

Understanding User Search Behavior of Humkinar Urdu Search Engine

Nazish Azam, Hafiz Muhammad Shafiq, Muhammad Amir Mehmood
Al-Khawarizmi Institute of Computer Science, UET, Lahore, Pakistan
University of Engineering and Technology, Lahore, Pakistan
{nazish.azam, hafiz.shafiq, amir.mehmood}@kics.edu.pk

Abstract

Search engines have become inevitable in the current digital information age. Different search engines such as Google, Bing, Yahoo, etc., provide access to the most relevant information present on the World Wide Web to users. These search engines not only require the infrastructure to crawl the World-Wide-Web regularly but also need a framework to gather user metadata to understand user search behavior for improving user experience. In addition, user metadata is required to perform business analytics and digital forensics. User information like IP address, location, type of device, response time, user website activity, etc., help us to know about user navigational pattern. In this paper, we present a user search behavior study of regional search engine called Humkinar Urdu Search Engine (USE) by integrating an open source web analytics application “Matomo”. We collect metadata of Humkinar users for about 35 months. Summary reports generated by the tool show different analyses which can help to effectively monitor the search engine. Furthermore, we present subjective test results and feedback to highlight the preferences of USE users. The analysis and survey can be used to improve the overall performance of Humkinar Urdu search engine in terms of ranking and personalization.

1. Introduction

In recent years, search engines have turned into a significant source of multi-domain data. Our knowledge source has moved from books and papers to web, predominantly because of the way that search engines give a wide variety of relevant data in a couple of seconds [1]. About 98.8% Internet users utilize search engines to get required information [2]. There are many search engines available in different languages for public, e.g., Google [3], Yahoo [4], Bing [5], Baidu [6], DuckDuckGo [7], and many others [8]. “Baidu” is specifically designed for Chinese region, “Yandex” is a well-known search engine in Russia and similarly, there are many other search engines available in different languages.

38% of all Americans use a search engine, 31% read news online, and 30% peruse the Internet just for entertainment. During this online activity, users leave “digital footprints” with their internet service provider

(ISP) or search engine, disclosing their interests [9]. Collection of user information is necessary as government agencies and parties in civil litigation regularly ask technology and communication companies to turn over user data. In Pakistan, out of around 205 million population, about 76% have mobile phone subscriptions, 37 million people are active social media users, and an estimated 22% of the population uses Internet [10]. Other than this, user information helps to improve user experience of website visitors.

Urdu Search Engine (USE) [11], named as “Humkinar”, is a practical step to encourage research in Urdu and facilitates such community who prefers to search and get information in Urdu. On the basis of above discussion, we have used a monitoring tool to make USE better with respect to design, development, content, and ranking. USE team needs to know what their visitors are doing on site, where do they click, what content they read and which links they follow. To attract more people on USE, it is required to make it perform efficiently by giving as much minimum delay as possible.

For website performance improvement, user behavior analysis is an important factor. It shows the interests of the user, and its engagement can be increased by upgrading most visited sections. For this purpose, a large variety of solutions are available as products or services, e.g., Matomo, AWStats, Elogic, Google Analytics, and many more [12]. In most cases, one has to append a small snippet of JavaScript in web pages where user monitoring is required [13]. Also, user activity analysis on a website helps to check the security of a website indirectly. Another important fact to keep in mind is that no one can find out about what your clients need except the clients themselves. So why not ask them? Our aim is to improve the user experience of incoming visitors, that is why we are analyzing user activities and their interests regarding USE.

In this paper, we describe the design, integration, and usage of our user tracking framework. Our main objective includes collecting user tracking details for performance betterment, ranking, and personalization of USE. We use an open source web analytics tool known as “Matomo” (formerly Piwik) [14]. For user survey, we made a questionnaire and got the feedback from 87 users. Our key findings in this study, for last 35 months are mentioned below:

- 23,022 people visit USE and total viewed pages are 117,439.
- Total searched queries are 54,710 out of which 15,694 are single word queries and the most searched query is “Pakistan”.
- About 84.06% of visitors belong to Pakistan and 24.4% used GNU/Linux OS.
- Average page load time of USE is 1.6403s, average network latency is 0.5116s, and average server serve time is only 0.0064 seconds.
- From user survey, we found that 70.1% of users know how to type in Urdu.
- From design point of view, 23% of users gave us 8 points showing a positive impression.
- 59.8% of users said that the design and features of USE are easy to use for searching and reading Urdu content.

