TrackThink: A tool for Tracking a Thought Process on Web Search

Kazuma Nagano
Nara Institute of Science and Technology
Ikoma, Nara, Japan
nagano.kazuma.na5@is.naist.jp

Yutaka Arakawa
Nara Institute of Science and Technology /JST PRESTO
Ikoma, Nara, Japan
arakawa@is.naist.jp

Keiich Yasumoto
Nara Institute of Science and Technology
Ikoma, Nara, Japan
yasumoto@is.naist.jp

UbiComp/ISWC'17 Adjunct , September 11–15, 2017, Maui, HI, USA
ACM 978-1-4503-5190-4/17/09.
https://doi.org/10.1145/3123024.3129267

Abstract
This paper proposes a tool for tracking a thought process in web searching because we considered that user's browsing history potentially contains the transition information of a thought process of how to collect information efficiently. In fact, we have few opportunities to look back own browsing way. Even if we check the enormous browsing logs, it is hard to convert them into sharable knowledge. Therefore, our tool measure the dwell time, the amount of scrolling and page contents and transition between pages for evaluating the importance of each page. In addition, our tool provides an editing function to summarize a thought process based on the importance of each page.

Author Keywords
thought process, browsing behavior, searching effectiveness, web search

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Currently, people are exposed to a lot of information. From among overflowing contents, we need to figure out the information that was suitable for the purpose. In a previous study[4] that investigated user characteristics on web

search, features such as better educational background and higher critical thinking abilities were found for users with good information retrieval skills. On the other hand, when the information retrieval skills of the user is low, there are problems that required information can not be found due to enormous sources, and that unreliable information is shared frequently by Social media etc.

Therefore, recently, curation media, in which users with high information retrieval skills compiled a lot of contents under a specific purpose, have attracted attention. By using such media, the user can acquire information of interest irrespective of his / her ability while reducing the state of excessive information and the load accompanying it. However, users who only use information gathered by curators do not know what kind of thought processes the provided contents were obtained. So it is not possible to improve their information retrieval skills. Also, the major topics of much prior research in information search area[2] are 1) conformity of content to query, 2) estimation of user's information request, 3) performance evaluation of search system. There are few studies focusing on the User thought process itself.

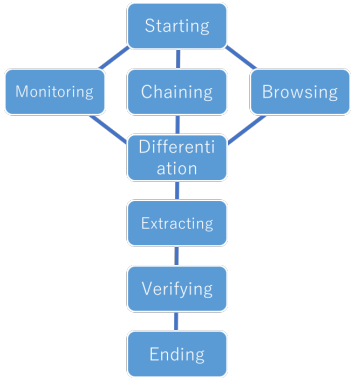

Figure 1: Ellis model overview

As reported by Su et al.[7], user's search histories might contain personal identifiable information. It is very dangerous to share all search history among users to peer into thinking process. Even if you try to edit the search history of the browser yourself to show to others , it is difficult to extract your own thinking process from the enormous URL list and convert it to shareable something. Therefore, for many users, thought processes on web search are tacit knowledge which is difficult to look back on and evaluate each other later.

In this research, as a new topic in the research on the information search field, we try to extract the user's thought process on Web search, which is tacit knowledge, from the browsing history and convert to the knowledge that can be provided as an "Experience Supplement" to others. In this paper, considering privacy information, we propose a tool which helps users to extract the own thought process from browsing history in order to look back and share it. As an initial method for it, Considering privacy information, this paper present an idea of tool for users to extract their thought process from browsing history and summarize it in one screen.

Related work

There are various forms of web search. As the diversity of Web content on has expanded, the high abstract query such as "Better presents to friends" or learning purpose has been carried out. this search is called Exploratory Search. As a definition of exploratory search, the survey paper[6] summarizes the features mentioned in the previous research. In this paper, And they introduce the Ellis model [3] as an existing information search model that most closely matches the features mentioned in it. Figure 1 outlines the Ellis model. The Ellis model is evaluating information retrieval skills through interviews with researchers and consists of the following eight information search processes.

- Starting (activities that form the information search)
- Chaining (following references in initial information sources)
- Browsing (semi-directed search)
- Differentiating (filtering and selecting sources based on quality and relevance)
- Monitoring (keeping track of developments in an area)

- Extracting (systematic extraction of material of interest from sources)

In this model, after the Starting process, Chaining, Browsing, Differentiation, Monitoring loops are repeated until you can obtain the information you want. Extracting, Verifying, the user terminates the search action with the Ending process.Also, in the Ellis model, the time series order of the above process is characterized as changing, which is consistent with Exploratory Search. In this research, paying attention to the user's thinking process at the time of Exploratory Search, which is closely related to the information retrieval skills of the user, we assume the Ellis model as the user's thinking process model.

