero CTR, mentioned earlier, was in total observed 171 times. As we see, more than 50% of the time, Domain A had a click-through rate of up to 10%, and for the rest of the time, there were different fluctuations.

It is necessary to pay extra attention to ten separate days with peak values we observe for Domain A: two days 100% – in February and April 2022; three days 50% – in November 2021 and May and June 2022; two days 33.33% – in January and May 2022; and, finally, two days with 25% – in April and June 2022. Such separate jumps of value (especially 100%) should not be considered when analyzing the trend of users' behavior and should be referred to as outliers in the overall statistics. The reason is that in most cases, such high CTR values are created not by a considerable number of clicks and impressions but, on the contrary – by very low ones. And this was the case for Domain A: 100% CTR is the ratio of one click and one impression; 50% CTR is the ratio of one click to two impressions; one click to three impressions resulted in 33.3% CTR; and one click to four impressions gave the 25% CTR. Thus, to sum up, these ten days are examples of very low users' interest in Domain A, also caused by a low presentation from the recommender system, which is Google Discover.

Overall, the distribution of the CTR for Domain A seems to be relatively normal. Every online store may face moments of increase and decrease in users' interest, caused by such factors as, for instance: lack of new offers, lack of discounts and promotions, financial crisis in the country or in the world, or any other factors from the global environment that may influence the activity of an online store. Moreover, the algorithms of Google Discover on how the system recommends the content have a significant influence on that. The same URL in a domain may appear in the recommendations today and disappear tomorrow because the system considers it irrelevant for users. In the end, as stated before, the situation with the CTR level eventually normalized, and the online store in Domain A has been functioning quite effectively in Google Discover up to the time these data were extracted from the system.
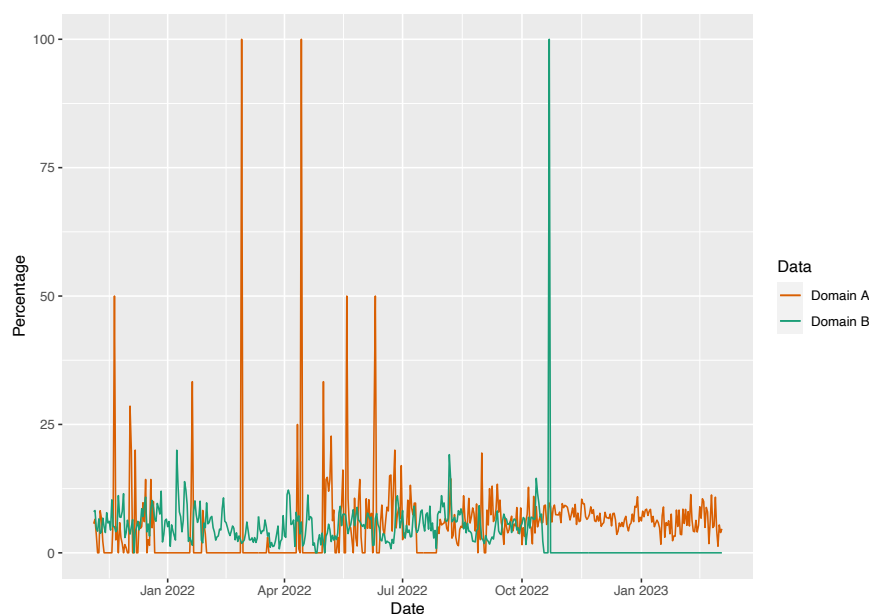

Fig. 2. Click-through rate values as a time series.

For Domain B, we observe a generally even distribution within these 17 months: there was only one significant increase in CTR in October 2022, when it reached 100%. And that is the same situation described for Domain A – when the ratio of 1/1 clicks to impressions resulted in such a CTR. Therefore, this value should not be considered when presenting the overall picture for Domain B. For the rest of the period, the CTR fluctuated between 0.77% and 19,98%. There

were two days when the CTR reached almost 20% (19.12% and 19.8%); 22 days showed the CTR from 10.06% to 14.51%). For the most time (295 days, around 60% of the time analyzed in the study), the CTR was between 2.01% and 9.73%. Finally, for 24 days, the CTR value was between 0.77% and 1.96%. For the rest of the time (141 days), we observe zero CTR. It is necessary to highlight, once again, not to cause confusion, the fact that the CTR value for Domain B was not just growing gradually day by day but was different every day of the analyzed period. That is also clearly seen in Figure 2.

