THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 3. Issue 1 (2023)
https://doi.org/ 10.62019/abbdm. v3i1.59

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/
ISSN (online): 2959-0809

Data-Driven Exploration of Web Browsing Habits: A Visual Analysis with BHVis

Abdul Qayoom, Shafiq ur Rehman, Muhammad Waqas, Muhammad Aoun, Umair Saeed, *Wu Yadong, Wang Song

Chronicle

Article history
Received: October 28, 2023
Received in the revised format: Nov,5 2023
Accepted: Nov 20, 2023
Available online: Nov 26, 2023

Abdul Qayoom is currently affiliated with School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, P.R. China, and Department of Computer Science, Lasbela University of Agriculture, Water and Marine Sciences, Baluchistan 90150, Pakistan.
Email: aqbuzdar@luawms.edu.pk

Shafiq Ur Rehman is currently affiliated with Mir Chakar Khan Rind University of Technology, Dera Ghazi Khan Pakistan, and Lasbela University of Agriculture, Water and Marine Sciences, Baluchistan 90150, Pakistan.
Email: shafiqbaloch@gmail.com

Muhammad Waqas is currently affiliated with School of Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
Email: m.waqas@std.uestc.edu.cn

Muhammad Aoun is currently affiliated with Department of Computer Science and Information Technology, Ghazi University, Dera Ghazi Khan, 32200, Pakistan
Email:muhammadaoun151@gmail.com

Umair Saeed is currently affiliated with Department of Computer Science, Bahria University, Islamabad Campus, Pakistan
Email: umairsaeedmixit@gmail.com

***Wu Yadong is** currently affiliated with School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, P.R. China and Sichuan University of science and engineering , Zigong, P.R. China,
Email: wyd028@163.com

Wang Song is currently affiliated with School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, P.R. China,
Email: wangsong@swust.edu.cn

***Corresponding Author**

Abstract

Browsing history is an eminent tool for internet users to keep track of the daily and specially visited webpages without bookmarking them. Traditional history tools have become insufficient to satisfy their purpose of addressing complex browsing patterns. Graphical interfaces are the critical solutions for these complexities. Therefore, visualising the browsing histories interactively can be helpful in this case. There are many built-in basic visualisation browsing icons for the ease of the users in different browsers, which help the user quickly access the visited pages. But these tools lack in-depth analysis of users' browsing habits. In this study, we propose a novel approach for visualising the browsing history data and browsing habits of the user. We named this system BHVis (Browsing Habits Visualisation), which visualises the user's web page visits using different visualisation methods and visual representations to make the browsing habits of the users explicable. The main contribution of this system is to enable users to gain insight into their browsing patterns and enable self-analysis and self-improvement. The system is evaluated based on different case studies. The study shows that this system enables users to gain insights into their web usage and monitor their web browsing habits.

Keywords Information Visualization; Browsing Habits; Interactive Visualization; Visual Analysis; Browsing History.

INTRODUCTION

The internet is inevitable in today's era of technology due to competence, inventiveness, socialisation, and globalisation, so web browsing is now very common for most people in the world. Browsing history records are a significant source of information about a person's individual and on-the-job interests (Bilezikjian et al., 2022). All top browsers only provide a basic implementation for web surfing and checking already accessed sites; existing built-in history implementations are obstinate in assisting individuals to find previously visited links (Won et al., 2009). Due to their simple interface, the web history entries have no charm for the user. However, it has been of greater interest to researchers to find different patterns based on temporal visit trajectories. An efficient examination of browser histories allows one to retrospectively study users' behaviour on the Web (Cernea et al., 2013).

Browser history is being used in different research and development applications. Tziortziotis, Nikolaos, et al. (2021) developed a model that uses web history for advertising campaigns. (A Mingxiao et al., 2019). used browsing history. For news recommendations. Sun, Wei, et al. (2019) proposed a point-of-interest intelligent search method based on the user's browsing history, while the authors of Pretorius (2017) proposed a study for user attributing based on the semantic properties of web browser history into digital forensics.

