Don't Judge by Looks: Search User Interface to Make Searchers Reflect on Their Relevance Criteria and Promote Content-Quality-Oriented Web Searches

Masaki Suzuki*
Yusuke Yamamoto
suzuki@design.inf.shizuoka.ac.jp
yusuke_yamamoto@acm.org
Shizuoka University
Hamamatsu, Shizuoka, Japan

ABSTRACT

Searchers often judge the webpage quality by the visual appearance, even for tasks where the content quality is more important such as health information. We propose a search user interface (UI) that makes searchers reflect on their webpage selection criteria and promote content-quality-oriented web searches regardless of visual appearances. The proposed UI visualizes how often searchers have browsed high-quality webpages during a search session. It does not explicitly tell the searcher that a high-quality webpage means high content quality; instead, the searcher is expected to reflect on their relevance criteria and improve their search behavior towards careful information seeking by considering the relationship between their page selection and the visualized score on the UI. We conducted an online user study to verify the effectiveness of the proposed UI, and the results indicated that participants without a university education tended to focus on the content quality of webpages and browsed higher-quality pages with the proposed UI than with a conventional UI.

CCS CONCEPTS

• **Information systems** → *Web searching and information discovery*; **Search interfaces**; • **Human-centered computing** → *User studies.*

KEYWORDS

web search, search interface, human factor, behavior change, critical information seeking

ACM Reference Format:
Masaki Suzuki and Yusuke Yamamoto. 2022. Don't Judge by Looks: Search User Interface to Make Searchers Reflect on Their Relevance Criteria and Promote Content-Quality-Oriented Web Searches. In *Conference on Information Technology for Social Good (GoodIT'22), September 7–9, 2022, Limassol, Cyprus.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3524458.3547222

*Both authors contributed equally to this research.

GoodIT'22, September 7–9, 2022, Limassol, Cyprus

ACM ISBN 978-1-4503-9284-6/22/09…$15.00
https://doi.org/10.1145/3524458.3547222

1 INTRODUCTION

Nowadays, it is common to learn about topics through web searches. Although web searches are convenient to obtain information, webpages often contain inaccurate information. Sillence et al. reported that more than half of the health information available on the web has not been verified by medical experts [17]. In addition, searchers tend to consider webpages with beautiful designs as credible information [10] and tend to prefer webpages that are easy to understand [4]. However, not all beautiful/comprehensive webpages provide accurate information. Conversely, unfashionable webpages can provide fruitful information. If searchers only focus on visually-attractive webpages, they are likely to obtain incorrect information and miss fruitful information. Therefore, information access systems are needed that encourage searchers to focus on the content quality of information rather than on the visual appearance.

Several studies have focused on supporting judgement on web information quality [5]. However, these systems are only effective in limited situations where searchers are strongly motivated to scrutinize information. Furthermore, several studies have reported that people do not frequently consider the credibility of information [13].

We propose search user interfaces (UIs) that makes the searcher reflect on the webpage selection criteria and promotes content-quality-oriented web searches regardless of the visual appearance. In this paper, we define that webpages with high content quality could contain data and evidence information about target topics and be authorized by experts. The proposed interfaces aim to visualize the page selection tendencies of searchers to promote self-reflection and support a more critical scrutiny of information. We considered two designs: ImplicitUI presents the content quality of the webpages implicitly, while ExplicitUI presents the content quality explicitly.

As shown in Figure 1(a), ImplicitUI presents a gauge next to a list of search results. The gauge represents how often searchers have browsed high-quality webpages during a search session by going up when the searcher browses high-quality webpages and going down when the searcher browses low-quality webpages. The change in the gauge score depends only on the content quality of the webpages, not the visual appearance. ImplicitUI does not explicitly tell the searcher that a high-quality webpage means high content

quality. The searcher is expected to reflect on their relevance criteria by considering the relationship between their page selection and the visualized score on the UI to improve their search behavior for more careful scrutiny of the information.

As shown in Figure 1(b), ExplicitUI presents two gauges next to a list of search results. One gauge represents how often searchers browse webpages with high content quality, and the other gauge represents how often searchers browse webpages with high visual quality webpages. In contrast to ImplicitUI, ExplicitUI explicitly reveals how often searchers browse webpages with high-quality information and with an attractive appearance. Therefore, the searcher can easily see whether they focus on content quality or visual quality in their web searches.

Through an online user study, this paper presents the followings:

- Participants without university education that used ImplicitUI tended to focus on the content quality of webpages and browsed higher-quality content than when using the conventional search interface.
- University-educated participants did not change their search behavior much even when they used ImplicitUI.
- ExplicitUI did not affect the search behavior and attitude of participants relative to the conventional search UI.

2 RELATED WORKS

Several studies have focused on the relationship between information quality and user behavior. Pothirattanachaikul et al. [15] analyzed the impact of document credibility and opinion on searchers. They reported that searchers spend more time searching and browsing when presented with webpages that conflict with their beliefs. Pogacar et al. [14] analyzed the behavioral logs of searchers to investigate the impact of the information quality in search results on search behavior and decision-making. Their results showed that, when the search results contain a large amount of wrong information, searchers are more likely to make wrong decisions. Kammerer et al. analyzed search behavior by manipulating the quality of search results to understand how searchers evaluate information about medical issues [6]. They found that searchers spend less time to evaluate the webpage quality the more that they believe the information to be credible.

