Towards Memorable Information Retrieval

Qiu, S.; Gadiraju, Ujwal; Bozzon, A.

DOI
10.1145/3409256.3409830

Publication date
2020

Document Version
Final published version

Citation (APA)
Qiu, S., Gadiraju, U., & Bozzon, A. (2020). *Towards Memorable Information Retrieval*. 69–76. Paper presented at In Proceedings of the 2020 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '20). https://doi.org/10.1145/3409256.3409830

Towards Memorable Information Retrieval

Sihang Qiu
Delft University of Technology
Delft, The Netherlands
s.qiu-1@tudelft.nl

Ujwal Gadiraju
Delft University of Technology
Delft, The Netherlands
u.k.gadiraju@tudelft.nl

Alessandro Bozzon
Delft University of Technology
Delft, The Netherlands
a.bozzon@tudelft.nl

ABSTRACT

Information overload is a problem many of us can relate to nowadays. The deluge of user generated content on the Internet, and the easy accessibility to a vast amount of data compounds the problem of remembering and retaining information that is consumed. To make information consumed more memorable, strategies such as note-taking have been found to be effective by augmenting human memory under specific conditions. This is based on the rationale that humans tend to recall information better if they have produced the information themselves. Previous works in online education have shown that conversational systems can improve learning effects. Although memorization is an important part of learning, the effect of conversation on human memorability remains unexplored. We aim to address this knowledge gap through an experimental study, by investigating human memorability in a classical information retrieval setup. We explore the impact of note-taking affordances and conversational interfaces on the memorability of information consumed by users. Our results show that traditional web search and note-taking have positive effects on knowledge gain, while the search engine with a conversational interface has the potential to augment long-term memorability. This work highlights the benefits of using note-taking and conversational interfaces to aid human memorability. Our findings have important implications on building information retrieval systems that cater to optimizing memorability of information consumed.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; • **Human-centered computing**;

KEYWORDS

Web Search, Memorability, Information Retrieval, Note-taking, Conversational Interface

ACM Reference Format:
Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Towards Memorable Information Retrieval. In *2020 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '20), September 14–17, 2020, Virtual Event, Norway.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3409256.3409830

ICTIR '20, September 14–17, 2020, Virtual Event, Norway

ACM ISBN 978-1-4503-8067-6/20/09…$15.00
https://doi.org/10.1145/3409256.3409830

1 INTRODUCTION

Information overload is a byproduct of the rapid development of information technology and the plethora of user generated content. By issuing a simple search query, an Internet user can access billions of relevant items from a search engine within seconds. The data deluge and a constant exposure to new information leads to the problem of remembering and retaining information during informational search sessions. Most popular search engines today are optimized to serve relevance related needs with respect to user queries. We believe that an unexplored opportunity lies in how information can be retrieved and presented to users, with an aim to improve the memorability of information consumed.

To improve human memorability, researchers in the field of experimental psychology have studied the "generation effect" [26]. By comparing memory for words, experiments revealed that humans could better recall information if they produced it themselves rather than if they received it. Based on the generation effect, prior studies have shown that note-taking, a simple way to re-produce received information, can improve human memorability, particularly for text-based learning and comprehension [7, 27]. However, the effects of note-taking in a classic information retrieval setup remain unexplored.

Prior studies in online learning have revealed that conversational systems can significantly improve learning outcomes [13, 16, 28]. As the goal of learning is to develop a deep understanding of some information, memorization is an important element [4, 15]. Although conversation can produce unique context linked with information, the effect of conversational systems on human memorability needs further exploration. A recent study has investigated the role of text-based conversational interfaces in online information finding tasks [22]. Authors demonstrated that a conversational interface could better engage online users. However, the question of whether improved user engagement through conversational interfaces leads to better memorability of information remains unanswered.

In this paper, we aim to fill this knowledge gap by proposing novel approaches to improve human memorability during information retrieval. We specifically focus on information retrieval activities carried out through the Web search using desktop browsers. Through rigorous experiments, we seek to address the following research questions.

> **RQ1**: How can human memorability of information consumed in informational web search sessions be improved?

Inspired by prior work in psychology and HCI, we propose novel search interfaces which (a) provide the affordance of note-taking to users, and (b) provide a conversational interface. We propose methods to quantify knowledge gain and long-term memorability of information consumed, and investigate the impact of the proposed

search interfaces on the memorability of information consumed. We conducted an online user study in a classical information retrieval setup. Results reveal that traditional Web interfaces with a note-taking affordance can benefit knowledge gain (up to 25% higher than other interfaces), while conversational interfaces have the potential to augment long-term memorability (7.5% lower long-term information loss). Our findings suggest that both note-taking and conversational interfaces are promising tools for augmenting human memorability in information retrieval.

RQ2: How does note-taking and the use of text-based conversational interfaces affect the search behavior of users?

We found that users leveraging conversational interfaces input more queries but opened links less frequently compared to users leveraging the traditional Web interfaces. In addition, the users of conversational interfaces tend to type notes themselves, while the Web users input significantly longer notes by copying content directly from the search engine result pages.