Due to the rush of websites in the internet world now, it's almost impossible for a user to remember the web links of most of the sites. As a result, users always rely on search engines to find the websites they need. However, users may unintentionally fall into adverse patterns that may be destructive for themselves or their efficiency (Carrasco et al.,. 2019). Excessive usage of non-work-related websites may divert users from their set goals and cause rapid behaviour changes over time. Web browsers keep history data, but they don't enable users to benefit from this data to understand their habits. A suitable model for the analysis of browser history and its correlated intrinsic user interactions is essential to providing a helpful and creative user experience (Khaksari, G. H. et al., 2011). Visual analysis assists users in interacting with computer applications through state-of-the-art visualisation tools, which make the data analysis process simple and efficient.

Visual illustrations of web browsing history help users learn their browsing habits (Menchen-Trevino et al., 2016). Interactive visual representations of web browsing history can assist users in analysing their behaviors. It can help them self-improve because visualisations can improve the time needed to inspect and analyse data substances (Reiss, S. P., & Eddon, G. 2005). Numerous approaches have been developed to visualise and analyse web browsing history data. In this paper, we present a novel approach for web history visualisation, and we present a prototype called BHVis; this system includes state-of-the-art interactive visual designs to make the history records more attractive and understandable for users. This system has two parts: to visualise user data interactively and to analyse user data to know the user's web surfing behaviors. The main contributions of the system are:

- This system offers a combination of interactive visualization designs enhanced with new features for visually aided information finding and sense-making from web history data.
- This system focuses on better categorization of visited web pages to make the results better and more accurate to assist users in finding visit patterns.
- This system facilitates the user with a visual analysis report for temporal visits and helps the user to understand web visiting habits and working behavior.

LITERATURE REVIEW

Both regular users and behavior analysts are interested to know what types of Web sites are visited by browser users. To do so, websites should be categorized, and single visits should be aggregated according to those categories (news/e-commerce/etc.).An eminent example of user activities visualization was developed by Reiss and Eddon (Van Kleek et al,. 2010). They created a tool named Webviz to monitor user behavior in real time. Webviz gathers data from a large number of users, monitoring the URLs that they access. Then, it summarizes this information by categories of Web sites (provided by the Open Directory hierarchy) and displays the result. Users can identify browsing patterns, trends, or peaks of unusual activities. (Van Kleek et al,. 2008) assumed that Web browsing trials reflect the interests of users and what they do in their daily lives.

These trails have the potential to help users in various ways, e.g., to keep track of how users spend their time. To exploit this potential, they developed an Eyebrows tool that provides quick access to the individuals' browsing activities and presents trends aggregated by various time intervals. Vartiainen et al. (2014) developed a solution called Rolling History for mobile devices with four navigation control directions and graphics acceleration hardware. RescueTime (Whittaker et al,. 2016) is a recently developed visualization tool offered commercially for users' facilitation to track their web visits, define objectives, block undesirable websites, and obtain results for getting or not to their aims. However, these systems can be unproductive due to dependency, which may lead to instability for users' behavior change efforts by abandoning the application and consequently disengaging with behaviors. Similarly, Kim et al. compared positive and negative framing to yield feedback and segregate desired and undesired behaviors.

meTime demonstrates the latest (past half hour) application and website usage in a stubborn, on-screen window. Whittaker et al. found that users condensed the time consumed on non-work activities when using mime, even though users did not have to set goals or define unproductive apps. MTVIS (Islam, M., & Jin, S. 2019) is a visualization tool developed for detecting fraudulent transactions in money transfers, which can be used in browsing history for analysis. E. Menchen-Trevino proposed Web Historian (Menchen-Trevino, E. 2016) as a Chrome extension to collect users' web browsing history data for studying actual world web-behavior. He introduced some impressive visual designs to enable users to explore the web using habits. Still, this tool was just introduced to collect user data as a part of the research project, so this tool did not entirely focus on user-oriented functionalities to provide behavior analyses. PopHistory (Carrasco et al,. 2013) is the latest tool developed for browsing habits, which incorporates animations to promote reflection on browser history habits. Through a formative study from participants, M. Carrasco et al. found that their tool is efficient in viewing and tracking trends in web usage histories compared to previous and traditional history list views.