Several researchers have focused on developing information access systems to promote careful web search behavior. Ennals et al. proposed a system called Dispute Finder, which highlights controversial sentences and presents both positive and negative opinions [5]. Schwarz et al. created a system to visualize the credibility of websites [16] that helps users evaluate the credibility of their search results. These studies focused on suggesting whether information was suspicious and providing information to judge the content quality.

Several studies have focused on evoking a critical attitude in the searcher instead of providing supplementary information for judging the content quality. For example, Yamamoto et al. proposed the Query Priming system, which inserts queries to evoke critical thinking during query completion/recommendation in a search system [19]. They found that the system caused searchers to modify their queries and visit more websites with data and evidence-based information. Yamamoto et al. also proposed Personalization Finder, which is a web browser extension to reveal the effects of web search personalization and promote careful web search practices [20]. Personalization Finder exposes the personalized and/or hidden search results so that searchers become aware that search engines provide them with a biased list of webpages according to their preferences. Bateman et al. [3] proposed Search Dashboard, which aims to make a searcher reflect on the search behavior. Their experiments indicated that Search Dashboard can help users modify their search behavior to improve search performance.

3 PROPOSED USER INTERFACES

In this study, we focused on two criteria for webpage quality: *content quality* and *visual appearance*. We defined the content quality as the extent of fruitful information provided by webpages in relation to a query. We defined the visual appearance as the attractiveness of a webpage to a searcher, which is related to aesthetics and readability. Our objective was to change the document selection criteria of searchers to search for webpages based on content quality regardless of visual appearance. We proposed two search UIs to achieve this objective: ExplicitUI and ImplicitUI. The proposed UIs are similar to Search Dashboard from Bateman et al. [3]. The main difference is that our proposed UIs aim to change the searcher's attitude and criteria for information selection to promote careful information seeking. Specifically, we adopted a method of visualizing the webpage quality to help users reflect on their browsing tendencies and to encourage more careful information searches. The differences between ExplicitUI and ImplicitUI are explained below.

ExplicitUI

As shown in Figure 2, ExplicitUI is a search UI with two gauges for visual appearance and content quality. The visual appearance gauge indicates the degree to which a searcher has viewed good-looking webpages on a list of search results. The content quality gauge indicates the degree to which a searcher has viewed webpages with high-quality content.

As shown in Figure 3, the gauge scores for ExplicitUI increase or decrease as searchers browse webpages on a list of search results. Specifically, when a searcher browses a webpage with high-quality content, the gauge score for the content quality increases by 10, and the gauge score for the visual appearance decreases by 10, and vice versa. The variation of the score was fixed to a minimum value of 0 and maximum value of 100. If the scores for the content quality and/or visual appearance of the webpages were over a predetermined threshold, the webpages were considered to have high-quality content and/or visual appearance. The two gauges explicitly suggest how often the searcher browses webpages with high-quality content and/or high-quality visual appearance. Therefore, searchers can intuitively determine whether they focus on content quality or visual appearance during web searches.

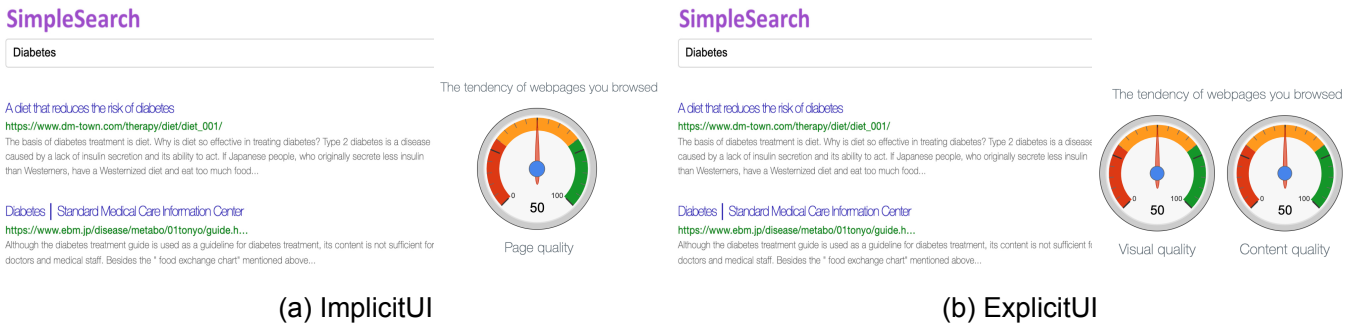


Figure 1: (a)ImplicitUI and (b) ExplicitUI.

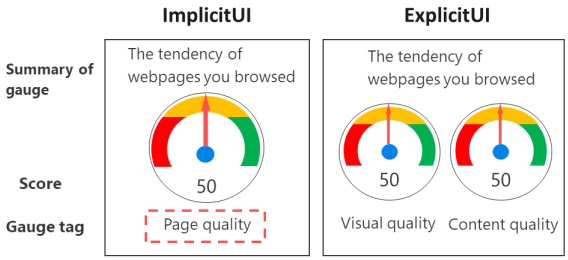


Figure 2: Difference between ImplicitUI and ExplicitUI.