2 RELATED LITERATURE

Augmenting Human Memory. Different theories for augmenting human memory have been studied in the field of psychology. The memory consolidation theory proposed by Müller and Pilzecker explained the processes to make information memory [18, 20]. The Atkinson-Shiffrin memory model shows that the long-term memory can be consolidated by repeatedly rehearsing short-term memory [1]. To study how the 'remembering information' relates to one's self, previous work has revealed that the memory could be enhanced if it relates to one's self-concept or an episode from one's life [6]. A prior study in experimental psychology has shown evidence of the existence of the "generation effect" [26]. Authors conducted experiments at the word-level to show that people could remember information better if the information was produced by themselves. A simple and direct application of the generation effect is the use of note-taking. Previous studies have shown that note-taking can improve human memorability in different scenarios [7, 10, 19, 27]. Intons-Peterson et al. examined the use of internal and external memory aids in experiments with 489 undergraduates. It was found that at least one external aid, i.e. taking notes, can effectively facilitate remembering [14]. Based on the findings of prior works, in this study we investigate how an external aid such as note-taking can affect the long-term memorability of users in informational search.

Aiding Memorability in Information Systems. Augmenting human memory has also been studied from an information systems standpoint. Many previous studies have used context as a key aspect to improve human memorability [9, 23]. The 'Remembrance Agent' is an automatic system which uses the role of context in memory to augment human memory, by listing documents related to the user's current context [23]. Blanc-Brude et al. have performed experiments to find the attributes (e.g. file name, time, title, location, size, etc.) that help memorability for a document search tool [5]. Previous works have also shown that many strategies, such as time-aware contextualization [8, 29], and optimizing recollection by generating analogies [24], have a positive effect on human memorability. Furthermore, a recent study built an application named 'ReflectiveDiary', to investigate how self-generated daily summaries can improve memorability [25]. Predictive methods have also been proposed to consolidate human memory in the workplace environment [3]. Since memorization is an essential element of the learning process [4, 15], we also examined relevant literature in online learning. Across multiple studies, conversational systems were found to be useful in facilitating learning effects [13, 16, 28] and in effectively improving user engagement in information retrieval tasks [22]. These previous works with regard to aiding memorability or improving learning effects in information systems are not directly applicable in the current information retrieval ecosystems. Inspired by these prior works, we propose novel search interfaces and design experiments to study human memorability in information retrieval.

3 METHOD

The goal of this study is to investigate whether note-taking and conversational interfaces can affect human memorability in informational web search sessions. To this end, we measure long-term memorability of information consumed by users.

3.1 Study Design

The taxonomy of human memory, which is rather complicated and detailed, has been developed for over a hundred years. Human memory can be classified into two big categories; short-term and long-term memory. Short-term memory only persists for seconds or minutes [1, 2, 12], while long-term memory can last for much longer [1]. In this study, we focus on improving the long-term memorability of information consumed by users in web search sessions. According to Ebbinghaus' curve and recent replication works [21]: the forgetting curve goes down slowly after 24 hours (people forget more than 60% within 24 hours, 70% within 2 days, and 80% within 30 days). It was found that fluctuations might appear at the 24-hour point. However, after 2 days, the forgetting curve becomes stable. Therefore, we choose 3-7 days as the time interval to measure user long-term memorability in this study.

The basic idea of measuring memorability in web information retrieval is to quantify how much information a user can remember at the end of an informational search session. Therefore, as shown in Figure 1, we first assign a topic and an information need to users, and ask the users to finish a "knowledge calibration" test (*pre-task test*) with 10 questions related to the topic. We use 10 topics and the corresponding questions from a previous work about analyzing knowledge gain in informational search [11], as listed in Table 1. Topics are randomly assigned to users. Through the pre-task test users can better understand different facets of the information need, and we can calibrate the background knowledge of users.

Next, users are directed to the search session, where they must spend at least 7 minutes searching about their assigned information need. As we can see from Figure 1, users are assigned any 1 of 4 different user interfaces. Half of the users use a Web interface to perform their search sessions, while the rest are assigned a conversational interface. Both Web and conversational interfaces have two conditions, i.e. with note-taking function enabled or disabled. In the Web interfaces, users leverage a Web search page that is similar to typical search engines. In the conversational interfaces, users are guided by a conversational agent through their session.

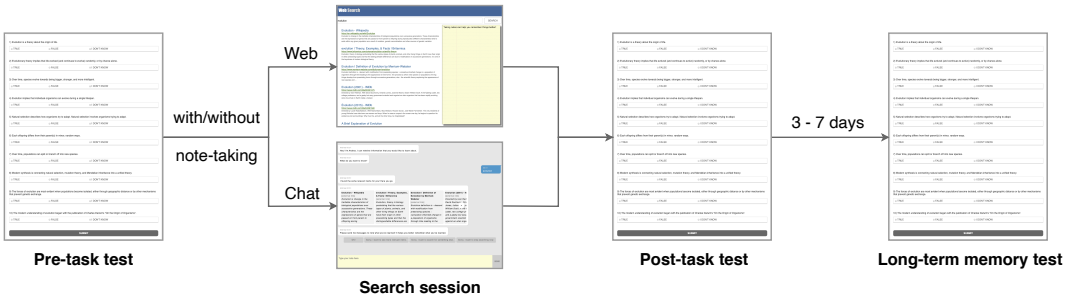

Figure 1: Workflow of our study. The *pre-task test*, the *search session* and the *post-task test* pertain to a single Human Intelligence Task (HIT) published on Amazon MTurk. The *long-term memory test* is deployed separately in a follow-up HIT.