Recent research that tried to visualize the browsing history of the user's Web browser has an application called " popHistory "[1]. In this research, the bubble chart is displayed in chronological order by animation. In addition, by changing the size of bubbles according to the number of visits by users, they aim to make it easier for users to discover pages with high importance in the browsing history, We report that the participants showed a positive response in comparison. However, this application displays sites for different search purposes as the same day bubble view. Therefore it becomes difficult to look back own thought process as the number of browsing increases. And even with pages of high importance, the user can't discover Web pages that are viewed only once.

In this paper, we adopt the method using Viewport time proposed by Legan et al. [5], to evaluate the importance of users of Web sites. They investigated the user's browsing patterns for 267,210 pageviews for 1,971 news articles posted on yahoo. They point out that we can't know to what extent the user has browsed the web page only by the staying time of each news article of the user. They define the dwell time for each display area as the Viewport time. Then, they categorize the viewing behavior pattern of the user by Markov model. They hypothesized that user engagement is higher enough that users read more Web pages. Then, based on the ViewPort time, they predicted action patterns for users' articles and proposed the following four evaluation methods at the engagement level.

- Bounce: the user that do not engage with the article and leave the page relatively quickly (Dwell Time < 10 sec)
- Shallow engagement: the user that decides to stay and read[1] the contents but reads less than 50% of it
- Deep engagement: the user that decides to read more than 50% of the content.
- Complete engagement: the user that entered a comment in the comment column of the article or replied

In this study, only Complete Engagement is defined in the form of Bi-model. However, in this research, since the evaluation target is not only the news article, it is necessary to define another evaluation method. Therefore, we decided to search for useful information on the purpose of exploratory search of the user, and showed the action showing the highest engagement.

- Complete engagement: The user visited the page again by tab switching and registered in Bookmark.

We use it as a method of evaluating the importance of Web pages.

[1] read is defined as the proportion of the article body (main article text) having a viewport time longer than 5 seconds.

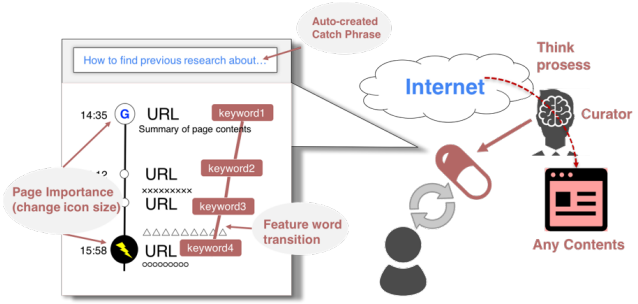

Figure 2: This is the application concept view. We aim for applications that provide a summarized single view as a supplement (red capsule) that allows users to read their own and others' thought process. It collects Web contents of the same purpose, evaluates the user's importance of web page, and visualizes the transition of the page contents.

Vision

The purpose of our research is to convert the user's thought process of collecting information into knowledge that can be shared and to provide others. Figure 2 shows the Outlook of our application. We aim to applications that can be brought closer to the higher level way of our information search by a glimpse of the human thought process with a high information retrieval skills. For example, in a Programming Class, Professor usually let students solve some problems. The students who are good at curation will find a correct answer immediately. And those students who totally don't understand programming may even don't know which keyword they need to search. By tracking the Think Process of the good students and share our summarized information to the other student, the quality of class can be improved a lot.

The focus of this research lies on 1) Tracking thought Process method, 2) Expression of the knowledge from which User can read the thought process, 3) Reward for the user to publish voluntarily a thought process as a supplement. An image of the application currently being created is shown in figure 2. After completing the application, we are planning an interactive experiment with users about 2) and 3). In this paper, we discuss 1). We consider how to track the user's thought process from the search history. As a result, we propose a tool that visualizes thinking process based on importance of each page and has editing function.

Data Collection

In order to respond to search the situation of the various users, we have created a smart phone browser application and a Google Chrome extension to collect data in each environment. By logging in with the same account, the user can share the browsing history collected in another environment. Figure 3 shows each application's view. These applications can record the browsing time of each page of the user, the scroll amount and speed, text contents, tab switching event, and Bookmark event. In this paper, only browse history, search keywords, Viewport time are used for evaluation, but other data to be collected will be analyzed for thinking process extraction in the future.