The authors did not analyze the opinions of internet users to conduct this study. Nor can we refer to any specific information about what happened in Domain B after October 2022, when the CTR value fell to 0% and remained at this level until the date when the data were extracted from Google Discover. It might be that Domain B faced a particular event that resulted in this online store being removed from Google Discover algorithms. It also might be that nothing in particular changed in the Domain, but still, it stopped appearing in Google recommendations. It is necessary to highlight here that the lack of data in Google Discover does not mean that the Domain is not attractive to internet users. Yet, the recommendations of Google no longer help this online store attract its potential customers.

As it was mentioned, for Domain A, the data on clicks in Google Discover are presented only for one country – Poland. And the average click-through rate for this country for all 485 days was 5.96%. Yet, for Domain B, we have data about clicks from 20 countries. So, Figure 3 provides us with the geographical distribution of the CTR value for Domain B from November 2021 to March 2023 (when the data were downloaded from Google Discover). Since Domain B is a Polish online store, it is unsurprising that Poland takes first place in this rating of countries, with the CTR being 5.05% for the 485 days. On the other side of this rating, we have Belgium, with the CTR of 0.49%. Of these 20 countries, eight (40%) have CTR below 1%; five (25%) have CTR from 1% to 2%; another five countries (25%) have CTR above 2% and up to 4%; finally, only two countries (10%) have the click-through rate above 4%. The median CTR value for all these countries is 1.27%.
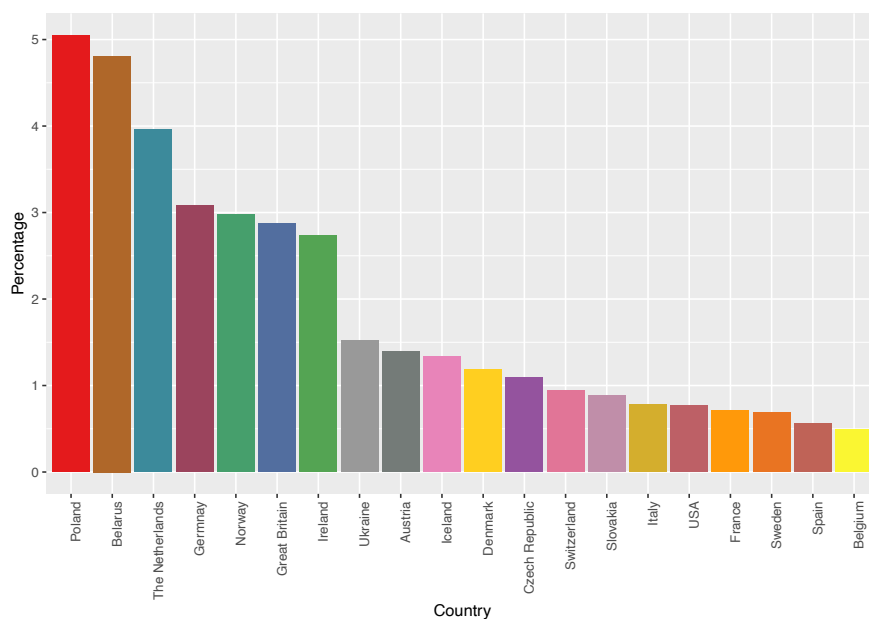

Fig. 3. Click-through rate values for Domain B by geographical source.

The authors assume that language settings are the reason why other countries (apart from Poland) appear in the data. This online store does not provide translation to any other language than Polish. Thus, it is most probable that Google Discover recommended Domain B to those users abroad, who have their Google accounts set in Polish, and who conducted their previous internet searches in Polish.

5. Discussion

This paper is dedicated to the exploration of Google Discover, a recommender system released by Google in 2018 with the task of offering personalized content to users. The authors' objective was to analyze the impact of Google Discover on two groups of its stakeholders – content publishers and internet users searching for the content. The authors also set two research questions that were answered with the results of the study. As the first step of the research, the authors reviewed the literature dedicated to Google Discover. It was revealed that until now (as of March 2023), not many works have been written about this system's activity. The sources where Google Discover is mentioned are mostly web pages from Google blog (like [2, 7, 15]) or other blogs (like [5]). There is only one research paper dedicated not to recommender systems in general but to Google Discover algorithms (the work of Lopezosa et al. [10]). This paper is the only one that appears when the key phrase "Google Discover" is applied in such research databases as Web of Science[5], Scopus[6], and Google Scholar[7]. This led the authors to the conclusion that there is a noticeable research gap that requires filling. And thus, to explore the mechanism of the recommender system, the authors collected data from the Google Search Console. As the subject for analysis, two domains were selected – two online stores operating in Poland which sell electronic devices. In this study, they are referred to as Domain A and Domain B. The authors have access to data from these two stores from November 2021 to March 2023 (when the data were retrieved). The data were downloaded to an Excel spreadsheet and included the total number of clicks, total number of impressions, and click-through rate values for each URL page within the two Domains.