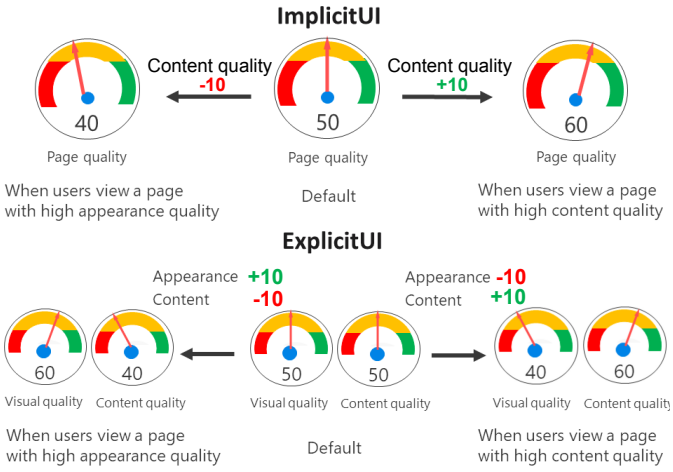


Figure 3: Rule of score changes for ImplicitUI and ExplicitUI. The red rectangle on ImplicitUI is just for explaining the UI and does not appear during actual use.

ImplicitUI

As shown in Figure 2, ImplicitUI has a single gauge that indicates the degree to which a searcher has viewed webpages with high-quality content on a list of search results. The gauge score increases only when the searcher views webpages with high-quality content. As shown in Figure 3, the score increases by 10 when a webpage with high-quality content is viewed and decreases by 10 otherwise. ImplicitUI regards a webpage to have high-quality content if the content quality is over a predetermined threshold.

Note that ImplicitUI does not explicitly suggest whether searchers have viewed webpages with high-quality content. The gauge is labeled as "page quality," not "content quality." A searcher using ImplicitUI cannot understand how the gauge score changes intuitively. To learn what page quality means, searchers have to associate their page selection with the gauge score as follows: (1) If searchers view webpages with high-quality content, the gauge score increases. (2) Even if searchers view webpages with good visual appearance, if the webpages have low-quality content, the gauge score decreases. Once a searcher learns the above rule, they can determine whether they have viewed a high-quality webpage. We designed ImplicitUI to make searchers reflect on their page relevance criteria and improve their search behavior to be more careful, regardless of visual appearance.

ImplicitUI is similar to ExplicitUI except that it has a single gauge with an ambiguous label. This study compared ImplicitUI and ExplicitUI to evaluate the effectiveness of a gauge that does not explicitly explain how its score changes.

3.1 Prediction of page quality

To change the scores on the gauges of ExplicitUI and ImplicitUI, we needed to know the content quality and visual appearance of each webpage on a list of search results. Several studies have proposed machine learning models to predict the content quality [1] and visual appearance [18]. Therefore, one approach would be to embed these models in our proposed search UIs. However, our aim was to study the effect of the UIs under the assumption that search engines can predict the content quality and visual appearance of webpages. For the sake of simplicity, we manually evaluated the

content quality and visual appearance of webpages in a corpus for our user study. Section 4 describes the details.

3.2 Hypotheses

We expect that the proposed UIs can encourage web searchers to critically scrutinize web information in the same manner which the information literacy researchers or librarians think is important [12]. Their search behaviors are expected to include: (1) spending more time searching, (2) browsing more webpages for comparison, and (3) checking evidence to support webpage content, such as the expertise of webpage authors and existence of valid references. Thus, we set the following hypotheses to verify the effectiveness of the proposed UIs:

- **H1** If searchers use the proposed UIs, especially ImplicitUI, they will spend more effort to select webpages on a list of search results than the conventional search UI.
- **H2** Searchers using the proposed UIs, especially ImplicitUI, will view more webpages with high-quality content.

Critical thinking is important to obtaining fruitful information via careful web searches [9]. Yamamoto et al. found differences in the critical thinking abilities and attitudes to web searches between searchers with and without university education [21]. Therefore, we set the following additional hypothesis:

- **H3** The effectiveness of the proposed UIs varies relative to educational background, where university education is assumed to represent critical thinking abilities.

4 USER STUDY

We performed an online user study to investigate the effects of ImplicitUI and ExplicitUI on search behavior and attitudes. The study was conducted in Japanese (on January 11 and 14, 2022). We recruited 159 participants from Lancers.jp, a Japanese crowdsourcing service. We excluded nine participants from the analysis because they did not complete all tasks, or they unintentionally used other search engines such as Google. Thus, we used the data from 150 participants. We paid the participants 200 JPY (approximately $2) for their time.

Each participant performed searches on two topics: HIV/AIDS and diabetes. The participants searched for webpages to learn about each of the above topics and improve their understanding. Before each search task, we asked the participants to answer quizzes to examine their comprehension of the topics. For HIV/AIDS, we used a quiz published by the Chugoku-Shikoku Regional AIDS Center [1]. For diabetes, we used a quiz published by the Japan Diabetes Association [2]. Each quiz had 10 questions with yes/no answers. On average, the participants had the correct answers number of 7.7 (SD = 1.1) for the HIV quiz and 6.4 (SD = 1.5) for the diabetes quiz. The results of the quizzes indicate that most participants did not have enough knowledge on the search topics.