Table 1: Topics and corresponding information needs (topics are re-used from [11]).

Topic	**Information Need**
Altitude Sickness	The users are required to acquire knowledge about the symptoms, causes and prevention of altitude sickness.
American Revolutionary War	The users are required to acquire knowledge about the 'American Revolutionary War'.
Carpenter Bees	The users are required to acquire knowledge about the biological species 'carpenter bees'. How do they look? How do they live?
Evolution	The users are required to acquire knowledge about the theory of evolution.
NASA Interplanetary Missions	The users are required to acquire knowledge about the past, present, and possible future of interplanetary missions that are planned by the NASA.
Orcas Island	The users are required to acquire knowledge about the Orcas Island.
Sangre de Cristo Mountains	The users are required to acquire knowledge about 'Sangre de Cristo' mountain range.
Sun Tzu	The users are required to acquire knowledge about the Chinese author Sun Tzu - about his life, his writings, and his influence to the present day.
Tornado	The users are required to acquire knowledge about the weather phenomenon that is called 'tornado'.
USS Cole Bombing	The users are required to acquire knowledge about the 2000 terrorist attack that came to be known as the 'USS Cole bombing'.

After the search session, users need to finish a *post-task test*. The questions shown in the *post-task test* are identical to the questions in the *pre-task test*, allowing us to measure user knowledge gain. To incentivize active search behavior during the search session, users were informed that an extra reward will be given depending on the number of correct answers in the *post-task test*. To elicit honest and genuine responses, users were also told that their accuracy in the *pre-task test* would not affect the reward.

Three days after the search session, we notify all the users who participated in our study and give them an opportunity to answer our *long-term memory test* within the next 4 days in return for an additional reward of 1 USD. The questions in the *long-term memory test* are identical to the *pre-task test*. By comparing the results of the *post-task test* to the *long-term memory test*, we can measure how much information users have retained or forgotten over this long-term period.

3.2 Measuring Memorability

Measuring knowledge gain. Similar to prior work in *search as learning* [11, 30], we measure the knowledge gain of users as the normalized difference in performance of users between the post-task and pre-task knowledge tests.

We use A_t ($t \in \{pre, post, long\}$) to denote the set of answers of the test t, and use $A_t^i \in A_t$ ($1 \leq i \leq 10$) to represent if the i^{th} question of the test t is correctly answered ($A_t^i = 1$) or not ($A_t^i = 0$) by the user. If a user chooses "I DON'T KNOW", we consider it as incorrect answer. For instance, if the 5^{th} question of the *pre-task test* is correctly answered by the user, then we assign $A_{pre}^5 = 1$; if the answer of the 7^{th} question of the *post-task test* provided by the user is incorrect, we assign $A_{post}^7 = 0$. Thus, the normalized knowledge gain can be calculated by using the following equation (where the max/min(*topic score*) means the maximum or minimum score among all the tests sharing the same topic, and the score of a test t can be calculated by $\sum_{i=1}^{10} A_t^i$).

$$knowledge\ gain = \frac{\sum_{i=1}^{10} A_{post}^i - \sum_{i=1}^{10} A_{pre}^i}{\max(topic\ score) - \min(topic\ score)} \quad (1)$$

Measuring long-term memorability. Similarly, we can also use *information gain* to measure the long-term user memorability, which can be calculated by the following equation.

$$information\ gain = \frac{\sum_{i=1}^{10} A_{long}^i - \sum_{i=1}^{10} A_{post}^i}{\max(topic\ score) - \min(topic\ score)} \quad (2)$$

Long-term memorability can also be measured using *information loss*. The *information loss* after the post-task test can be quantified by the number of questions which are correctly answered in the *post-task test* but incorrectly answered in the *long-term memory test*. Thus, it can be calculated by the following equation.

$$information\ loss = \frac{\sum_{i=1}^{10} A_{post}^i - \sum_{i=1}^{10} A_{long}^i \cdot A_{post}^i}{\max(topic\ score) - \min(topic\ score)} \quad (3)$$

3.3 User Interfaces

Addressing **RQ1**, we designed Web and conversational interfaces to support informational search sessions, with an optional note-taking

functionality. Both the Web and conversational interfaces use the Bing Search API [1] for sending search query requests and receiving search results (relevant web pages).

The Web interface is designed according to the typical user interface of popular search engines, as shown in Figure 2 (a). The Web interface consists of two main components — a textarea for entering search queries, and a rectangular frame for displaying search results. During the search session, users need to type search queries in the textarea at the top of the page. Users can either click the "SEARCH" button or press the "Enter" key on the keyboard to issue the search query asking for 10 relevant items (Web pages), and then the sever will respond with a list of search results. The search results include 10 items with their titles, links and snippets, which are shown under the text area, occupying the most part of the Web interface. Since each query fired only requests for 10 relevant items, the Web interface only shows 10 search results at a time. Each item is clickable. To prevent users from jumping to other pages or applications, once the user clicks an item, an embedded browser will pop up to show the content of the corresponding item (Web page). To retrieve more items, users can click the "NEXT PAGE" button to send a query asking for the next 10 relevant items, or click the "PREVIOUS PAGE" button to go back.