Different visualization methods, charts(Herman et al,. 2018), graphs(Zhu et al,. 2021), scatter plots(Li, G et al., 2020), and trees(Pei, W et al., 2018) are used to visualize the extensive textual data and understand the insights of data. Bvis(Xu, L et al., 2021) contains novel visualization methods for traffic states and pattern analysis. Many new visualization tools, such as LogCanvas , have a graphical user interface to support researchers for collaborative web searches. All the above visualization methods are beneficial to make the history items understandable. Still, the core dimension that all previous research has not wholly addressed is to make the visualizations simple and detailed to enable the user to self-analyze online behavior as a part of daily life. Our visual designs are unique from all previously developed prototypes; our system can cover all user visit information and is the most straightforward presentation to assist. Our system allows users to find different kinds of temporal similarities and behavioral differences based on the available data.

DATA RETRIEVAL AND PROCESSING

The system queries and processes the data directly from a web browser (Google Chrome browser) [24]. The data returned by the browser contains basic user visit details and visit counts of each unique URL. This visit data is then processed based on three different types of filtrations. First, filtration is done concerning visit time. The data is grouped into different visit time spans, from all-time data to hourly data like all visits, monthly visits, weekly visits, daily visits, and hourly visits. Second, filtration is done concerning website categorization. Each website is categorized in a specific type according to its domain name and content. We used twelve main categories in this system; if some website doesn't fall into any category, it goes to a general category named "other," which means it is an uncategorized list.

Third, filtration is done concerning visit counts. All visit counts of similar websites are added and grouped to show visually to the user.After data processing, color encoding is essential for the visual presentation of data. Each category in the system has a pre-defined color, and the same color is used everywhere for the categories and websites belonging to those categories. The second group of colors is used for months; each month is encoded in a different color for user facilitation.

Bhvis: System Implementation

BHVis is implemented using the d3.js JavaScript library. This system is intended to facilitate individual users' understanding of online visiting behavior. Users can install this system as a Google Chrome extension , available for demonstration on the Chrome web store. The core of the system is to get the user's web history data from Chrome, process it accordingly, and then visually present it to the user using different charts and illustrations.

- *Visual Designs*

Visit Timeline ArcView fig-1 is the first and foremost visual design in our system, which shows the visit items in a timeline, where inner arcs and visited websites present time in the shape of months, weeks, days, and hours are shown by small colored arcs. Each small arc presents a single website, and its color is according to the color of the category to which this item belongs. This ArcView timeline is the leading interactive control design for filtering data by selected period. Other related visual designs are updated as the user applies time filtration to this visual design.

Category Bubble Chart fig-2 **is** the second visual design in our system, which shows the list of categories containing the visited websites. The bubble size shows the number of websites in that category; legends are also provided for users to understand the categories' color encoding. The information about which category occupies what percentage of your visits is essential for the user's understanding, so this visual design also provides this information to assist the user. On the bubble mouse over, the user can see the name of the category and the number of websites belonging to that category. This chart is also helpful for users to find the most and least visited websites.

Top Sites Radial Bar Chart Fig. 3 is the third visual design in our system. In this chart, we present the top ten visited sites of the user in a novel circular visual design, where the color of the upper layer of the circle is of the category to which the website belongs. Legends are used to show the website name and category (name and color are shown). On the circle mouse over, the user can view the number of visits to each website. This chart is helpful for users' understanding of essential visit habits.

Month-wise Visit Radar Chart fig-4 presents web visits belonging to different categories grouped by month. The polygon area is a pattern of user visiting behaviors in a timeline. This chart is handy for users to view and track their monthly visiting habits.

Day-wise Websites Stream view fig-5 presents the frequency of visited websites in a timeline where the axis shows the time as single day slots and the axis depicts the number of websites accessed. The tides of the stream presented by the colors of categories show the number of websites visited by each category on a single day. This chart is fundamental for users to view the visit frequencies on different days, enabling users to detect the busiest and most free days and to view the patterns of visits on websites of different categories.

Day-Wise Visits Sunburst fig-6 lets users view the number of visits on each website per day under some specific category. The sunburst's inner level depicts the category of websites; the second level depicts the visit months of the relevant category; the third level presents the day of the month, and the fourth level shows the websites and the number of visits to those websites. This visual design is helpful for the user to know about daily habits and the days containing the most and least minor visits, which are helpful for the self-improvement of the user by analyzing visiting patterns.