[1] https://www.aids-chushi.or.jp/
[2] https://www.nittokyo.or.jp/

Table 1: Participant allocation. UE means "university educated".

	Educational Background	
UI	Not UE	UE
Control	18	30
ExplicitUI	25	25
ImplicitUI	24	28

4.1 Design and procedures

We adopted a 2 × 2 between-subjects design to examine the effects of two factors: the educational background and UI. The educational background factor had two levels: university-educated and not university-educated. The UI factor had three levels: ExplicitUI, ImplicitUI, and Control. Participants in the Control group used a search UI without gauges displaying the quality of browsed webpages. Table 1 presents the random assignment of the 159 participants to the three groups.

The user study comprised the following procedure: The first was a registration step. After participants agreed to the consent form on the crowdsourcing website Lancers.jp[3], they moved to our experimental website prepared by our laboratory. We randomly allocated the participants to each UI. Second, each participant answered quizzes to measure their comprehension of the two topics: HIV/AIDS and diabetes. Third, each participant performed two search tasks about HIV/AIDS and diabetes on our experimental search system, which displayed a fixed list of search results about each topic. We prepared the list before the user study. The participants were limited to not issuing queries and could only search the fixed list for information to learn about the topics. Note that the search process was not time-limited. Once a participant felt that they learned well a search topic, they stopped the search task. After the two search tasks, the participants answered questions about their attitudes during the search tasks as well as their gender, age, and educational background. The questions about their attitudes asked the participants how much they valued the content quality and visual appearance of the webpages on a five-point scale.

4.2 List of search results

In the user study, for each search task we presented the participants with a fixed list of 50 search results, similar to common search engines such as Google. As shown in Figure 4(a), we then displayed the search results with high-quality visual appearance and high-quality content alternately. The search results were sorted by the estimated score of visual appearance and content quality. In addition, we allowed the participants to learn whether the visual appearance/content quality of each search result was high or low on the search list, as shown in Figure 4(b). Although searchers cannot get such information on the quality information in actual situations, we used this setting so that the participants could intuitively judge the visual appearance/content quality of each search result without visiting the results page.

When the participants clicked on a search result, an archived version of the corresponding webpage was displayed. To measure

