Privacy-Aware Remote Information Retrieval User Experiments Logging Tool

Hanyu Li, Hongyu Lu, Songhao Huang, Weizhi Ma, Min Zhang*, Yiqun Liu, and Shaoping Ma
Department of Computer Science and Technology, Institute for Artificial Intelligence,
Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China
{lihanyu20, luhy16, mawz14}@mails.tsinghua.edu.cn, huangsonghao@sjhhxm.com
{z-m, yiqunliu, msp}@tsinghua.edu.cn

ABSTRACT

User behaviors and experiences are the fundamental parts of information retrieval systems, but are often difficult to collect, bringing challenges to both applications and research. Recently, researchers have been exploring more fine-grained user behavior than simple clicks, such as time patterns, mouse/scroll patterns, etc., with their own specific laboratory experimental platforms. However, the lack of public available toolkits for logging user behaviors and experiences leads to difficulties on field study of remote user experiments in real scenarios. In this work, we propose a Privacy-Aware Remote User Logging Tool for remotely collecting user behaviors and explicit experience feedback, with a special care for user privacy. With this tool, participants can conduct the user experiments remotely without time and location constraints, giving researchers the possibility to observe users' more natural behaviors and experiences.

CCS CONCEPTS

• **Information systems** → **Information retrieval**.

KEYWORDS

User behavior logging, Privacy protection, Experimental toolkit

ACM Reference Format:
Hanyu Li, Hongyu Lu, Songhao Huang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2021. Privacy-Aware Remote Information Retrieval User Experiments Logging Tool. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3404835.3462793

1 INTRODUCTION

Online interactive information systems, such as search engines and recommender systems, play an increasingly important role in human's daily life. The goal of these information systems is to understand and satisfy the information needs of users. Due to its subjective and implicit nature, user behaviors are widely used as the implicit feedback for their information needs and experiences. Therefore, user behaviors and subjective experiences are keys to build information systems, and have received much research attentions.

The conduct of user experiments, also known as user study, in information retrievals mostly relies on the collection of user behaviors and experience feedback. To this end, researchers usually build their own specific experimental platforms. However, besides of the waste of effects, the lack of unified approaches and tools also brings much risks about the reliability of the results. Meanwhile, current experimental tools concern less for user privacy issues, behavior data, such as queries and browsing, may face the risks of leakage. Moreover, the less flexibility and stability limits the experiments to simple and controllable environment, usually laboratory.

Therefore, a unified, well-designed and privacy-aware recording tool is very important. Through it, the collection of user behaviors and experiences can be consistent across different works, benefiting the generalizability and reliability of research results. Meanwhile, the user experiments can thus be conducted remotely, which is valuable for collecting users' more natural behavior and can also accommodate the time and location constraints that now exist due to COVID-19.

In this work, we propose a Privacy-Aware Remote User Logging Tool for remotely collecting user behaviors and explicit experience feedback, with a special care for user privacy. It automatically collects user's online interaction behaviors, such as mouse movements, time sequences, keyboard inputs, etc., as well as the page contents, with high flexibility and stability. Meanwhile, we design multiple functions to protect users' privacy, including back-checking and encryption mechanisms. After recording, the system will not directly upload the data until the user actively looks back and selects the records.

The tool including user interfaces implemented as the Chrome plug-in for collection and storage of data, and experimenter interfaces implemented as the server for account management and data reception. The tool has several unique features, such as versatility, scalability, anti-interference, and privacy protection. To the best of our knowledge, this is the first publicly unified recording tool for remote user experiments in information retrieval scenario. Source code is available at https://github.com/THUIR/RUS-toolkit

* Corresponding Author

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462793

Table 1: Interaction types collected by the recorder

Behavior Records	
Mouse Events	move, stay, click sequence of (event, timestamp)
Scrolling behavior	scroll (pos_from, pos_to, time)
Statistical Measures	click_cnt, dwell_time, revisit_cnt max_scroll_interval, etc.
Page Records	
Content	html, snapshot
Meta information	url, type, title, etc.
Context Records	
Time	start time, end time
Location	longitude, ip
Device	device, screen size

2 RELATED WORK

User understanding and modeling is one of the major tasks of information retrieval systems. User's behavior patterns have inspired large number of ranking algorithms and evaluation methods. For example, users' clicking on the item have been used as implicit feedback of document relevance [1, 5] and user experiences [6]. The further investigation on user clicking behavior motivate various click models [3, 4]. In addition to the click behavior, other behaviors, like mouse movement [10], scroll information [7], and gaze [2, 8] are also investigated for inferring user experiences.