As for the perspective of the publishers of content (i.e., the owners of Domains A and B), which is processed by Google Discover, we have come up with a few observations. Firstly, the content can be divided into six groups by type: news, products, guides, landings, categories, and promotions. For Domain A, we have information only about the first three types (9 URLs with news, 9 with guides, and 14 URLs with products); for domain B we have data for all six groups, but three of them are prevailing by the number of ULRs (261 with products, 238 with news, 9 with guides). Secondly, we can state that the most frequently posted news topics for both Domains are sales, promotions, and discounts for various electronic devices. Thirdly, both Domains frequently publish detailed guides on how to choose and use electronic devices. This content group is the one that attracts attention when considering users' behavior.

The behavior of the users of Domain A and B was analyzed by such indicators as the number of specific URLs that appeared in Google Discover, the number of clicks for each URL, the number of times (impressions) when each URL was shown to a user by the recommender system, and the click-through rate for each URL. The comparative analysis was performed for three content types (news, products, and guides) since information about them is available for both Domains. For Domains A and B, the CTR (ratio of clicks to impressions) for the URLs with guides was the highest (7.7% and 7.16%, respectively), and then, for both Domains as well, for the news (5.04% and 5.61%) it was higher than for the products (4.7% and 4.57%). But was this high click-through rate value caused by users' behavior or by the "behavior" of Google Discover algorithms? Or was it also the content publisher who earned clicks for their content? The answer to that question would be "all at once".

As we know, Google Discover works with the data about users' likes and dislikes, preferences, links clicked and saved, location and language settings, etc. The system not only shows the users personalized content based on their interests, but it also allows them to follow specific topics to see more similar content in their feed or to save articles, videos, and other content for later viewing. Google Discover is constantly evolving and can show its recommendations not only in search but, for instance, in Google Maps[8] [6]. If a user saves a URL with a guide on how to choose a better smartphone, in the future, Discover may show related articles about smartphones and guides on similar topics (devices, hardware, etc.). For a publisher, getting the content in Google Discover is not different from a regular SEO procedure performed to promote a website in search engines. The content should be rather outstanding,

written with an understanding of the audience, and be placed on a website that is attractive for a user to visit. Moreover, it is reasonable and valuable for publishers to check their Discover performance in Google Search Console [17]. Finally, one more thing about Google Discover is that apart from considering users' preferences, it will aim to introduce some new topics and ideas for a user to see, prioritizing high-quality content from authoritative sources.

Therefore, when, for example, a guide from Domain A is recommended to a user by Google Discover, and this URL is clicked by this user, the algorithm will remember to start offering this user more URLs with guides (from Domain B and other domains). This will attract the user's attention to more content of Domain B, but also of other sources. Depending on what interests the user later, Domain B will be getting more, less, or the same number of clicks. When to this picture the activity of all other users is added, URLs of Domain B gain or lose their CTR in the system. At a particular moment (as it did happen to Domain B), Google Discover might remove a website from its algorithms, considering it no longer relevant enough for the users.

5.1. Limitations of the study

The authors see three major limitations of the conducted study. Firstly, there is quite a little research conducted about Google Discover and its algorithms. Most sources found are internet blogs, which provide a commercial view of the recommender system, not the scientific one. More research papers dedicated to Google Discover would make possible a comparison of different methodological approaches to its "behavior" analysis, as well as a comparison of study results with drawing extended conclusions about how Google Discover works for websites and its users.

Secondly, the authors were limited by the subjects of research since they have access to Google Search Console only for two domains with Discover data. Moreover, these domains represent only one country – Poland. With a chance to access data about more websites in the country dedicated to the same topic, we would be able to get a broader picture of the types of content in Poland that is taken to Google Discover, about the types of content that users in Poland prefer, and also about whether Polish content is of interest for users outside the country. Furthermore, with access to data from a few countries, a higher level of comparison could be possible, and that would lead the authors to provide a better understanding of Google Discover algorithms and the factors they depend on.

Finally, the third limitation comes from the Google Discover itself. The code behind its algorithms is changed frequently. There is no option of gaining access to it or finding out which version it is. Neither can we know what settings each of the code editions bring to this recommendation system. Thus, we cannot state that we know for sure what technical configurations drive the "decisions" of Google Discover. And our analysis is performed in a black-box way, where we see the result of the recommendation system activity, not seeing how its processes run. What we can do, and what we perform in this paper – is to analyze the "behavior" of Google Discover on the examples of content that was already processed by Discover and was recommended by it to the users, or not.