The remainder of the paper is organized as follows: Section 2 describes the related work. In Section 3, some tools are discussed which are applied to USE. Section 4 presents design and implementation. In Section 5, we discuss the results obtained from the tool. Next, we present a user survey of USE showing the feedback of users in Section 6. Finally, Section 7 summarizes the whole discussion.

2. Related Work

Famous search engine Google has developed a web analytics application named as Google Analytics. In article [15], a case study has been done using Google Analytics showing prominent features, literature review, real life application of the software and guidelines for the first time users of Google Analytics. Another article states that all search engines track user behavior and recent development shows that search engines try to integrate results from different collections into their results to guide their users for relevant results [16]. This is how users can be guided to quality content based on personalization functionality. In another paper [17], the authors have proposed a new ranking algorithm for user-oriented web page ranking. They did it by tracking the user’s time spent on web page and compare it with Google’s PageRank algorithm. The study made in [18], shows that the authors used AWStats and Google Trends to visualize the statistics comprising of number of unique visitors, page views, keywords, origin of search, and geographic trends.

Eye-tracking analysis of user behavior in WWW search engine has been done which investigates how user interacts with result pages, browsing pattern and views [19]. A quantitative study has been made to explore that how the behavior of the Google users can help web masters to improve their techniques to be in top results on Google [20]. Search engines capture

users’ activities in the search log, which is stored at the search engine server. An interface is proposed and developed by [21] which acts as a layer between Google and the searcher. This framework captures users’ queries before redirecting them to Google.

For large volume of user data, an intelligent system is required to analyze the user behavior and show trend prediction. Discovery of user information allows web based organizations to predict user access pattern and helps in future developments [22]. A methodological framework was proposed in the study [23], which predicts purchase behavior of websites audiences. Instead of targeting individual user interests and activities, they profile websites audiences.

Web server logs provide information like traversal from one page to another, storing user IP address and all the related information. In [24], a study has been done in which authors have found different statistics such as most visited web-pages, user IPs with most visits, and type of errors users have to face, etc., using *WebLogExpert* tool. Similarly in [25], authors have used both web client data as well as web server logs to build an automated data mining and recommendation system for web usage via KNN classification method. User click stream data was obtained via web client and other information such as IP address, user name, server name, etc., were obtained from web server logs.

The analysis of user behavior also helps in building a better recommendation system for users while searching on website. For this purpose, [26] has proposed a new method through semantic enhancement by analyzing web access logs. The

Table 1: Comparison of Google Analytics and Matomo

Feature	Google Analytics	Matomo
Vendor	Google	Matomo
Edition	Single	Self/Cloud hosted
Installation	No	Easy to install
User interface	Easy	Easy
Link to website	Addition of tracking ID	Addition of JavaScript
Addition of plugin	Not allowed	Allowed
Number of users	Limited	Unlimited
Re-marketing integration	Google Ads	None
Data freshness	Not guaranteed	All time
Data	Limited	Unlimited

authors have built three models for this purpose, two of them are for domain knowledge of website and third one is an ontology based model. They have shown that their proposed method enhances the web-page recommendation system and performs better than the most advanced web mining methods, i.e., PLWAP-Mine. Furthermore, [27] has examined web-server logs to find the number of visitors and their behavior to enhance the usability of an educational website. For this

analysis, the authors have used *logExpertLite* tool and found different statistics such as total hits, users, bandwidth usage, unique IPs, etc., for 5 days of the week. In this study, they have discussed how to increase the accessibility and usability of a website from these metrics.