- **Interactions**

BHVis system offers complete interactivity to users. All visual designs are equipped with all related interactivity functions, e.g., showing tooltips on the mouse containing information details. The system's first four visual designs are interconnected and updated as the user applies the filtration by time, category, or website. These interactivity options assist users in having a complete insight into history data.

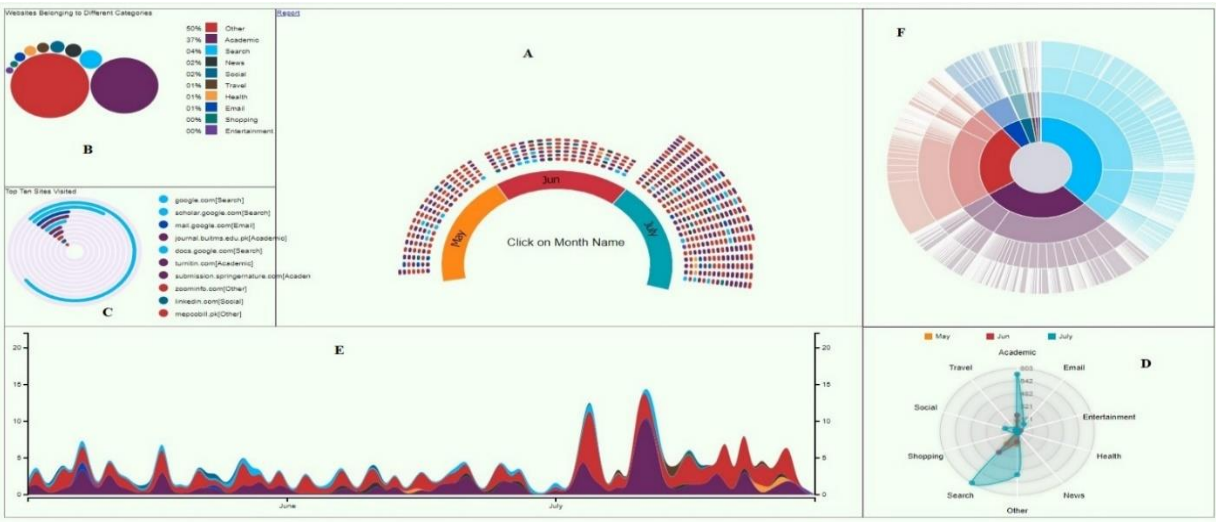


Figure 1.
Main dashboard of BHVis, (A) Shows visit timeline ArcView, (B) shows website categorization bubble

chart, (C) shows the top visited sites in a Radial Bar Chart, (D)shows month-wise categorical visits in Radar Chart, (E) shows daily categorical websites in Stream View Timeline, and (F) shows the daily categorical number of visits to each website.

