Google Discover data-driven study of user activity on e-commerce platforms

Journal:	*Aslib Journal of Information Management*
Manuscript ID	AJIM-03-2024-0247.R2
Manuscript Type:	Research Paper
Keywords:	Google Discover, recommendation service, click through ratio, recommendation system, website efficiency, search engine

Google Discover data-driven study of user activity on e-commerce platforms

Abstract

Purpose

This research focuses on the analysis of the recommendation algorithms employed by Google Discover, utilizing data from two e-commerce platforms operating in Poland.

Design/methodology/approach

The study uses the information obtained from Google Search Console over a period of 17 months. The examination of Google Discover focuses on the number of displays, clicks, and click-through rate, from the viewpoints of content publishers and web users.

Findings

The results suggest that user engagement positively influences a website's efficiency in Google Discover, yet the algorithm also considers variables such as the popularity of similar content on other websites, user location, and content update frequency. Thus, a website may be excluded from Discover despite a substantial click count.

Originality/value

There is a lack of studies on how Google Discover is perceived by users, based on real data. We offer a quantitative perspective, which has not yet been done. This study offers a brief overview of the history and evolution of Google Discovery, an analysis of data used to show the perception of the service, and two unique perspectives on the recommendation service - from the point of view of the users and content publishers.

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

Introduction

Google Discover, integrated into Google Search, functions as a search recommendation service offering users individual streams containing articles, news, videos, and other various material types (Thurman, 2011). Introduced in 2018, it has gained popularity as a service for mobile users to explore and read websites (Corby, 2018). Operating as an algorithmic system, Google Discover considers various factors, including user data such as history of search, geographic whereabouts, and past activity on the platform, to recommend pages that are both related and engaging. Google Discover demonstrates the capacity to gradually discover a user's preferences and areas of interest (Seo & Zhang, 2000). The algorithm examines how users behave while they use the platform, improving content recommendations in response to their comments (Alyari & Jafari Navimipour, 2018). This iterative process enables Google Discover to enhance its accuracy in anticipating the kinds of content that a user is probably to interact with over time.

To maintain such a high degree of individualization, Google Discover depends on a range of data sources. For instance, it assesses users' search history to identify their past interests and utilizes location data to provide contextually relevant information. The platform takes into account user engagement with the webpages, such as the articles they have clicked (J. Liu et al., 2010) as well as the duration of them reading these articles (Kellar et al., 2005). The algorithm powering Google Discover employs "reinforcement learning", a machine learning technique that enables the recommendation system to learn and adjust progressively, aiming to maximize content suggestions for individual users (Wiering & van Otterlo, 2012). This entails instructing the machine learning model to optimize the reward signal, where user engagement with the platform's content serves as the primary measure.

Theoretical background

Foundations of recommendation systems

In scientific literature, recommendation systems (also called "recommender systems") related to Google Discover are characterized by their emphasis on personalization, where user data like browsing history and preferences are pivotal in tailoring content. These systems utilize sophisticated content analysis to categorize various media types, while employing advanced machine learning algorithms, including neural networks, content and collaborative filtering, to predict user preferences. A crucial aspect is the user feedback loop, which continually refines recommendations based on user interactions. In a traditional interaction paradigm, the user indicates their preference for an item (for instance, rating a movie), and the system uses this information to model the user's preferences and generate new recommendations (Ghori et al., 2022). Additionally, these systems strive to maintain a balance between diversity and novelty in content, ensuring that users are exposed to a wide range of topics. They are also context-aware, taking into account factors like location and time. Lastly, these systems navigate the complex landscape of privacy and ethical considerations, ensuring the responsible use of data and avoiding the creation of echo chambers.

The personalization of online content, such as that seen in Google Discover, relies heavily on algorithms. These algorithms analyze individual user data, like browsing habits and preferences, to customize content. The core of this approach is "collaborative filtering", a technique that recommendation systems employ extensively. It forecasts user preferences by aggregating data from various users (Su & Khoshgoftaar, 2009). The success of these systems in enhancing user engagement is well-documented, with Google Discover, for example, employing user data (like clicks and reading time) and device information to tailor content feeds, thereby improving user engagement (Covington et al., 2016). These personalization algorithms, integrating machine learning techniques, continuously evolve to improve content relevance based on both implicit and explicit user feedback (Ricci et al., 2022). Despite the popularity of collaborative filtering, this method has some drawbacks. For instance, there are many concerns about data privacy and the creation of “filter bubbles”, where users see only content that echoes their views (Pariser, 2011). It is also stated that this method is unable of including user and item information into the modeling process; it also faces problems with the cold start (which is a rather frequently mentioned problem of recommendation systems) and sparse data, since it relies a lot on historical data (past interactions and behaviors of users with items within the system) of users (Li et al., 2024).

In contrast to the approach of collaborative filtering, recommendation systems may be based on content filtering. Content-based approach gathers previous knowledge provided by users about a particular item, or the information provided about this item by content creators (Bhatia, 2024). In systems like Google Discover, content analysis is vital for understanding and categorizing various media forms, from text to videos. This process, driven by complex algorithms and machine learning, examines elements like keywords and sentiment (Campbell et al., 2015; B. Liu, 2012). Google’s algorithms, employing techniques such as computer vision and natural language processing (NLP), categorize content into themes matching user interests (LeCun et al., 2015). Importantly, these systems evolve with user interaction, refining their understanding of user preferences to adjust recommendations (Koren et al., 2009). However, the challenge lies in achieving accurate content categorization while maintaining user privacy (Zhu et al., 2017).

The third possible method of building recommender system models is the hybrid system. Such recommender systems combine two or more methods for better performance, terminating eliminating the flaws of each method. For instance, some strategies include joining collaborative filtering to make recommendations based on information equations between users and content-based filtering. In general, there are seven hybridization mechanisms applied to build recommendation systems: weighted, mixed, switching, feature combination, feature augmentation, cascade, and meta-level (Fahrudin & Wijaya, 2024).

In recommendation systems like Google Discover, machine learning algorithms are central. They predict future preferences based on past user behavior, with deep learning, a subset of machine learning, playing a significant role (LeCun et al., 2015; Ricci et al., 2022). Deep learning's ability to process vast unstructured data helps Google Discover tailor personal content feeds. However, concerns about transparency and potential biases in these algorithms remain (Burrell, 2016).

In Google Discover, the user feedback loop is essential, where the system refines content recommendations based on user interactions (Ricci et al., 2022). Metrics like click frequency and engagement duration inform the system's understanding of user preferences (Covington et al., 2016). This feedback loop continually adjusts recommendations, improving accuracy over time (Koren et al., 2009). Yet, it raises privacy and data security questions (Zhan et al., 2010).

To combat the echo chamber effect and over-specialization, diversity and novelty are crucial in maintaining user engagement in systems like Google Discover (Castells et al., 2022; Nguyen et al., 2014; Yang et al., 2023). Balancing relevance and diversity is achieved through algorithms that explore new content while exploiting known user preferences (Kunaver & Požrl, 2017). The challenge is to provide engaging, varied content without overwhelming users (Ziegler et al., 2005).

Integrating factors like location and time into recommendation systems enhances content relevance (Adomavicius et al., 2022). For instance, Google Discover might suggest location-relevant news or evening leisure activities, demonstrating how contextual factors influence recommendations (Baltrunas & Ricci, 2009). This approach requires sophisticated algorithms for real-time data analysis and predictive modeling. While context-awareness improves user engagement, especially for mobile users, it also poses privacy and security challenges due to the need for personal and sensitive data (Zheng et al., 2014; Zhuang et al., 2011).