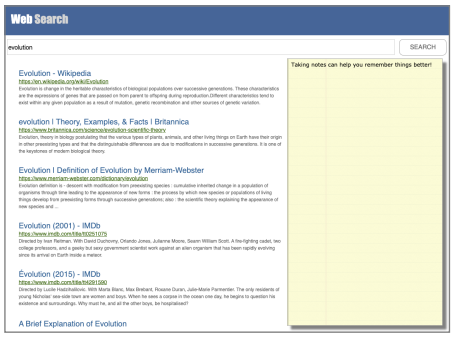

Figure 2: Web search interfaces with the note-tasking function enabled. The yellow notepad becomes invisible if the note-taking function is disabled.

Furthermore, as shown in Figure 2 (b), to enable the function of note-taking, a notepad is embedded on the right side of the Web interface. The notepad can be enabled or disabled depending on the experimental condition. On the notepad, we leave a sentence "tasking notes can help you remember things better" to encourage users to take notes during the search session. All the on-page activities including querying, browsing (clicking) items, and note-taking are automatically logged for user behavior analysis.

The conversational interface uses the same search engine as the Web interface. However, the search workflow is guided by a text-based conversational agent, as shown in Figure 3. The logic of the conversational interface for web search is designed as follows:

1) Greetings. The conversational agent opens the conversation with the user and then asks the user to provide a search query. The conversational agent sends the greetings to initiate the search session.

[1] https://www.customsearch.ai/

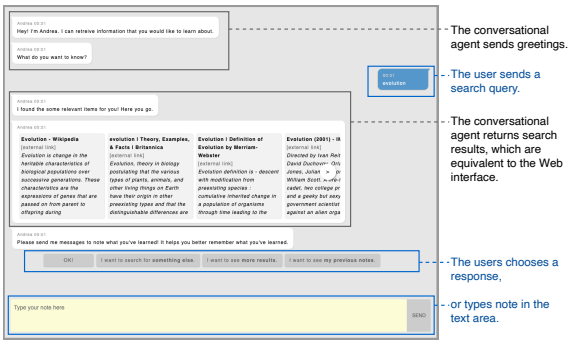

Figure 3: Conversational search interface.

- Hey! I'm Andrea. I can retrieve information that you would like to learn about.
- What do you want to know?

Note that we assign a gender-neutral name ('Andrea') to the conversational agent, to avoid potential biases. Andrea is a name commonly used for both males and females around the world.

2) Search. After the user provides the agent with a search query, the conversational agent uses Bing Search API to retrieve results. To make the conversational interface comparable to the Web interface, the agent also shows 10 relevant items at a time. However, on the conversational interface, all the content is presented within chat bubbles to replicate typical conversational interfaces [17]. As we can see from Figure 3, the relevant items are listed horizontally in a chat bubble, where the user can scroll horizontally to view them. Also, each item in the chat bubble is clickable and linked to the embedded Web browser.

3) Response selection. The conversational agent provides the user with four options after the search results have been displayed. The four options correspond to taking notes, showing more results, entering a new query, and showing previous notes, respectively. However, if the note-taking function is disabled, the agent only presents two options — showing more results and entering a new search query. If a user chooses to **take notes**, the message that the user sends to the agent will be recorded and integrated with previous notes (if any) from the user in the search session. If the user chooses **show more results**, the next 10 relevant items will be displayed to the user with a new chat bubble. The functionality is equivalent to that of the Web interface. The conversational interface does not provide an option to show previous items, since users can easily find previous items by viewing the conversation history. If the user chooses to **input a new query**, the agent goes back to *step 2 Search* to re-start the search process. Finally, all the previous notes can be shown in a chat bubble if the user chooses to see the notes by using the **show previous notes** option.

4 EXPERIMENTS

4.1 Experimental Conditions

In this study, we use two user interfaces (Web and Conversational) with a note-taking function either enabled or disabled to address our research questions. This results in four experimental conditions.

Chat w/ note: the conversational interface with note-taking. In this experimental condition, users are redirected to a conversational interface, where the searching process is guided by a conversational agent — Andrea. In addition, the note-taking function is enabled, meaning users can take notes by sending messages to Andrea.

Chat w/o note: the conversational interface without note-taking. In this experimental condition, users are redirected to an ordinary conversational interface, where the searching process is also guided by Andrea, but the note-taking function is disabled.

Web w/ note: the Web interface with note-taking. In this experimental condition, users are redirected to a custom Web search interface to complete the search session. A notepad is visible on the right side of the Web interface where users can type their notes.

Web w/o note: the Web interface without note-taking. In this experimental condition, users are also redirected to a custom Web search interface to complete the search session. However, the notepad is hidden and disabled. This experimental condition represents the most typical search engines nowadays.