Traditional studies on user behavior and experience utilize the user experiment methodology (also known as user study) [11–13]. Researchers usually build their own experimental platforms and recruit participants to laboratory environments. Although user behaviors and experiences feedback are collected in most of these works, there are no unified recording tools proposed, to the best of our knowledge.

To help researchers better conduct user studies, adaptable to changing scenarios and web pages, we design a unified general user behavior and experience recorder. Through it, experimenters can design user studies without caring how behaviors and feedback are captured and recorded. The studies can also be extended to remote and more natural scenario.

3 SYSTEM DESIGN

3.1 Objectives and Design Considerations

Our proposed recorder has two main functions, the first one is collecting user behaviors, and the second one is collecting user's explicit experience feedback. In this section, we describe the objectives when we design the recorder.

Behavior Collection: For user behavior collection, the tool needs to be able to automatically record user various behaviors, and maintain a high accuracy and stability. Moreover, to avoid its impacts on user natural behavior and experience, the recording needs to be non-disturbing and simple enough. For remotely recording, there should be reliable data transmission function between the local storage of front end and the database of the back end.

Recorders meeting these objectives can extend traditional laboratory user experiments to remote experiments in users' natural scenarios, allowing the collection of more realistic behaviors and experiences. To this end, we design various functions and mechanisms which will be fully demonstrated in the following sections.

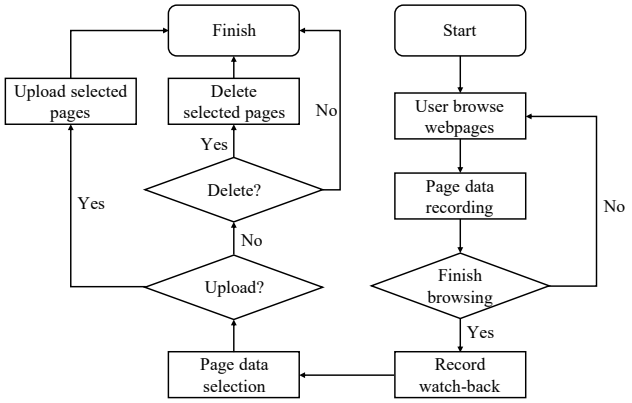

Figure 1: Interaction Mechanism. The flow of user participants interact with the recorder.

Feedback Collection: Different experiments may have different requirements on the experiences and questions to collect. Merely the implicit behavior data may not be able to reflect the user experience[9]. In order to make the recorder more flexible, the feedback content and collection objects should be configurable. Therefore, based on different configuration, the experimenters can customize the description of the user feedback as well as additional data attributes for certain kinds of websites(e.g. search engine result pages).

Privacy Protection: Besides of the function objectives, we also consider the responsibility of our proposed recorder in terms of *privacy*. Recently, user privacy issue in machine learning and information retrieval scenarios gains more and more attention. Existing user experiment platforms can hardly guarantee the privacy of participants because all logged user behavior is compulsively uploaded. We consider to make the collection more controllable for the participants. They can check the results of collections, and decide what can be uploaded, including which behaviors, which pages, which time.

Taking the above objectives into consideration, we design the recorder system and its interaction mechanism, as described in the following sections.

3.2 System Functions

The functions of the system involve two main parts: behavior collection, aiming to record user various interaction behaviors, and feedback collection, aiming to record users' explicit feedback for their experiences.

Behavior Collection: As shown in Table 1, we collect user behaviors in several types, as well as the corresponding page content. Specifically, the record is based on the web page session level, so a record will be saved after the web page is closed or redirected.

- *Behaviors*: Users interact with the web page through mouse events, including *moving*, *staying*, *clicking*, etc. These events are recorded as sequence, as well as time information, i.e.