[3] https://www.lancers.jp/

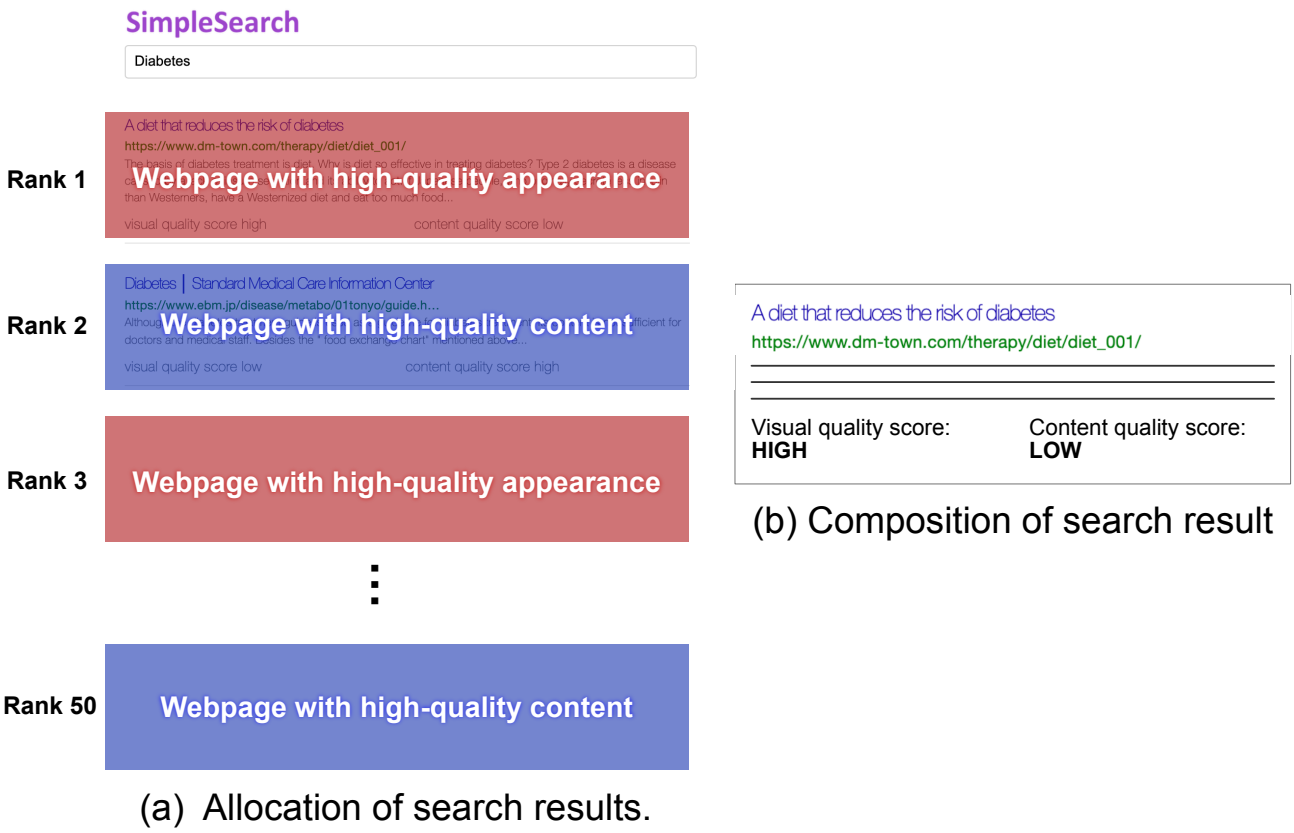


Figure 4: Composition of search results list. (a) Allocation of search results. The red and blue search results indicate high-quality visual appearance and content, respectively. (b) Composition of web search results.

the browsing time on each webpage, we embedded a JavaScript code in the archived webpages. We also disabled hyperlinks in the webpages so that participants could not view webpages other than those displayed in the search results list. Thus, we were able to measure the browsing time on only the webpages in the search results list.

4.3 Webpage quality

Before the user study, we used Lancers.jp to ask 100 crowdworkers to evaluate 1000 webpages on HIV/AIDS and diabetes for their quality. We collected the 1000 webpages by using Microsoft Bing Web Search API[4] with the queries "HIV" and "diabetes." Only the Japanese search results were collected. The crowdworkers evaluated the collected webpages from two perspectives: visual appearance and content quality. Each crowdworker evaluated ten randomly allocated webpages, and we paid 40 JPY (approximately 0.50 USD) as compensation. The workers evaluated each webpage by answering the following four questions:

- Q1: How good do you feel is the design of the webpage? (1: Very bad to 5: Very good)
- Q2: How easy do you feel the webpage is to read? (1: Quite difficult to 5: Quite easy)
- Q3: How much data and evidence from public institutions (government ministries, medical institutions, academic organizations) does the webpage contain? (1: No citation, 2: At least one citation from a public institution, 3: More than three citations from a public institution)
- Q4: How much expertise do you feel the website author has? (1: No info. about the author exists, 2: Info. about the author exists, but the author is not an expert, 3: Info. about the author exists, and the author is an expert)

The Q1 and Q2 were to evaluate the visual appearance, and the Q3 and Q4 were to evaluate the content quality. The visual appearance score was the sum of the Likert scores for Q1 and Q2. The content quality score was the sum of the Likert scores for Q3 and Q4.

After the evaluation, we calculated the final scores for the visual appearance and content quality of each webpage by averaging the workers' scores. A webpage was defined as high quality if the content quality or visual appearance score was higher than the overall average score.

5 RESULTS

Here, we describe our analysis of the collected behavioral data and post-task questionnaires.

[4]https://docs.microsoft.com/ja-jp/azure/cognitive-services/bing-web-search/

We employed a generalized linear mixed model (GLMM) [2] to analyze the behavioral logs and the questionnaire responses. A GLMM can separate the main effect of an intervention from random effects due to individual differences in the participants and tasks. In addition, a GLMM allows for more accurate analysis of small-scale data than methods using frequentist statistics [7]. The GLMM is becoming an increasingly established method of modeling user behavior in the information retrieval and human–computer interaction fields as an alternative to the traditional ANOVA [8].

We assumed that the dwell time on the search engine results pages (SERP) followed a Weibull distribution [11] and that the number of page views followed a Poisson distribution. We assumed that the comprehension score of the topic and attitude score towards information seeking followed a normal distribution. We used the 95% highest density interval (HDI) to measure the effects of the UI and university education factors. The 95% HDI represents the possible range of a parameter. If the HDI does not contain zero, the parameter is considered effective.

5.1 Search Behavior

We compared the dwell times on SERP to analyze how long participants spent selecting search results with each UI. Table 2 compares the ImplicitUI and ExplicitUI groups with the Control group; the 95% HDI of the coefficients for the UI, educational background (EB), and interaction contained zero. This is equivalent to not rejecting the null hypothesis and indicates that the UI did not affect the dwell time on SERP.

We compared the maximum browsing times of the webpages to evaluate the effects of a high-quality visual appearance or content. Table 2 compares the maximum page browsing times of the ImplicitUI, ExplicitUI, and Control groups; the 95% HDI of the coefficients for the UI, EB, and interaction contained zero. This indicates that the UI did not affect the maximum dwell time on a webpage.

We evaluated how often the participants viewed webpages with a high-quality visual appearance or content (i.e., page views). Table 2 presents the results. For ImplicitUI and ExplicitUI, the 95% HDI of the coefficients for the UI, EB, and interaction contained zero for webpages with a high-quality visual appearance. This indicates that the UI did not affect the number of page views for webpages with a high-quality visual appearance.

In contrast, the comparison between ImplicitUI and Control for webpages with high-quality content did not contain zeros in the 95% HDI of the coefficients for interaction. This result is equivalent to rejecting the null hypothesis. Because the interaction was confirmed, we performed a simple main effect analysis, which indicated that ImplicitUI significantly affected both participants with and without a university education ($p < 0.05$). The marginal effects are shown in Fig. 5(a). The ImplicitUI group tended to view more pages with high-quality content than the Control group when the participants were not university-educated (6.71 vs. 3.61). In contrast, university-educated participants in the ImplicitUI group viewed slightly fewer pages with high-quality content than the Control group (4.04. vs 5.28).