5.2. Contribution of the study

The contribution of the study is twofold. First, the conducted literature review presents the most relevant research and discussions about Google Discover, clearly showing what was done in that area and, what is more important, what was not. The research gap was defined, and there is a large area of possible studies on this topic. Second, the study of Google Search Console data of the two Polish domains, despite the abovementioned limitations, provides some understanding of how Google Discover algorithms support web content publishers in promoting their work.

5.3. Avenues for future research

Directions for further research of Google Discover recommendation algorithms arise, first of all, from the limitations of this study discussed above. They lead to a suggestion that the author

would need to get access to the Google Search Console data for more websites, at least within one country (Poland). Then it would be possible to obtain an understanding of the behavior and interests of Polish web users and of how they influence the "behavior" of Google Discover. One more extension that would enable better study results would be to acquire data from a larger time scale. It might allow us to catch the moment when a website was added to the algorithm of Discover and track the changes in impressions for this website up to the moment when (and if) this site is removed from the Google Discover algorithm. And then, we would be able to relate these changes to the changes in contents provided by this particular website.

References

1. Alyari, F., Jafari Navimipour, N.: Recommender systems. Kybernetes. 47 (5), 985–1017 (2018)
2. Corby, K.: Discover new information and inspiration with Search, no query required, https://blog.google/products/search/introducing-google-discover/, Accessed: March 27, 2023, (2018)
3. Giomelakis, D., Veglis, A.: Investigating Search Engine Optimization Factors in Media Websites. Digital Journalism. 4 (3), 379–400 (2016)
4. Haim, M., Graefe, A., Brosius, H.-B.: Burst of the Filter Bubble? Digital Journalism. 6 (3), 330–343 (2018)
5. Hamilton, A.: How to Succeed in Google Discover, https://www.searchenginejournal.com/google-discover/361142/, Accessed: March 27, 2023, (2020)
6. iTrust: What is Google Discover? An analysis of what is yet to be discovered, https://itrust-digital.com/google-discover/, Accessed: March 27, 2023, (2022)
7. Jasti, V.: A new home for Web Stories on Discover, https://blog.google/web-creators/web-stories-discover-search/, Accessed: March 27, 2023, (2020)
8. Kellar, M., Watters, C., Duffy, J., Shepherd, M.: Effect of task on time spent reading as an implicit measure of interest. Proceedings of the American Society for Information Science and Technology. 41 (1), 168–175 (2004)
9. Liu, J., Dolan, P., Pedersen, E.R.: Personalized News Recommendation Based on Click Behavior. In: Proceedings of the 15th International Conference on Intelligent User Interfaces. pp. 31–40. Association for Computing Machinery, New York, NY, USA (2010)
10. Lopezosa, C., Guallar, J., Santos-Hermosa, G.: Google Discover: entre la recuperación de información y la curación algorítmica. Scire: representación y organización del conocimiento. 28 (2), 13–22 (2022)
11. Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: A survey. Decis Support Syst. 74 12–32 (2015)
12. McKelvey, F., Hunt, R.: Discoverability: Toward a Definition of Content Discovery Through Platforms. Soc Media Soc. 5 (1), 2056305118819188 (2019)
13. Rader, E., Cotter, K., Cho, J.: Explanations as Mechanisms for Supporting Algorithmic Transparency. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–13. Association for Computing Machinery, New York, NY, USA (2018)
14. Seo, Y.-W., Zhang, B.-T.: Learning User's Preferences by Analyzing Web-Browsing Behaviors. In: Proceedings of the Fourth International Conference on Autonomous Agents. pp. 381–387. Association for Computing Machinery, New York, NY, USA (2000)
15. Thakur, S.: Feed your need to know, https://blog.google/products/search/feed-your-need-know/, Accessed: March 27, 2023, (2017)
16. Thurman, N.: Making 'The Daily Me': Technology, economics and habit in the mainstream assimilation of personalized news. Journalism. 12 (4), 395–415 (2011)
17. Toonen, E.: Google Discover: What it is and how to get your content in, https://yoast.com/google-discover/, Accessed: March 27, 2023, (2022)
18. Wiering, M., Otterlo, M. eds: Reinforcement Learning: State-of-the-Art. Springer Berlin, Heidelberg (2012)
19. Zou, L., Xia, L., Ding, Z., Song, J., Liu, W., Yin, D.: Reinforcement Learning to Optimize Long-Term User Engagement in Recommender Systems. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2810–2818. Association for Computing Machinery, New York, NY, USA (2019)