3. Tools

There is a large variety of web monitoring tools available on the Internet like AWStats, eLogic, Google Analytics, ShinyStats, Webalizer, and many others. Here, first, we provide a brief description of Google Analytics and Matomo. Next, we discuss the rationale behind our choice of analytical platform for studying user behavior of USE.

3.1. Google Analytics

It is a service based solution which is provided by Google to track traffic of a website. Free version is perfect for small companies and provides multiple data collection options across websites. Enterprise version is required for integration with Google BigQuery, Salesforce, advanced analysis, and access to raw data. A maximum of 200 number of views per property can be utilized while enterprise solution gives limit to 400 numbers. In order to use it, one just needs a Google account and has to append a small JavaScript code provided by Google Analytics in the footer of web pages. Google Analytics Spreadsheet add-on is available to access and manipulate data using Google spreadsheet. Native re-marketing is done with Google Ads. Google Ads, AdSense, and Search Console are used for native data on-boarding [28].

3.2. Matomo

Matomo (formerly Piwik) is an open source web analytics platform which provides detailed insights about user activities and their engagement on a website. Real-time data updates can be received containing detailed view of visitors and their activities. It also provides row evolution feature which allows to compare current and past metric data for various reports. Page transitions can be seen through it which help to view what visitors did before, and after viewing a specific page. The dashboard of this platform is customizable and can be extended by adding a wide variety of widgets and plugins. Major advantage of this tool is that one has complete control over it as this can be installed on web server side. Using Matomo APIs, data accessibility is easy. Advance reports can be collected by adding manual queries in the database. Adding custom dimensions and settings is another feature provided by Matomo. It gives privacy protection by not sharing user data with advertising companies. It uses database for

archival and storage. Data formatter is used to format the data in presentable format [14]. Many other features of this tool are discussed later in this paper.

3.3. Comparison

Table 1 provides a brief comparison between Matomo and Google Analytics. Although Google Analytics is easy to use and there is no need for any type of installation, but being a search engine website, USE should own the complete user data, privacy and web hosting. Also, there are bandwidth and user limitations while using Google Analytics services. Moreover, it is not allowed to customize available plugins. Due to such restrictions, we have to use Matomo that is an open-source solution and easily customizable

4. Design and Implementation

In this section, first, we briefly discuss USE, its major components, and features. After that, we provide brief description about hosting and dashboard customization of Matomo. Finally, in the end, we discuss integration of tool with USE along with data acquisition and rendering.

4.1. Urdu Search Engine

USE is an Urdu language search engine which can be accessed at www.humkinar.com.pk. USE is comprised of three major components: Cloud Infrastructure (CI), Information Retrieval (IR), and Search Management (SM). CI is responsible for incremental web crawling services, development, testing and deployment of the work. On the other hand, IR performs linguistic and textual analysis on raw content while SM deals with building of indexes for available documents and apply ranking algorithms to present meaningful results to the user. Figure 1 presents a workflow diagram for USE. It has a distributed crawler that crawls and indexes web documents continuously. Customized ranking algorithms are being used to display most relevant and trending results to the user. An adaptable web interface is developed to serve results according to the query of user. For indexing and search solutions, “Apache Solr” is used by USE. Primary source of information storage and retrieval is Apache Hadoop framework. USE has developed their own filters for checking language, age, size and profanity of the documents. It has its own developed summary module to present summarized result according to the query of the user. Another major achievement of USE is that it has given SMS facility to users so that they can get latest and updated news by using SMS facility through their smartphones.

To keep all the above mentioned functionalities safe and updated, there is a dire need to monitor all the activities on USE. Unique requirements of USE include self-