The interplay between personalization and privacy is a critical issue. While personalization enhances user experience, it often involves collecting extensive user data, raising privacy concerns (Kobsa, 2007). Ethical data usage is crucial, especially given the risks like identity theft and privacy breaches (Yu, 2016). Google Discover, for instance, tailors content based on detailed user data, necessitating robust data protection measures in line with regulations like GDPR (Kamarinou et al., 2017). Additionally, there’s the issue of filter bubbles, where highly personalized systems might limit exposure to diverse views (Pariser, 2011). Recommendation algorithms must therefore balance data use with ethical practices and content diversity, avoiding echo chambers while ensuring a well-rounded user experience (Bozdag, 2013).

Apart from the possible echo chambers, the other potential problem of recommendation systems is the so-called cold start. When a new user enters the system, the recommendation task is complicated by the insufficiency of information about this user and their preferences. The same happens when a certain new item is inserted into the system (Latrech et al., 2024).

Quite a lot of research have been recently dedicated to the Food Recommendation Systems (FRS) that analyze users’ preferences and behavior to provide personalized, flavorful, and health-conscious food recommendations. Such FRSs are used in websites and applications offering food recipes, healthy diet, etc. Such systems consider the recommendation process as just one step, in which user’s historical data is used to suggest food they might prefer (L. Liu et al., 2024). The weakness of this one-step process is the necessity to train the model again when the user’s preferences change. In contrast to this model of recommendation, the researchers discuss the systems based on a multi-step process - the Interactive Recommendation Systems (IRS). These systems dynamically learn user preferences and update all the time, thus meeting the changing needs of the users (Zhou et al., 2020). While one-step recommendation systems are suitable for immediate recommendations, the multi-step ones would work for the users who consistently search for recommendations of a particular type. Such applications as TikTok and Instagram are based on IRS.

Google Discover recommendation system

In 2017, Google LLC introduced Google Feed, a content discovery tool initially accessible only in the Google application (Thakur, 2017). In 2018, Google moved it to the Google main page on smartphones, enabling readers to receive tailored suggestions for content without opening the app. The system also provided “Topic Cards” presenting the recent information linked to specific topics. In the same year, Google transformed Google Feed into Google Discover, giving it a refined appearance, and incorporating new properties, such as “Topic Channels” allowing readers to follow certain themes and get recommendations for customized content. The platform keeps continuously evolving and improving, with ongoing introductions of new functions and upgrades. An example of this is the implementation of the “Web Stories” function (Jasti, 2020).

Google Discover employs NLP methods for exploring and categorizing information according to subjects, objects, and attributes. It also utilizes user engagement metrics to ensure a constant improvement in the relevance of its recommendations. Google Discover provides recommendations for articles, videos, photographs, and various multimedia types sourced from a diverse range of websites. These recommendations are displayed in an aesthetically pleasing and intuitive manner, offering users the flexibility to customize the suggested content based on their preferences. Unlike algorithms on social media platforms like Facebook and Twitter, Discover recommends material from a more diverse array of resources, considering users’ search history and interests. The computational structure of Google Discover is intended to be more open, allowing users to personalize their stream and provide comments on the presented material. Instead of relying only on engagement metrics, Google Discover algorithm gives priority to classifying and displaying information based on subjects and objects, offering users a broader selection of content. Overall, Google Discover delivers a considerable number of visitors and interactions potential for web content owners.

Google Discover plays a crucial role for both web users and media outlets, offering a tailored experience that improves the process of finding relevant content while expanding the range of online information (McKelvey & Hunt, 2019). For users, Google Discover provides access to diverse and relevant content according to their specific goals, past searches, and web activities (Lu et al., 2015). The recommendation of content aligned with a user's interests reduces the effort and time required to locate relevant information, establishing Google Discover as a powerful instrument to improve user experience. Website owners, leveraging Google Discover algorithms, gain the chance to expose their material to users who might not have otherwise noticed it, therefore expanding their audience. This strategic use of machine learning increases exposure and generates more website traffic, boosting interaction and revenue for publishers (Giomelakis & Veglis, 2016). Furthermore, the recommendation system's capability to offer content aligned with users' interests is believed to result in longer user dwell durations and better engagement rates, which are important markers of the relevance and quality of the content (Zou et al., 2019). Previous studies have highlighted the significance of user engagement metrics in refining recommendation systems (Yi et al., 2014). Metrics such as depth of browse and bounce rate offer deeper insights into user preferences and behaviors, which are instrumental in enhancing the accuracy of content recommendations.

Researchers highlight that Google Discover has recently grown into a significant origin of traffic for newspapers. In the light of recent studies, "one news domain witnessed over 30% of its total web traffic originating from Discover, while other publishers reported that Google Discover drove more traffic than organic search in some months" (Hamilton, 2022). It underscores the increasing significance of the recommendation system in facilitating finding and consuming content. However, like each platform reliant on an algorithm to give recommendations for content, Google Discover faces issues related to biased and manipulated recommendations. Critics question Google Discover algorithm's transparency and how it affects the content that users can access (Haim et al., 2018). In response to these worries, Google LLC has undertaken initiatives to make the algorithm more transparent, aiming to guarantee users are presented with a variety of content (Rader et al., 2018). For example, they have added an updated function that lets users comment on the suggested content – a step intended to improve the relevance and accuracy of the algorithm.

Google Discover offers its users various options of engagement with the presented content. Users can share specific content with others, watch videos, read the entire article by clicking on a card, or save it for later use. Furthermore, users can actively contribute to the posts shown by expressing their opinions, aiding in the continuous refinement of recommendations. To ensure the quality and adherence to specific criteria, Google Discover policies have been established. These standards serve as guidelines for web content publishers, outlining the types of content that Google Discover should feature. For instance, the content needs to be unique, of high quality, and free from information that could be misleading or deceptive. Additionally, these policies incorporate directives for user data protection and privacy. Google employs machine learning algorithms to customize recommendations of content for readers while simultaneously prioritizing reader privacy and maintaining information security. By customizing their activity and interest preferences, users can also have control over the content that is presented to them.

Publishers can use various tactics to enhance the probability of their material being featured in Google Discover. Such tactics include: (1) producing high-quality content that adds value for readers; (2) creating mobile-friendly material; (3) selecting concise headings that precisely represent the content; (4) concentrating on subjects that interest users; (5) establishing a robust online presence by email newsletters, social media, and SEO; (6) utilizing structured data markup to assist Google in understanding their content; (7) systematically assessing content performance in Google Discover and making necessary adjustments to the strategy. Implementing these tactics can increase the likelihood of publishers displaying their content highlighted in Google Discover, thereby obtaining a broader viewership.

A short analysis of different papers and reports, as well as the study conducted by Lopezosa et al. (2022), confirm that there is currently a gap in scientific works dedicated to Google Discover.

In the scientific peer-reviewed literature, we encounter only one paper examining Google Discover by Lopezosa et al. (2024). The paper examines the effect of Google Discover on web traffic and its impact on both high-quality and sensationalist content. It employs a qualitative methodology and includes semi-structured interviews with experts from Brazil, Spain, and Greece to explore Google Discover's role and effectiveness. Lopezosa et al. emphasize Google Discover's varied impact across different countries, its influence on content quality, and the strategies media outlets use to enhance their visibility on this platform.