Participants in our experimental study were recruited from Amazon Mechanical Turk (AMT). Since popular crowdsourcing platforms generally support custom task design based on HTML, CSS and Javascript, we design the conversational interface purely based on HTML/CSS/Javascript. The conversational interface can be directly presented on the default task page of crowdsourcing platforms without any re-directions. The code for our text-based conversational interface along with all the data will be made available to the community to facilitate further research[2]. We published online tasks with the aforementioned four experimental conditions on AMT. The Human Intelligence Task (HIT) published on AMT only contained the *pre-task test*, search session and the *post-task test*. The *long-term memory test* was not included in the HIT batches. We used the notification function provided by AMT, to send the link of the *long-term memory tests* to workers after three days. The Web page of the *long-term memory test* was set up on our own server. We recruited 35 online crowd workers per condition from AMT, as the users of our search systems. Each worker was assigned a random topic from Table 1. The experiment was approved by the ethics committee of our institute, and we did not collect and store any identifiable data of human subjects.

4.2 Quality Control

The minimum time for each search session was set to 7 minutes (users were not allowed to proceed to the next stage before 7 mins). Apart from incentivizing genuine search behavior through attached rewards for performance in the *post-task test*, we took additional measures to ensure reliable behavior. The timer stops if a worker temporally leaves the page (for instance, switching to other tabs or programs). Furthermore, we use an embedded browser to enable workers to open and browse the search results on our own task page, instead of opening a new tab. Considering the effects of learning bias, we add an extra Javascript code to record the unique AMT Worker ID on our server, to prevent a worker from executing our HIT multiple times. We restricted participation by using the default qualification type, "Overall HIT approval rate is greater than 95%" provided by AMT to further ensure high worker quality. In addition, we manually inspected users' answers to exclude any potentially unreliable users. We exclude users if they:

(1) Enter no queries during the search session;
(2) Always select the same option — either 'YES' or 'NO' in *pre-/post-task test* or *long-term memory test*.

Due to the criteria we defined, 8 workers were manually excluded in our experiments.

4.3 Worker Reward

Upon the task completion, we immediately reward each worker with 2 USD. After three days (72 hours), we bonus workers according to the number of correct answers given in the *post-task test* (0.01 USD per correct answer). In the notification message corresponding to the bonus, we requested workers to participate in our *long-term memory test* by providing a link to the test page. The Web page of the *long-term memory test* is set up on our own server instead of AMT. We incentivized workers to complete the *long-term memory test* with an additional reward of 1 USD on completion. For the next three days (i.e., until 7 days after their search session), we sent a notification every 24 hours to those workers who did not finish *the long-term memory test* yet.

4.4 Evaluation Metrics

We measure the user knowledge gain, long-term memorability, search time, and user behavior including number of queries, browsing frequency and the length of notes that users take (where applicable) while completing the HITs.

Knowledge Gain and Long-term Memorability. User knowledge gain is calculated using Equation 1. The long-term memorability is measured using (i) information gain, calculated using Equation 2, and (ii) information loss, calculated using Equation 3.

Search time. We recorded how long each user spends on the search session, which is the length of the time period starting from when the user submits the answers of the *pre-task test*, until the worker clicks the "NEXT" button to proceed to the *post-task test*. The "NEXT" button becomes visible only after 7 minutes, enforcing a minimum search time of 420 seconds.

Search Behavior. We also analyze user behavior during the search sessions to better understand how user behavior relates to the memorability of information consumed. To this end, we focus on:

1) Number of queries. It represents how many queries a user sends to search engine through either Web or conversational interfaces;

2) Browsing frequency. This is the frequency of a user opening a link and using the embedded Web page browser to view the content of the search results.

3) Length of notes. It represents the number of characters written in the notes provided by the user.

5 RESULTS

After excluding unreliable workers, the four experimental conditions — **Chat w/ note**, **Chat w/o note**, **Web w/ note**, and **Web w/o note**, we are left with 32, 34, 33, and 33 unique valid users respectively. Furthermore, the four conditions had 14, 11, 15 and 16 users who returned for the *long-term memory test* respectively.

[2]https://sites.google.com/view/memorableir

5.1 Memorability Analysis

Knowledge gain. *The Web interface with note-taking can significantly improve the knowledge gain in comparison to the conversational interface conditions, while the conversational interface without note-taking shows no positive impact on the knowledge gain of users.*

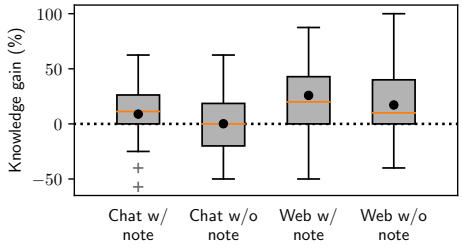

Figure 4: Knowledge gain of users across the four interfaces.