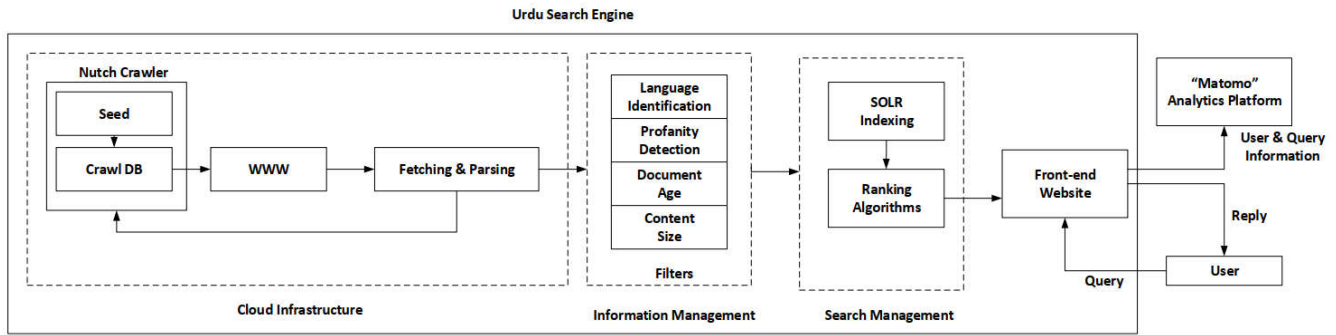


Figure 1: Architectural diagram of Humkinar USE

hosted tool so that it can have total control. Based on these requirements, a monitoring tool is designed for debugging, user behavior analysis, trends, ranking, personalization, and security checking. The next section briefly describes the design and implementation of the tool developed for USE.

4.2. Self-Hosting of Matomo

In our case, we use “self-hosted” approach to install Matomo on our web-server. Before its installation, it is required to make sure that you have a web server, shared hosting or dedicated server. If web server is not available then “Cloud Hosted” Matomo can be used for user analytics. By fulfilling all requirements, we successfully integrated version 3.7.0 of Matomo with

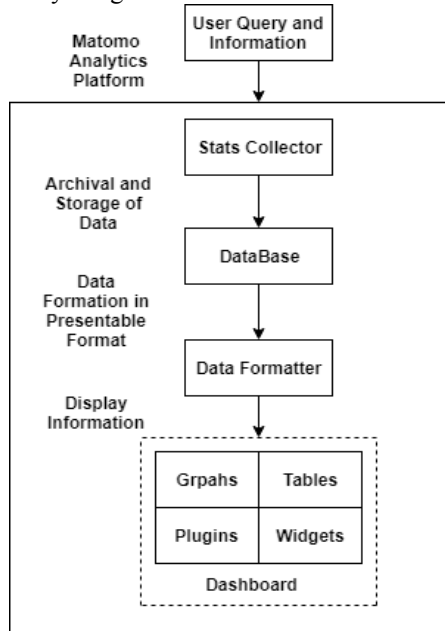


Figure 2: Matomo structure

USE. It has a user-friendly graphical interface which is also customizable. We customized different plugins according to the requirements.

4.3. Dashboard Customization of Matomo

After providing login credentials, dashboard of Matomo can be accessed and there we have quick links to various sections of the analytics tools. The real-time section shows two subsections namely “real-time IP” and “searches”. This is a custom plugin that shows only the summary of currently active IP addresses and searches made. Dashboard is the main analytics section of Matomo which can be customized according to the requirements. Different metrics can be used to track user behavior like evolution over the period, reports, device type, operating system, top searches, best performing pages, visitor logs, out-links etc. Default analytics features of Matomo are somehow limited in their usage. For example, default location provider of Matomo identifies the location of a user based on the language they use which is not very accurate. To tackle this problem, we added GeoIP2

Table 2: Yearly based analytics of Humkinar (October 24, 2016 - October 01, 2019)

Attributes	2016	2017	2018	2019
Total visits	859	4,560	9,661	7,942
Unique visits	244	2,104	3,640	5,390
Total page views	6,948	22,267	71,896	16,328
Total search keywords	3,916	11,101	33,926	5,767
Bounce rate	23%	46%	42%	63%
Total outlinks	227	1,379	7,642	5,747

Table 3: General statistics

Attributes	Values
Total visits	23,022
Unique visits	11,378
Average page load Time	1.6403s
Average time spent by visitor	14 min 21s
Total page views	117,439
Total searches	54,710
Total outlinks	14,995

(PHP version) which uses GeoIP2 database and MaxMind’s PHP API to find accurate location of the user. Another custom analytics feature was added in Matomo which helps us to record the document position. This position is then used for ranking of search results in Humkinar. Similarly, instead of using default

reports, we have used custom reporting APIs, not limited in usage, to get our desired information in JSON or other formats.