In response to this identified gap, by a lack of scientific studies on Google Discovery, we have set the goal of this study as follows: to investigate the influence of Google Discover on both online readers and online content creators, utilizing the collected information from the service. Our study aims to advance knowledge of Google Discover from two angles: (1) from the web publishers point of view, by investigating the possible traffic directed to a webpage through the recommendation service, and (2) from the user's point of view, by measuring the user's level of interest when using the recommendation service. To achieve the goal of the paper, we have formulated two research questions:

1. How does the perception of online content among users influence the efficiency of web content creators in Google Discover?

2. What criteria does Google Discover utilize to decide what type of material it recommends to its users?

Methodology

The data utilized in this research was obtained from the Google Search Console (GSC) tool (Authors, 2023). GSC, a service developed by Google, enables web administrators and approved operators to observe and evaluate the performance of their webpages in Google search results (Authors, 2024). Accessing data in GSC involves meeting specific conditions. First, a domain name must be operational and undergo verification with GSC. Verification methods include an HTML link or tag being added to the domain's server or the source code header respectively. Second, the domain must report visitors from the search engine, e.g. from natural search results or inclusion in Google Discover stream. After these prerequisites are fulfilled, web administrators and approved operators can access a variety of statistics detailing the domain's visibility in search results, encompassing information on search terms, displays, clicks, and click-through rate (CTR). Analyzing this data, webmasters can learn more about how their website displays in search results and make adjustments to enhance its performance and visibility.

The authors of this research gained access to the GSC accounts of two e-commerce websites that are consistently displayed in the Google Discover feature., The owners of these domains have not granted consent to publish domain names, therefore, for this study, we have anonymized them and will refer to them as Website1 and Website2. Both websites represent Polish internet e-commerce stores specializing in a range of consumer electronics and offering valuable information about these items to their clients. Website1 sells household appliances, whereas Website2 offers computers, consoles, TV & audio, and mobile devices.

Table I contains statistics regarding the popularity and use of these both websites among Internet users. n Poland there are about 30 million active online users. From the data given in the table, we see that these two websites cover only about 20% of the Polish internet population. However, it is important to notice that there is a significant difference in range between both websites. Website1 has a range of 1% of Polish internet users which is a relatively good score, but Website2 is a leading online store in the consumer electronics sector - covering around 20% of internet users in Poland.

<<< INSERT TABLE I ABOUT HERE >>>

To collect data on visits displayed in the Discover section of GSC, we have taken the following steps. Initially, we accessed the GSC account and navigated to the “Performance” tab. Subsequently, we chose the “Discover” tab, which provided details about the number of clicks, displays, and CTR that each website obtained via Google Discover. Within GSC, this data can be sorted by period, pages, country, and presence in Discover. Additionally, we clicked on individual tabs in the chart to obtain more thorough details regarding particular content featured in Google Discover. To compile the data, we clicked the “Export” option, exporting the data into an XLSX file.

The downloaded file contained a list of particular URLs displayed in Google Discover, accompanied by the number of clicks and impressions, along with the CTR for every webpage. Additionally, it provided an overall number of impressions, clicks, and CTR for each country and date. By a click, we understand the action of clicking a link to a certain website, while the number of impressions means the frequency of a link to a certain website being displayed to users. The ratio of clicks to impressions forms the click-through rate, i.e. CTR. The CTR is one of the key factors in the e-commerce business (Pahor et al., 2022). In online advertising, search engines, recommendation systems, human-computer interaction, movies, and many others, the click-through rate (as well as its analysis and prediction) is of great commercial value, since it can rank the items (i.e., web pages) returned to a user and helps to further maximize the number of clicks (Wu et al., 2022). The higher the CTR, the better, as it indicates that an advertisement content or a webpage title and description are well-matched to the expectations of potential customers.

The analytical approach in this study is based on the user interaction with the recommendation system. The set of data offered by Google Search Console belongs to the group of user interactions that can be measured on a website. Among such metrics there are: the number of clicks, the number of impressions, the CTR, the origin of traffic, the list of URLs, and the date of occurrence in Google Discover. Despite the fact that it is a rather limited type of data, using the URL addresses we were still able to analyze the type of content that is of most interest for the users.

Results

The data obtained from GSC for Website1 consists of 32 unique webpages featured in Google Discover, resulting in 2 337 clicks, 39 233 impressions, and a CTR of 5.96%. The whole traffic originated from a single country. For Website2, the retrieved statistics encompass 514 unique webpages displayed in Discover, 127 189 clicks, 2 561 639 impressions, and a CTR of 4.97%. Clicks for Website2 originated from 21 countries. While gathering data, we observed a slight disparity between the dataset that can be downloaded and the data depicted on plots within the GSC service, with the figures on the plot being a little larger than those in the downloaded dataset.

Analyzing the data from Website1 and Website2, we have examined it from two points of view: that of the material publisher (i.e., the website creator) and that of the material reader (i.e., the website's user). The results of the examination (presented below) show us (1) what kind of content (on what topics) the owners of these two websites publish most frequently, and (2) which topics (among the offered) attract the most attention of the users of the two analyzed websites.

Publisher perspective on content analysis

We analyzed the URLs obtained from both websites to discern the nature of the content presented on these pages. Each of the two websites is an online store. Through analysis of the content, we categorized the content into six types: (1) news (primarily blog articles featuring news content; (2) products (pages showcasing goods or services that are offered for sale on the internet shop); (3) guides (extended posts providing in-depth information about a specific topic, typically longer than standard news items; (4) landings (pages designed for promotional objectives to showcase products or services; (5) categories (webpages that list goods related to a specific group); (6) promotions (webpages containing discounts on offers).

Website1

The major themes within the listing of new content are promotions and sales, with a special emphasis on LEGO kits discounts and special offers during Black Friday and Cyber Monday. In the webpages categorized as “product” content, the major topics include gym and fitness gear, appliances for the kitchen, and LEGO sets are a few examples.

The major theme for guides is purchasing diverse household equipment, including irons, electric toothbrushes, men’s shavers, water filter pitchers, and lawnmowers. Additionally, there are posts focusing on particular goods, like hair dryers or automatic toothbrush heads. There is also some advice on the proper maintenance of a coffee maker and a post highlighting the advantages of relaxing in the water. This content suggests that e-commerce is tailored to deliver valuable details for Polish customers intending to buy household goods and appliances.

Website2

In this e-commerce, the primary themes covered in the service include (1) sales and promotions on computer hardware and electronics (notebooks, monitors, accessories, and PC equipment from different producers); (2) unique occasions and sales, such as Cyber Monday, Black Friday, and Black Weeks; (3) new releases in the technology and gaming industries, featuring headphones, graphics cards, and gaming chairs; (4) holiday gift guides and inspirations, for events like Saint Nicholas Day and Christmas; (5) professional evaluations and recommendations for electronic items. Other subjects included computer cooling systems, networking hardware, and Wi-Fi extenders.

Regarding product URLs, the most frequently occurring device types were smartphones (mostly Xiaomi and Samsung), smartwatches (mainly Huawei and Samsung), headphones (with Arctis Nova Pro Wireless, H9 Inzone, and Razer Viper V2 Pro being the most popular), laptops (mostly from Huawei and Apple), desktop computers (Acer being the only one in the list), tablets (Huawei only), and various other devices.

The main themes in the catalog of guides were technology-related guidance and suggestions, including topics such as selecting a soundbar, processor, graphics card, power bank, USB drive, memory card, and headphones., or maximizing the capabilities of Nvidia GeForce RTX graphic cards.