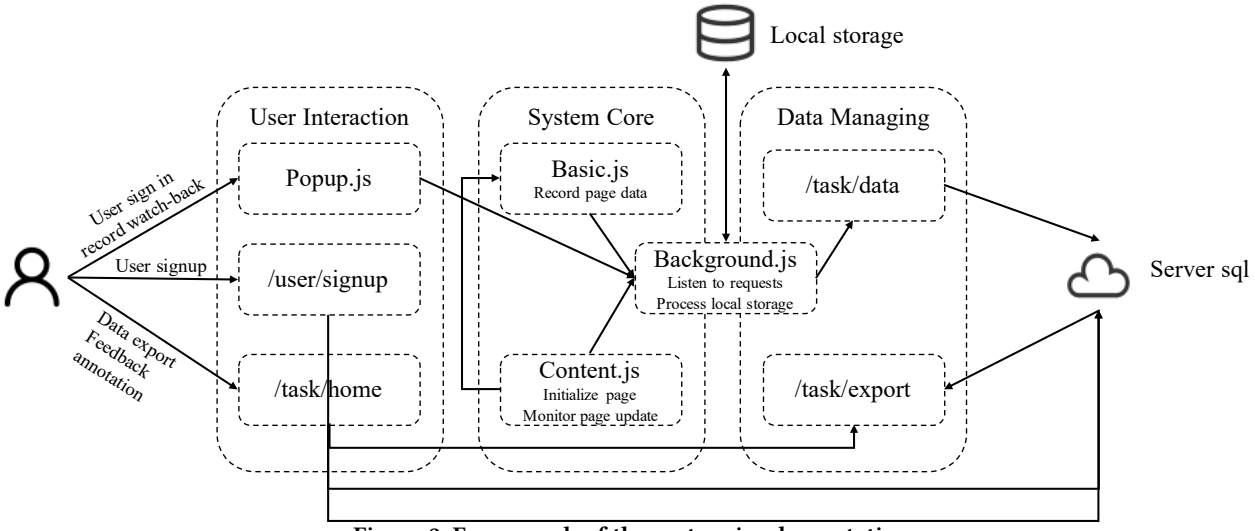

Figure 2: Framework of the system implementation

[(event type, timestamp)]. Meanwhile, user *scrolling* behaviors are also recorded for re-constructing the interaction and linking the mouse event to the page content. Besides of original behavior data, we also generate various behavior statistical metrics to measure its patterns, including number_of_click, dwell_time, max_scroll_interval, etc.

- *Page Content*: To understand user's behavior and experience, the content information of the web page is important. To cover most of the details, we record the *html* content, as well as some extracted important information, such as *url*, *type*, *title*, etc.
- *Context*: Besides of above information, we also record user's context information, such as the *starting time*, the *devices*, the *location*, the *screen size*, the *internet situation*, the *referrer* of the page, etc. Such context information is essential for conducting remote experiment, and can give us more comprehensive and dynamic understanding of user behavior and experience.

Feedback Collection: Besides of user behavior, we also provide the function of collecting user's explicit feedback for their experiences. Recently, various user's subjective experience are studied, such as user satisfaction for the session, user preference for recommended pages, user perceived quality and usefulness of the search results, and are found essential for improving the information system. To meet these potential experimental needs, we design the feedback collection function. Specifically, experimenters can make the configuration for the feedback questionnaires, and the rule for collection, e.g. for the "google.com", show the questionnaire "search experience". The feedback questionnaires will be filled by users in their user homepage of the back-end.

3.3 Interaction Mechanism

Figure 1 shows the interaction mechanism of users view. At the beginning, when users visit the web browser and choose to turn on the experimental recorder, their behaviors (as well as page content and content information) will be recorded automatically, and stored locally after user finishing a web page. User's interaction in single page constructs one record in the data.

At any time, users can click the extension to start the selection phase. The pages and recorded data will be shown as sequence, users can check the content and make two decisions: upload or not, delete or not. If users decide to upload or delete one record, the data will be removed from local storage.

4 SYSTEM IMPLEMENTATION

The overview of the system implementation is shown in Figure 2, including the interfaces for user interaction and data managing, as well as the system core. In this section, we describe the detailed implementation in terms of front-end and back-end. The communication between two parts are conducted by POST web request.

Front-End Part:Web browser is the main way of web users to use information retrieval systems, such as search engine and recommender systems. In research, most of previous user experiments are conducted in web browser. Therefore, we choose to build our recorder via web extension (based on google chrome kernel).