4.4. Integration and Data Acquisition

After installation of Matomo on USE platform, a script is provided by Matomo that we append at the footer of those web pages that should be monitored. It logs all activities being carried out on the frontend and sends to back-end monitoring server. For USE, it includes information such as entered queries, click events, number of new and recurring users, IP, browsers information etc. Figure 2 shows a high level view of work-flow diagram for user monitoring at USE. Client enters a query on search engine and information about user and his query is stored in Matomo stats collector. This data is then sent to database for archival and storage. Data formatter converts the received data into presentable format and passes it to web dashboard. User is not disturbed at all in the whole process and he sees only search results on frontend of USE as a reply. Furthermore, in this study, we have analyzed data of October 24, 2016 to October 01, 2019.

5. Results

In this section, we present our findings for user behavior monitoring on USE with Matomo. First section describes yearly based statistics of Humkinar. Then we discuss other metrics like visitor browser, device type, event logs etc. After that, we discuss about the metrics that are very important for search engine websites such as searched keywords, clicks, user Geo-location, and website performance for different sections etc. Table 3 shows general statistics of USE.

5.1. Yearly Based Analytics

Table 2 shows statistics for year 2016 (start from 24 October), 2017, 2018, and 2019 (up till 01 October). For each year, we are presenting attributes and their respective values. Attributes include total visits, unique visits, total page views, total search keywords, and total out-links. The statistics show that total number of visits is increasing every year, i.e., in 2016 total visits were counted 859 and in 2019 total visits count is 7,942. It can be seen that bounce rate is increasing every year as the users are increasing. The reason is that as USE is not only a search engine but a portal as well and provides latest content on its home page. Hence, it is quite obvious that some users just visit USE to read the latest content and leave the page after reading. Overall, these statistics show that USE is getting more attention year by year.

5.2. Visitor Browser

Information about the visitor browser is really supportive for solving the browser inconsistencies. Designers need to keep in mind that cross browser testing is necessary to avoid the most common problems [29]. Hence on the basis of this point, we obtained the information about it to avoid any cross-browser inconsistency. We found that 55.89% of visits are from Chrome browser, so USE developers should pay more attention to this for display of USE. Other browsers include Firefox, Opera, Safari, and others. More than 15 different types of browsers and their types are found in our record while tracking the users of USE, e.g., Mobile Safari, Chrome Mobile etc.

5.3. Device Type

We observe that more than 80% of the users use desktop/laptop to visit USE. Other devices include smartphone, tablet and phablet. This information is really helpful as it suggests to improve the site visibility with respect to desktop devices. Device type information helps to make the website responsive with respect to different screen sizes. It is also possible to show more on large screens and less on small screens.

5.4. Event Logs

These type of logs provide two levels of information, user queries and corresponding clicks on search results. It can be used to know user interest on the

Table 4: Number of unique searches for different tabs

Tab Name	Number of unique searches
Web	3,892
Books	872
Islam	1,148
News	1,035
Poetry	966
Sports	270
Videos	559
Wikipedia	266
Famous websites	187

website e.g., most clicked results and corresponding queries, images, tabs visit etc. Keeping this information in mind, further changes can be made in these sections of website to attract more users. In event logs section, a sample shows that 0.1% of visits contain search term “Pakistan” and clicks on Urdu Wikipedia outlink.

5.5. Site Search Keywords

Matomo also provides searched keywords information for each user. We observe that a total of

54,710 queries are searched and "Pakistan" keyword is at top. We also analyzed the length of searched keywords i.e., how many are single word, two words and so on. Most users search single word query on USE and their total count is 15,694. Similarly, for two-word, three-word, and four-word queries, we have frequency values of 2,036, 1,069 and 548 respectively.