Tracking Thought Process

the following information is extracted from the data collected by our applications.

(1) User engagement with Web Pages

We evaluate the user engagement on the page based on the method of [5] above. Using the four evaluation levels, Circular thumbnail image obtained from the page metadata is displayed in deferent size. and Since the page of Bounce can be estimated as a page that did not contain useful information, Display only the URL on the screen.

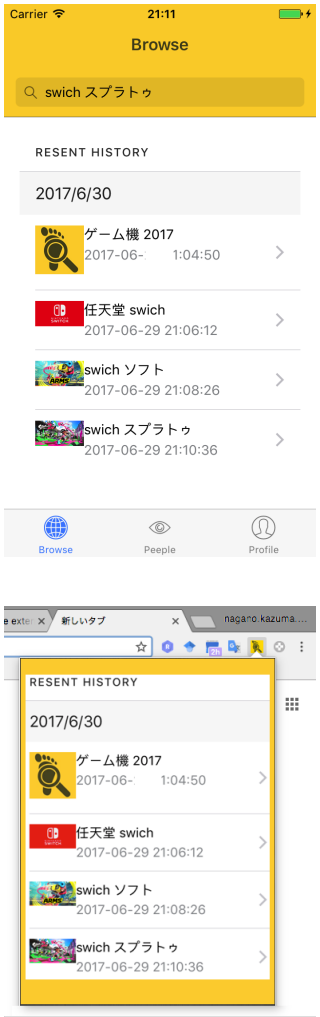

Figure 3: Two types applications for Data collection. These applications can be used on Windows, Mac OS, iOS and Android

(2) Content browsing intention
We extract the starting position of the keyword search from the search history as the start point of the user's Exploratory Search.Based on the Ellis model, we define the following as endpoints: 1) new keyword search using search engines, 2) access to bookmarked pages, 3) address bar based on extrinsic information, We grouped pages visited by the time the endpoint occurred for the same viewing purpose. The user also can aggregate the group of pages classified into different groups on the tool into one. The group name assumes that the user adds a title indicating the intention of browsing the content.

(3) Tracking of feature words of content
In order to help the user remember contents when editing and to facilitate other users to read thought process, we extract some feature words from web page text. In this paper, TF-IDF, which is one of the most popular feature words extraction methods, was used as an initial method. Document Frequency is used Japanese Wikipedia articles as documents.

Visualizing the Thought Process

Using our methods, Figure 4 shows the edit view visualized the information extracted from the data collected by our application. Actually, this views were created based on user A's search history. The extracted feature words by TF-IDF on the right side are arranged in descending order of importance from the right. they will help users remember the content. Icon size reflecting user engagement makes it easy to look back on important pages from a huge list. Based on these information, the user selects the page viewed for the same purpose from the pages grouped by the search keywords on the left side and gives a title. And the user can remove history he or she does not want to share. In addition, the user color readily distinguishable words making others easier to read thought processes.

A result view in which user A edited his own thought process is shown in figure 5. The extracted feature words are used as the next search keyword. And thought process loops of Chaining, Browning, Differentiation, Monitoring in the Ellis model are seen. In addition, we can read that there was an answer to the title question on the last page.

Conclusion and Future Work

In this paper, we proposed an initial method of tracking users' thinking process in Web search and visualizing collected data. Using the proposed application, we collected the dwell time, scroll amount and its speed, contents text, and extract the user's importance of each page and feature words. By grouping and visualizing this information, we proposed a tool that the user can easily convert the browsing history into the thinking process.

As future research, we will compare the method of extracting thinking processes of various users based on the method of this paper, and reveal methods to reflect user 's thinking process and minimize the user' s editing effort. In addition, we will build a platform that can share thought process to others, and propose these knowledge personalized to each user as a "Experience supplement" at an appropriate timing. The performance of these applications will be confirmed by the experiment improving the user's retrieving skills.

Acknowledgements

This research was supported in part by JST CREST (JPMJCR16E1) and CICP (Creative and International Competitiveness Project) in NAIST (Nara Institute of Science and Technology).