User perspective and behavior analysis

To explore the behavior of users on Website1 and Website2, we considered the subsequent markers: (1) quantity of content (expressed as "content pieces", representing the total number of distinct webpages associated with a website that is displayed in Google Discover); (2) clicks on every unique URL; (3) click-through rate (the rate of clicks to displays for each unique URL).

Website1

Table II shows the click count for 32 webpages in Website1, providing information on the impression count for each content type and their respective CTRs. The content in Website1 is categorized into three distinct groups, and the table is organized in ascending order based on their CTR values.

<<< INSERT TABLE II ABOUT HERE >>>

In Website1, it is evident that the guides receive the highest level of user attention. Despite having approximately 1.5 times more specific URLs with products (14) compared to guides (9), the guides exhibit a higher click-through rate of 7.7% as opposed to 4.7% for the products. Website1 offers users informative guides covering a broad spectrum of topics related to household items. This suggests that users find significant value in informative guides, even when they may come across products available for purchase on other websites that may offer better deals, prices, quality, or a combination of these factors.

It is crucial to consider that the data analysis was conducted using a content recommendation system. The analyzed clicks refer to users' activities with URLs to the content recommended by Google Discover during a specific time. Following these interactions, users were presented with fresh content recommendations from Website1, with their past clicks and the overall number of clicks made by other users taken into consideration. In other words, clicks on engaging content likely contributed to a higher number of clicks on specific content URLs.

Website2

In Table III, we present the data for 514 webpages in Website2, including the number of clicks, impressions, and the click-through rate for each content category. The content in Website2 is categorized into six groups, organized based on the CTR.

<<< INSERT TABLE III ABOUT HERE >>>

For Website2, a comparable situation to Website1 is observed – of all the six types of content, webpages categorized as guides, have the highest CTR, despite having a relatively lower number of content pieces (only nine). Interestingly, the "Guides" content type is not remarkable in terms of the number of clicks and impressions when compared to "News" and "Products". The latter types have significantly more clicks (51 519 and 72 681, respectively) and views (917 528 and 1 591 508). Nevertheless, the number of clicks (2 166) and impressions (30 233) places "Guides" in the first position. It appears that the guides offered by Website2 are more appealing to its customers than the goods available for purchase. The CTR for Website2 is 4.70%, having 261 webpages in the service, as well as for Website1, which has 14 webpages.

An analysis of click-through rates (CTR) across different product categories reveals significant insights into how content features impact user engagement (Table IV). Categories with a limited number of products, such as laptops (7 products, 7.94% CTR) and consoles (5 products, 6.27% CTR), exhibit higher CTRs. This suggests that a curated selection may reduce user choice overload, facilitating quicker decision-making and increasing engagement. In contrast, categories with an extensive range of products, like smartphones (105 products, 5.14% CTR) and graphics cards (34 products, 4.33% CTR), show moderate CTRs, indicating that an abundance of options might overwhelm users and dilute their engagement.

Furthermore, the frequency of content updates appears to influence user interest and CTR. Frequently updated categories such as smartphones and graphics cards maintain user attention through regular new releases but may suffer from diluted CTRs due to market saturation. Conversely, categories with less frequent updates, like action cameras (2.51% CTR) and drones (3.81% CTR), demonstrate lower engagement, potentially stemming from outdated content or a lack of novelty. These findings suggest that optimizing product selection, enhancing content freshness, and improving user navigation and personalization are crucial strategies for increasing CTR and overall user engagement in e-commerce platforms.

To draw conclusions about user behavior, we conducted a comparative analysis of the results for Website1 and Website2, and the findings are illustrated in Figure 1 and Figure 2. Figure 1 compares Website1 and Website2 in terms three content of types (products, news, and guides) presented in Google Discover for both websites. It becomes evident that in both scenarios, “Guides” content gathered significantly more interest from users compared to the other two content types, with a small variation between the websites: the CTRs are 7.7% for

Website1 and 7.16% for Website2. Furthermore, user interest in “News” content prevailed in Website2 – CTR is 5.61%, while in Website1 it was 5.04%. Lastly, “Products” content is in the third position, with nearly equal scores for both websites – 4.7% for Website1 and 4.56% for Website2.

<<< INSERT FIGURE 1 ABOUT HERE >>>

The results obtained provide a general understanding of users’ behavior when using the examined Website1 and Website2. We can only make recommendations about users’ motivations based on the information that is currently available about these websites, as we did not study users’ opinions. In today’s web industry, numerous stores offer computers, mobile phones, and a variety of home electrical equipment. Online customers rarely face the challenge of not finding a deal that interests them and fits their preferences or budget. However, in addition to making purchases, users sometimes seek reliable comparisons, comprehensive instructions, and other information about the devices they already own or plan to purchase. This is why well-written guides and up-to-date news from various tech areas may hold greater value for users, as not all online retailers offer them.

Figure 1 depicts the division of click-through rate values for Website1 and Website2 over time, spanning November 2021 through March 2023. The illustration reveals a relatively uneven fluctuation in the CTR level for Website1 during the considered period. There were 171 days (out of the 485 studied) where the CTR was recorded as 0%. These zero levels prevailed in November and December 2021, remaining relatively consistent in the first quarter of 2022. Later, there were a few days with 0% CTR in several months, including a 16-day stretch in August 2022. In contrast to the time from November 2021 to April 2022, these 16 days are less significant for the inquiry because it may be argued that such less “successful” days could occur for any online store. In summary, 13 instances of CTR values between 20% and 50% were noted throughout the investigation period. 36 days had a CTR ranging from 10% to 16.98%, 249 days had a CTR ranging from 2.25% to 9.76%, and 14 instances had a CTR ranging from 0.78% to 1.79%. A total of 171 observations were made of the zero CTR previously described. Website1 had a CTR of up to 10% for over 50% of the time, with distinct fluctuations during the remaining period.

It is crucial to highlight ten specific days when values are at their highest observed for Website1: two days with 100% CTR in Q1 2022; three days with 50% CTR in Q4 2021 and Q2 2022; two days with 33.33% CTR in Q1 2022; and two days with 25% CTR in Q2 2022. These isolated spikes in value, particularly the instances of 100% CTR, should be considered anomalies in the data and not indicative of the typical pattern of users' behavior. They occur because very few clicks and impressions are often required to reach such high CTR levels. For Website1, the observed 100% CTR is the result of one click and one display; the 50% CTR is based on one click to two displays; the 33.3% CTR comes from one click to three displays, and the 25% CTR is derived from one click to four impressions. Therefore, these ten days are cases of quite minimal user interest in Website1, coupled with a limited presence in Google Discover.

Overall, the distribution of the CTR for Website1 appears to follow a rather regular pattern. Any internet retailer might experience fluctuations in customer interest due to various reasons, such as a lack of new products, absence of sales and incentives, global or national financial crises, or other external factors influencing an online retailer's operations. Additionally, the algorithms determining how Google Discover recommends content play a significant role in these fluctuations. URLs within a domain may be part of the recommendations in one day and absent in another day as the algorithm deems them less relevant for readers. Ultimately, as mentioned earlier, the issue with the CTR level gradually normalized, and the online shop in Website1 has been operating efficiently in Google Discover up to the point when we collected these data from the system.