Front-end is responsible for user interaction and management of the local storage. We implement the extension popup interfaces for sign-in, record switching, back-checking functions. After turning on the recording, the logging will automatically be activated. Note that, the record of user behaviors will not be immediately uploaded to the back-end. Instead, it will be temporarily stored in the local storage of the chrome extension for privacy concerns.

As shown in Figure 2, the extension are build by several javascript files. Background.js runs in the background and listens to the requests from other scripts. Basic.js defines the behavior recording function, as well as the message sending function (send the behavior data from newly closed web page to the local storage). Content.js listens to the update of web pages, and correspondingly update or initialize the recorder state. Popup.js responds to the user's interactions on the extension.

Back-End Part: The back-end is build as online server and powered by Django written by Python. The back-end is build for experimenter, and involves two parts of functions and interfaces.

First is the management of experimental tasks, including the

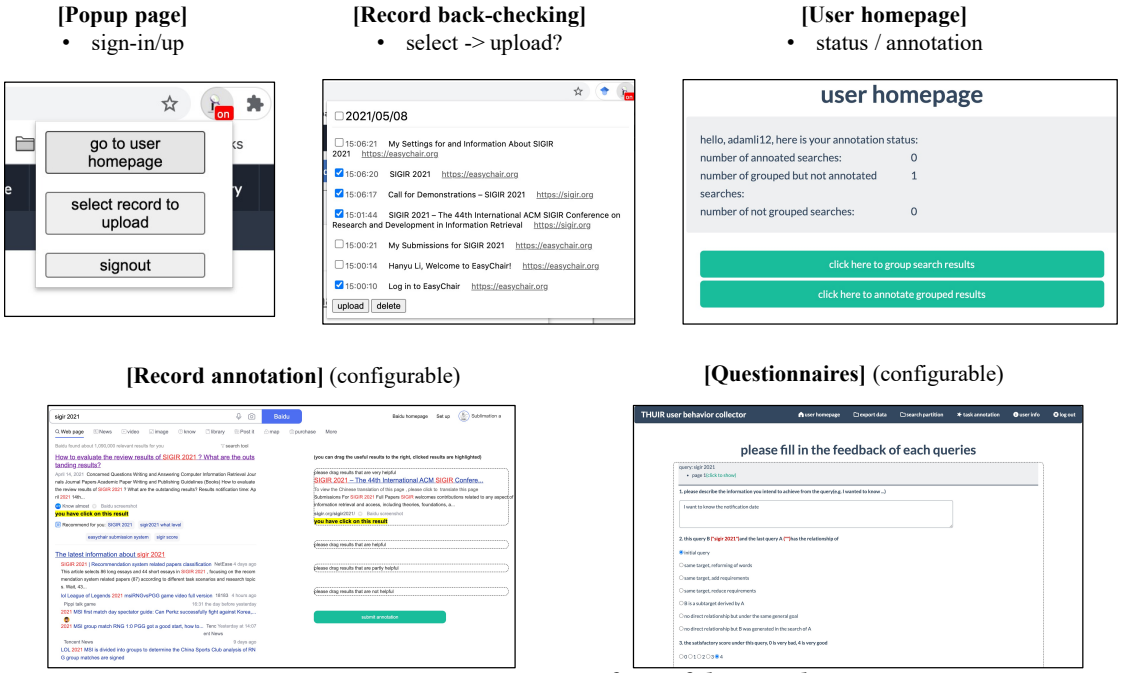

Figure 3: Some main interfaces of the recorder

configuration and state monitoring of the specific experiment. Experimenters can initialize or update the configurations, and checking the received record data. Second is the management of user accounts. Experimenters can track the processes and states of participants.

The front-end part and back-end part are completely decoupled, which means that the user participants can complete the experiment without knowing the saving and uploading of data, as well as the management (done in back-end). The communication between front-end and back-end includes the account events (sign-up/sign-in), and the record data transmission.

5 DEMONSTRATION AND APPLICATION

The screen recording of the demo is submitted with the paper. We use an user experiment in information search scenario as an example to demonstrate the functions implemented by the recorder. Some of the main interfaces are shown in Figure 3.