Figure 4 presents the knowledge gain of users across the four interface conditions. The average knowledge gain of users corresponding to the conversational interfaces is 4.4%, while that of the Web interfaces is 21.5%. Particularly, the knowledge gain of the Web interface with note-taking function enabled (Web w/ note) is 25% higher than the conversational interface with note-taking function disabled. Since the distributions of knowledge gain follow normal distributions (verified by the Shapiro-Wilk tests for normality), we use independent t-tests ($\alpha = 0.05$) to find the significant differences between user interfaces. We found three pairs having a p-value less than 0.05 (Chat w/ note vs Web w/ note p=0.030, Chat w/o note vs Web w/ note p=9.7e-4, and Chat w/o note vs Web w/o note p=0.031). After Holm-Bonferroni correction, the knowledge gain of Web w/ note is still significantly higher than Chat w/o note. Results suggest that note-taking is a useful tool for improving knowledge gain, aligned with findings from previous studies. However, the conversational interface revealed no specific advantage over the traditional web interface in facilitating knowledge gain.

Long-term Memorability. *Results revealed no significant difference across interface conditions with regard to long-term information gain (computed using information gain). However, conversational interfaces exhibit the potential to reduce long-term information loss.*

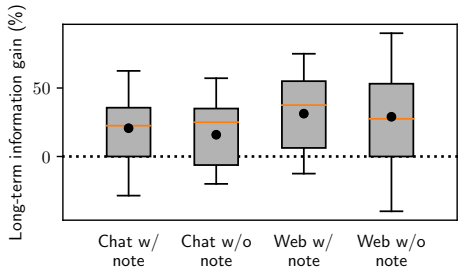

Figure 5: Long-term memorability (using information gain) across the four interfaces.

As shown in Figure 2, the average long-term information gain of users across all user interfaces is actually higher than the average knowledge gain observed. This is due to the subset of users who returned to complete the *long-term memory test* (these users had relatively higher knowledge gain scores). We also found that the long-term information gain is significantly correlated to knowledge gain according to Pearson correlation coefficient testing ($p < 0.05$ except Chat w/o note). The distributions of long-term information gain also follow normal distributions (verified by the Shapiro-Wilk test for normality). However, we found no significant difference between long-term information gain across the four interface conditions by independent t-tests. This suggests that the Web interface and note-taking show no positive effect on long-term memorability, although they can effectively improve knowledge gain.

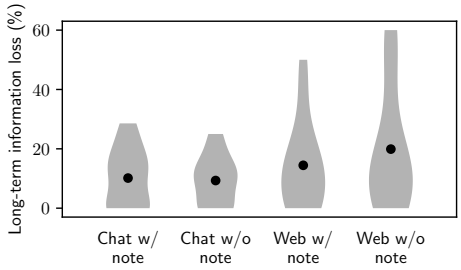

Figure 6: A violinplot of long-term memorability (information loss) across the four interfaces.

The long-term information loss was calculated to further analyze long-term memorability. The distributions of information loss across four interfaces are shown in Figure 3. We found the average information loss of users corresponding to conversational interfaces is (9.8%), which is 7.5% lower than that of the Web interfaces (17.3%) with a small p-value ($p = 0.06$, independent t-tests). Furthermore, the maximum information loss among the 25 users who use conversational interfaces is 28%, while that of the 31 users using Web interfaces is 60%. These results indicate that the conversational interface has the potential to improve user long-term memorability.

5.2 Search Time Analysis

Search Time (in seconds). *We found no significant difference in the average search time of users across the four interface conditions.*

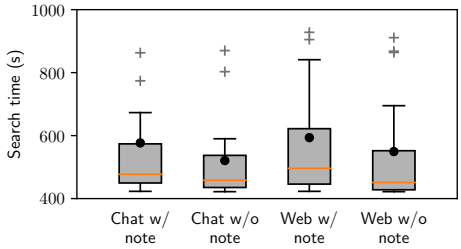

Figure 7: Search time (s) across the four interface conditions.

We measured the time that the user spends on the search session for each experimental condition. The average search time across all user interfaces is 559 seconds. As the distributions of search time do not follow normal distribution (according to Shapiro-Wilk tests), we used Mann-Whitney U tests ($\alpha = 0.05$) to compare the search time across four user interfaces. Although the average search time of the user interfaces with note-taking, for both conversational and Web interface, is slightly higher than the user interfaces without note-taking, this was not found to be statistically significant.

5.3 Worker Behavior Analysis

The worker behavior during the search session is analyzed using three measurements, i.e. the number of queries, the browsing frequency, and the length of notes.

Table 2: Mean and standard deviation ($\mu \pm \sigma$) of the number of queries, the browsing frequency, and the length of notes across the four user interface conditions.

Interfaces	Number of queries	Browsing frequency	Length of notes
Chat w/ note	9.56 ± 5.23	0.47 ± 1.09	348.68 ± 457.15
Chat w/o note	9.71 ± 8.66	0.44 ± 1.01	/
Web w/ note	3.76 ± 3.16	1.82 ± 2.43	1004.58 ± 1431.63
Web w/o note	4.64 ± 5.66	2.09 ± 1.96	/

Number of Queries. *The users corresponding to conversational interfaces tend to send more queries on average (ask more questions to the conversational agents), while the users corresponding to the web interfaces input significantly fewer queries.*

In terms of the number of queries, we found that users using conversational interfaces generally send more queries than the users who use Web interfaces (2.3 times more queries). We applied Mann-Whitney U tests ($\alpha = 0.05$) and Holm-Bonferroni correction to discover significant differences across conditions with respect to number of queries fired. Results of significant testing revealed that note-taking had no impact on the number of queries. However, the conversational interfaces significantly increase the number of queries entered by users, compared to the traditional Web interfaces (Chat w/ note vs Web w/ note p = 3.5e-6; Chat w/ note vs Web w/o note p = 2.6e-5; Chat w/o note vs Web w/ note p = 1.6e-5; Chat w/o note vs Web w/o note p = 9.1e-5). Moreover, a manual investigation of the search histories show that users using conversational interfaces tend to use questions as queries. This suggests that the users using a conversational interface tend to retrieve information by frequently posing questions to the agent as expected.