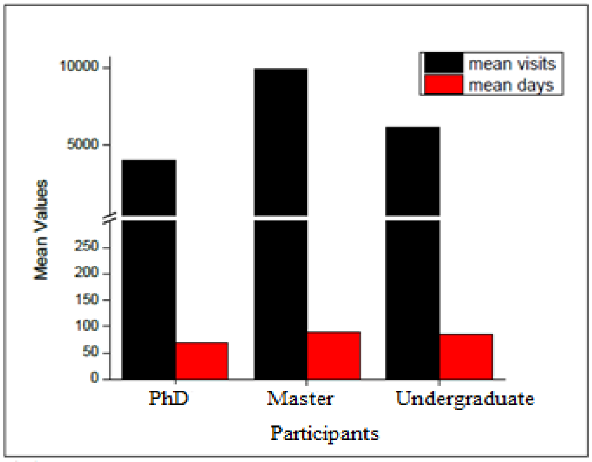
Self-Analysis

We added a self-analysis page for a specific group of users to verify the importance of visual designs used in this system. This analysis page is added for academic users only. The visual design used to show analysis results is shown in Fig. 3. This page shows the user's visit results in three different measurements. First, we divided the visits into work and non-work, where the former shows the valuable and productive visits. In contrast, the latter shows unproductive or useless visits compared to academic users' domains. Second, we divided the visits into early mornings, afternoons, evenings, nights, and late nights. We showed user visit percentages for every part of the day as work and non-work visit percentages. Third, we showed the user two bar charts to enable him to view the number of websites visited from each category and the number of visits from each category.Comparison of these two charts lets the users know about the most visited category because sometimes users have accessed more websites from a category but fewer visits and vice versa. The logic behind the work and non-work categories we used for academic users are as follows: we added academic, news, email, and search categories to the work group while others to the non-work group, as we discussed already; this analysis page is developed for academic users only to enable them to self-analyze their web visiting behavior. While designing this page, we kept in mind to analyze three main factors (F) of the user's web-using behavior: F1, to what extent the user uses the web; F2, the productivity or work attitude of the user; F3, is the user's visiting habits. In the next section, we will present a case study to analyze these three factors of user online behavior. This case study will help to verify the importance and usefulness of our **Figure 2.** (A) Shows the working behavior of the user, (B) shows the visit radar at different times of day, (C) shows the day visits in a tabular form showing total visits and the working habits measured in percentages, (D) Shows the number of websites accesses from all categories and (E) shows the number of

Figure 2.

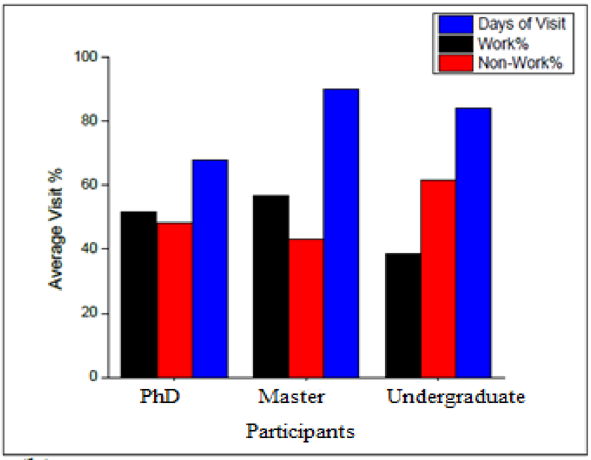
Case Studies

This study serves two purposes: (a) to provide real examples to showcase the usefulness of our system and (b) to show the self-analysis user data as evidence of behavior analysis. This study will focus on the real-world demonstration of three factors of behavior analyses we discussed in the previous section. We collected the self-analysis result page data from fifteen participants who are students at our university. We divided the participants into three different categories (PhD, Master, and Undergraduate students) to provide a comparative study of behavior analysis. We recruited five volunteer participants from each category, and they voluntarily shared their web visit results page data with us for research purposes. Among these participants, most were aware of the web history data, but only some knew about the importance of history data. Before collecting results page data, we briefed all the participants about the importance of browsing history data for understanding their online behaviors. The BHVis prototype system was added to Chrome browsers on all participants' machines. We will present three cases here to demonstrate the previously discussed factors of web browsing behavior analysis.