For Website2, we see a relatively balanced distribution over this period, with a single notable spike in CTR in October 2022, reaching 100%. Similar to Website1, this spike is associated with a one-to-one ratio of clicks to impressions, making it an outlier that should not be used to analyze the overall trend for Website2. Throughout the remaining time, the CTR fluctuated between 0.77% and 19.98%. There were two days with almost 20% CTR (19.12% and 19.8%); 22 days exhibited a CTR from 10.06% to 14.51%. For the longest duration (295 days, roughly more than half of the analyzed period), the CTR remained between 2.01% and 9.73%. Ultimately, within 24 days, the CTR ranged from 0.77% to 1.96%. For the remaining 141 days, non-CTR was observed. It is crucial to notice that the CTR values for Website2 did not exhibit a steady expansion day by day but rather varied daily throughout the examined period, as seen in Figure 2.

<<< INSERT FIGURE 2 ABOUT HERE >>>

As mentioned earlier, this study did not involve the assessment of online users' opinions. There is also no specific information available regarding the events in Website2 after October 2022, when the CTR scores dropped to 0%. It is possible that Website2 encountered a specific incident leading to its exclusion from Google Discover service. It is also possible that nothing was altered within the website, yet it ceased to feature in Google recommendations. It is crucial to emphasize that the absence of an e-commerce website in Google Discover should not be considered as a lack of appeal to internet users. However, the online store may no longer attract potential customers through Google recommendations.

As previously mentioned, the data on clicks for Website1 are shown only for one country, with a mean CTR of 5.96% over the 485 days. However, for Website2, statistics are available in 20 countries. Figure 3 illustrates the geographical breakdown of the CTR values for Website2 spanning November 2021 through March 2023. Given that Website2 is a Polish internet retailer, Poland leads the list with a CTR of 5.05% over the 17 months. On the opposite end, Belgium has the lowest CTR of 0.49%. Eight of these twenty countries have a CTR lower than 1%, five have a CTR between 1% and 2%, another five have a CTR higher 2% and lower than 4%, while only two have a CTR exceeding 4%. The average value of CTR for all regions is 1.27%.

<<< INSERT FIGURE 3 ABOUT HERE >>>

The findings imply that the presence of other countries in the data can be attributed to language settings. This online store operates only in Polish language. Therefore, it can be suggested that Google Discover suggested Website2 to individuals with Google settings set to Polish or those who carried out their prior online searches in the Polish language, regardless of their actual location.

Discussion

This paper undertakes an examination of Google Discover, a recommendation service introduced by Google in 2018, designed to provide personalized content to online users. The study aims to assess the influence of Google Discover on two categories of stakeholders, content creators and internet users seeking specific content. To achieve this, in the study we formulated two research questions, the answers to which are derived from the results. In the initial phase of the study, we conducted a literature review focused on Google Discover. The findings disclosed a shortage of recent publications addressing the recommendation system's activities. References to Google Discover are mostly found on web pages from Google blogs, such as (Corby, 2018; Jasti, 2020; Thakur, 2017), or other blogs like (Hamilton, 2022). Notably, there are only two studies specifically exploring Google Discover algorithms, authored by Lopezosa et al. (2022) and Lopezosa et al. (2024). These publications stand out as the only two results when searching the keyword "Google Discover" in scientific databases such as Web of Science or Scopus. Such discovery led to the identification of a significant research gap that needs to be filled. To examine the mechanism of the recommendation service, we obtained data from the GSC, focusing on two domains: e-commerce stores functioning in Poland and specializing in electronic equipment, referred to as Website1 and Website2. The study covers data retrieved spanning November 2021 through March 2023. The collected data, comprising the aggregate counts of clicks, impressions, and click-through rates for every URL inside Website1 and Website2, were gathered in a spreadsheet file for analysis.

When considering the content creators (the owners of Website1 and Website2), whose material is handled by Google Discover, several observations have been made. Firstly, the information they present can be categorized into six types: news, products, guides, landing pages, categories, and promotions. For Website1, data is available for only three types (9 webpages with news, 9 with guides, and 14 webpages with products), while for Website2, data covers all six groups, with three dominating in terms of the number of URLs (261 with products, 238 with news, and 9 with guides). Furthermore, it is noted that the majority of popular news subjects for both websites are deals, discounts, and promotions on a range of electronic devices. Lastly, both websites consistently offer comprehensive guidance on choosing and utilizing domestic and sports equipment. This particular content group seems to attract the most attention from users.

To analyze user behavior on Website1 and Website2, we considered several indicators: (1) the count of unique webpages appearing in Google Discover, (2) the count of clicks on every webpage, (3) how many times the recommendation system presented each URL to a user (referred to as “impressions”), and (4) the CTR for each webpage. We conducted a comparative analysis for three content categories (news, products, and guides) as data on them is accessible for both websites. For Website1 and Website2, the CTRs for webpages with guides were the highest (7.7% and 7.16%, respectively). Additionally, for both websites, the CTRs for news (5.04% and 5.61%) were higher than for products (4.7% and 4.57%). However, it is essential to analyze whether this high CTR was a result of user behavior, the effect of Google Discover algorithms' work, or the efforts of content producers in gaining clicks to access their material. A combination of these elements would be the answer to this.

Google Discover is known for utilizing user data, including preferences, location, language settings, visited and bookmarked links, as well as likes and dislikes. This recommendation system not only delivers the content customized to user interests but additionally enables them to follow particular themes for more similar information in their feed or save materials for later viewing. Google Discover continuously evolves and extends its suggestions beyond search, reaching platforms like Google Maps (iTrust, 2022). For example, if a user saves webpages with a guide on choosing a newer laptop, Discover may subsequently offer similar posts about laptops and guides on related topics (PCs, devices, etc.). From a publisher's perspective, appearing in Google Discover involves similar principles to standard SEO practices for website promotion in search engines. The content ought to be noteworthy, tailored to audience preferences, and hosted on an appealing website for users to explore. Additionally, publishers find it beneficial to monitor their Discover performance in GSC (Toonen, 2022). Consequently, Google Discover not only considers user preferences but also strives to present new themes and ideas, giving priority to reliable information from credible sources.

For instance, if Google Discover presents a guide from Website1 to a reader and the user visits that URL, the algorithm will take note of this interaction. Subsequently, the algorithm will start recommending more guide URLs to the same user, not only from Website1 but also from Website2 B and other domains. This continuous exposure aims to capture the user's interest in additional content from various sources. Depending on user engagement, Website2 may experience an increase, decrease, or maintenance in the number of clicks. When aggregated with the actions of other users, the CTR for URLs from Website2 will fluctuate within the system. At a certain point, as observed in the case of Website2, Google Discover may take a website from its algorithm if it deems the content no longer useful to readers.

The results of this study (all the data we possess about two websites) allow, first of all, to answer the question of what is happening inside Google Discover algorithms - through the behavior of users of Website1 and Website2. However, it might be reasonable to answer also the questions “why” and “how” for the analyzed websites. Why do we observe this particular behavior of users: paying more attention to the comprehensive guidance on choosing and utilizing various equipment than to purchasing this equipment? As we observe the online market of domestic appliances and other devices (the local one, even not taking into consideration international deliveries), we see a very wide offer with prices suitable for various budgets. Thus, we can claim that there is no shortage of the goods of this type to buy online, and Website1 and Website2 hold no monopoly in selling them. Yet, it is more difficult to make a decision about which particular device to buy. Its price, quality, ease of use, warranty period, compatibility with other devices - these are a few of many factors that would be taken into consideration by the users. And if they require assistance in setting together all the criteria they consider important - they might be interested in reading one or a few guides to finally make up their mind.