Browsing frequency. *Users in the conversational interface conditions tend to retrieve information by viewing snippets rather than by frequently opening links.*

The browsing frequency represents the frequency with which a user opens the links of search results. We found that note-taking has no significant impact on the browsing frequency of users according to Mann-Whitney U tests, while users using Web interfaces depict a significantly higher frequency of browsing search results (Chat w/ note vs Web w/ note p = 0.0013; Chat w/ note vs Web w/o note p = 4.9e-05; Chat w/o note vs Web w/ note p = 9.0e-04; Chat w/o note vs Web w/o note p = 3.8e-05). Our results suggest that the users using Web interfaces open the links more frequently, while the users of conversational interfaces tend to obtain information from snippets. This behavior of users in conversational interfaces can potentially be explained by their reluctance to break the coherence of conversation by opening links.

Length of Notes. *The users corresponding to web interfaces input significantly longer notes by copy-pasting content directly from the source, while the users in the conversational interface conditions type shorter notes by themselves.*

As for the length of notes, the users in the Web interface conditions input significantly longer notes compared to the users in the conversational interface conditions (p=0.022, Mann-Whitney U tests). A manual inspection reveals that users of web interfaces prefer copying content from the search results and pasting it to the notepad, while the users of conversational interfaces tended to type information themselves. Prior work has revealed that generating information by onself (notes), can aid long-term memorability. The fact that users in the conversational interface conditions indulged in generating notes themselves is promising and should be explored in future work.

Worker Behavior and Long-term Memorability. We investigated the linear relationship between users' search behavior and the memorability of the information consumed across the four interface conditions.

We performed Pearson correlation coefficient testing ($\alpha = 0.05$) to find the potential correlation between long-term memorability and all the worker behavior measurements. Although no statistical significance was found after Holm-Bonferroni correction, here we report the pairs whose p-value is less than 0.2. We found that the information loss has negative correlations with the number of queries and the length of notes, for users using a conversational interface with note-taking ($R = -0.46, p = 0.10$ and $R = -0.43, p = 0.13$ respectively). This indicates that the greater the number of queries or the longer notes that a user inputs, the less information the user tends to forget. As for users using a Web interface with note-taking, we found the information loss has positive correlations with the number of queries and the the browsing frequency ($R = 0.48, p = 0.07$ and $R = 0.58, p = 0.02$ respectively), indicating that a higher frequency of querying and browsing can potentially lead to information loss on a Web interface.

6 DISCUSSION

Implications. Our findings in this study reveal that users employing conversational interfaces in informational search sessions exhibit a different search behavior compared to traditional web search: they rely primarily on text-based conversation, resulting in a significantly higher frequency of issuing queries but significantly lower frequency of opening SERP (search engine results page) links. This can potentially explain the relatively lower knowledge gain corresponding to users in the conversational interface conditions, since these users appear to consume information by means of viewing titles and snippets rather than opening links and exploring SERPs in detail. In contrast, our results indicate that note-taking in the traditional web interface can significantly increase user knowledge gain. We found that users employing conversational interfaces have

the potential to better retain information consumed (conversational interfaces were found to reduce long-term information loss). This is possibly due to the fact that conversational interfaces can generate unique context connected to the information during the search session. Our inspection of users' notes also corroborate that users using conversational interfaces tend to generate the information by themselves rather than copying content from sources (Web users' preference). These findings suggest that both note-taking and conversational interfaces can be promising tools towards achieving memorable information retrieval.

Limitations and Future Work. In this work, we found that using note-taking and conversational interfaces could enhance human long-term memory, and the users tended to exhibit different subjective perceptions. Therefore, to what extent the note-taking with different perceptions can improve (or probably reduce) information retrieval performance needs further exploration.

We found that only around half of the users returned for our *long-term memory test*, which is typical of such experiments. Our results show that the users with a relatively higher *post-task test* scores were more willing to return and participate in our *long-term memory test*. It should be noted that this participation bias presents a threat to the representativeness of our findings. In our imminent future research on memorable information retrieval, we will explore whether a higher user engagement relates to a better user memorability of information consumed.

7 CONCLUSIONS

This work presents a first exploration of how human memorability can be improved in information retrieval. To this end, we proposed novel search interfaces and quantified long-term memorability. We designed user interfaces with note-taking affordances and text-based conversational agents for informational search. We found that traditional Web interfaces and note-taking can improve user knowledge gain significantly, while conversational interfaces have the potential to benefit long-term memorability.