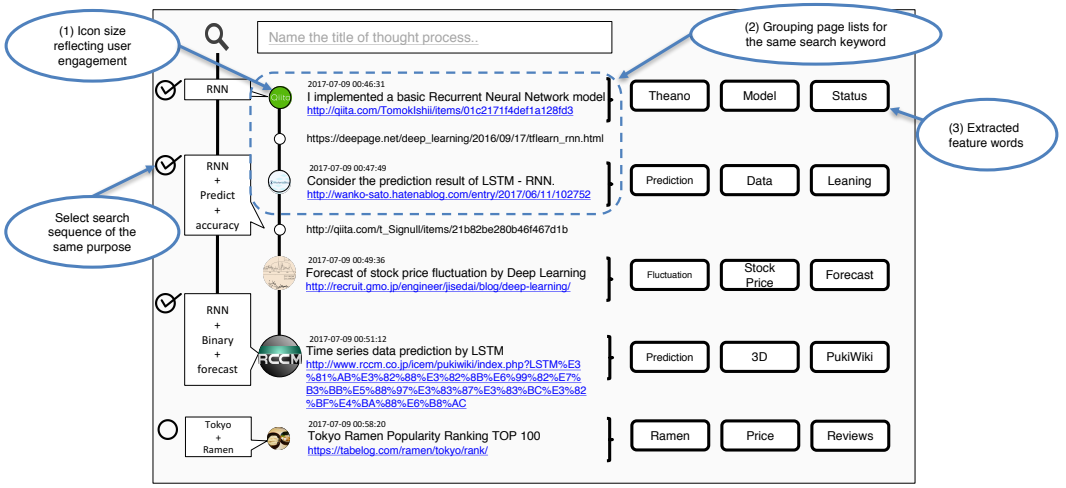

Figure 4: TrackThink: Edit view

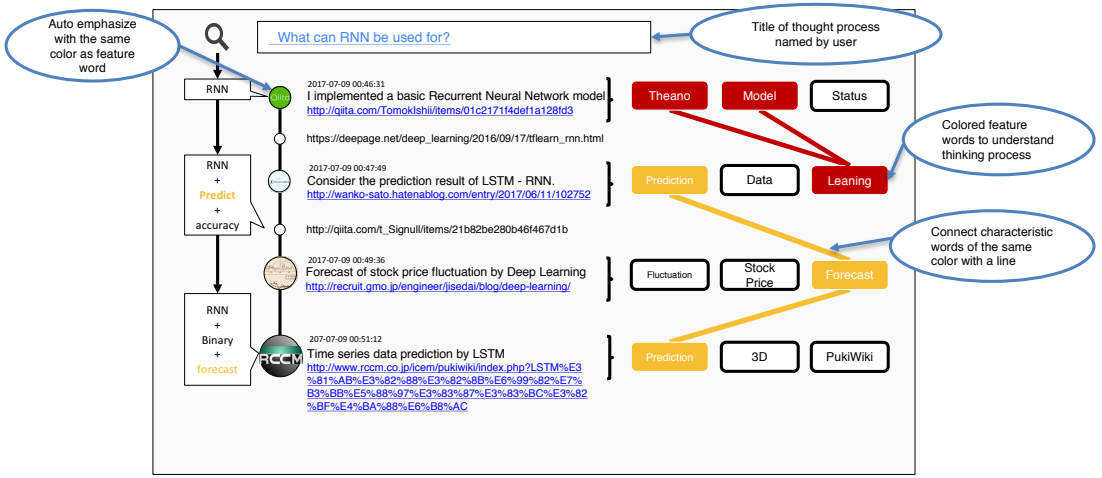

Figure 5: TrackThink: Result view

REFERENCES

1. Matthew Carrasco, Eunyee Koh, and Sana Malik. 2017. popHistory: Animated Visualization of Personal Web Browsing History. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2429–2436.
2. W Bruce Croft, Donald Metzler, and Trevor Strohman. 2010. *Search engines: Information retrieval in practice*. Vol. 283. Addison-Wesley Reading.
3. David Ellis, Deborah Cox, and Katherine Hall. 1993. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation* 49, 4 (1993), 356–369.
4. Emi Ishita, Yosuke Miyata, Shuichi Ueda, and Keiko Kurata. 2017. A structural equation model of information retrieval skills. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 317–320.
5. Dmitry Lagun and Mounia Lalmas. 2016. Understanding user attention and engagement in online news reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 113–122.
6. Emilie Palagi, Fabien Gandon, Alain Giboin, and Raphaël Troncy. 2017. A Survey of Definitions and Models of Exploratory Search. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*. ACM, 3–8.
7. Jessica Su, Ansh Shukla, Sharad Goel, and Arvind Narayanan. 2017. De-anonymizing web browsing data with social networks. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1261–1269.