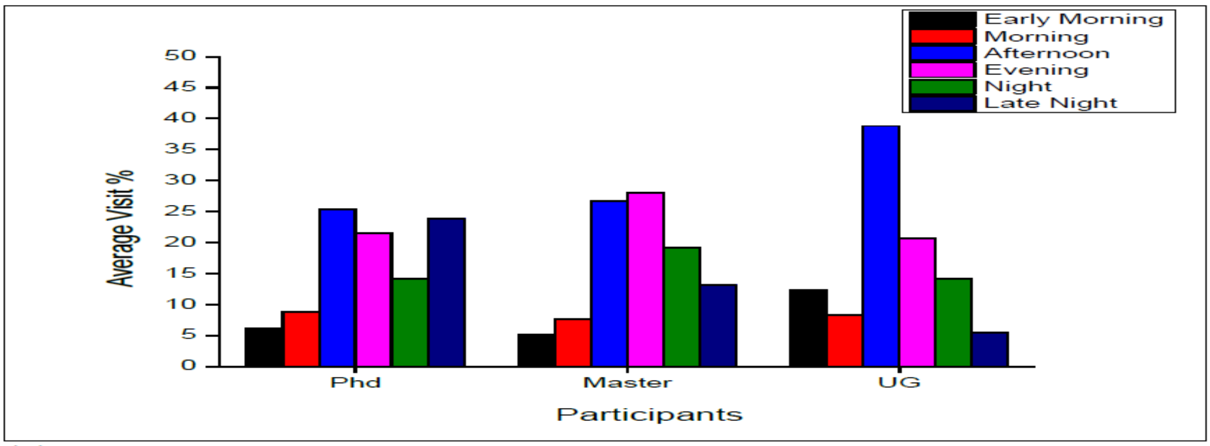
(a) Web Usage

Figure 4.



(b) Productivity

Figure 5.



(c) Habits

Figure 6.

Case-1 Web Usage

Chart (a) shows average visits about the average days of visits of all three types of participants. This chart is a comparative study of different categories of users, which helps understand the web using behavior factors. This comparative study shows that according to collected data, an average Master's degree student spends more time on the web than Undergraduate and Ph.D. students and Ph.D. students spend the least time on the web compared to the other two categories of participants. These results are beneficial for studying the behavior comparison of different participants. Users who use the web more can be categorized as less social and less outgoing, and others can be more socially outgoing. Extensive web usage is also bad for health; this habit can keep users away from book readings and collaborative learning.

Table 1.
Average of all visits related data collected from undergraduate, masters and PhD participants

Participants	Total Visits	Work	None-work	Days of Visit	Early Morning	Morning	Afternoon	Evening	Night	Late Night
U_{avg}	6188	39	61	84	12	8	38	21	14	6
M_{avg}	9866	57	43	90	5	8	27	28	19	13
P_{avg}	4033	52	48	68	6	9	25	22	14	24

Case-2 Browsing Habits

Chart (c) shows average visits at different times of the day of each group of participants. The analysis in this chart depicts the comparison of users' browsing habits. According to test data results, afternoon and evenings are the parts of the day where participants spend more time on the web. According to the chart, undergraduate and Ph.D. students' favorite time for web browsing is the afternoon, while for master students, it's the evening. These statistics are handy to define the good and bad habits of the users. Users who use the web more in the mornings, afternoons, and evenings can be categorized as those with good habits and vice versa. Because the night is the time to spend with friends and family, eat dinner, and go out late, the night is the best time for sleeping, and early morning is best for having morning walks and refreshing activities instead of using gadgets or browsing the web.

Case-3 Productivity/Working Attitude

Chart (b) compares the average productivity/work attitude of all three types of participants. This comparative study shows the work and non-work time percentages along with days of visits, which helps understand the working behavior of users on the web. This chart shows that master's degree students, on average, are more productive than Ph.D. and undergraduate students. Therefore, this chart defines a comparative pattern of productivity between different groups. This analysis is enough to define the productivity factor of web-using behavior. Another significant result is the relevancy of this chart with chart (a); we can see that the more web usage, the more productivity. It means the people who spend less time on the web use it for entertainment and other non-useful things, but those who use it more spend more time on productive web pages according to their relevant professions. The productivity of humans is an essential aspect of life, so if someone spends more time on some tasks, it should result in being more productive; otherwise, it wastes time and resources, which can be an overhead instead of being beneficial. In the three cases described above, we showed that our method of visualizing browsing history data is valuable and advantageous to studying users' online behavior and detecting the behavioral patterns of different users to understand their web-using habits. The above analysis shows that these three factors (web usage, browsing habits, and productivity) are critical to studying the online or digital behavior of the users, and a visualization system is an essential aid for such studies. These results are also beneficial in validating the history data visual designs and self-analysis result designs we used in our prototype to demonstrate our concept.