5.6. Website Tabs Usage & Search Statistics

As USE has many sections (tabs) e.g., web, news, poetry, books, etc., here we present the usage distribution of each section. Obtained statistics show that most people visit the home page of USE with about 24% share. Other top visited sections are web, poetry, Islam, news and videos tabs with a share of 13.5%, 11%, 3.2%, 1.9% and 1.1% respectively. These statistics also indicate the interest of users on USE at section level. It also suggests which section should be further improved to increase user engagement. Similarly, we also collect information about number of search queries for different tabs. Table 4 shows unique search statistics in different tabs of USE. We have mentioned the number of unique search keywords for each tab. Out of total searches, 9,195 searches are unique keywords.

5.7. Visitor Log

To analyze the user behavior, we made a visitor log displaying its profile and details as each and every minor information is important to be logged. Table 5 shows the user-level details of a sample visitor. It has

Table 5: Visitor profile attributes

Attributes	Values
IP address	66.249.93.88
Visitor profile ID	1362e2e13b0b8819
Browser type	Chrome mobile
OS type	Android 6.0
Device type	Smartphone, Motorola
Location	United States
Total time spent	3min 34s
Number of actions	5
page views	1

different attributes about the visitor like IP address, user ID, browser type, Geo-location etc. A sample taken from record shows that a user from the United States with IP address 66.249.93.88 visits USE through Android 6.0 using chrome mobile browser in Motorola Smart-phone. He spends 3min 34s on USE and performs 5 different actions. He finds 1 item of his choice and redirects to the respective link. His actions include www.humkinar.com.pk/Poetry, www.punjnud.com and some other outlinks.

5.8. Website Performance Statistics

The performance monitoring of our website with respect to page load time, network latency, and server serve time is also calculated. As it is not affordable to overlook the significance of website load speed because clients who are baffled by a slow page speed are probably going to leave the site. This is why it is important to improve the website load time to enable clients to get where they're speeding up. We found that average page load time of USE is 1.6403s, average network latency is 0.5116s, and average server serve time is only 0.0064s.

5.9. Others

We find that 40 different versions of operating systems like Windows, Linux, Ubuntu, Android, iOS, etc., are used to visit the USE. By analyzing these statistics, we observe that Linux is the most used Operating System (OS) with 24.4% of the total users. We also observe that USE visitors belong to more than 50 different countries with Pakistan at the top position with 84.06% share. Other countries include United States, Australia, India, Saudi Arabia etc. These properties may seem less important but they actually guide the developers to avoid any limitations in their website. Another important information about the user is to find the channel type from where he/she is accessing the site. In our case, we found three channels, i.e., search engine, websites, and social network. It means that users are visiting USE through other search

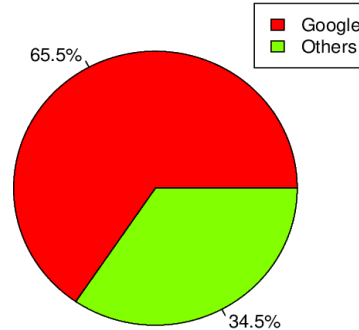


Figure 3: Search platform preference for Urdu content

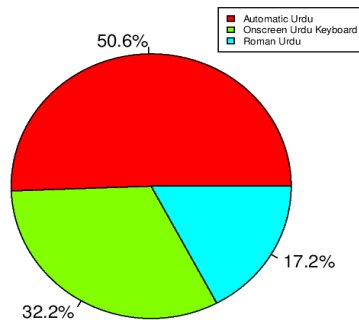


Figure 5: Urdu typing methods

engines, from some website redirection, or from any social network like Twitter, Facebook etc.

6. User Survey for Humkinar Urdu Search Engine

In this section, we discuss the user survey results and feedbacks regarding USE. To observe the user behavior and interest on Humkinar, we conducted a survey in which different questions regarding the features and search results of Humkinar were asked. We got a total of 87 responses from both males and females subjects. Out of the 87 users, 69% were males and 31% were females. Most of them belong to the age of 20-30 as majority of the subjects were students. We asked them to fill the questionnaire by visiting Humkinar and checking the features and functionalities step by step and answer the questions accordingly. It was necessary to ask them about Urdu typing experience as Urdu typing is the key functionality for our search feature. Most of them answered Yes, i.e., 70.1%, while 29.9% answered in No, which shows that majority of users already know how to type Urdu. Figure 3 shows that 65.5% of the users said that they use Google to find Urdu documents while remaining 34.5% use other platforms to search Urdu content.