When the user participants launch our logging tool (extension) for the first time, they will be directed to fill out an user information form to register a new account. They will be informed that all data will be stored locally before they select and upload the data themselves. After that, the participants can follow the instructions to start the experiment. Researchers can design customized record setting by adjusting the js files for several predefined types of webpages. Similarly, explicit feedback questionnaires can also be conveniently modified in Django files. Participants can simply log into the extension and browse the web as usual. At any time, the participants can view their browsing history, and select the history by page or by date. By clicking the home button in the pop-up page, they can view the annotation status and start to annotate.

https://www.youtube.com/watch?v=YFqYYcx-W-Q

Researchers can get the behavior data and explicit feedback of the subjects from the database on the back-end server.

6 CONCLUSION AND FUTURE WORK

To address the problem of inconsistent user behavior and feedback collection methods used by current user experiments in information retrieval scenario, we propose a unified, well-designed and privacy-aware user logging tool. It can remotely record user various interaction behaviors without disturbance, and provides flexible feedback recording configuration. Researchers can easily apply our tool to different user experiments, and even extend them from laboratory to remote. Furthermore, we design multiple mechanisms to protect user privacy, including back-checking strategy.

To the best of our knowledge, this is the first publicly unified recording tool proposed for remote user experiments in information retrieval scenario, and can is believed to benefit user understanding research. As for future work, we consider to incorporate more fine-grained types of user behaviors, for example, gesture and gaze. We also plan to improve the generalizability of the proposed tool by adapting to more platforms.

ACKNOWLEDGMENTS

We'd like to thank Jiaxin Mao, Fan Zhang and Zeyang Liu for their essential contribution. This work is supported by the National Key Research and Development Program of China (2018YFC0831900), Natural Science Foundation of China (Grant No. 61672311, 61532011) and Tsinghua University Guoqiang Research Institute. This project is also funded by China Postdoctoral Science Foundation (2020M670339) and Dr Weizhi Ma has been supported by Shuimu Tsinghua Scholar Program.

REFERENCES

[1] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 19–26.

[2] Antti Ajanki, David R Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction* 19, 4 (2009), 307–339.

[3] Olivier Chapelle and Ya Zhang. 2009. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*. ACM, 1–10.

[4] Georges E Dupret and Benjamin Piwowarski. 2008. A user browsing model to predict search engine click data from past observations.. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 331–338.

[5] Thorsten Joachims, Laura Granka, Bing Pan, Helene Hembrooke, and Geri Gay. 2017. Accurately interpreting clickthrough data as implicit feedback. In *ACM SIGIR Forum*, Vol. 51. Acm, 4–11.

[6] Diane Kelly and Jaime Teevan. 2003. Implicit feedback for inferring user preference: a bibliography. In *Acm Sigir Forum*, Vol. 37. ACM, 18–28.

[7] Dmitry Lagun, Chih-Hung Hsieh, Dale Webster, and Vidhya Navalpakkam. 2014. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 113–122.

[8] Yixuan Li, Pingmei Xu, Dmitry Lagun, and Vidhya Navalpakkam. 2017. Towards Measuring and Inferring User Interest from Gaze. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 525–533.

[9] Mengyang Liu, Yiqun Liu, J. Mao, Cheng Luo, Min Zhang, and S. Ma. 2018. "Satisfaction with Failure" or "Unsatisfied Success": Investigating the Relationship between Search Success and User Satisfaction. *Proceedings of the 2018 World Wide Web Conference* (2018).

[10] Zeyang Liu, Jiaxin Mao, Chao Wang, Qingyao Ai, Yiqun Liu, and Jian-Yun Nie. 2017. Enhancing click models with mouse movement information. *Information Retrieval Journal* 20, 1 (2017), 53–80.

[11] Jiaxin Mao, Yiqun Liu, N. Kando, Min Zhang, and S. Ma. 2018. How Does Domain Expertise Affect Users' Search Interaction and Outcome in Exploratory Search? *ACM Transactions on Information Systems (TOIS)* 36 (2018), 1 – 30.

[12] F. Zhang, J. Mao, Yiqun Liu, W. Ma, M. Zhang, and Shaoping Ma. 2020. Cascade or Recency: Constructing Better Evaluation Metrics for Session Search. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020).

[13] Yukun Zheng, J. Mao, Yiqun Liu, M. Sanderson, Min Zhang, and S. Ma. 2020. Investigating Examination Behavior in Mobile Search. *Proceedings of the 13th International Conference on Web Search and Data Mining* (2020).