Finally, how do the users of Website1 and Website2 perform their behavior? As we see in the online activity metrics of the websites, values for mobile version usage are rather high. Apart from the fact that today Internet users tend to interact with various websites mostly via mobile devices as well as to use mobile apps for various tasks, the key is the fact that Google Discover is available for users only on mobile devices. Once a user starts interacting with Google Discover, the recommendation algorithm begins learning from the user about their preferences. With each subsequent interaction the recommendations are more and more customized. Once the users' preferences are met, they become more and more engaged into further interactions with the system.

Contributions

We believe that this study makes significant contributions to the studies about Google Discover in particular and recommendation systems in general. The theoretical contribution is the conducted literature review which serves as a comprehensive summary of the most pertinent studies and discussions concerning Google Discover, illustrating what was carried out in this field and what was not. The research gap was identified, emphasizing the considerable potential for investigations regarding this subject. Secondly, the practical contribution of our research is the examination of GSC data from two Polish websites. Besides the acknowledged constraints, it offers valuable insights into how Google Discover algorithm is assisting online content providers in promoting their websites and webpages. Moreover, we believe that since Website1 and 2 differ quite significantly in size, a comparison of their results in Google Discover provides additional insight into the "attitude" of the algorithm toward websites with different user activity. Unlike the studies conducted previously, our research focuses not on general analysis of suggestions provided by Google Discover within a certain period of time or on the specificity of GD algorithms, but one the behavior of the recommender systems towards two particular websites, with a particular type of content, over time. The authors assume that with further deeper analysis of content published by the two websites during the explored period, and with correlation of it with the results of GD recommendation, website owners (content publishers) may be able to draw extra conclusions on even the minor reactions and changes that take place withing the recommendation algorithm.

Google estimates that there are 800 million users that use Google Discovery feed every month (Corby, 2018), and statistics show that the market share of Google search engine worldwide is significantly higher than of any other engine (StatCounter, 2024). The recommendation system behind the feed is able to provide the users with content on any topic they might be interested in. Summing up – we believe that the conclusions we have made might be of value and of practical use for website owners who want to promote their websites and would like to engage such a widely used recommendation system for it.

Limitations

We acknowledge five primary limitations of the study. First, there is a scarcity of studies on Google Discover and its algorithm in the academic literature. The available sources are mostly internet blogs, providing a more commercial perspective rather than a scientific analysis of the recommendation system. A broader range of research papers on Google Discover would allow for the comparison of alternative methodological approaches to analyze its behavior. This, in turn, could facilitate a more comprehensive understanding of how Google Discover functions for webpages and their visitors.

Second, the study is constrained by the limited scope of studied subjects, as access to Discover data in GSC is available for only two websites. Additionally, these e-commerce stores pertain to just one country, Poland. Access to data from more domains within the same country, covering similar subjects, could provide additional insights into the types of material preferred by users in Poland, the content that makes its way to Google Discover, and whether local content is appealing to consumers from abroad. Expanding the dataset to include multiple countries would enhance the comparative analysis, leading to a deeper comprehension of the variables and algorithms behind Google Discover.

Third, our study can be considered a short-term study because we only have data from a 17-month period, from November 2021 to March 2023. This timeframe may not capture longer-term trends, seasonal variations, or shifts in user behavior and preferences. A recommendation would be to obtain data from a longer period of time, which could show whether content recommendations have a significant impact on user behavior and preferences.

Fourth, is the lack of detailed user engagement metrics such as depth of browse and bounce rate. GSC data only provides metrics on clicks, impressions, and CTRs for content. Additional metrics would be important for understanding the depth of user interaction with the content and could potentially influence the accuracy and prioritization mechanisms of the recommendation system.

The final limitation is derived from the nature of Google Discover itself. The code responsible for the algorithm undergoes frequent modifications, and there is no accessible means to obtain or identify the current version. The technical specifications controlling the "decisions" of Google Discover remain unknown, as it is not transparent how each change in the code contributes to the recommendation system. Consequently, the evaluation is conducted in a black-box manner, where the outcomes of the recommendation system's activity are observed without insight into the internal operations. The study focuses on examining the "behavior" of Google Discover based on the samples of content that have been by now processed by Discover and suggested (or not) to the readers.

Future work

The constraints outlined in our study suggest several directions for additional investigation of Google Discover recommendation algorithm. First, we could obtain data from GSC for a more extensive range of domains in one language. Further, it could be valuable to analyze the behavior and topics of interest of internet users, to understand how they impact the functioning of Google Discover. Future research should consider incorporating additional data sources that provide detailed user engagement metrics. Another suggested continuation that might enhance the outcomes of the study could be to obtain data over a much longer time frame. It would enable us to capture the moment when a domain becomes part of the Discover service and to monitor the shifts in visibility for that domain. Finally, we could connect these modifications with changes in the content offered by the website to its users.

References

Authors (2024). *Aslib Journal of Information Management*

Authors (2023). *Information Systems Development*

Adomavicius, G., Bauman, K., Tuzhilin, A., & Unger, M. (2022). Context-aware recommender systems: From foundations to recent developments. In F. Ricci, L. Rokach, B. Shapira, & P. Kantor (Eds.), *Recommender Systems Handbook* (pp. 211–250). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_6

Alyari, F., & Jafari Navimipour, N. (2018). Recommender systems. *Kybernetes*, *47*(5), 985–1017. https://doi.org/10.1108/K-06-2017-0196

Baltrunas, L., & Ricci, F. (2009). Context-based splitting of item ratings in collaborative filtering. *Proceedings of the Third ACM Conference on Recommender Systems*, 245–248. https://doi.org/10.1145/1639714.1639759

Bhatia, V. (2024). DLSF: Deep learning and semantic fusion based recommendation system. *Expert Systems with Applications*, *250*, 123900. https://doi.org/10.1016/j.eswa.2024.123900

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 205395171562251. https://doi.org/10.1177/2053951715622512

Campbell, J. C., Hindle, A., & Stroulia, E. (2015). Latent Dirichlet allocation: extracting topics from software engineering data. In *The Art and Science of Analyzing Software Data* (Vol. 3, pp. 139–159). Elsevier. https://doi.org/10.1016/B978-0-12-411519-4.00006-9

Castells, P., Hurley, N., & Vargas, S. (2022). Novelty and diversity in recommender systems.

In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 603–646). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_16

Corby, K. (2018). *Discover new information and inspiration with Search, no query required.* Google. https://blog.google/products/search/introducing-google-discover/

Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for Youtube recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198. https://doi.org/10.1145/2959100.2959190

Fahrudin, T., & Wijaya, D. R. (2024). New custom rating for improving recommendation system performance. *Journal of Big Data*, *11*(1), 91. https://doi.org/10.1186/s40537-024-00952-3

Ghori, M. M., Dehpanah, A., Gemmell, J., Qahri-Saremi, H., & Mobasher, B. (2022). Does the user have a theory of the recommender? A grounded theory study. *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, 167–174. https://doi.org/10.1145/3511047.3537680

Giomelakis, D., & Veglis, A. (2016). Investigating search engine optimization factors in media websites. *Digital Journalism*, *4*(3), 379–400. https://doi.org/10.1080/21670811.2015.1046992

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? *Digital Journalism*, *6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Hamilton, A. (2022). *How to Succeed in Google Discover*. Search Engine Journal. https://www.searchenginejournal.com/google-discover/361142/

iTrust. (2022). *Google Discover: The ultimate guide*. https://itrust-digital.com/google-discover/

Jasti, V. (2020). *A new home for Web Stories on Discover*. Google. https://blog.google/web-creators/web-stories-discover-search/

Kamarinou, D., Millard, C., & Singh, J. (2017). Machine learning with personal data. In *Data Protection and Privacy: The Age of Intelligent Machines* (pp. 89–114). Hart Publishing.