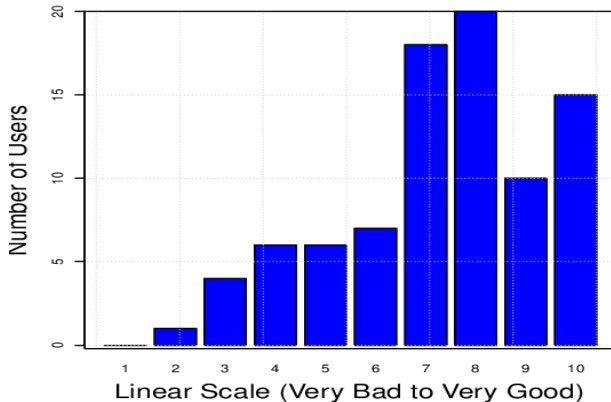


Figure 4: Subjective test results – Humkinar design

From the design and features point of view, we prepared a separate section containing questions related to design view only. To get the overall feedback about design from a user, we used 1-10 linear scale range, i.e., 1 shows very bad and the number goes on to 10 showing very good. Figure 4 shows the chosen values by users regarding design of Humkinar. Majority users, i.e., 23% chose scale value 8. 71.3% users voted that they like the color scheme and presentation of Humkinar frontend. Humkinar uses Nafees Nastaleeq Urdu font and 97.7% users liked its rendering style and readability. For Urdu typing, Humkinar provides three typing methods: 1) Automatic Urdu Typing 2) On-screen Urdu Keyboard

3) Roman Urdu Typing. Figure 5 provides division of users based on the Urdu typing methods. For search results, an individual section was made to ask search result questions for different tabs of Humkinar. 59.8% users said that it is easy to find their required results using this platform, 26.4% selected the option of "Very Easy", and 13.8% of the users found it difficult to search Urdu content using Humkinar.

Overall, the feedback was satisfying as majority of the responses were positive. We also got comments from each and every user at the end of questionnaire and many useful suggestions were given by them, e.g., add more sections like cooking, health, horoscope, currency rates, biography page for famous personalities etc. Some of them proposed that we should also add voice search option to find query results. We can conclude that the overall survey feedback was good enough to implement new functionalities in Humkinar for the ease of users and to make it more adaptable.

7. Conclusion

In this study, we analyzed Urdu Search Engine (USE) user behavior and obtained different statistics. For this purpose, we have used open-source solution "Matomo" and customized it according to our requirements. With this tool, we have analyzed last 35 months user search behavior on USE. For this interval, our findings show that USE is visited 23,022 times and total page views are 117,439. Total searched queries are 54,710, top query is "Pakistan" and most search queries are single word query (15,694). About 84.06% visitors belong to a single country, i.e., Pakistan and most of them used Chrome browser (55.89%) with Linux (24.4%) OS. While loading the USE website, total load time is only 1.6403 seconds. By incorporating click information of visitor for search query, we updated ranking algorithm of search results. Further, we presented user survey results, total 87 participants, regarding USE design, content, and features. It was found that 65.5% users use Google to search Urdu content. 71.3% users liked the interface of USE. Overall feedback is agreeable and it is helpful for us to improve the quality of USE with respect to design, features, and content. In future, we plan to use "Matomo" stack personalization" to implement personalization feature in Humkinar for enriched user experience.

8. Acknowledgement

This research work was funded by Higher Education Commission (HEC) Pakistan and Ministry of Planning Development and Reforms under National Center in Big Data and Cloud Computing.