5.2 Attitude towards information seeking

We compared how much the participants focused on content quality and visual appearance with different UIs to evaluate their attitudes towards careful information seeking. Table 2 presents the results. For ImplicitUI, ExplicitUI, and Control, the 95% HDI of the coefficients for the UI, EB, and interaction regarding the visual appearance contained zero. This indicates that the UI did not affect the importance of the visual appearance.

In contrast, the results for the ImplicitUI and Control groups regarding the content quality did not contain zeros in the 95% HDI of the coefficients for interaction. Because the interaction was confirmed, we performed a simple main effects analysis, which indicated that ImplicitUI significantly affected only participants without a university education ($p < 0.05$). The marginal effects are shown in Figure 5(b). The ImplicitUI group tended to focus more on content quality than the Control group if they did not have a university education (4.17 vs. 3.60).

6 DISCUSSION

6.1 Hypothesis verification

We analyzed the dwell times on SERP and webpages to verify **H1** regarding the amount of time the participants spent during web searches. The results did not confirm that the UI had any effect on these two indicators. We assumed that users using ImplicitUI would spend more time browsing SERP because they would try to search for fruitful information with high-quality content more carefully. However, we did not observe such results. This may be because the quality scores of the webpages were directly displayed on SERP to help participants intuitively understand whether each webpage (search result) had a high-quality visual appearance or high-quality content. Therefore, the participants may not have had to make much effort to determine which webpages were of high quality, such as checking the summary of the search results and the webpages carefully. Therefore, we concluded that the UI did not affect the SERP dwell time.

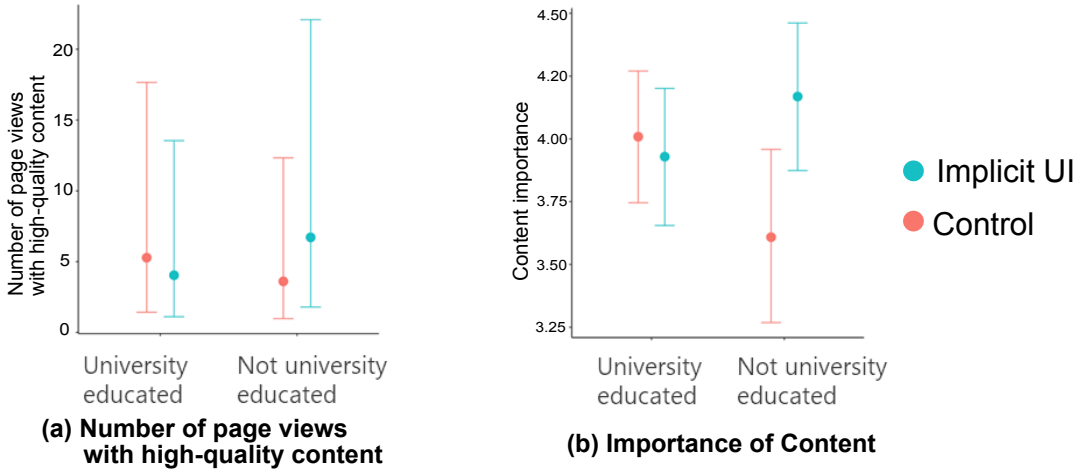
We assumed that the ImplicitUI group would focus on high-quality content more and thus would spend more time browsing webpages with high-quality content. However, we did not observe such results. This may be because the score was displayed on the search results, which is similar to the above discussion on the SERP dwell time. Participants in the ImplicitUI group who intended to view webpages with high-quality content could have judged the content quality based on the score provided on SERP. We concluded that the UI did not affect the dwell time on web pages. Therefore, we believe that **H1** was not verified.

We analyzed the number of page views to verify **H2**. Participants in the ImplicitUI group tended to view more pages with high-quality content than the Control group if they did not have a university education. In addition, the results on information seeking criteria indicated that the ImplicitUI group focused more on the content quality than the Control group if they did not have a university education. These results indicate that participants without university education in the ImplicitUI group placed more importance on content quality.

We assumed that university-educated searchers using the proposed UIs would have the advantage in critical thinking and thus

Table 2: GLMM results comparing the UI groups. Numbers represent the median and interval of 95% HDI. Numbers that do not contain zero in the 95%HDI are in bold.