Acknowledgments. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

REFERENCES

[1] Richard C Atkinson and Richard M Shiffrin. 1968. Human memory: A proposed system and its control processes. (1968).
[2] Alan D Baddeley, Neil Thomson, and Mary Buchanan. 1975. Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior* 14, 6 (1975), 575–589.
[3] Seyed Ali Bahrainian and Fabio Crestani. 2018. Augmentation of Human Memory: Anticipating Topics That Continue in the Next Meeting. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 150–159. https://doi.org/10.1145/3176349.3176399
[4] John B Biggs. 1987. *Student Approaches to Learning and Studying. Research Monograph*. ERIC.
[5] Tristan Blanc-Brude and Dominique L. Scapin. 2007. What Do People Recall About Their Documents?: Implications for Desktop Search Tools. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 102–111. https://doi.org/10.1145/1216295.1216319
[6] Gordon H Bower and Stephen G Gilligan. 1979. Remembering information related to one's self. *Journal of research in personality* 13, 4 (1979), 420–432.
[7] Dung C Bui, Joel Myerson, and Sandra Hale. 2013. Note-taking with computers: Exploring alternative strategies for improved recall. *Journal of Educational Psychology* 105, 2 (2013), 299.
[8] Andrea Ceroni, Nam Khanh Tran, Nattiya Kanhabua, and Claudia Niederée. 2014. Bridging temporal context gaps using time-aware re-contextualization. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 1127–1130.
[9] Tangjian Deng, Liang Zhao, Ling Feng, and Wenwei Xue. 2011. Information Re-finding by Context: A Brain Memory Inspired Approach. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM '11)*. ACM, New York, NY, USA, 1553–1558. https://doi.org/10.1145/2063576.2063799
[10] Gilles O Einstein, Joy Morris, and Susan Smith. 1985. Note-taking, individual differences, and memory for lecture information. *Journal of Educational psychology* 77, 5 (1985), 522.
[11] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. 2018. Analyzing knowledge gain of users in informational search sessions on the web. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 2–11.
[12] E Bruce Goldstein. 2014. *Cognitive psychology: Connecting mind, research and everyday experience*. Nelson Education.
[13] Bob Heller, Mike Proctor, Dean Mah, Lisa Jewell, and Bill Cheung. 2005. Freudbot: An investigation of chatbot technology in distance education. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 3913–3918.
[14] Margaret J Intons-Peterson and JoAnne Fournier. 1986. External and internal memory aids: When and how often do we use them? *Journal of Experimental Psychology: General* 115, 3 (1986), 267.
[15] David Kember. 1996. The intention to both memorise and understand: Another approach to learning? *Higher Education* 31, 3 (1996), 341–354.
[16] Annabel Latham, Keeley Crockett, David McLean, and Bruce Edmonds. 2012. A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education* 59, 1 (2012), 95–109.
[17] Panagiotis Mavridis, Owen Huang, Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2019. Chatterbox: Conversational interfaces for microtask crowdsourcing. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 243–251.
[18] James L McGaugh. 2000. Memory–a century of consolidation. *Science* 287, 5451 (2000), 248–251.
[19] Catherine Houdek Middendorf and Therese Hoff Macan. 2002. Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology* 87, 2 (2002), 293.
[20] Georg Elias Müller and Alfons Pilzecker. 1900. *Experimentelle beiträge zur lehre vom gedächtniss*. Vol. 1. JA Barth.
[21] Jaap MJ Murre and Joeri Dros. 2015. Replication and analysis of EbbinghausâĂŹ forgetting curve. *PloS one* 10, 7 (2015).
[22] Sihang Qiu, Ujwal Gadiraju, and Alessandro Bozzon. 2020. Improving Worker Engagement Through Conversational Microtask Crowdsourcing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[23] Bradley Rhodes and Thad Starner. 1996. Remembrance Agent: A continuously running automated information retrieval system. In *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology*. 487–495.
[24] Christopher Riederer, Jake M Hofman, and Daniel G Goldstein. 2018. To put that in perspective: Generating analogies that make numbers easier to understand. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 548.
[25] Rufat Rzayev, Tilman Dingler, and Niels Henze. 2018. ReflectiveDiary: Fostering Human Memory Through Activity Summaries Created from Implicit Data Collection. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018)*. ACM, New York, NY, USA, 285–291. https://doi.org/10.1145/3282894.3282907
[26] Norman J Slamecka and Peter Graf. 1978. The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and Memory* 4, 6 (1978), 592.
[27] Virpi Slotte and Kirsti Lonka. 1999. Review and process effects of spontaneous note-taking on text comprehension. *Contemporary Educational Psychology* 24, 1 (1999), 1–20.
[28] Donggil Song, Eun Young Oh, and Marilyn Rice. 2017. Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th international conference on human system interactions (HSI)*. IEEE, 78–82.
[29] Nam Khanh Tran, Andrea Ceroni, Nattiya Kanhabua, and Claudia Niederée. 2015. Back to the past: Supporting interpretations of forgotten stories by time-aware re-contextualization. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 339–348.
[30] Ran Yu, Ujwal Gadiraju, Peter Holtz, Markus Rokicki, Philipp Kemkes, and Stefan Dietze. 2018. Predicting user knowledge gain in informational search sessions. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 75–84.