Kellar, M., Watters, C., Duffy, J., & Shepherd, M. (2005). 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. https://doi.org/10.1002/meet.1450410119

Kobsa, A. (2007). Privacy-enhanced personalization. *Communications of the ACM*, *50*(8), 24–33.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30–37. https://doi.org/10.1109/MC.2009.263

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems – A survey. *Knowledge-Based Systems*, *123*, 154–162. https://doi.org/10.1016/j.knosys.2017.02.009

Latrech, J., Kodia, Z., & Ben Azzouna, N. (2024). Twit-CoFiD: a hybrid recommender system based on tweet sentiment analysis. *Social Network Analysis and Mining*, *14*(1), 123. https://doi.org/10.1007/s13278-024-01288-9

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, M., Ma, W., & Chu, Z. (2024). KGIE: Knowledge graph convolutional network for recommender system with interactive embedding. *Knowledge-Based Systems*, *295*, 111813. https://doi.org/10.1016/j.knosys.2024.111813

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

Liu, J., Dolan, P., & Pedersen, E. R. (2010). Personalized news recommendation based on click behavior. *Proceedings of the 15th International Conference on Intelligent User Interfaces*, 31–40. https://doi.org/10.1145/1719970.1719976

Liu, L., Guan, Y., Wang, Z., Shen, R., Zheng, G., Fu, X., Yu, X., & Jiang, J. (2024). An interactive food recommendation system using reinforcement learning. *Expert Systems with Applications*, *254*, 124313. https://doi.org/10.1016/j.eswa.2024.124313

Lopezosa, C., Giomelakis, D., Pedrosa, L., & Codina, L. (2024). Google Discover: uses, applications and challenges in the digital journalism of Spain, Brazil and Greece. *Online Information Review*, *48*(1), 123–143. https://doi.org/10.1108/OIR-10-2022-0574

Lopezosa, C., Guallar, J., & Santos-Hermosa, G. (2022). Google Discover: entre la recuperación de información y la curación algorítmica. *Scire: Representación y Organización Del Conocimiento*, *28*(2), 13–22. https://doi.org/10.54886/scire.v28i2.4796

Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems*, *74*, 12–32. https://doi.org/10.1016/j.dss.2015.03.008

McKelvey, F., & Hunt, R. (2019). Discoverability: Toward a definition of content discovery through platforms. *Social Media + Society*, *5*(1), 205630511881918. https://doi.org/10.1177/2056305118819188

Nguyen, T. T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. *Proceedings of the 23rd International Conference on World Wide Web*, 677–686. https://doi.org/10.1145/2566486.2568012

Pahor, S., Kopič, D., & Demšar, J. (2022). *Feature embedding in click-through rate prediction*. http://arxiv.org/abs/2209.09481

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3173677

Ricci, F., Rokach, L., & Shapira, B. (2022). Recommender systems: Techniques, applications, and challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 1–35). Springer US. https://doi.org/10.1007/978-1-0716-2197-4_1

Seo, Y.-W., & Zhang, B.-T. (2000). Learning user's preferences by analyzing Web-browsing behaviors. *Proceedings of the Fourth International Conference on Autonomous Agents*, 381–387. https://doi.org/10.1145/336595.337546

StatCounter. (2024). *Search Engine Market Share Worldwide*. https://gs.statcounter.com/search-engine-market-share/all/

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, *2009*, 1–19. https://doi.org/10.1155/2009/421425

Thakur, S. (2017). *Feed your need to know*. Google. https://blog.google/products/search/feed-your-need-know/

Thurman, N. (2011). Making 'The Daily Me': Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism*, *12*(4), 395–415. https://doi.org/10.1177/1464884910388228

Toonen, E. (2022). *Google Discover: What it is and how to get your content in*. Yoast. https://yoast.com/google-discover/

Wiering, M., & van Otterlo, M. (Eds.). (2012). *Reinforcement Learning* (Vol. 12). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-27645-3

Wu, Y., Hu, Y., Xiong, X., Li, X., Guo, R., & Deng, S. (2022). Graph relation embedding network for click-through rate prediction. *Knowledge and Information Systems*, *64*(9), 2543–2564. https://doi.org/10.1007/s10115-022-01714-4

Yang, D., Wang, Y., Shi, Z., & Wang, H. (2023). Toward topic diversity in recommender systems: integrating topic modeling with a hashing algorithm. *Aslib Journal of Information Management*. https://doi.org/10.1108/AJIM-01-2023-0019

Yi, X., Hong, L., Zhong, E., Liu, N. N., & Rajan, S. (2014). Beyond clicks: dwell time for personalization. *Proceedings of the 8th ACM Conference on Recommender Systems*, 113–120. https://doi.org/10.1145/2645710.2645724

Yu, S. (2016). Big privacy: Challenges and opportunities of privacy study in the age of big data. *IEEE Access*, *4*, 2751–2763. https://doi.org/10.1109/ACCESS.2016.2577036

Zhan, J., Chia-Lung Hsieh, I-Cheng Wang, Tsan-Sheng Hsu, Churn-Jung Liau, & Da-Wei Wang. (2010). Privacy-preserving collaborative recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(4), 472–476. https://doi.org/10.1109/TSMCC.2010.2040275

Zheng, Y., Mobasher, B., & Burke, R. (2014). CSLIM: contextual SLIM recommendation algorithms. *Proceedings of the 8th ACM Conference on Recommender Systems*, 301–304. https://doi.org/10.1145/2645710.2645756

Zhou, S., Dai, X., Chen, H., Zhang, W., Ren, K., Tang, R., He, X., & Yu, Y. (2020). Interactive recommender system via knowledge graph-enhanced reinforcement learning. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 179–188. https://doi.org/10.1145/3397271.3401174

Zhu, T., Li, G., Zhou, W., & Yu, P. S. (2017). Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *29*(8), 1619–1638. https://doi.org/10.1109/TKDE.2017.2697856

Zhuang, J., Mei, T., Hoi, S. C. H., Xu, Y.-Q., & Li, S. (2011). When recommendation meets mobile: contextual and personalized recommendation on the go. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 153–162. https://doi.org/10.1145/2030112.2030134

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. *Proceedings of the 14th International Conference on World Wide Web - WWW '05*, 22. https://doi.org/10.1145/1060745.1060754

Zou, L., Xia, L., Ding, Z., Song, J., Liu, W., & Yin, D. (2019). Reinforcement learning to optimize long-term user engagement in recommender systems. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2810–2818. https://doi.org/10.1145/3292500.3330668

Table I

Online activity metrics for both websites

Metric	Website1	Website2
Years of activity	17	21
Average monthly visits	271 000	6 244 000
Average monthly unique visitors	161 431	2 227 000
Average visit duration	2:38	4:35
Average pages per visit	3.66	5.19
Bounce rate**	53.91%	49.17%
Device distribution		
Desktop	56.49%	62.03%
Mobile	43.51%	37.97%

* *source: Similarweb, period April to June 2024*

**bounce rate is a percentage of visitors that view only one page on the website before leaving

Table II

The quantity of clicks and impressions for each type of content displayed in Google Discover for Website1

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 %

Table III

Number of clicks and impressions for each content type listed in Google Discover for Website2