	ImplicitUI			ExplicitUI		
Behavior index	UI	EB	Interaction	UI	EB	Interaction
Dwell time on SERP	0.05 [-0.33, 0.44]	0.12 [-0.55, 0.34]	0.25 [-0.36, 0.84]	0.21 [-0.20, 0.62]	0.12 [-0.55, 0.34]	0.31 [-0.31, 0.91]
Maximum dwell time on page (with high-quality visual appearance)	-0.16 [-0.67, 0.36]	-0.16 [-0.73, 0.42]	0.36 [-0.43, 1.14]	-0.04 [-0.58, 0.49]	-0.16 [-0.73, 0.42]	0.13 [-0.68, 0.91]
Maximum dwell time on page (with high-quality content)	-0.14 [-0.61, 0.34]	-0.40 [-0.93, 0.16]	0.46 [-0.29, 1.20]	0.08 [-0.43, 0.57]	-0.40 [-0.93, 0.16]	0.34 [-0.46, 1.11]
Page views (with high-quality visual appearance)	-0.23 [-0.67, 0.20]	-0.18 [-0.66, 0.33]	0.50 [-0.16, 1.16]	-0.22 [-0.65, 0.23]	-0.33 [-0.73, 0.07]	0.46 [-0.19, 1.16]
Page views (with high-quality content)	-0.22 [-0.56, 0.14]	-0.33 [-0.73, 0.07]	**0.73 [0.17, 1.27]**	-0.16 [-0.55, 0.20]	-0.33 [-0.73, 0.07]	0.56 [-0.01, 1.12]
	ImplicitUI			ExplicitUI		
Post-task questionnaire	UI	EB	Interaction	UI	EB	Interaction
Importance of visual appearance	0.01 [-0.62, 0.55]	$6.21e^{-3}$ [-0.60, 0.61]	$7.00e^{-3}$ [-0.87, 0.90]	-0.39 [-0.88, 0.26]	$6.21e^{-3}$ [-0.60, 0.61]	0.47 [-0.45, 1.20]
Importance of content quality	-0.08 [-0.49, 0.30]	-0.40 [-0.85, 0.04]	**0.64 [0.05, 1.28]**	-0.17 [-0.59, 0.22]	-0.40 [-0.85, 0.04]	0.56 [-0.06, 1.19]



(a) Number of page views with high-quality content

(b) Importance of Content

Figure 5: (a) Page views (with high-quality content) for each task and (b) importance of content quality according to educational background and UI (error bar indicates 95% confidence interval).

would browse webpages more frequently with high-quality content than the Control group. However, such results were not observed. Kusumi et al. stated that accurate information evaluation requires a critical attitude and critical thinking skills (e.g., language and reasoning skills) [9]. We assumed that the participants with a university education would be likely to have both the skills and attitude to use ImplicitUI to search for high-quality content more carefully. However, the user study results indicate that the university-educated participants already had a sufficiently critical thinking attitude, so ImplicitUI did not strongly affect their search behavior as we expected.

In contrast, ImplicitUI had a significant effect on the page views of webpages with high-quality content and information-seeking attitude of participants without a university education. We assumed that participants without a university education would not have sufficient critical thinking skills and attitude for ImplictUI to affect their search behavior. However, the results showed that ImplicitUI caused the participants without a university education to focus on the content quality rather than the visual appearance. We concluded that ImplicitUI can improve the attitude towards critical information seeking, and we believe that **H2** was verified. In addition, we believe that **H3** was verified because the effect of the UI differed depending on the educational background.

Our results indicate that the ImplicitUI group tended to place more importance on the content quality than the Control group, especially for participants without a university education. However, we did not observe this tendency in the ExplicitUI group. Compared to ExplicitUI, ImplicitUI does not explain the definition of high-quality webpages. The gauge in ImplicitUI is ambiguous. Therefore, the users of ImplicitUI could not find the definition without associating their search behavior with the change in the gauge score. We consider this association process to be important for the searcher to reflect upon and improve their attitude towards critical information searches. Therefore, ambiguous UIs may be an effective approach to help searchers reflect upon and review their search behavior and attitude. Designing an efficient and usable information access system is important. However, we believe that providing

opportunities for searchers to develop better information-seeking criteria is also important.

6.2 Limitations

Based on the results of this study, we identified three issues that should be addressed. The first issue was the estimation of the content quality and visual appearance of the webpages. In this study, we used a crowdsourcing service to estimate the webpage quality before the user study. For practical application of the proposed UIs, an automatic page quality estimation technique instead of manual annotation is necessary, such as that of Bahri et al. [1].

The second issue was the visualization of the quality scores. In the user study, our experimental search system showed the scores for the content quality and visual appearance along with each search result. Our intention was so that participants could intuitively judge the quality of the search results in terms of our two target criteria of visual appearance and content. However, this score visualization may have affected the participants' behavior and attitudes. To examine the substantial effects of the proposed UIs, we plan to conduct user studies where no quality information appears on SERP.

The third issue was the experimental design. We conducted an online user study using crowdsourcing and analyzed the behavior logs and questionnaire answers. However, this was not enough information to analyze the actual thoughts of the participants during the search tasks. For such a qualitative analysis, we will require a laboratory study. In addition, the persistence of the effects of the proposed UI also needs to be analyzed. Our results indicate that ImplicitUI encourages searchers without a university education to perform web searches oriented towards content quality. However, it is unclear whether searchers retain this effect after using the UI. Therefore, we have to analyze whether the proposed UI permanently enhances the critical thinking and attitude of its users.

7 CONCLUSION

We proposed search UIs that visualize the quality of browsed webpages to encourage searchers to reflect on their criteria for more careful information searches. To evaluate the effectiveness of the proposed UIs, we conducted an online user study using a Japanese crowdsourcing service. The results indicated that searchers without a university education and using ImplicitUI tended to (1) focus on the content quality of webpages and (2) view more webpages with high-quality content. In contrast, we did not confirm the above tendencies for searchers with a university education. ImplicitUI does not explicitly define high-quality webpages, which forces searchers to associate their search behavior with changes in the gauge score on the UI. We consider this association process important for improving the critical attitude and thinking skills for information searches.