CONCLUSION

In this work, we proposed novel visual designs for browser history data visualization. This system enables users to self-analysis and self-improvement through visual aids and statistics. We introduced a BHVis prototype for interactive exploration of the browsing history of the users and a self-analysis method for the users to enable them to study and improve their web browsing behaviors. In the future, this concept can be more broadly used to enhance the system for different professionals and organizations. This concept can be implemented in a client-server architecture to design a network traffic monitoring system that can analyze the web using the productivity and behaviors of the network users of any organization.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

" 'RescueTime." [Online]. Available: https://www.rescuetime.com/.
"Data Driven Documents." [Online]. Available: https://d3js.org/.
"Google Chrome." [Online]. Available: https://www.google.com/chrome/.
"Chrome Web Store." [Online]. Available: https:// chrome.google.com/webstore/category/extensions.
An, M., Wu, F., Wu, C., Zhang, K., Liu, Z., & Xie, X. (2019, July). Neural news recommendation with long-and short-term user representations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 336-345).
Bilezikjian, M., Tang, J. C., Begole, J. B., & Yankelovich, N. (2002, April). Exploring web browser history comparisons. In CHI'02 Extended Abstracts on Human Factors in Computing Systems (pp. 828-829).
Carrasco, M., Koh, E., & Malik, S. (2017, May). pophistory: Animated visualization of personal web browsing history. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2429-2436).
Cernea, D., Truderung, I., Kerren, A., & Ebert, A. (2013, February). WebComets: A Tab-Oriented Approach for Browser History Visualization. In *GRAPP/IVAPP* (pp. 439-450).
Herman, I., Melançon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. IEEE Transactions on visualization and computer graphics, 6(1), 24-43.
Islam, M., & Jin, S. (2019, November). An overview of data visualization. In *2019 International*

Conference on Information Science and Communications Technologies (ICISCT) (pp. 1-7). IEEE.
Khaksari, G. H. (2011). Direct manipulation of web browsing history. International Journal of Computer Science & Information Technology, 3(5), 185.
Li, G., Tian, M., Xu, Q., McGuffin, M. J., & Yuan, X. (2020, April). Gotree: A grammar of tree visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).
Menchen-Trevino, E. (2016). Web historian: Enabling multi-method and independent research with real-world web browsing history data. IConference 2016 Proceedings, 10, 16611.
Pei, W., Wu, Y., Wang, S., Xiao, L., Jiang, H., & Qayoom, A. (2018). Bvis: urban traffic visual analysis based on bus sparse trajectories. Journal of Visualization, 21, 873-883.
Pretorius, S., Ikuesan, A. R., & Venter, H. S. (2017, November). Attributing users based on web browser history. In 2017 IEEE Conference on Application, Information and Network Security (AINS) (pp. 69-74). IEEE.
Reiss, S. P., & Eddon, G. (2005, September). Visualizing what people are doing on the Web. In 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC'05) (pp. 305-307). IEEE.
Shaikh, z. A., qayoom, a., rehman, s. U., khan, a. A., ousmane, b., makar, s. V., ... & Chapin, a. L. Mtvis: a framework for visual analysis and exploration of mobile money transactions.
Sun, W., Cong, Y., Liu, X., & Li, C. (2019, May). A Point of Interest Intelligent Search Method based on Browsing History. In GISTAM (pp. 189-192).
Tziortziotis, N., Qiu, Y., Hue, M., & Vazirgiannis, M. (2021, July). Audience expansion based on user browsing history. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
Van Kleek, M., Moore, B., Xu, C., & Karger, D. R. (2010). Eyebrowse: Real-time web activity sharing and visualization. In CHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 3643-3648).
Vartiainen, E., Roto, V., & Kaasalainen, J. (2008, June). Graphical history list with multi-window support on a mobile web browser. In 2008 Third International Conference on Internet and Web Applications and Services (pp. 121-129). IEEE.
Whittaker, S., Kalnikaite, V., Hollis, V., & Guydish, A. (2016, May). 'Don't Waste My Time' Use of Time Information Improves Focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1729-1738).
Won, S. S., Jin, J., & Hong, J. I. (2009, April). Contextual web history: using visual and contextual cues to improve web browser history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1457-1466).
Xu, L., Tolmochava, T., & Zhou, X. (2021). Search history visualization for collaborative web searching. Big Data Research, 23, 100180.
Zhu, J., Ran, J., Lee, R. K. W., Choo, K., & Li, Z. (2021). AutoChart: A dataset for chart-to-text generation task. arXiv preprint arXiv:2108.06897.