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%

Table IV

Number of clicks and impressions for each product type listed in Google Discover for Website2

Product type	Content pieces	Clicks	Impressions	CTR
Smartphone	105	12008	233485	5.14%
Smartwatches & Smartbands	29	3884	78418	4.95%
Headphones	16	1880	38224	4.92%
Laptops	7	3296	41519	7.94%
Graphics Cards	34	45918	1060020	4.33%
Processors	12	981	18138	5.41%
Motherboards	6	235	6372	3.69%
Consoles	5	484	7718	6.27%
Drones	7	1183	31081	3.81%
Action Cameras	4	189	7535	2.51%
Total	225	70058	1522510	4.60%

Figure 1

Click-through rate numbers for the same content type displayed in Google Discover

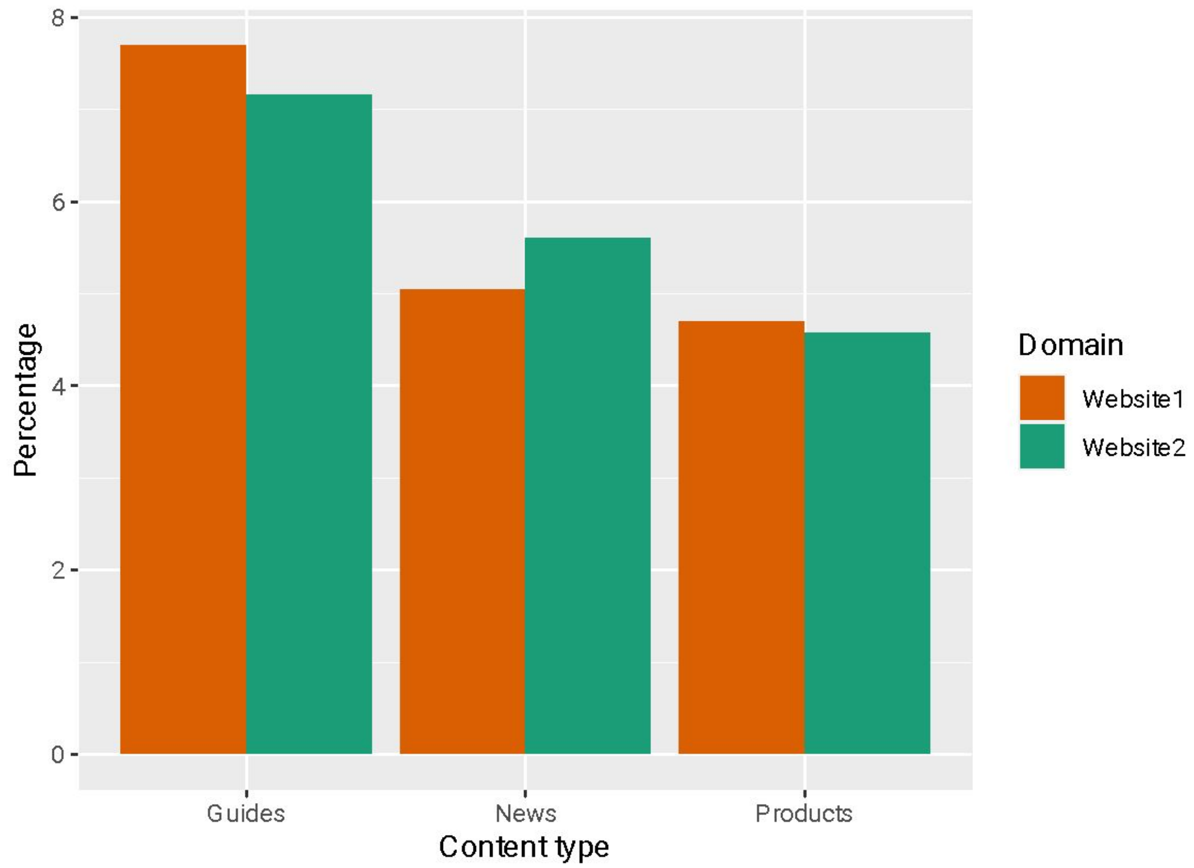

Figure 2

Time series of CTRs for two Domains

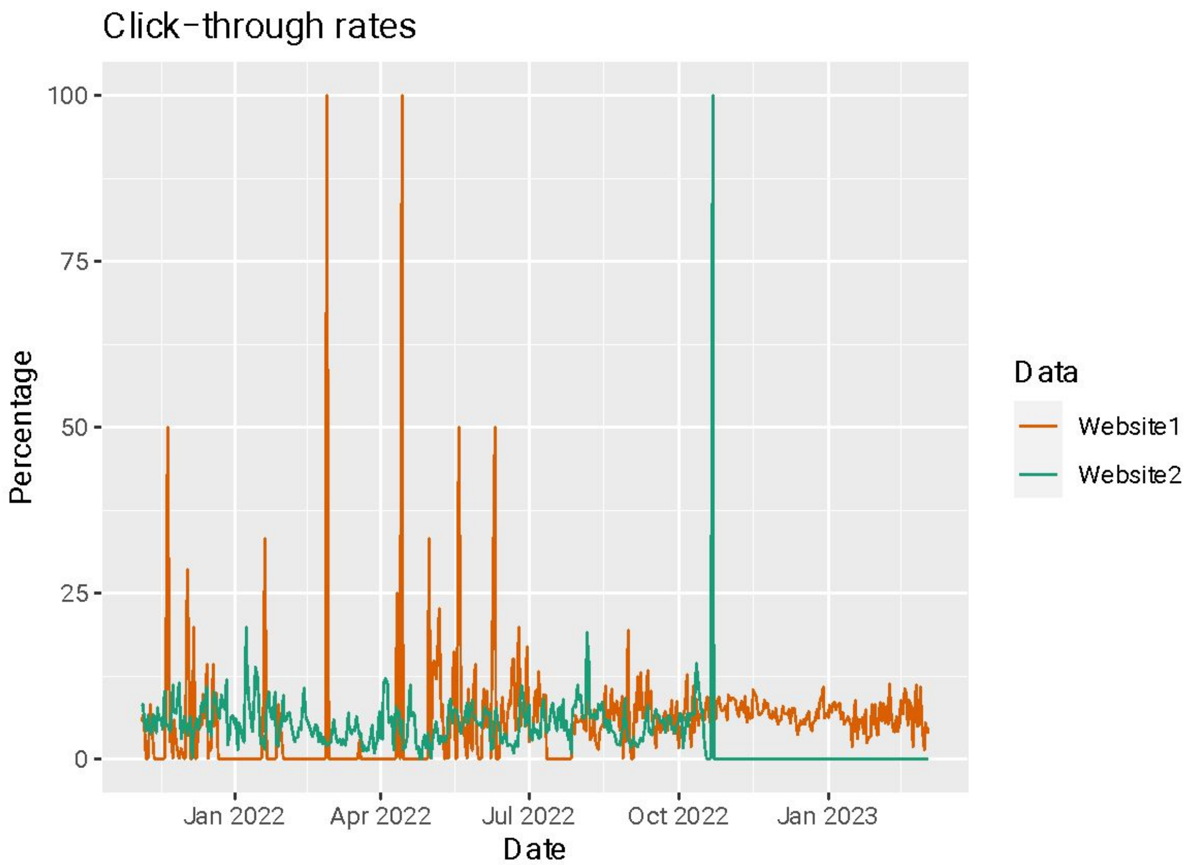

Figure 3

CTR values for Website2 by countries