ACKNOWLEDGMENTS

The work was supported in part by the Grants-in-Aid for Scientific Research (18H03244, 18H03554, 18H03775, 22H03905) from the MEXT of Japan.

REFERENCES

[1] Dara Bahri, Yi Tay, Che Zheng, Cliff Brunk, Donald Metzler, and Andrew Tomkins. 2021. Generative Models Are Unsupervised Predictors of Page Quality: A Colossal-Scale Study. In *Proc. of WSDM 2021*. 301–309.
[2] D. Barr, R. Levy, C. Scheepers, and H. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 3 (2013), 255–278.
[3] Scott Bateman, Jaime Teevan, and Ryen W. White. 2012. The Search Dashboard: How Reflection and Comparison Impact Search Behavior. In *Proc. of CHI 2012*. 1785–1794.
[4] Charles L. A. Clarke, Eugene Agichtein, Susan Dumais, and Ryen W. White. 2007. The Influence of Caption Features on Clickthrough Patterns in Web Search. In *Proc. of SIGIR 2017*. 135–142.
[5] Rob Ennals, Beth Trushkowsky, and John Mark Agosta. 2010. Highlighting Disputed Claims on the Web. In *Proc. of WWW 2010*. 341–350.
[6] Yvonne Kammerer, Ivar Bråten, Peter Gerjets, and Helge I. Strømsø. 2013. The Role of Internet-Specific Epistemic Beliefs in Laypersons' Source Evaluations and Decisions during Web Search on a Medical Issue. *Computers in Human Behavior* 29, 3 (2013), 1193–1203.
[7] Matthew Kay, Gregory L. Nelson, and Eric B. Hekler. 2016. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proc. of CHI 2016*. 4521–4532.
[8] Jaewon Kim, Paul Thomas, Ramesh Sankaranarayana, Tom Gedeon, and Hwan-Jin Yoon. 2017. What Snippet Size is Needed in Mobile Web Search?. In *Proc. of CHIIR 2017*. 97–106.
[9] Takashi Kusumi, Rumi Hirayama, and Yoshihisa Kashima. 2017. Risk Perception and Risk Talk: The Case of the Fukushima Daiichi Nuclear Radiation Risk. *Risk Analysis* 37, 12 (2017), 2305–2320.
[10] Gitte Lindgaard, Cathy Dudek, Devjani Sen, Livia Sumegi, and Patrick Noonan. 2011. An Exploration of Relations between Visual Appeal, Trustworthiness and Perceived Usability of Homepages. *ACM Transactions on Computer-Human Interaction* 18, 1 (2011).
[11] Chao Liu, Ryen W. White, and Susan Dumais. 2010. Understanding Web Browsing Behaviors through Weibull Analysis of Dwell Time. In *Proc. of SIGIR 2010*. 379–386.
[12] Marc Meola. 2004. Chucking the checklist: A contextual approach to teaching undergraduates Web-site evaluation. *portal: Libraries and the Academy* 4, 3 (2004), 331–344.
[13] Meredith Ringel Morris, Jaime Teevan, and Katrina Panovich. 2010. What Do People Ask Their Social Networks, and Why? A Survey Study of Status Message Q&A Behavior. In *Proc. of CHI 2010*. 1739–1748.
[14] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. 2017. The Positive and Negative Influence of Search Results on People's Decisions about the Efficacy of Medical Treatments. In *Proc. of ICTIR 2017*. 209–216.
[15] Suppanut Pothirattanachaikul, Takehiro Yamamoto, Yusuke Yamamoto, and Masatoshi Yoshikawa. 2019. Analyzing the Effects of Document's Opinion and Credibility on Search Behaviors and Belief Dynamics. In *Proc. of CIKM 2019*. 1653–1662.
[16] Julia Schwarz and Meredith Morris. 2011. Augmenting Web Pages and Search Results to Support Credibility Assessment. In *Proc. of CHI 2011*. 1245–1254.
[17] Elizabeth Sillence, Pam Briggs, Lesley Fishwick, and Peter Harris. 2004. Trust and Mistrust of Online Health Sites. In *Proc. of CHI 2014*. 663–670.
[18] Ou Wu, Yunfei Chen, Bing Li, and Weiming Hu. 2011. Evaluating the Visual Quality of Web Pages Using a Computational Aesthetic Approach. In *Proc. of WSDM 2011*. 337–346.
[19] Yusuke Yamamoto and Takehiro Yamamoto. 2018. Query Priming for Promoting Critical Thinking in Web Search. In *Proc. of CHIIR 2018*. 12–21.
[20] Yusuke Yamamoto and Takehiro Yamamoto. 2020. Personalization Finder: A Search Interface for Identifying and Self-Controlling Web Search Personalization. In *Proc. of JCDL 2020*. 37–46.
[21] Yusuke Yamamoto, Takehiro Yamamoto, Hiroaki Ohshima, and Hiroshi Kawakami. 2018. Web Access Literacy Scale to Evaluate How Critically Users Can Browse and Search for Web Information. In *Proc. of WebSci 2018*. 97–106.