9. References

- [1] Mike Cafarella and Doug Cutting. Building nutch: Open source search. *Queue*, 2(2):54, Jan 2004.
- [2] Daniel C. Fain and Jan O. Pedersen. Sponsored search: A brief history. *Bulletin of the American Society for Information Science and Technology*, 32(2):12–13, 2006.
- [3] Google. www.google.com/ (visited on 30 Sep, 2019).
- [4] Yahoo. www.yahoo.com/ (visited on 30 Sep, 2019).
- [5] Bing. <https://www.bing.com/> (visited on 30 Sep, 2019).
- [6] Baidu. <https://www.baidu.com/> (visited on 30 Sep, 2019).
- [7] Duckduckgo. <https://duckduckgo.com/> (visited on 30 September, 2019).
- [8] Alex Chris. Top 10 search engines in the world, 2018. <https://www.reliablesoft.net/top-10-search-engines-in-the-world/> (visited on 30 September, 2019).
- [9] Jayni Foley. Are google searches private-an originalist interpretation of the fourth amendment in online communication cases. *Berkeley Tech. Law Journal*, page 447,2007.
- [10] DataReportal Follow. Digital 2019 pakistan (january 2019) v02, Feb 2019.
- [11] Humkinar urdu search engine. <https://www.humkinar.com.pk/> (visited on 30 September, 2019).
- [12] Rick Tansun. 10 web analytics tools for tracking your visitors, Mar 2009.
- [13] Web analytics: Why they matter — siteimprove (en).
- [14] Free web and mobile analytics software. <https://matomo.org/> (visited on 30 September, 2019).
- [15] Suraj Chande. Google analytics -case study. 01 2015.
- [16] Dirk Lewandowski. Search engine user behaviour: How can users be guided to quality content? 28:261–268, 01 2008.
- [17] Songhua Xu, Yi Zhu, Hao Jiang, and Francis C. M. Lau. A user-oriented webpage ranking algorithm based on user attention time. In *AAAI*, 2008.
- [18] Francesco Brigo, Simona Lattanzi, Michael O Kinney, Nicola Luigi Bragazzi, Laura Tassi, Raffaele Nardone, and Oriano Mecarelli. Online behavior of people visiting a scientific website on epilepsy. *Epilepsy & Behavior*, 90:79–83, 2019.
- [19] L. A. Granka, T. Joachims, and Geri Gay. Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '04*, pages 478–479, New York, NY, USA, 2004. ACM.
- [20] Bartomeu Riutord Fe. User behaviour on google search engine. *International Journal of Learning, Teaching and Educational Research*, 28:104–113, 2014.
- [21] Fadhilah Mat Yamin. Interfacing google search engine to capture user web search behavior. *International Journal of Electronic Commerce Studies*, 4(1):47–62, 2013.
- [22] Xiaozhe Wang, Ajith Abraham, and Kate A. Smith. Intelligent web traffic mining and analysis. *Journal of Network and Computer Applications*, 28(2):147–165, 2005.
- [23] Saar Kagan and Ron Bekkerman. Predicting purchase behavior of website audiences. *International Journal of Electronic Commerce*, 22(4):510–539, 2018.
- [24] N. Goel and C. K. Jha. Analyzing users behavior from web access logs using automated log analyzer tool, 2013
- [25] D.A. Adeniyi, Z. Wei, and Y. Yongquan. Automated web usage data mining and recommendation system using k-nearest neighbor (knn) classification method. *Applied Computing and Informatics*, 12(1):90 – 108, 2016.
- [26] T. T. S. Nguyen, H. Y. Lu, and J. Lu. Web-page recommendation based on web usage and domain knowledge. *IEEE Transactions on Knowledge and Data Engineering*, 26(10):2574–2587, Oct 2014.
- [27] N. Kaur and H. Aggarwal. Web log analysis for identifying the number of visitors and their behavior to enhance the accessibility and usability of website. *International Journal of Computer Applications*, 110, 2015.
- [28] Google Analytics Solutions - Marketing Analytics & Measurement. <https://www.google.com/analytics/> (visited on 01 October, 2019).
- [29] Nepal Barskar and C.p. Patidar. A survey on cross browser inconsistencies in web application. *International Journal of Computer Applications*, 137(4):37–41, 2016.